Extremal graph theory and Ramsey theory

Class by Yuval Wigderson

Notes by Sanjana Das

July 7–25, 2025

Contents

1 July 7, 2025 **3**
- 1.1 A magic trick . . . 3
- 1.2 Extremal graph theory . . . 7
- 1.3 Extremal numbers . . . 7

2 July 8, 2025 **10**
- 2.1 The Erdős–Stone–Simonovits theorem . . . 11
- 2.2 The lower bound . . . 14
- 2.3 Bipartite graphs . . . 15

3 July 10, 2025 **19**
- 3.1 Review . . . 19
- 3.2 Lower bounds for 4-cycles . . . 20
- 3.3 Lower bound for $K_{3,3}$ . . . 23
- 3.4 Hypergraphs . . . 26

4 July 11, 2025 **28**
- 4.1 KST for hypergraphs . . . 28
- 4.2 Supersaturation . . . 33

5 July 14, 2025 **35**
- 5.1 Supersaturation . . . 35
- 5.2 Erdős–Stone–Simonovits . . . 38
- 5.3 Exact extremal numbers . . . 41

6 July 15, 2025 **44**
- 6.1 Stability . . . 47

7 July 17, 2025 **50**
- 7.1 Fermat's last theorem . . . 51
- 7.2 Ramsey's theorem . . . 54

8 July 18, 2025 **58**
- 8.1 Lower bounds . . . 58
- 8.2 Hypergraph Ramsey . . . 60
- 8.3 Points in convex position . . . 63

9 July 21, 2025 **65**
9.1 Graph Ramsey numbers . . . 65
9.2 Ramsey numbers of trees . . . 66
9.3 Complete bipartite graphs . . . 68
9.4 Ramsey numbers vs. average degree . . . 70
9.5 Degeneracy vs. large Ramsey numbers . . . 71

10 July 22, 2025 **73**
10.1 Ramsey numbers of bounded-degree graphs . . . 73
10.2 Overview of greedy embedding . . . 74
10.3 An embedding algorithm . . . 75
10.4 Multiple colors . . . 78
10.5 Monotone subsequences . . . 78

§1 July 7, 2025

Our lecturer is Yuval, and our TA Is Bryce, who will be running the problem sessions 4:30–5:30 in this room (or you can stretch into where we have lunch, which is probably better since we'll be working in groups). Yuval knows we're coming from many separate backgrounds. He doesn't have any concrete agenda for what we'll cover — there's a few things he thinks we really should get to, but other than that he'll adjust things based on how things go and what topics are interesting to us. So please tell him what's too hard or easy, or interesting or boring. We can send Yuval an email (you can find his email by googling his name), message on Whova, talk to Bryce, or also if you want to give anonymous feedback, there's a google form at `tinyurl.com/pcmiussfeedback` (of course, if you're comfortable saying things non-anonymously that'd be nice, but feel free to use this).

§1.1 A magic trick

We'll start with a magic trick (Bryce is the assistant). We need 10 integers between 1 and 100:

$$57, 53, 72, 13, 7, 32, 94, 24, 2, 89.$$

Bryce outputs $57 + 32 = 89$. So apparently, the sum of these two numbers equals that number.

This was a bit boring, because Yuval was hoping to get more numbers involved, so let's fire 89 and replace it with 90. Then Bryce gives

$$53 + 94 = 57 + 90.$$

We've probably seen enough magic, so Bryce now sits down.

What's the magic trick? The magic trick is that this always works.

Theorem 1.1

For any $a_1, \ldots, a_{10} \in [100]$, there exist disjoint nonempty $S, T \subseteq [10]$ such that $\sum_{i \in S} a_i = \sum_{j \in T} a_j$.

Notation 1.2. We write $[n] = \{1, \ldots, n\}$.

In words, this says that for any ten numbers between 1 and 100, I can find two disjoint nonempty subsets with the same sum. We saw two examples — we gave Yuval ten numbers, and Bryce did some magic and found in the first case two numbers with the same sum as one other number, and in the second case two sets of two numbers with the same sum.

Student Question. *Do the a_i have to be distinct?*

Answer. They don't have to be. But if they're not, then you're done for a silly reason — if two are equal, you can take S to be one instance of that number and T another. For the magic trick it's more impressive to make them distinct (otherwise you get $34 = 34$, which is not very impressive), but the theorem is true either way.

Proof. We're going to look at all possible sums — we'll define $a_S = \sum_{i \in S} a_i$ for all nonempty subsets $\emptyset \neq S \subseteq [10]$. Then there's two simple observations. First, there are 1023 nonempty subsets of [10] — there's 2^{10} subsets including $\emptyset$. So we've defined 1023 integers by doing this.

But second, for every S, I can get an upper bound on this number a_S — by definition we have

$$a_S = \sum_{i \in S} a_i \leq |S| \cdot 100 \leq 1000$$

(because each of the numbers is between 1 and 100, and $|S| \leq 10$).

So I've defined 1023 numbers and they're all between 1 and 1000, so by the pigeonhole principle, two of them have to be equal.

So far, we've shown there exist some $S \neq T$ such that $a_S = a_T$ (because we have 1023 numbers and only 1000 options for each). So I've found two *distinct* nonempty sets with the same sum. This wasn't what we promised — we promised *disjoint*, which is stronger. But luckily, there's an easy fix — if they overlap, I can throw away the intersection. For example, I could write

$$53 + 94 + 24 = 57 + 90 + 24.$$

Once I have equal sums with some intersection, I can throw away the intersection and keep the equality. $\square$

Student Question. *Why can't you get a better bound by just using* $1 + \cdots + 100$*?*

Answer. That would be more than 1000 — it'd be like 50000. It's important that we have 10 here. You can get a *slightly* better bound, though.

The point of this example was to show us what extremal combinatorics is — this is one of Yuval's favorite instances of extremal combinatorics.

In general, extremal combinatorics is about questions like:

Question 1.3. How many things can you have, or how large can a thing be, if it satisfies a certan property?

Extremal means you want something to be as big or small as possible, and combinatorics means we're working with discrete structures.

Question 1.4. What's the general phenomenon going on here? Can we do this with less than 10 numbers? Or what happens if instead of 10 numbers we have 1000?

Let's make a definition:

Definition 1.5. Define $\mathsf{magic}(n)$ be the maximum $M \in \mathbb{N}$ such that for all $a_1, \ldots, a_n \in [M]$, there exist disjoint nonempty $S, T \subseteq [n]$ such that $\sum_{i \in S} a_i = \sum_{j \in T} a_j$.

What this says is, suppose I want to play this magic trick and let the audience pick n numbers. What's the biggest possible range I can allow them to pick these numbers in such that no matter how they pick the numbers, I can always do this magic trick — I can always find two sets with equal sums?

Before we talk more about this function, this is an instance of extremal combinatorics — we're defining the maximum of something subject to some sorts of constraints. When we want to analyze a function like this, we want lower and upper bounds that are ideally as close as possible. What does it mean to prove a *lower* bound on this quantity? This means guaranteeing to you that some M works. For example, the proof from earlier tells you that $\mathsf{magic}(10) \geq 100$ — because what I proved to you is that 100 works in this definition.

What does it mean to prove an upper bound? Proving an upper bound means showing me that some M does *not* work. That boils down to finding n numbers in some range that *don't* have this property; that proves $\mathsf{magic}(n)$ is less than whatever that range was.

So here, proving a lower bound consists of a proof like the one we saw; proving an upper bound consists of finding an example.

Student Question. *Is there any interest in finding the minimum instead of maximum?*

Answer. It'd probably just be 1, because it's vacuously true for 1. In extremal combinatorics, these constraints are generally one-sided — the empty or complete object trivially satisfies it, and you want to go as far as possible in the other direction (the one that makes sense).

Let's talk a bit more about this function. First, we claim the same proof shows

$$\mathsf{magic}(n) \geq 102.$$

Why? The main thing is in the statement where we say $a_s = \sum_{i \in S} a_i \leq |S| \cdot 100 \leq 1000$. If we replace 100 with 102, then we get $10 \cdot 102 = 1020$, which is still less than 1023. And if you think through this in general:

Claim 1.6 — We always have $\mathsf{magic}(n) \geq \lfloor (2^n - 1)/n \rfloor$.

(We may or may not need a floor here.)

That's the lower bound. Can anyone come up with an upper bound? An upper bound means an example of some set of integers $a_1, \ldots, a_n$ in some range, where you do not have this distinct sums property.

Claim 1.7 — We have $\mathsf{magic}(n) < 2^{n-1}$.

The idea is that for any set of powers of 2, their sum is uniquely determined by the powers of 2 appearing in that sum (every number has a unique binary representation). So if you take 2^0, 2^1, 2^2, …, that'll be a bunch of numbers in this range with the property that you can't find two sets with the same sum. So this gives an upper bound (it's strict because it shows this number does *not* work as M, so the maximum M is strictly less).

Can you improve either of these bounds? One suggestion is to only consider sets of size at most 9 (for example). Our argument was a bit wasteful — there's no point in taking $S = [10]$, because there'll never be anything disjoint from that. So we can put a 9 here and do a tiny bit better. This will get you somewhere, but not to the state of the art.

The state-of-the-art lower bound is:

Theorem 1.8 (Erdős–Moser)
We have $\mathsf{magic}(n) \geq \Omega(2^n/\sqrt{n})$.

Before that, to back up a bit, these two bounds are quite close together — they're both on the scale of 2^n, and just off by a factor of n. (We'll see many things in this class that are much further.) But we'd like to close this gap.

This notation might not be familiar to everyone, so let's define it (we'll use it throughout the course):

Definition 1.9. We write $f \geq \Omega(g)$ to mean there exists some constant $c > 0$ such that $f(n) \geq cg(n)$ for all sufficiently large n.

So $f \geq \Omega(g)$ means I have a lower bound, up to some absolute constant factor that I'm not specifying. This is a close cousin of something we're more likely to have seen, big-O notation (which is basically just the reverse):

Definition 1.10. We write $f \leq O(g)$ to mean $g \geq \Omega(f)$.

Erdős and Moser, in the 1960s, came up with a really beautiful argument that got something along these lines of 2^n, but instead of dividing by n, they divided by $\sqrt{n}$. Their proof will be an optional problem on the homework. The amazing thing is it's probabilistic.

This program is called probabilistic and extremal combinatorics. The question has no probability anywhere — it's purely deterministic. But in many instances in combinatorics, probabilistic tools turn out to be very powerful. We won't see this proof in class, but we'll see other instances where probability comes into play.

To close out this story, Yuval will tell us what the absolute state of the art is:

Theorem 1.11 (Dubroff–Fox–Xu, Bohman)
We have
$$\left(\sqrt{\frac{2}{\pi}} - o(1)\right) \cdot \frac{2^n}{\sqrt{n}} \le \mathsf{magic}(n) \le 0.22002 \cdot 2^n.$$

Both of these are along the same lines as what we had here. The upper bound is still some specific constant times 2^n, it's just that the specific constant is better (we had $2^n \cdot 1/2$; the upper bound is due to Bohman, who found an explicit construction that does a bit better than powers of 2 and gets a better constant). The lower bound is a recent result of Quentin Dubroff, Jacob Fox, and Max Xu. Again, the shape of it is like Erdős–Moser's lower bound of $2^n/\sqrt{n}$. They just got a better explicit constant, the funny $\sqrt{2/\pi}$. For more notation:

Notation 1.12. We write $o(1)$ to mean something that tends to 0 as $n \to \infty$.

So this means for very large n, you can get arbitrarily close to $\sqrt{2/\pi}$.

The upper bound is from 15 or 20 years ago, and the lower bound is from 2 years ago. Yuval was hoping to tell us all three of these people are here, but unfortunately none of them are.

Student Question. *Is there a reason π shows up?*

Answer. Yes, ask Yuval about it later.

Student Question. *Are there any explicit algorithms that are better than quasipolynomial to find S and T?*

Answer. Yuval isn't sure.

Just to end off this section, Yuval wants to tell us this isn't just some weird thing no one cares about. There's a major conjecture by Erdős:

Conjecture 1.13 (Erdős distinct sum conjecture) — We have $\mathsf{magic}(n) \ge \Omega(2^n)$.

So Erdős predicted we shouldn't need the n or $\sqrt{n}$. This is wide open, and Erdős offered 500 dollars for it. Erdős was a famous mathematician from the last century. One of his great talents was coming up with amazing problems, and he liked to assign monetary values to some of them. In these monetary values, 500 is quite big (when he made the offer, 500 was worth more; he has an old paper in some conference in Japan where he offers some number of yen, or the equivalence in US dollars). That's to indicate this problem is very difficult; it's been wide open for decades.

Student Question. *If you solved one of these problems, it'd change your career path; what's the estimate for how much?*

Answer. Much more than this. Erdős passed away in the 90's; for decades Ron Graham continued to

> give them out, but then he passed away too, and at the moment no one is likely giving them out. So you shouldn't solve this just for the 500 dollars because you will likely not get it, but it'd still be a big deal.

The reason Yuval is bringing this up isn't just to wow us with big numbers. Realistically, we're not likely to solve the conjecture, because it's probably very difficult. But there's an amazing phenomenon in extremal combinatorics where we get problems that are very easy to describe, at the complete forefront of human knowledge. We'll see many examples this course, as well as breakthroughs on problems that were previously intractable. Lots of these breakthroughs too are such that once you know what you're doing, they're not very hard — you just need one clever idea. One thing Yuval loves about this area is how easy it is to see the forefronts of human knowledge, and how you often only need one clever idea to get beyond it. Yuval does encourage us to think about this problem (and if we do get something, many people would be thrilled).

§1.2 Extremal graph theory

This class is about extremal graph theory and Ramsey theory; we'll start with extremal graph theory.

What's extremal graph theory? It's the same kind of thing, but now we're working with graphs. Here are two facts:

> **Fact 1.14** — If an n-vertex graph has no cycles, then it has at most $n-1$ edges.

This is more commonly known as the fact that every tree on n vertices has exactly $n-1$ edges.

> **Fact 1.15** — Every n-vertex planar graph has at most $3n-6$ edges (for $n \geq 3$).

(If you don't know what this is, don't worry; we're not going to talk about it at all.)

Both of these are instances of extremal graph theory statement. Generally, I give you a natural property of graphs (some constraint on the graph) and ask how many edges it can have subject to that constraint. In these examples, both of these are the strongest possible upper bounds, so we understand everything.

So that's what extremal graph theory is about, and what we'll be talking about for a while.

§1.3 Extremal numbers

The main type of extremal question we'll be studying, the most classical one, is where the constraint we'll impose is forbidding some fixed subgraph.

> **Definition 1.16.** We say G is H-free (where G and H are graphs) if H is not a subgraph of G.

So G is H-free if you cannot find H as a subgraph of G.

> **Definition 1.17.** The extremal number is defined as
>
> $$\mathrm{ex}(n, H) = \max\{e(G) \mid G \text{ is an } n\text{-vertex } H\text{-free graph}\}.$$

So the extremal number is the most number of edges you can have in a graph on n vertices without H as a subgraph.

This notion was first studied by Mantel in 1907, who figured out the extremal number of a triangle (i.e., K_3). Proving a lower bound means giving an example of a graph with many edges and no triangle. To prove an upper bound on this quantity, you'd show me that some number of edges *can't* work — that if I give you any graph with m edges, as long as m is at least some specific value, it's guaranteed to have a triangle.

To get started, let's try to think about lower bounds. Can anyone come up with a good construction for a graph with many edges and no triangle? Any tree is going to work, and gives $\mathrm{ex}(n, K_3) \geq n - 1$. But we can do better. Another example is a grid, which will have roughly $2n$ edges. We can still do even better — you can take a complete bipartite graph, and this is actually going to be the right answer.

Claim 1.18 — For any a and b with $a + b = n$, we have $\mathrm{ex}(n, K_3) \geq ab$.

Proof. Take a complete bipartite graph — take two blobs with a and b vertices (respectively), and plop down all possible edges between them.

This can't have a triangle because two vertices would have to be on the same side by Pigeonhole, and we don't have any edges inside a side. □

This holds for any $a + b = n$, but there's a best choice, which is making them as equal as possible. This best choice is to make them as equal as possible, because of the AM–GM inequality — you have

$$ab \leq \left(\frac{a+b}{2}\right)^n = \frac{n^2}{4}.$$

Equality holds if and only if they're equal. If n is odd you can't do this, but you can take them to be one above and one below. So we always have

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

You can see that for large n, this is quite a bit more than our previous examples — those grow linearly in n, while this grows quadratically.

Can anyone think of a better example? You cannot:

Theorem 1.19 (Mantel 1907)

We have $\mathrm{ex}(n, K_3) = \lfloor n^2/4 \rfloor$. Moreover, the unique extremal example is $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

What that means is that this number of edges is the most edges you can have in an n-vertex graph without a triangle; and if you have that number of edges, this is the only example you can come up with. So any graph with $\lfloor n^2/4 \rfloor$ edges which is not this one (the complete bipartite graph) is guaranteed to have a triangle.

We'll prove this in a minute (hopefully). But before we do, we'll generalize it. Yuval doesn't really know why Mantel was thinking about this question, but in the 1940s Turán for other reasons came up with the same question, but more general; and this is what really kicked off the whole field of extremal graph theory.

Question 1.20. What is $\mathrm{ex}(n, K_r)$?

So rather than a triangle, we're asking for any complete graph (on r vertices).

Generalizing the bipartite example, can anyone come up with a reasonable lower bound? What we should do is divide our world into $r - 1$ blobs. We'd like to have each have size $n/(r-1)$, but because of divisibility issues, we really have floors and ceilings. And we put all edges between the blobs, and none inside. This will be K_r-free for the same reason as before — for any r vertices, two of them must be in the same blob, so there won't be an edge between them. And if you want to maximize the number of edges with this property, you should make all the blobs as equal as possible, again for AM-GM-type reasons.

Definition 1.21. The Turan graph, denoted $T_{r-1}(n)$, is the complete $(r-1)$-partite graph on n vertices where all blobs have size $\lfloor n/(r-1) \rfloor$ or $\lceil n/(r-1) \rceil$.

How many edges does it have? With Mantel you could write down the number of edges as a nice explicit formula. Here you can also do that, but it's annoying because of divisibility issues, so:

Definition 1.22. We define $t_{r-1}(n)$ as the number of edges $T_{r-1}(n)$ has.

It's a homework problem to figure out what this is. It's not that bad when $r-1 \mid n$ — then we have

$$t_{r-1}(n) = \binom{r-1}{2}\left(\frac{n}{r-1}\right)^2. = \left(1 - \frac{1}{r-1}\right)\frac{n^2}{2}$$

(we have $\binom{r-1}{2}$ of these complete bipartite pieces, and the number of edges in each piece is the product of blob sizes).

This formula is only valid when $r-1 \mid n$, but for *all* n we have

$$t_{r-1}(n) = \left(1 - \frac{1}{r-1} + o(1)\right)\frac{n^2}{2}.$$

So this is still the right formula; you just have a small error (which is negligible for large n).

That's merely the construction; hopefully you can guess what Turán's theorem says.

Theorem 1.23 (Turán 1941)
We have $\mathrm{ex}(n, K_r) = t_{r-1}(n)$. Moreover, the graph $T_{r-1}(n)$ is the unique extremal example.

Student Question. *Unique up to isomorphism?*

Answer. Yes.

Of course, you see that the case $r = 3$ of this theorem is precisely Mantel's theorem; so this is a strict generalization of Mantel's theorem.

We have 10 minutes, so we have time to prove it. This is an amazing theorem with a huge number of proofs; over this week there'll be many on the homework assignments. Mantel's theorem has even more. But we'll just see one in this class, because Yuval wants to get to more theorems rather than just proving Turán over and over again. But there's lots of nice ones, so we should look at the homework.

Before we prove this, we'll make an observation about $t_{r-1}(n)$ (he'll write down a crazy formula and then explain it to us).

Fact 1.24 — We have $t_{r-1}(n) = t_{r-1}(n-r+1) + (r-2)(n-r+1) + \binom{r-1}{2}$.

How do we prove this formula? What you're going to do is think about what happens if I take one vertex from each blob and just delete these vertices. There's two observations. I started with $t_{r-1}(n)$ edges, and after I remove one vertex from each blob, I end up with another Turán graph on $n-r+1$ vertices, so the number of edges I have is $t_{r-1}(n-r+1)$. I deleted some edges; how many? Between the deleted (red) vertices I deleted exactly $\binom{r-1}{2}$. And between the red and undeleted vertices, each undeleted vertex is adjacent to exactly $r-2$ red vertices (all the ones except the one in its part). So the number of edges between red and white vertices is exactly $(r-2)(n-r+1)$.

Proof of Turán's theorem. We're going to do induction on n, but an unusual induction — instead of going up by 1 every step, we'll go up with steps of size $r-1$. So we'll prove the statement for n based on the statement for $n-r+1$. This means you actually need $r-1$ different base cases, rather than just 1 — the base cases are when $n = 1, 2, \ldots, r-1$. Why? The theorem is not very interesting if n is this small, because if I give you a graph with fewer than r vertices, no matter what it is, it won't have K_r as a subgraph. So the extremal number is $\binom{n}{2}$ if $n < r$ (because you take the complete graph). And this Turán graph is a complete graph on n vertices (and of course this is the unique extremal example).

For the inductive step, to prove an equality for the extremal number, we need both a lower and upper bound. We've already seen a lower bound — it comes from the explicit example — so now we need an upper bound. For this, let G be an n-vertex K_r-free graph with the maximum number of edges. So we just take any example with the maximum number of edges. We're going to prove first that the number of edges is at most $t_{r-1}(n)$, and then that if equality holds it must be our one example.

First note that $K_{r-1} \subseteq G$ — i.e., G must contain some copy of K_{r-1}. Why? Otherwise I can just add any missing edge, and I have no hope of creating a K_r by accident, so G can't have the maximum number of edges (since I can just add an edge and maintain K_r-freeness).

So let K be a copy of K_{r-1} in G, and let $F = G \setminus K$. So I'm deleting all the vertices from K.

And now we're just going to think about how many edges live in this picture. First of all, how many edges does F have? We have

$$e(F) \le t_{r-1}(n-r+1).$$

This is where we're using the induction hypothesis — the reason is that if I remove these vertices, I get a graph F on a smaller number of vertices that also can't have a copy of K_r. So by induction, the number of edges F has must be at most the maximum number of edges for $n-r+1$ vertices, which is this.

And we know $e(K)$ exactly — it's $\binom{r-1}{2}$ (because K is a copy of the complete graph).

The final step is, how many edges can there be between F and K (which we denote by $e(F,K)$)?

> **Claim 1.25** — Each vertex in F is adjacent to at most $r-2$ vertices in K.

Proof. Otherwise we'd have some vertex in F which is adjacent to all $r-1$ vertices in K, which would give a copy of K_r in the original graph. □

This means the number of edges across is at most $(r-2)(n-r+1)$ (where $n-r+1$ is the number of vertices in F). And now we've learned

$$e(G) = e(F) + e(K) + e(F,K) \le t_{r-1}(n-r+1) + \binom{r-1}{2} + (r-2)(n-r+1).$$

And if you recall our magic formula, that's exactly $t_{r-1}(n)$ — it's exactly what we're looking for.

We're not quite done — weve only proved the equality $\text{ex}(n, K_r) = t_{r-1}(n)$. But we also need to prove the Turán graph is the unique extremal example. However, since it's 2pm, we won't do that. But the point is if equality holds, every single one of these inequalities is an equality; that tells you a lot about the structure of the graph, and you can induct. □

§2 July 8, 2025

Rob Morris is giving a talk today (3:15–4:15, in the other end of the building in the big room, the Bison conference room). We're encouraged to attend it. (Rob is an incredible speaker.)

On the homework, there are recommended exercises and optional problems. Everything is optional; if you don't want to work on any of the problems that's fine, but a bad idea (it's hard to learn). But the point is so we have a sense of which problems are more important to make sure we're keeping up, the exercises. The problems are usually in a different direction or harder. We're encouraged to look at them, but we don't have to. Even with the exercises, we shouldn't feel like we have to solve all of them; but it's a good idea to look at all of them and spend a bit of time thinking about them, so we have some idea of what goes into them.

Tell Yuval if we have any suggestions; there is also a feedback form at tinyurl.com/pcmiussfeedback.

As a reminder of what we talked about yesterday, we proved Turán's theorem:

Theorem 2.1

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

The extremal number of a graph H (here, K_r) is the most number of edges you can have in a graph on n vertices without a copy of H. The right-hand side is the number of edges of $T_{r-1}(n)$ — the graph on n vertices where you split them into $r-1$ blobs of sizes as equal as possible, and draw all edges between blobs and none inside blobs. We have

$$t_{r-1}(n) = \left(1 - \frac{1}{r-1} + o(1)\right)\frac{n^2}{2}$$

($o(1)$ means something that tends to 0 as $n \to \infty$; when n is divisible by $r-1$ you don't need the $o(1)$ error term). One useful way of thinking about this is that it's also

$$\left(1 - \frac{1}{r-1} + o(1)\right)\binom{n}{2}$$

(all we've done here is that $\binom{n}{2} = \frac{n^2}{2} - \frac{n}{2}$, which is asymptotically basically the same as $\frac{n^2}{2}$; so if we're already okay with a bit of noise, we can switch between them). The advantage of writing it this way is that $\binom{n}{2}$ is a natural benchmark — a graph on n vertices has anywhere between 0 and $\binom{n}{2}$ possible edges. So Turán says that if you have a K_r-free graph, the fraction of possible edges you can have is at most roughly $1 - 1/(r-1)$.

§2.1 The Erdős–Stone–Simonovits theorem

Already in Turán's original paper (Yuval hasn't read it because it's in Hungarian, but he has been told this), he points out that he's solved this for K_r, and asks the natural question:

Question 2.2. What's $\operatorname{ex}(n, H)$ for other graphs H?

This was from the 1940s, when people thought about graphs a bit differently. Turán had solved the K_4 case; K_4 is the graph of the tetrahedron, so he suggested to study what happens for the graphs of other platonic solids. That turns out not to be the most interesting direction, but of course he didn't know that at the time. In general, what can we say about this extremal function?

Yuval's first reaction is this is probably going to be very difficult, because it'll probably depend a lot on the structure of H. But the amazing thing (which we'll talk about for the rest of the week) is that that's actually not true — there's a very simple answer!

Depending on who you ask, this is either called Erdős–Stone or Erdős–Stone–Simonovits, and was either proved in 1946 or 1966. Erdős–Stone in 1946 proved a special case, which sounds like a very niche one. For 20 years no one noticed that this special case immediately resolves the general one (with no extra work); this was observed in a paper with Simonovits in 1966.

This is also sometimes called the fundamental theorem of extremal graph theory — this is the most fundamental question in extremal graph theory, and the theorem tells you the answer.

Theorem 2.3

For any graph H, we have

$$\operatorname{ex}(n, H) = \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right)\binom{n}{2}.$$

Intuitively you'd expect the value of $\operatorname{ex}(n, H)$ to depend on the actual structure of H in some weird, complicated way. But that's not true — it's completely controlled by a single property of H, its chromatic number.

Definition 2.4. The chromatic number of a graph, denoted $\chi(H)$, is the minimum number of colors needed to color the vertices such that adjacent vertices receive different colors.

And this theorem says that parameter of the graph completely determines the extremal function.

Student Question. *Do we know the exact extremal graph?*

Answer. This is open in general. For many H we know the exact extremal structure, but in general it's not known. There's a famous conjecture called the Simonovits product conjecture that kind of predicts what the structure should be (we're not going to state this; it's complicated).

Student Question. *How hard is it to compute the chromatic number?*

Answer. If I draw you a graph, you can do it pretty quickly. In general, computing the chromatic number is very difficult (NP-hard). For our purposes we don't care — we'll be talking about graphs like the 5-cycle, where you can just look at it and say its chromatic number is 3.

One other remark, that's very important: If H is bipartite (meaning that $\chi(H) = 2$), then what does this say? We have

$$1 - \frac{1}{2 - 1} = 0,$$

so this tells you $\operatorname{ex}(n, H) = o(1) \cdot \binom{n}{2} = o(n^2)$ (this means that $\lim_{n\to\infty} \operatorname{ex}(n, H)/n^2 = 0$). So what's going on? With triangles, we saw Mantel's theorem, where the extremal number is basically $n^2/4$ — some positive constant times n^2. So for a triangle, this limit is not 0. In fact, ESS says that for any H which is *not* bipartite, the actual extremal number is some specific positive constant times $\binom{n}{2}$, in the limit. But for bipartite graphs, the thing that's supposed to be the main term is 0, and all that remains is what was supposed to be the error term.

That says two things. First, it's highly non-obvious that this is true, that

$$\lim_{n\to\infty} \frac{\operatorname{ex}(n, H)}{n^2} = 0.$$

For triangles, or even any odd cycle (which has chromatic number 3), the extremal number is roughly $n^2/4$. If you take a complete bipartite graph, it doesn't have any odd cycles and has roughly $n^2/4$ edges. So that's a graph with quadratically many edges and no odd cycles. But if you're interested in any even cycle, like C_4 or C_6 or C_{100}, you can't construct a graph with quadratically many edges and no copy of that even cycle. This is quite weird — you can avoid every even cycle with quadratically many edges, but you can't even avoid a single even cycle. This is a highly non-obvious fact; we'll prove it in a little bit. It's a special case of ESS, but as we'll see this week, it's almost fully equivalent. The proof of ESS goes by proving this thing for bipartite graphs, proving a suitable generalization, and then using that to prove ESS.

One other thing we'll say about this, which we'll come back to, is that for non-bipartite graphs this tells you what the answer is. For bipartite graphs it doesn't — it tells you it's subquadratic but doesn't tell you the actual growth. That's the major question in extremal graph theory — to understand the behavior of the extremal function for bipartite graphs. Yesterday we saw $\text{ex}(n, C_3) = \lfloor n^2/4 \rfloor$; as we'll see in a moment, it turns out that

$$\text{ex}(n, C_4) = \Theta(n^{3/2})$$

(where $\Theta(n^{3/2})$ means we have both upper and lower bounds of the form $cn^{3/2}$ — it's the intersection of Ω and O notation). For C_5, ESS tells us that

$$\text{ex}(n, C_5) = \left(\frac{1}{4} + o(1)\right) n^2,$$

but in fact we know that it's actually exactly $\lfloor n^2/4 \rfloor$ for n large (we may get to this next week). For the 6-cycle, we'll talk a little about this; it turns out that the answer is

$$\text{ex}(n, C_6) = \Theta(n^{4/3}).$$

(In fact, for the 4-cycle we know exactly what the implicit constant is; for the 6-cycle that actual constant is still a major open question.) For C_7, it's the same story, and

$$\text{ex}(n, C_7) = \lfloor n^2/4 \rfloor$$

for large n. And $\text{ex}(n, C_8)$ is wide open — no one knows. This is one of the biggest open problems in the field. We have an upper bound and a lower bound, but they don't match.

> **Student Question.** *When we say n large, do you mean $n \geq 7$ or $n \geq 2^{100}$?*
>
> **Answer.** When we prove it, we'll prove it for $n \geq 10^9$. For C_5 we know it's false for $n = 8$ and true for $n = 9$, or something like that; so the truth is 'not very large.'

The odd cycle case is easy, so again we know $\text{ex}(n, C_9) = \lfloor n^2/4 \rfloor$ (as we'll be shocked to hear). But very strangely, we actually know the answer for C_{10} — we have

$$ex(n, C_{10}) = \Theta(n^{6/5}).$$

And C_{10} is the last even cycle we know the answer for. So this is weird — for even cycles we know C_4, C_6, and C_{10}, but not C_8 or C_{12} or C_{14} or anything larger; this is an unusual quirk of nature. So there's a lot we don't know; it's a major open problem to learn more about this.

> **Student Question.** *Is $n^{5/4}$ a natural guess for C_8?*
>
> **Answer.** Looking at these exponents, that makes sense. We do know an upper bound of $n^{5/4}$, and for all even cycles we have an upper bound that matches this pattern you observed. What's missing is the lower bound. Seven years ago Yuval would've said anyone is certain these are tight. These days people are less certain; David Conlon is a firm believer the answer for C_8 is not $n^{5/4}$ but something smaller. Yuval is agnostic; lots of people think he's right, and lots of people think he's wrong.

> **Student Question.** *Are there lots of bipartite graphs this is open for?*
>
> **Answer.** Yes, it's open for pretty much everything. We'll see two bipartite graphs for which it's known, but for most it isn't.

As a bit of a tangent, here are two famous conjectures, in sort of opposite directions:

Conjecture 2.5 (Erdős–Simonovits) — For every bipartite H, there exists a rational number $\alpha \in [1, 2)$ such that $\mathrm{ex}(n, H) = \Theta(n^\alpha)$.

This conjecture says that whatever's going on, there's some exponent — some rational number α — where that's the answer.

They also conjectured some sort of weird converse:

Conjecture 2.6 — For every rational α, there exists some bipartite H such that $\mathrm{ex}(n, H) = \Theta(n^\alpha)$.

This is a much crazier conjecture — actually if you pick your favorite rational number, there is a bipartite graph that achieves that. Experts are basically certain that one of these conjectures is true and one is false. Surprisingly, the *second* is almost certainly true and the first is almost certainly false. This is super weird. The reason we're bringing this up is someone asked if there are other bipartite graphs for which we know the answer. The answer is yes, but not natural ones. There's a huge collection of rational numbers now where we know how to prove there is some H that achieves them; so we know the answer for that H. But they're not the natural bipartite graphs you'd write down yourself.

Student Question. *Why do we think the first one is false?*

Answer. There are many analogous questions in closely related fields for which the analogous statement is not true. So if it's true, it's true for weird reasons related to the specifics of bipartite graphs; that's possible but seems unlikely, since we have general phenomena elsewhere where this doesn't happen.

§2.2 The lower bound

Proving ESS will take us basically to the end of this week. But one direction is super easy.

Proposition 2.7

For every H, we have $\mathrm{ex}(n, H) \geq t_{\chi(H)-1}(n)$.

In particular, as we've seen earlier,

$$t_{\chi(H)-1}(n) = \left(1 - \frac{1}{\chi(H)-1} + o(1)\right)\binom{n}{2}.$$

ESS tells us that $\mathrm{ex}(n, H)$ is *equal* to something of this form; the lower bound is easy, and the hard part is the upper bound.

Proof. We claim that the Turán graph $T_{\chi(H)-1}(n)$ is H-free. So we're claiming this specific graph doesn't have a copy of H; that's good enough for the lower bound because this graph has $t_{\chi(H)-1}(n)$ edges, so the *maximum* number of edges you can have is at least this.

How do we prove this? Suppose it's false. What does that mean? In this Turán graph, remember we have $\chi(H) - 1$ blobs. And somehow, we have a copy of H living in this picture. But if you think about what this means, because the Turán graph has no edges inside the blobs, in this copy of H I've drawn, I know I definitely don't have an edge between anything in the same blob. So this gives me a way of coloring the vertices of H with $\chi(H) - 1$ colors — give everyone in the first blob the first color, everyone in the second blob the second color, and so on. By construction, two vertices with the same color don't have an edge. So this gives a coloring of H with $\chi(H) - 1$ colors, which is impossible (since $\chi(H)$ is the least number of colors you can do this with). □

So the lower bound is easy — the only example we've seen so far also works for this more general problem.

Student Question. *Is the Turán graph the extremal example then?*

Answer. Not in general. In some cases yes — for example, for all odd cycles. On the homework we'll see some instances where the Turán graph is not extremal — sometimes you can add a few extra edges to the Turán graph to get a graph with a few extra edges that still has no H. The Simonovits product conjecture says that's the best you can do — that the true extremal example is obtained from the Turán graph by fiddling a bit. But that's wide open in general. (Specific examples will be on the homework.)

To stress this point, here we've proved a lower bound $\mathrm{ex}(n, H)$ is at least this specific value. It's *not* true that we have an upper bound $\mathrm{ex}(n, H) \le t_{\chi(H)-1}(n)$. We'll prove an upper bound of the same asymptotic shape, but not with this specific value (that's not true in general).

§2.3 Bipartite graphs

Now let's talk about bipartite graphs. Our next goal for a little bit is going to be proving at least $\mathrm{ex}(n, C_4) = \Theta(n^{3/2})$, and hopefully the statement that

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{n^2} = 0$$

in general (but we'll warm up with C_4).

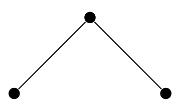
In fact we'll make things a bit easier to start out with:

Proposition 2.8
Let G be an n-vertex d-regular graph with $d \ge 2\sqrt{n}$. Then $C_4 \subseteq G$.

What does this say in contrapositive? If I give you an n-vertex d-regular graph that is C_4-free, then $d < 2\sqrt{n}$. So this gives an upper bound on the degree of a regular graph that doesn't have C_4 as a subgraph. That's almost the same thing as proving an upper bound on the extremal number. Why? For any G, the *average* degree of G is exactly $2e(G)/n$ (this is called the handshaking lemma or something). So the average degree of a graph is the same information as the number of edges. In particular, for a regular graph, knowing its degree is the same as knowing the number of edges. So this says if I have a *regular* C_4-free graph, then its number of edges is at most $n^{3/2}$. So this is the same as the upper bound $\mathrm{ex}(n, C_4) = \Theta(n^{3/2})$ when restricted to regular graphs.

We'll prove this first because it's a bit easier to think about, but then afterwards we'll see this regular assumption is not really necessary and get the general case.

Proof. The proof will be by double-counting — we'll count the number of paths of length 2 (or $K_{1,2}$s, or cherries — because it's supposed to look kind of like two cherries glued at their stem) in two ways.



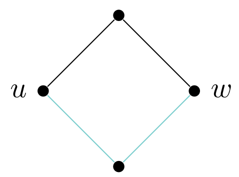
First, we can figure out *exactly* how many of these there are — we have

$$\#\text{cherries} = \sum_{v\in V} \binom{\deg(v)}{2}.$$

Why? If we imagine how this guy can live in the graph, let's begin by specifying the top vertex (that gives the sum over v). Then the other two guys must be neighbors of my chosen v. And how many ways are there of picking that? I need two neighbors, and I have $\deg(v)$ in total, so that's $\binom{\deg(v)}{2}$.

And I assumed the graph is d-regular, so $\deg(v) = d$ (this is where the assumption is convenient), so this is exactly $n\binom{d}{2}$. That's one way of counting this object.

The other way of counting this object is the inverse. In the first count, we started by looking at the top guy and seeing where it can go. For the next count, we'll instead start by thinking about where the bottom two guys can go. The observation for this is that for any $u, w \in V(G)$, they can be the outer vertices of at most one cherry. That's because if we had another one, then we'd get a 4-cycle.



So by summing over all u and w, we get

$$\#\text{cherries} \leq \sum_{u,w \in V(G)} 1 = \binom{n}{2}.$$

If we compare what we've done, we had some mysterious quantity and counted it in two ways. On one hand it's exactly $n\binom{d}{2}$, and on the other hand it's at most $\binom{n}{2}$. That tells us

$$n\binom{d}{2} \leq \binom{n}{2}.$$

And now we can work out what this means in symbols — this says

$$\frac{nd(d-1)}{2} \leq \frac{n(n-1)}{2}.$$

Then some stuff cancels, and we learn that $d(d-1) \leq n-1$. But we assumed $d \geq 2\sqrt{n}$; the 2 is massive overkill, but it's sufficient — if $d \geq 2\sqrt{n}$, then there's no way $d(d-1)$ is less than $n-1$ (it's actually close to $4n$, which is way bigger than $n-1$). $\square$

So to summarize, the point of this proof was counting these cherry objects in the graph. We know a lot about these objects, and can look at them in two ways. On one hand we know their *exact* count, by looking at the central vertex. (Even if the graph isn't regular, the $\sum_{v \in V} \binom{\deg(v)}{2}$ expression is correct.) On the other hand, the C_4-free assumption says that if I fix the outer two vertices, I can't have too many of them. Then we compared these bounds and did a bit of manipulation, and got our result.

Now we'll remove the regular assumption and prove a more general statement, where instead of C_4 we do all bipartite graphs. Before that, we need a useful analytic inequality.

Definition 2.9. For $x \in \mathbb{R}$, define

$$\binom{x}{r} = \frac{x(x-1)(x-2)\cdots(x-r+1)}{r!}.$$

If x is an integer then this is exactly what $\binom{x}{r}$ is; if x is not an integer then $\binom{x}{r}$ has no combinatorial meaning, but we just define it as this.

The lemma is a special case of Jensen's inequality. (We won't state it in full generality; you can see it on the homework.)

Remark 2.10. One of the people in this program is Matt Jenssen. They're different people — his name has two s's, and the Jensen of this theorem died a long time ago. But Matt Jenssen is great to talk to, and you should talk to him. There's also a thing in probabilistic combinatorics called Janson's inequality. Janson is still alive; he's a professor in Sweden, and also great to talk to, but is not at this conference.

Lemma 2.11 (Jensen's inequality)

Let $r \geq 1$ be an integer, and let $x_1, \ldots, x_n \geq 0$ be integers. Assume that $\frac{1}{n}\sum x_i \geq r$. Then

$$\sum_{i=1}^{n} \binom{x_i}{r} \geq n\binom{\frac{1}{n}\sum x_i}{r}.$$

This is an instance of convexity (which is what Jensen is about). It's saying if I want to minimize $\sum \binom{x_i}{r}$ (given their sum), then I should take all the x_i's to be equal. Concretely, the value of this sum is at least what would happen if I replaced all the x_i's by their average.

The proof is on the homework.

Theorem 2.12 (Kövári–Sós–Turán 1954)

For all $s \leq t$, we have $\mathrm{ex}(n, K_{s,t}) \leq O(n^{2-1/s})$, where the implicit constant may depend on s and t.

So KST gives an upper bound on the extremal number for any complete bipartite graph — it tells us for a complete bipartite graph with parts of size s and t, the extremal number is subquadratic. In fact, it's less than n^2 by some polynomial factor $n^{-1/s}$.

Student Question. *Is this an equality?*

Answer. This is only an upper bound. Often people write $\mathrm{ex}(n, K_{s,t}) = O(n^{2-1/s})$ to mean that we just have an upper bound, and Yuval also sometimes does this; but in this course he'll try to use $\leq$ for clarity. This statement explicitly means

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

(These big O's and stuff are really confusing the first few times you see them, but then it clicks and becomes easy.)

Student Question. *Do we need n large enough?*

Answer. You can say that, or you could make this constant $C_{s,t}$ so ridiculously large that it's trivially true when n isn't large enough.

The proof is really similar to what we saw for C_4, so we'll do it somewhat quickly.

Proof. Let G be an n-vertex graph with at least $Cn^{2-1/s}$ edges, where C is some constant we won't specify (we'll see at the end that if C is large enough then we win). Assume for contradiction that G is $K_{s,t}$-free. Our goal is to show that if a graph has $Cn^{2-1/s}$ edges then it must have a $K_{s,t}$ (that's sufficient to prove an upper bound on the extremal number), so we'll assume for contradiction that it doesn't have a $K_{s,t}$, and then get a contradiction.

Student Question. *This is a small thing, but are we not concerned with the fact that if C is too big, things start breaking because there's more edges than there could be?*

Answer. It's still true, but for trivial reasons — if I tell you a graph on 4 vertices has less than a million edges, then that's true, it's just true for silly reasons.

Let d be the average degree of G, which is

$$d = \frac{2e(G)}{n} \geq 2Cn^{1-1/s}.$$

We're again going to be interested in the number of $K_{1,s}$ in G. So we're counting not cherries, but 'generalized cherries' (or 'stars') where we have s leaves.

We'll again do this in two ways. On one hand, we have

$$\#K_{1,s} = \sum_{v \in V(G)} \binom{\deg(v)}{s},$$

for the same reason as before (we sum over all options for the central thing, and then we need to choose s things out of its neighbors). Before, we knew an exact equality for this, because we assumed the graph was regular. Now we don't, but we can use Jensen's inequality, which tells us this is at least

$$\#K_{1,s} \geq n\binom{d}{s}.$$

(We apply Jensen with the x_i being the degrees of our vertices; Jensen tells us the sum of these degrees choose s is at least n times their average (which is d) choose s.) Note that in our earlier proof, we didn't need an equality — all we needed was a lower bound, because we'd compare it with an upper bound — and luckily Jensen still gives us that, so it works out.

On the other hand (as before), we have

$$\#K_{1,s} \leq \sum_{u_1, \ldots, u_s \in V(G)} (t-1).$$

This is for the same reason as before — if I tell you these outer vertices are some specific $u_1, \ldots, u_s$ and ask how many stars there are that use those specific vertices as their leaves, then there can't be t of them. Why? If I had t different stars all using these outer vertices, I'd get a copy of $K_{s,t}$ in my graph, and I assumed I don't have that. So any s vertices can be the outer vertices of at most $t-1$ stars. (This is the same argument as before, just appropriately generalized.) And this is $(t-1)\binom{n}{s}$, because I have $\binom{n}{s}$ terms in the sum.

Comparing these bounds, we learn that

$$n\binom{d}{s} \leq (t-1)\binom{n}{s}$$

(because they're a lower and upper bound for the quantity we're interested in, respectively). Now we can expand out what it means. It's kind of annoying, but it says

$$\frac{nd(d-1)(d-2)\cdots(d-s+1)}{d!} \leq \frac{(t-1)n(n-1)\cdots(n-s+1)}{s!}.$$

Again some terms cancel — the $s!$'s cancel, and an n also cancels. And then we have some stupid annoying terms with a bunch of $-$'s. But remember the regime we're interested in is where s is a constant (like 4) and

n is huge (like a billion); then we have something like $(10^9 - 1)(10^9 - 2)(10^9 - 3)(10^9 - 4)$, which is basically $(10^9)^4$. So we can basically just ignore all these $-$'s (you can make this rigorous), and we roughly get

$$d^s \leq (t-1)n^{s-1}$$

(we only have $s - 1$ n's on the right-hand side, because one of them cancelled). And we said $d = 2Cn^{1-1/s}$, s the left-hand side becomes $2^s C^s n^{s-1}$. So we've shown

$$2^s C^s n^{s-1} \leq (t-1)n^{s-1}.$$

We have n^{s-1} on both sides, and I get to pick C; so I can pick C large enough that $C^s > t-1$, and this is a contradiction. (We can make C large enough that the left-hand side is much larger, like t^{10^9}, and that lets us sweep the small errors from ignoring the $-$'s under the rug.) □

The reason we presented the special case for C_4 and the d-regular case beforehand is that it captures all the main ideas. The only new things needed for the general case are the magic of Jensen's (which says the d-regular case is really the worst — we win even more if it's not d-regular) and being a bit more fiddly with what s and t are doing (if we have too many stars on the same outer s things we get a $K_{s,t}$, just as if we had two cherries on the same outer things we'd get a C_4).

Next class we'll talk a bit about lower bounds, and then move to hypergraphs (which are the same thing but more hyper).

§3 July 10, 2025

The cross-program activity today, from 4:30–6:30, is by Dave Auckly; it'll be a building party, where you're going to build graphs and things on stage with strings or something like that (Paul doesn't know quite what he's going to do, but it's a lot of fun). It's in the Bear Conference Room (the dining area). As a consequence, the problem session has moved from 4:30 to 3:15 (for today only).

Student Question. *We talked about bounding complete bipartite graphs from below. Do we have a lower bound?*

Answer. Great question; that's what we will be talking about today.

Yuval will also mention there was a mistake on Tuesday's homework, where he wrote the wrong definition of how Zykov symmetrization works. In general, feel free to reach out if we think he did something wrong. Also, Paul pointed out on the first homework, the thing about magic probably needs to be slightly changed.

§3.1 Review

Let's get started. Last time, we proved the KST theorem:

Theorem 3.1 (Kővári–Sós–Turán)
For all $s \leq t$, we have $\operatorname{ex}(n, K_{s,t}) \leq O(n^{2-1/s})$.

Our proof never used $s \leq t$; the other thing is also true, but it's a weaker result (if $s > t$ it's better to put a t instead of an s). The proof is by double-counting stars using Jensen's inequality.

One immediate consequence is, this was only about complete bipartite graphs, but it immediately implies something about *all* bipartite graphs. If $H_1 \subseteq H_2$, then there's some relation between their extremal numbers — we have

$$\operatorname{ex}(n, H_1) \leq \operatorname{ex}(n, H_2).$$

The point is if we have a graph with no H_1, it certainly also has no H_2. So the maximum is over a larger set for $\mathrm{ex}(n, H_2)$ than $\mathrm{ex}(n, H_1)$.

The reason this is useful is that *every* bipartite graph is a subgraph of *some* complete bipartite graph — just take your bipartition and add in all the missing edges.

Corollary 3.2

For every bipartite H, there exists some $s \in \mathbb{N}$ such that $\mathrm{ex}(n, H) \leq O(n^{2-1/s})$.

(Here s is just the number of vertices in one side of the graph.)

This is because H is a subgraph of some bipartite graph $K_{s,t}$, so

$$\mathrm{ex}(n, H) \leq \mathrm{ex}(n, K_{s,t}) \leq O(n^{2-1/s}).$$

The reason to bring this up is that it immediately tells you $\mathrm{ex}(n, H) = o(n^2)$. To remind us what this means, it's the same as saying

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n, H)}{n^2} = 0.$$

That's because we have an upper bound that's of some smaller order than n^2, so when we divide by n^2, this ratio is upper-bounded by $n^{-1/s}$, which tends to 0 as $n \to \infty$.

The point is we stated this as a consequence of ESS, and we said it's a highly surprising and non-obvious consequence, and we've now proved it.

And as Yuval mentioned (and as we'll see on Friday or more probably Monday), the ESS theorem actually follows from an appropriately generalized version of this statement! So it's a special case, but it's in some sense the full case.

§3.2 Lower bounds for 4-cycles

Now we've talked about upper bounds; so we'll now talk about lower bounds. The first example we saw was $\mathrm{ex}(n, C_4) \leq O(n^{3/2})$, and on Tuesday we stated (but didn't prove) that this is the right order — there's a matching lower bound, up to constants.

Theorem 3.3 (Klein 1938)

We have $\mathrm{ex}(n, C_4) \geq \Omega(n^{3/2})$.

Concretely, we'll prove something like $\mathrm{ex}(n, C_4) \geq n^{3/2}/64$.

Student Question. *How come this result comes from 1938, but...*

Answer. If you'll remember, Turán's theorem is from 1941, and KST is from 1954; 38 is less than both those numbers. So what's going on? She didn't write a paper but communicated this fact to Erdős, who wrote it in a paper of his. He was interested in something else, which naturally led him to think about what we'd now call $\mathrm{ex}(n, C_4)$. Erdős later wrote he had the opportunity to invent extremal graph theory at this moment but didn't see what a beautiful field it was, so he just solved this problem and left it at that; it was only after Turán's theorem that EGT became a field of math, and Erdős felt a bit sad about this because he was so close.

Proof. Basically what we need to do is construct a graph on n vertices with no C_4 and this many edges. First of all, we'll assume for now that $n = 2p^2$ where p is a prime (later we'll figure out how we get rid of this assumption, but it'll make our lives easier for the moment).

There's primes coming up, so we'll need to use something about primes. There's no algebra prerequisite for this class and the construction is algebraic, so we won't go through all the details, but we'll give the ideas. We'll work with the field $\mathbb{F}_p$, consisting of the integers mod p (the integers $0, \ldots, p-1$ where addition and multiplication happen mod p). If you've never seen the word 'field' before, don't worry; it just means the rules of addition, multiplication, and division work the same way as in $\mathbb{R}$ (the important thing is you can divide — this is why primality is important).

We will not actually be working with $\mathbb{F}_p$, but rather in $\mathbb{F}_p^2$. What's that? Well, if you want to think in abstract algebra language, it's the vector space of dimension 2 over this field. But really it's pairs (x, y) where x and y are integers mod p. But the point is this looks like the real plane, and behaves like the Euclidean plane in a lot of ways (except that it's different because it's finite and has this weird mod thing).

So we'll let $\mathcal{P}$ be the set of points of $\mathbb{F}_p^2$, which concretely is

$$\mathcal{P} = \{(x, y) \mid x, y \in \mathbb{F}_p\}.$$

And we'll let $\mathcal{L}$ be the set of lines $\ell_{m,b}$ in $\mathbb{F}_p^2$, where $\ell_{m,b}$ is the line $y = mx + b$. (You should be thinking about this geometrically in the sense of the lines we're used to, but formally it's just the set of points (x, y) that satisfy this equation.)

(We're not going to include the vertical lines; it doesn't really matter whether we do or not.)

Our goal is to define a graph, and so far we haven't gotten close to that. Our graph will be a bipartite graph with two parts $\mathcal{P}$ and $\mathcal{L}$. And what are the edges? We'll connect a point (x, y) to a line $\ell_{m,b}$ if and only if this point lies on that line — in other words, we say

$$(x, y) \sim \ell_{m,b} \quad \text{if and only if} \quad y = mx + b$$

(that's what it means for that point to be on that line).

So we've defined a graph. First, let's do some numerology. We have $|\mathcal{P}| = p^2$ and $|\mathcal{L}| = p^2$ (we have p choices for m and p choices for b). So the total number of vertices is $2p^2$, which is what we promised; that's good.

The harder question is, how many edges does this thing have? That's a bit weird to think about, but the easiest way of thinking about it is, what's the degree of some line $\ell_{m,b}$ — if we fix some vertex in $\mathcal{L}$, how many vertices is it adjacent to? The answer is p. Why? That's the same as how many points are on that line. And there's exactly p points on the line, because if you fix the value of x, that tells you a unique value for y.

And once we know the degree, it's easy to see that the number of edges is

$$\#\text{edges} = p^3,$$

because you can count the number of edges by taking the p^2 vertices in $\mathcal{L}$ and saying each gives you p edges. And that's great because $p^3 = (n/2)^{3/2}$ (because we said $n/2 = p^2$). The important thing is that's osme constant times $n^{3/2}$; that's good, because that's how many we're shooting for.

The final thing is, why doesn't this have a C_4? In order to have a C_4, you'd need two points on the left and two lines on the right; but any two lines can only have one intersection. In more detail, we need to think about all the ways a C_4 could live in the picture, but because the graph is bipartite, there's only one — if we have two points and two lines such that both points lie on both lines, i.e., those two lines intersect in two points. And from Euclidean geometry we know that can't happen. If you don't like that (which you shouldn't, because this isn't anything like Euclidean space), it's not hard to convert the formal proof you know in Euclidean setting to a formal proof that also works here.

So to summarize, a C_4 would correspond to two lines intersecting at two points, and that's impossible (a formal proof of why is in the notes, but hopefully this is at least believable).

We're not done — why? We assumed $n = 2p^2$, but what if n is something else? There's a ton of ways of dealing with that. The easiest is to use a useful fact from number theory:

Theorem 3.4 (Bertrand's postulate)
For any integer $x \in \mathbb{N}$, there exists a prime p such that $x \leq p \leq 2x$.

(Yuval thinks this is a stupid name because it's not a postulate or axiom, it's a theorem — a fact that's true that we have a proof of.)

The exact statement doesn't super matter; the point is you can always pick a prime of whatever size you want, up to a constant factor. So given *any* n, let $n' \leq n$ be of the form $2p^2$; the point is that by Bertrand's postulate, we can assume $n' \geq n/4$ or something. Why? What it means that $n = 2p^2$ is that $p = \sqrt{n/2}$; so if we apply Bertrand with $x = \sqrt{n/8}$, we can always find a prime in the appropriate range, so that $2p^2$ is less than n but not a ton less. And then we can add dummy vertices — we can take this construction for n' and add a bunch of isolated vertices that don't do anything. Certainly doing that won't create a C_4; and we'll get the number of vertices up to n. The number of edges is now $c(n')^{3/2}$, but n and n' are the same up to a constant, so if we don't care about the exact constant factor, this will also work.

(We're not doing this argument in a ton of detail; there's more in the notes, but it's not super interesting — the point is you can always massage things to be a prime if you want.) ☐

When you first see this, your instinct is, that's really cool but where did primes come from? This is a natural combinatorial question that doesn't see primes, so why are primes showing up? Maybe your first instinct should be, cool construction, but there must be others that don't. . . But that's not the case! Every known construction is kind of like this, in a vague sense. A much deeper fact, which we'll state informally:

Theorem 3.5 (Füredi)
If n is of a form such that 'such a construction' exists, then this construction is the *unique* extremal graph for $\mathrm{ex}(n, C_4)$.

There's a specific kind of construction, called the *polarity graph of a finite projective plane.* If n is of the right size that a finite projective plane of the right size exists (specifically, it turns out that $n = p^2 + p + 1$ for a prime p), then this is the *unique* extremal graph.

So what this says is that if n is of a form that can accomodate this type of construction, that construction is the best, and it's the only best one. So in some sense, the answer is no — there are not other constructions that don't see the primes. There are plausibly ones that do *okay*, but in some sense it'd be an accident if those existed, because the *true* best one (when it exists) does see the primes. So it is not a coincidence that primes show up!

Student Question. *Where did $p^2 + p + 1$ come up?*

Answer. Yuval is not telling us. 'Such a construction' isn't exactly what we did. What was wasteful in our construction is that our graph was bipartite, so there are tons of missing edges not inside our parts. There's a way to do something similar but a bit different that's not wasteful in this way; that's called a polarity graph of a finite projective plane, and that's the best construction.

Student Question. *Where does $n^{3/2}/64$ come from?*

Answer. The constants are in the notes, but that's just Yuval being lazy; you can get a much better constant than 64.

Student Question. *What's the constant from the good construction?*

Answer. The true leading constant is 1/2. On the homework we saw a construction with leading coefficient 1/2. This construction gives a lower bound that's also with leading coefficient 1/2 (and when n is of this specific form, we know all the lower-order terms too).

One other remark is that this immediately implies that for every $t \geq 2$, we have

$$\mathrm{ex}(n, K_{2,t}) = \Theta(n^{3/2}).$$

Why? The upper bound is KST. And the lower bound is just the observation that if $H_1 \subseteq H_2$ then $\mathrm{ex}(n, H_1) \leq \mathrm{ex}(n, H_2)$. If you reverse this observation, it says

$$\mathrm{ex}(n, K_{2,t}) \geq \mathrm{ex}(n, C_4) \geq \Omega(n^{3/2}).$$

So for all these complete bipartite graphs $K_{2,t}$, we now know the full story.

§3.3 Lower bound for $K_{3,3}$

What's next? We've done $K_{2,t}$ for all t, so the next one is $K_{3,3}$.

Theorem 3.6 (Brown)
We have $\mathrm{ex}(n, K_{3,3}) \geq \Omega(n^{5/3})$.

This also matches the KST upper bound (since $2 - 1/3 = 5/3$).

We did Klein's proof pretty sketchily; we'll do this one a lot more sketchily, because making it work involves some boring algebraic manipulations.

Proof. Now we'll work in $\mathbb{F}_p^3$. Our vertex set of G will be all the points of $\mathbb{F}_p^3$ (this time the graph isn't bipartite), and we join $(x, y, z) \sim (x', y', z')$ if and only if

$$(x - x')^2 + (y - y')^2 + (z - z')^2 = 1.$$

This is again a graph we're allowed to define.

It's easier to think about this in terms of the neighborhood of a vertex — that's the 'unit sphere' around that vertex (all the points at distance 1). This doesn't make sense because we're not in Euclidean space, so there's no distance and no spheres; but that's the formal equation for this. (This whole equation is in $\mathbb{F}_p$; otherwise there'd be a type error.)

Let's think a tiny bit about this. Here $n = p^3$. What's the degree of a vertex? Actually proving this is a bit annoying, but if you think about it, I have a 3-dimensional space, and for every vertex its neighborhood is a sphere. So the degree of every vertex is roughly p^2. (You can prove this if you want; it's not very hard. But the intuition is that the neighborhood of every vertex is like a 2-dimensional object, and a 2-dimensional object over $\mathbb{F}_p$ should have roughly p^2 points, just as a 3-dimensional one has p^3 points.) So the number of edges is going to be roughly $p^5/2$ (since it's the degree times n, and we've double-counted every edge); the point is that's $\Omega(n^{5/3})$, because $n = p^3$ and the number of edges is on the order of p^5.

The final thing is, why doesn't this have a $K_{3,3}$? The real proof is to write some equations and work things out. But the right intuition is to think Euclideanly. What does it mean to have a $K_{3,3}$? That means we have three vertices and three common neighborhoods. If we take two vertices, their common neighborhood is the intersection of two unit spheres, which is a circle. Now when we bring in a third, the intersection of a circle and another unit sphere is at most two points. So any three vertices will have at most 2 common neighbors; this implies there's no $K_{3,3}$.

That proof is BS; the real proof is that you work with the equations. And a thing we're hiding is this is only true if $p \equiv 3 \pmod 4$ (or maybe 1 mod 4). The reason is that this intuition isn't the truth; the truth is you have to work it out. □

Student Question. *Do you need a completely different construction for $p \equiv 1 \pmod 4$?*

Answer. You could replace the 1 with -1. But also, you don't really care because there will be a prime $p \equiv 3 \pmod 4$ close to n as well. Generally, as long as your construction works for *some* primes, you're happy (it doesn't really matter which ones).

Student Question. *Could you also think about this as a bipartite thing where you require some degree-2 polynomial?*

Answer. Yes, you could think of it as a bipartite thing where both sides are copies of $\mathbb{F}_q^3$, and the rule for drawing an edge is still this equation (where (x, y, z) and (x', y', z') come from the two sides). But it matters what degree-2 polynomial you pick — even this one doesn't work for all primes, just ones of a specific value mod 4.

Student Question. *Are there other polynomials that work?*

Answer. Yes. In some sense, these things are sort of unique in that the graphs you end up getting from different polynomials end up being more or less isomorphic; so it turns out to not really matter. But that's not obvious.

It's kind of obvious what to do next, right? We've done $K_{3,3}$, which implies that for all $t \geq 3$ we have

$$\mathrm{ex}(n, K_{3,t}) = \Theta(n^{5/3})$$

(once we have a construction for $K_{3,3}$ we get a construction for all $K_{3,t}$ by the same reason).

So now we turn to $K_{4,4}$. By the intuition we just discussed, the obvious thing is to work over $\mathbb{F}_p^4$, and pick some degree 3 polynomial intelligently. Then all the numbers work out and all you need to do is show the graph has no $K_{4,4}$. But this is still wide open!

Conjecture 3.7 — We have $\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s})$.

People are certain if this is true, it should be true for that reason. But even $\mathrm{ex}(n, K_{4,4}) = \Theta(n^{7/4})$ is a major open question.

Student Question. *What's the best lower bound?*

Answer. The one coming from $K_{3,3}$, that if you don't have $K_{3,3}$ you also don't have $K_{4,4}$.

In fact, these days lots of people are starting to think this might not be true. The reason is because of what we've discussed, we think the only construction should be of the form described, where you do something algebraic over 4D space using degree-3 polynomials. But no one's been able to come up with such a polynomial, and now there are deep theorems saying a big class of polynomials you might try don't actually work. So now there's a consensus forming that this might not be true. But that'd also be very difficult to prove.

What is known:

Theorem 3.8

For any s and t sufficiently large with respect to s, we have $\mathrm{ex}(n, K_{s,t}) = \Theta(n^{1-1/s})$.

So we don't know it for $K_{4,4}$, but we do know it for $K_{4,t}$ where t is super large. 'Super large' for 4 is actually not that large — we know it for $K_{4,7}$. There's a weird coincidence where there's a 7 and 4 on both sides of the equation; the 4 is not a coincidence, but 7 is. (This is the smallest value for 4 that's known.)

What is the general result for super large? THe first result in this direction:

Theorem 3.9 (Kollar–Rónyai–Szábo 1996)
$t > s!$ works.

For their result, they showed that $K_{4,25}$ worked; that's what we mean by super large.

A few years later, they modified the construction to get from s to $s-1$:

Theorem 3.10 (Alon–Rónyai–Szábo)
$t > (s-1)!$ works.

(Noga Alon is now here; he's done a ton of amazing stuff in the field.)

For big s you don't really care; but for small s this is significant (e.g., for $s = 4$ it gets you down from 25 to 7).

For a while this was best-known. A couple of years ago there was a major breakthrough:

Theorem 3.11 (Bukh)
$t \geq 100^s$ works.

The 100 is not important; the point is that this grows exponentially, in contrast to $s!$ growing super-exponentially.

This is where we're stuck; in particular, for $K_{4,4}$ we have no clue. The approach Yuval sketched *is* the right approach — you want to work over $\mathbb{F}_p^s$ and take degree-$(s-1)$ polynomials and use them to define your graph. But finding polynomials you can say anything useful about is very difficult. KRS and ARS used 'norm polynomials' from finite fields and heavy algebraic geometry machinery. That's as far as anyone was able to get by being really good at algebra. Bukh's amazing insight was that instead you can pick a *random* polynomial. Then analyzing that is still really hard, so you have to still do lots of AG. But by not trying to force a good choice and instead using a random one, you get a better result. But any approach like this seems completely doomed for $K_{4,4}$, or certainly $K_{10,10}$.

Student Question. *So with picking a random polynomial, the hope is to show there's some polynomial that does work?*

Answer. Yes. We'll see more probabilistic arguments of this type later.

One more lower bound thing: So far we've only been dealing with complete bipartite graphs, but the question works for any bipartite graph. So far we've seen one upper bound $\text{ex}(n, H) \leq O(n^{2-1/s})$ where s is the number of vertices on the smaller side of H — this is just saying H is a subgraph of $K_{s,t}$. This is basically the best general-purpose upper bound.

Every really good lower bound uses algebra in some way. But if you take your favorite bipartite graph, it's kind of hopeless to think about what algebra is good at controlling it. So there's really only one general-purpose lower bound.

Theorem 3.12
We always have $\text{ex}(n, H) \geq \Omega(n^{2-1/m_2(H)})$.

Definition 3.13. The 2-density of H is defined as

$$m_2(H) = \max_{F \subseteq H} \frac{e(F) - 1}{v(F) - 2}.$$

So it's basically the number of edges divided by the number of vertices (with some stupid -1's and -2's shown in); but you want to not just do this for your entire graph, but all subgraphs (and take the best one).

Student Question. *Is this true for all H or just bipartite ones?*

Answer. The argument works for all H, but for any non-bipartite H this is a bad bound, because we know $\text{ex}(n, H)$ is quadratic — for example, $K_{n/2,n/2}$ has no copy of H and has at least $n^2/4$ edges. So the truth is at least $n^2/4$, which is much bigger than this. So this is still true (and the same argument works), but it's uninteresting; it's only interesting for bipartite graphs.

Student Question. *What's the 1-density or k-density?*

Answer. Yuval doesn't want to tell us. The 1-density is when you remove the -1 on the top and replace the -2 with -1 on the bottom; k-density usually isn't defined for graphs.

Yuval won't show us the proof (it'll be on the homework). The proof is probabilistic. (We will get to probabilistic methods eventually...) Rather than being super smart and coming up with a good graph, you'll simply pick a graph at random (in an appropriate way), and show you can make sure the graph doesn't have any copies of H and has plenty of edges, and that gives you a lower bound. This crazy quantity just comes out of optimizing that argument — this argument can give you many different bounds, and this quantity comes out of using the best possible numbers.

§3.4 Hypergraphs

Finally, one topic we'll start today and discuss more next class:

What's a graph? We didn't define it here, but if you took a graph theory class, you were probably told that:

Definition 3.14. A graph is a pair $G = (V, E)$ where V is a finite set and E is a collection of unordered pairs of elements of V.

So a graph is a bunch of vertices and edges, where an edge is a pair of vertices.

What's a hypergraph? It's the same thing, except we're not going to do pairs.

Definition 3.15. A k-uniform hypergraph (abbreviated k-graph) is a pair (V, E) where E consists of unordered k-tuples (of elements of V).

So a 3-graph is a bunch of vertices, and a bunch of *triples* of those vertices, which we call edges (or sometimes we call them *hyperedges*; Yuval will try to do that in this class, but will probably screw this up and call them edges at some point).

This definition should always specialize the graph definition when you plug in $k = 2$.

Student Question. *Are directed hypergraphs a thing people study?*

Answer. They are, but not in this class.

Student Question. *Are we only including size-k edges, or can you have smaller ones?*

Answer. We require every edge to have size *exactly* k. (In simplicial topology you'll encounter a similar definition where you also allow things of smaller size; we won't do that.)

Remark 3.16. Everyone in this field gets lazy with the prefix 'hyper' (and call things edges and subgraphs instead of hyperedges and subhypergraphs); Yuval will try to use 'hyper' but may forget.

(Hypergraphs only have hyperedges; they don't have any notion of edge that is not this one.)

We'll also try to use script letters for hypergraphs.

Definition 3.17. For a k-graph $\mathcal{H}$, its extremal number $\mathrm{ex}(n, \mathcal{H})$ is the maximum number of hyperedges in $\mathcal{G}$ where $\mathcal{G}$ is a n-vertex $\mathcal{H}$-free k-graph.

So this is the same as before — how many edges can I put in a k-graph on n vertices without obtaining a copy of $\mathcal{H}$?

Why are we talking about this now? One is it's a very interesting topic in its own right — it's a natural generalization of what we've studied so far. It turns out to have a ton of strange features you don't see at all in the graph case (which we'll talk about). The other reason is there's tons of ways of proving Erdős–Stone, but the prettiest goes through hypergraphs. Tomorrow we'll prove a version of KST for hypergraphs, and then we'll see how to go from graphs to hypergraphs and back. That's exactly what we were saying when we said KST sort of captures everything needed to prove the general case of ESS — it's the generalization of KST to hypergraphs that we use to prove ESS.

Rather than stating KST for hypergraphs today, we'll just make one more definition:

Definition 3.18. The complete r-vertex k-graph, denoted $K_r^{(k)}$, is the k-graph on r vertices where every k-set is a hyperedge.

The complete graph is a graph where every possible edge is an edge; so the complete k-graph is the k-graph where every possible hyperedge (i.e., every k-set) is a hyperedge.

The annoying thing about hypergraphs is that they're way harder to draw than graphs, but the complete 3-graph on 4 vertices (the smallest nontrivial example) looks like a tetrahedron, where every possible face is included as a hyperedge (there's 4 possible faces, and all 4 are included as a hyperedge).

Obviously the first thing we did about extremal numbers of graphs was Mantel's theorem, the extremal number of a triangle. This is the obvious generalization (a triangle, just one uniformity higher). Turán in his 1941 paper asked for the extremal number of this hypergraph. But no one knows what it is!

Conjecture 3.19 (Turán) — We have $\mathrm{ex}(n, K_4^{(3)}) = (\frac{5}{9} + o(1))\binom{n}{3}$.

This is the biggest open problem in extremal hypergraph theory. Why 5/9? There's a specific construction that gives 5/9.

Student Question. *Are there other graphs for which we do know the asymptotics?*

Answer. There are many, and you can ask later. But not only do we not know this, but also for all $r > k \geq 3$, we don't know $\mathrm{ex}(n, K_r^{(k)})$. So it's not just that this is a weird one we don't understand; we don't understand the extremal number of *any* complete hypergraph in uniformity above 2. Erdős offered 1000 dollars to resolve this, and 500 to get even one instance of (r, k) for which you can do it.

Student Question. *Is there an upper bound?*

Answer. Yes. For $K_4^{(3)}$, the best upper bound is due to Razborov, and gives a bound of $(0.56166 + o(1))\binom{n}{3}$. Razborov (for a different problem) introduced a technique called flag algebras, which is beyond the scope of this course. It lets you do many things, but one is that it lets you get a computer to solve a finite approximation to things like this and prove good upper bounds, by outputting a huge matrix which is PSD (which serves as a formal proof of a bound like this). Flag algebras have had amazing success in a variety of problems, but for this problem it's believed they won't get the truth.

§4 July 11, 2025

A few announcements: The first weekend is coming up. We're free to do whatever we want. There's a lot of things to do. First, there's a free bus around Park City. At least today it goes every 15 minutes. You can walk down to 7-11, or if you go outside this door and a little up, there's a free shuttle that takes you down to the 7-11 bus station (it's a city shuttle, and everything is free). When you come back to the Canyon Village transit hub, the shuttle is waiting in front, so you can go from the bus to the shuttle. (The shuttle doesn't go straight down; it goes to other hotels.)

Park City is a wonderful town; it's easily walkable, there's lots to see. It's an old mining down. There are places to eat, there are places to see and shop on Saturdays and Sundays. There's also escape rooms. And they have karaoke nights and puzzle nights and so on. Also, Irena is organizing a hike tomorrow morning; we'll start around 9 and it'll be moderate and will go up the hill. Irena decided against a sunrise time, so we're starting only at 9. If you're interested in climbing, Irena posted her husband's email address (he's not on Whova).

There's also lots of things to do in Westgate — swimming, tennis courts, game rooms. You can also rent things from Dena from our office for the weekend (games, balls of various shapes, rackets). You're encouraged to hang out. Rooms and hallways are narrow, but there are beautiful open areas.

There is a cross-program 3:15–4:15 in the big Bison room. The speaker is Janos Pach, who is a great speaker.

Paul will add a few words. If anyone is interested in classical music, there is the Park City music festival, which has been running 40 years. They have free concerts on Mondays 6:30–8. It's in City Park, which is — you go near Main Street and left towards the mountains a bit. You can bring your food and sit out. Next Monday it's the Amadeus Trio (piano, violin, cello); you sit out in the park and listen to the music.

However, we have another PCMI event on Monday; it's a movie night. Irena doesn't know what the movie will be; usually it's a nice math movie. Last year they went to Oppenheimer.

Yuval reminds us that we can submit feedback. Someone did; the suggestion was (a) hints on the homework in footnotes, and (b) more homework on Tuesdays and Fridays, because there's no class on Wednesday and weekends. It's too late to implement for this Tuesday, but not today; so today's homework will be longer. It's mostly longer on the optional problems, so if you want more meaty stuff to chew on you can. But you shouldn't get scared that now the homework is longer and you should do all of it. Yuval thinks it's definitely a good idea to take the weekend to not think about math if you want.

§4.1 KST for hypergraphs

Yesterday we defined k-uniform hypergraphs and their extremal numbers, and mentioned the amazing fact we know very little about them for $k \geq 3$. On today's homework you'll see some of the best known bounds.

But one thing we *do* know is an analog for hypergraphs of the KST theorem, which is what we'll talk about now.

Definition 4.1. A k-graph is called k-partite if its vertex set can be split into k parts and each hyperedge contains exactly one vertex from each part.

('Edge' always means *hyperedge*, but Yuval will try to say 'hyperedge.')

This is the obvious generalization of a bipartite graph — that's a graph where you can split the vertex set into two parts such that every edge contains one vertex from each part. The obvious generalization to uniformity k is you split into k parts, and each edge contains exactly one vertex from each.

KST talked about complete bipartite graphs; so as an analog, there's an obvious generalization of complete k-partite k-graphs.

Definition 4.2. The complete k-partite k-graph, denoted $K^{(k)}_{s_1,\dots,s_k}$, has parts of sizes $s_1, \dots, s_k$, and all possible hyperedges across.

The deal is K stands for 'complete' (because graph theorists don't know how to spell), the superscript (k) records the uniformity, and the subscripts indicate the sizes of the parts. Here we have k parts, so I need to tell you k numbers $s_1, \dots, s_k$. So you have k parts of these specified choices, and every choice of one vertex from each of the parts gives you a hyperedge (that's what makes it complete).

For $k = 3$, you'll have three parts of sizes s_1, s_2, and s_3; and no matter how I pick a vertex from each of the three parts, it'll give a hyperedge.

And what does the hypergraph analog of KST say? It's going to be an upper bound on the extremal number of such a thing.

Theorem 4.3 (Erdős 1965)
For all $s_1 \le s_2 \le \cdots \le s_k$, we have

$$\mathrm{ex}(n, K^{(k)}_{s_1,\dots,s_k}) \le O(n^{k-1/s_1 s_2 \cdots s_{k-1}}).$$

In KST, we stated it as saying $s \le t$; that's without loss of generality. Here we have k numbers, and again we might as well assume they're nondecreasing.

Again, this extremal number is saying, how many hyperedges can I have in a k-graph which doesn't contain a copy of this thing? A k-graph on n vertices can have anywhere between 0 and $\binom{n}{k}$ hyperedges (and $\binom{n}{k} \approx \frac{n^k}{k!}$ for large n, which is on the order of n^k). KST for graphs told us the extremal number of any complete bipartite graph is *subquadratic* — it's $n^{2-\text{something}}$. Similarly, hypergraph KST will say this extremal number is smaller than the trivial upper bound (which is n^k) by something in the exponent.

In the graph case, this exactly recovers KST — we just have two numbers, so we get $n^{2-1/s}$. And in uniformity k, you multiply the $k-1$ smallest numbers.

The proof is by a smart induction on k (the uniformity). The base case will be $k = 2$, which is KST, which we've already done. In general, we'll prove it by induction on the uniformity. As you can tell by how annoying it is to write down the bound, the proof is a pain to write on the board, because there's lots of parameters. So Yuval won't show us the full proof, but will show the first instance of it, which captures all the difficulty — we'll just consider uniformity 3 and the case $s_1 = s_2 = s_3$, where we want to show

$$\mathrm{ex}(n, K^{(3)}_{s,s,s}) \le O(n^{3-1/s^2}).$$

We're only going to prove this; you can easily convince yourself that the same proof works regardless of what three numbers we put here, and the proof will use in it the case for graphs, which should give you a general idea of how the inductive argument works (if you want to prove it for uniformity 5, you'll reduce to uniformity 4).

Proof. Let $\mathcal{G}$ be an n-vertex 3-graph with $e(\mathcal{G}) \geq Cn^{3-1/s^2}$ (where C is some big constant), and assume for contradiction that it is $K^{(3)}_{s,s,s}$-free. (Just as with KST, we have a hypergraph with many hyperedges, and we're going to assume it doesn't have a copy of what we're looking for, and we'll get a contradiction at the end of the day; C is some constant we'll pick later. If we get a contradiction out of this, we'll have proved the claimed thing.)

The first thing is, in the proof of KST, we made use of this fact that the sum of degrees in a graph is twice the number of edges (equivalently, the average degree is $2e(G)/n$). We'll need a version of this for 3-graphs:

Definition 4.4. The codegree of distinct vertices v and w is

$$\operatorname{codeg}(v,w) = \#\text{hyperedges containing both } v \text{ and } w.$$

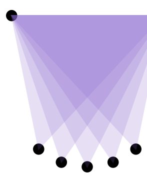
Fact 4.5 — We have $\sum_{v\neq w} \operatorname{codeg}(v,w) = 3e(G)$.

Why? Each edge uvw contributes to the sum exactly 3 times (it's counted once when we pick vw, once for uv, and once for uw).

Student Question. *Are we summing over distinct vertices?*

Answer. Yes — we don't want to define the codegree for a single vertex, so they have to be distinct for this to make sense.

We're going to try to follow the proof of KST as closely as we can, so this is good. The next thing we did in KST was count the number of stars $K_{1,s}$ in two different ways. That's what we'll do again, but we have to figure out the right notion of 'star.' So let X be the number of copies of $K^{(3)}_{1,1,s}$ in G. This is the complete 3-partite hypergraph where two parts have size 1; so it's a bunch of triangles all containing the same fixed pair.



It's not completely obvious that this is the *right* generalization of 'star,' but it's at least a plausible one, and it turns out to be a good one to make the proof work.

As before, we'll count this in two different ways.

First, we claim that

$$X = \sum_{v,w} \binom{\operatorname{codeg}(v,w)}{s}.$$

Why? Let's sum over all the options for the pair of vertices on the top; that's v and w. Then we need to pick s edges containing both v and w; the number of such edges is exactly $\operatorname{codeg}(v,w)$, and we need to pick s edges out of that, so the number of choices is exactly $\binom{\operatorname{codeg}(v,w)}{s}$.

Now we want a lower bound on this; so the inequality we should apply is Jensen. That tells us

$$X \geq \binom{n}{2}\binom{\frac{1}{\binom{n}{2}}\sum_{v,w}\operatorname{codeg}(v,w)}{s}.$$

That's because $\binom{x}{s}$ is a convex function; so Jensen says we can get a lower bound by replacing all terms in the sum with their average. The number of terms in the sum is $\binom{n}{2}$; and $\frac{1}{\binom{n}{2}}\sum_{v,w}\operatorname{codeg}(v,w)$ is the average.

Now, we know what this sum is, so let's plug that in; this is exactly equal to

$$\binom{n}{2}\binom{3e(\mathcal{G})/\binom{n}{2}}{s}.$$

Now the question is, how big is this? First, $\binom{n}{2}$ is some constant times n^2 (it's basically $n^2/2$). We assumed $e(\mathcal{G}) \geq Cn^{3-1/s^2}$, so $3e(\mathcal{G})$ is basically also of this form. We're dividing it by basically n^2, so the 3 in the exponent basically becomes a 1. And the final thing is that $\binom{x}{s}$ is basically just like x^s. So we're going to get

$$X \geq cn^2 \cdot \left(Cn^{1-1/s^2}\right)^{1/s}.$$

(We can make the constant c sufficiently small to absorb the constant-factor error in approximating $\binom{x}{s}$ by x^s.) To elaborate, we did a bunch of steps here; c is going to absorb every constant we throw away in what follows. We have $\binom{n}{2} \approx n^2$ and $\binom{x}{s} \approx x^s$. And $e(\mathcal{G}) \geq Cn^{3-1/s}$, and we're dividing by $\binom{n}{2} \approx n^2$; we don't care about the constants like 3.

Student Question. *Does this constant c depend on s?*

Answer. Yes, but that's okay — as with graph KST, the implicit constant will depend on $s_1, \ldots, s_k$ and k (we just need it to not depend on n).

Now let's try to upper-bound X. In KST, we were considering stars; and what we just did is the analog of choosing the center vertex and summing over the choices for the outer vertices. The other thing we did was choose the outer vertices and show that there's not many choices for the center vertex because there's no $K_{s,t}$. Here the analog is as follows.

Definition 4.6. Given distinct vertices $u_1, \ldots, u_s \in V(\mathcal{G})$, we define $G(u_1, \ldots, u_s)$ to be the *graph* (not a hypergraph) with vertex set $V(\mathcal{G}) \setminus \{u_1, \ldots, u_s\}$ and edge set

$$\{vw \mid vwu_1 \ldots u_s \text{ form a copy of } K^{(3)}_{1,1,s} \text{ with } v,\, w \text{ as the singletons}\}.$$

What is this graph? We choose $u_1, \ldots, u_s$ to be all the vertices on the bottom, and we make a pair vw an edge if and only if they form this picture (where v and w are these special vertices and $u_1, \ldots, u_s$ are the outer vertices).

The goal is to prove this by induction on k; this is good because we've defined a graph, and we know things about graphs. So we want to use the KST theorem on this graph.

We claim that this graph avoids a certain subgraph. Which one?

Claim 4.7 — The graph $G(u_1, \ldots, u_s)$ is $K_{s,s}$-free for any $u_1, \ldots, u_s$.

Why? If we had a $K_{s,s}$ in this graph, then we'd have three groups of s vertices, and the definition of this graph would ensure every time I pick a triple from one of the vertices in each of the three parts, that would be a hyperedge. So if this graph had $K_{s,s}$, the hypergraph would have $K^{(3)}_{s,s,s}$, and we assumed it does not.

This is great because our goal was to apply induction; so we conclude that

$$e(G(u_1, \ldots, u_s)) \leq O(n^{2-1/s}).$$

The final thing is to relate this back to what we were counting. Remember we were counting copies of $K_{1,1,s}$ (that's the parameter we called X). What's another expression for what X equals? We can write

$$X = \sum_{u_1,\ldots,u_s} e(G(u_1, \ldots, u_s))$$

(where we sum over distinct vertices $u_1, \ldots, u_s$) — given a copy of $K^{(3)}_{1,1,s}$ in the hypergraph, if we call u_1, $\ldots$, u_s its outer vertices, then it corresponds to one edge in this graph (so there's a bijection between edges in these graphs over all $u_1, \ldots, u_s$ and copies of $K^{(3)}_{1,1,s}$ in our graph).

And we can upper-bound this by

$$X \leq \binom{n}{s} \cdot O(n^{2-1/s}) \leq c' n^s n^{2-1/s} = c' n^{2+s-1/s}$$

(using the observation that every term in this sum is $O(n^{2-1/s})$, and we have $\binom{n}{s}$ terms in the sum, since we need to pick s distinct vertices).

Now we have a lower bound and upper bound on the same object, so we want to compare them — before, we showed that

$$X \geq cn^2(Cn^{1-1/s^2})^s = cC^s n^{2+s-1/s}.$$

This is good because it's the same exponent. (This might look like a miracle, but of course this isn't how you come up with the poof — you run through this argument and then *make* them the same thing, and that tells you the right thing to assume about $e(\mathcal{G})$.)

What's the contradiction? Remember that c and c' are just absolute constants depending on s (while C is something we have the power to choose). So now we pick C super large. Then this is going to be bigger than the upper bound (because they're the same except for some constant junk, and by picking C super big we can beat whatever constant junk we've accumulated). So that completes the proof. □

Student Question. *Is it ever worth tracing through this to see how large C has to be?*

Answer. Some people have done this. Especially in the graph case people do care about these constants. In the hypergraph case we know so little that people aren't really interested in the constants.

Student Question. *For $k = 4$, would you define a 3-graph instead?*

Answer. Yes. Now we'd count copies of $K^{(4)}_{1,1,1,s}$. The first part of the argument goes the same way (with codegrees for triples). For the second part, we take s-tuples of vertices $u_1, \ldots, u_s$, construct a 3-uniform hypergraph in a similar way, and do the same thing. (The ideas are the same; there's just more notation.)

Just one final observation:

Corollary 4.8

If $\mathcal{H}$ is a k-partite k-graph, then $\text{ex}(n, \mathcal{H}) = o(n^k)$.

This is the same as in the graph case, where once we proved KST, we discovered *all* bipartite graphs have subquadratic extremal number. Similarly, all k-partite k-graphs have extremal number $o(n^k)$ (where n^k is the trivial bound) for the same reason — any k-partite k-graph can be found inside a complete one, and we just proved that complete ones have this bound on their extremal number.

Student Question. *Do we know lower bounds?*

Answer. Yes, and they're kind of the same type — we know that if you fix $s_1, \ldots, s_{k-1}$ and let s_k be super-duper huge, then we do have a matching lower bound (up to the constant). But for example, it's a major open problem to determine $\text{ex}(n, K^{(3)}_{2,2,2})$. This is called Erdős's box problem (if you want. you can think about why it's called 'box,' this isn't obvious). We showed a bound like $n^{11/4}$, and the best-known lower bound has some exponent other than 11/4 (it's better than n^2 — n^2 is easy, and you can do better than this by being clever); closing this gap is a major open problem. But we do know that 11/4 is tight for something like $K^{(3)}_{2,2,100}$.

Student Question. *Is it conjectured that the upper bound is right?*

Answer. Just as with the graph case, for a long time there was consensus it should be, and now that's changing. But people don't really have a good guess for what the exponent should be otherwise.

§4.2 Supersaturation

We're now going to pivot like crazy to a completely different topic (not hypergraphs). Today is a mishmash of these two different topics; these are the two ideas we need to complete the proof of the ESS theorem on Monday (one input is hypergraph extremal numbers, which we've just done, and the other is this).

Supersaturation is a super general phenomenon in many parts of extremal combinatorics. We say if something is big, it contains a something — e.g., Turán says if your graph has many edges, it has a triangle (or K_r). Supersaturation is a general phenomenon that says in many instances, if you're even a tiny bit bigger than what I just told you, you have not just *one* object of the kind you're looking for, but a *lot*. So far we've talked about finding just one triangle (or one K_r). But supersaturation says once you just have a teeny bit more edges, you actually have a ton of triangles (or a ton of K_r's).

Theorem 4.9 (Erdős–Simonovits 1983)
For every $k \geq 3$ and every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds for all sufficiently large n: If G is an n-vertex graph with

$$e(G) \geq \left(1 - \frac{1}{k-1} + \varepsilon\right)\binom{n}{2},$$

then G contains at least $\delta\binom{n}{k}$ copies of K_k.

(Both ε and δ are real numbers, which you should think of as very close to 0.)

If we didn't have the $+\varepsilon$, then this would be the number of edges in the Turán graph. Turán's theorem implies if you have that many edges, you have at least one copy of K_k. But supersaturation says you don't just get one, but a ton of them. What's a ton? Any graph on n vertices has between 0 and $\binom{n}{k}$ copies of K_k. And here this is saying you get a *constant fraction* of all the possible copies you could hope for! So you just need to be a tiny bit over the number of edges you need to guarantee 1, and you guarantee a whole ton.

This is quite weird and surprising, but it's true, as we'll prove. To prove it, we'll need a lemma.

Student Question. *Can we say anything about how big δ is?*

Answer. Yes, but Yuval won't tell us (ask later). The proof will give some δ but it's not optimal; the optimal one is known. In fact, Razborov invented flag algebras to solve this for triangles.

We'll state this lemma in greater generality than we'll need it, because the proof works equally well (but that means we need more parameters).

Lemma 4.10

Let $0 < \alpha < \beta < 1$ be real numbers, and let $m \geq 2$ be an integer. Let G be an n-vertex graph with $n \geq m$ and $e(G) \geq \beta\binom{n}{2}$. Then the number of sets $M \subseteq V(G)$ with $|M| = m$ and $e(M) \geq \alpha\binom{m}{2}$ is at least $(\beta - \alpha)\binom{n}{m}$.

This is a mouthful, but let's walk through it. We have a big graph, and I'm assuming it has at least $\beta\binom{n}{2}$ edges (a β-fraction of all possible edges). Now I'm interested in sets M of size exactly m where these sets have a lot of edges. What's the 'a lot' I'm targeting? These sets M have anywhere between 0 and $\binom{m}{2}$ possible edges; I want ones which have at least an α-fraction. And the answer is a lot — it's a constant fraction of all m-sets ($\binom{n}{m}$ is the total number of m-sets).

This is a mouthful, but informally, this says that if my graph has many edges, then a bunch of subsets also have many edges. And I can pick α and β, so I can make them really close, so that these sets I have are *nearly* as dense as the global graph.

Proof. The proof will consist of two magic identities. The first is that

$$\binom{n-2}{m-2}e(G) = \sum_{|M|=m} e(M).$$

Why? Both sides are counting the same thing — the number of choices of an edge and an m-set containing it. On the right-hand side we're summing over all m-sets; on the left-hand side we first pick an edge, and now you have to pick the remaining $(m-2)$ vertices, which can be done in $\binom{n-2}{m-2}$ ways.

Let's give some names for the things we care about — let $\mathcal{M}_0$ be the set of M such that $e(M) < \alpha\binom{m}{2}$, and let $\mathcal{M}_1$ be the set of M such that $e(M) \geq \alpha\binom{m}{2}$. We want to understand how many sets are in $\mathcal{M}_1$ (that's what the statement of the lemma is about).

So continuing this computation, we'll split this sum as

$$\sum_{|M|=m} e(M) = \sum_{M\in\mathcal{M}_0} e(M) + \sum_{M\in\mathcal{M}_1} e(M).$$

(The second sum is the ones I like; the first is the ones I don't like.) Now I want to get an upper bound on this. For the first sum, there are $|\mathcal{M}_0|$ terms in the sum, and we can upper-bound every term by $\alpha\binom{m}{2}$ by definition. For the second sum, there are $\mathcal{M}_1$ terms in the sum, and we can upper-bound the terms by $\binom{m}{2}$. So this is at most

$$|\mathcal{M}_0|\,\alpha\binom{m}{2} + |\mathcal{M}_1|\binom{m}{2}.$$

Now let $x = |\mathcal{M}_1|/\binom{n}{m}$, so that $1 - x = |\mathcal{M}_0|/\binom{n}{m}$. (So we're defining x as the *fraction* of m-sets that are in $\mathcal{M}_1$.) Then we can rewrite this as

$$\binom{m}{2}\cdot(\alpha\,|\mathcal{M}_0| + |\mathcal{M}_1|) = \binom{m}{2}\binom{n}{m}(\alpha(1-x)+x) = \binom{m}{2}\binom{n}{m}(\alpha + (1-\alpha)x).$$

(This is just rearranging — we defined x so that this is true.)

The reason we write it like this is that our goal is eventually to get a bound on x. The good news is I have a crazy inequality starting with $\binom{n-2}{m-2}e(G)$, which I understand, then there's a $\leq$, and in the end we end up with something involving x. So what we learn is that

$$\alpha + (1-\alpha)x \geq \frac{\binom{n-2}{m-2}}{\binom{m}{2}\binom{n}{m}}e(G).$$

And now, I had an assumption on $e(G)$, so I had better use that at some point; the right-hand side is at least

$$\frac{\binom{n-2}{m-2}\binom{n}{2}}{\binom{m}{2}\binom{n}{m}} \cdot \beta$$

(since $e(G) \geq \beta\binom{n}{2}$). Now comes the magic — this big fraction is just 1! Why? It's the same bijective magic — both the numerator and denominator count the same thing, where I start with n things, pick m out of them, and pick 2 out of those. I can either pick m and then 2 out of the m, or I can first pick the special 2 and then pick the remaining $m-2$ out of the remainder. So these two things count the same thing, which means they're equal, and their ratio is 1.

So we learn that $\alpha + (1-\alpha)x \geq \beta$, which says

$$x \geq \frac{\beta - \alpha}{1 - \alpha} \geq \beta - \alpha.$$

And returning to how everything was defined, x was the fraction of sets I like; so I've proven that's at least $\beta - \alpha$, which is what we wanted. ☐

Student Question. *Why aren't we including the $1-\alpha$ in the lemma?*

Answer. Just because we don't need it (we have proven a stronger statement with the $1-\alpha$, but we don't need it).

We're out of time, so we won't prove the supersaturation theorem, but it's actually a fairly simple consequence of this lemma (we'll prove it on Monday).

§5 July 14, 2025

We have 35–38 of us. If you know any of your fellow USS members who have issues with coming to the lecture, let Irena or Paul know, and they can reach out. But we have pretty good attendance from a historical perspective.

Tonight we have two conflicting things. One is more widely popular (the movie), and the other more narrowly (classical music). We're near Kimball Junction. There's a shuttle, which has to be scheduled and holds 10–11 people. There's also a public bus, which picks up at the 7-11. It's free. You take either the 101 or 10X. The museum is off to the right on Main Street (after Park Street). One street over there's Sullivan, and there's City Park, and the concert is here. For the movie, you get tickets in advance from Dena, or at the door as long as you're wearing your badg. The movie is Science Fair; it's a Sundance film winner. Paul doesn't know how long it is, but it starts at 6. The concert is at 6:30. You'll see a band stand area, and you can sit on the grass and eat or do whatever you want. Paul will try to be out there at 5 and schedule the shuttle a bit later. If you're going to the movie, the bus is probably the best thing to do.

§5.1 Supersaturation

As a tiny sense of where we're going, the main theorem of extremal graph theory is ESS; that's what we'll prove today. After that, today and tomorrow, we'll do a bit more EGT. Starting Thursday we'll do the other topic of the course, which is Ramsey theory (a new topic, though related). There might be some stuff in the notes we won't get to.

On Friday, we stated the supersaturation version of Turán's theorem:

Theorem 5.1 (Erdős–Simonovits)
Let $k \in \mathbb{N}$ and $\varepsilon > 0$. There exists $\delta > 0$ such that the following holds for all large n: Suppose that

$$e(G) = \left(1 - \frac{1}{k-1} + \varepsilon\right)\binom{n}{2}.$$

Then there are at least $\delta\binom{n}{k}$ copies of K_k.

So we're assuming our graph has at least the density for Turán's theorem, plus a bit more. Turán's theorem says we have at least one K_k; supersaturation says we get a lot more. The important thing is $\binom{n}{k}$ is the total number of possible K_k's we could have. So as soon as we're just a tiny bit above the point at which we might not have any, we have a *constant fraction* of all possible ones.

Last time we stated and proved a useful lemma:

Lemma 5.2
For $\alpha < \beta$ in $(0, 1)$, if $e(G) \geq \beta\binom{n}{2}$, then at least $(\beta - \alpha)\binom{n}{m}$ sets $M \subseteq V(G)$ of size m with $e(M) \geq \alpha\binom{m}{2}$.

So G has a β-fraction of all possible edges, and I'm asking, what fraction of vertex sets of size m also have a lot of edges, where 'a lot' will be $\alpha\binom{m}{2}$? The important thing is $\alpha < \beta$, but it can be as close to β as you like — maybe I have 37% of all edges in the global graph, and I'm asking how many size-m subsets have 36% of all edges. And the answer is a constant fraction of all possible sets.

The proof was a couple of applications of some magic identities, where some crazy binomial coefficients were equal. All of those were for the same reason, where we were counting the same thing in two ways (and we'll see another of those soon).

Proof of supersaturation. In the supersaturation theorem, we're given ε. We want to eventually produce δ; importantly, it's allowed to depend on ε and k, but not anything else. We want to apply the lemma, so we need to pick α and β.

First recall that the Turán function grows like

$$t_{k-1}(m) = \left(1 - \frac{1}{k-1} + o(1)\right)\binom{m}{2}.$$

This means we can pick some *fixed* m (that depends on k and ε, but nothing else) such that

$$t_{k-1}(m) < \left(1 - \frac{1}{k-1} + \frac{\varepsilon}{2}\right)\binom{m}{2}.$$

(The fact that $t_{k-1}(m)$ grows like this means that the $o(1)$ can be made arbitrarily small; so I can find some fixed m, which is an absolute constant, where this $o(1)$ thing is less than $\varepsilon/2$. Importantly, m depends on k and ε, but not anything else.)

So that's m; now we need to say what α and β are. It's obvious what we should pick β to be — we're trying to prove a statement where we assumed a bound on $e(G)$, and in the lemma we're assuming $e(G) \geq \beta\binom{n}{2}$, so we should pick β to be the thing in our assumption. So we pick

$$\beta = 1 - \frac{1}{k-1} + \varepsilon \quad \text{and} \quad \alpha = 1 - \frac{1}{k-1} + \frac{\varepsilon}{2}$$

(so both are of the same form, and α is just a tiny bit less than β — you should think of ε as very small).

Now we've set up things such that we can apply the lemma. We're allowed to assume n is very large; and the number of edges in G is at least $\beta\binom{n}{2}$ by assumption; so we're exactly in the setup of the lemma, and we can apply it. So the lemma implies that the number of m-sets M with $e(M) \geq \alpha\binom{m}{2}$ is at least

$$(\beta - \alpha)\binom{m}{2} = \frac{\varepsilon}{2}\binom{m}{2}.$$

Now I have a ton of sets M that satisfy this property; what can I do with it? Every such M has strictly more than $t_{k-1}(m)$ edges (by how we set everything up and chose α). And what's the point of having more than $t_{k-1}(m)$ edges? That tells me that there's a K_k. So M contains a copy of K_k. (That's Turán's theorem — it tells me that if I have a graph, here the induced subgraph on M, and its number of edges is strictly more than this Turán number, then it must have a K_k.)

So that's really good — the goal at the end of the day was to find many copies of K_k, and I just generated a machine that can produce a ton of copies of K_k (I have many such sets, and each one contains a copy of K_k). So what we find is that — let's first write something wrong. We might say that the number of copies of K_k is at least

$$\frac{\varepsilon}{2}\binom{n}{m}$$

(because that's how many sets M I have, and each generates a K_k). That's wrong, because I overcounted. (You can tell it can't possibly be true because this can be way bigger than $\binom{n}{k}$, which is the maximum number of K_k's there could possibly be.) The point is that each K_k might be counted many times; so we need to divide to account for this overcounting.

How many times can I have counted a K_k? It's $\binom{n-k}{m-k}$. Why? I'm counting these sets M, and for any given K_k, how many m-sets can it be counted in? I need to pick all the other vertices in M; so I'm picking $m - k$ vertices out of $n - k$ options. So certainly this is the worst I could have overcounted by.

So the number of K_k's is at least

$$\frac{\varepsilon}{2}\frac{\binom{n}{m}}{\binom{n-k}{m-k}}.$$

We're about to be done — my goal is to tell you G has many K_k's, and I just got a lower bound. So all I need to do is make sure it's at least what I claimed (if I pick δ appropriately).

Here comes the magic — this is equal to

$$\frac{\varepsilon}{2}\cdot\frac{\binom{n}{k}}{\binom{m}{k}}.$$

This is again a magic identity of binomial coefficients. Why? If you cross-multiply, we want to show $\binom{n}{m}\binom{m}{k} = \binom{n-k}{m-k}\binom{n}{k}$. And both count the number of ways to start with n things, pick m inside, and then pick k inside that (on the right we're starting by picking the k, and then the remaining $m - k$).

Now we'll just write this in a slightly different form, as

$$\frac{\varepsilon}{2\binom{m}{k}}\binom{n}{k}.$$

The reason is our goal was to get at least $\delta\binom{n}{k}$. And now we've gotten something times $\binom{n}{k}$, so let's just define

$$\delta = \frac{\varepsilon}{2\binom{m}{k}}.$$

And we're done. Crucially, δ depends only on ε, m, and k; and m itself depends only on ε and k; so δ is just some constant depending on ε and k. □

Student Question. *Is this the best choice of δ?*

Answer. No. In fact, the optimal δ is known, but it's very difficult. But this is all w need here.

This is how lots of supersaturation theorems are proved in combinatorics. Supersaturation is a very general phenomenon where you have a theorem guaranteeing one instance of the object you care about, and you can often boost it to get several. There are many ways of proving it, but this is one of the most common. We already have a machine saying in some instances we can get one, and we apply it to many subinstances. Here the trick was we applied Turán's theorem all over, and we knew many of the sets had enough edges to apply Turán, so we did. And the numbers worked out (and they often do, but not always — there are supersaturation theorems that are true but can't be proved in this way, but this is one that is true and can be proved in this way).

§5.2 Erdős–Stone–Simonovits

Let's prove ESS, which is our main goal.

Theorem 5.3 (ESS)
For any graph H, we have

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

We need to prove a lower and upper bound. We already proved the lower bound on the first day — the Turán graph itself gave a lower bound. So all that remains to do is prove the upper bound.

We'll state the Erdős–Stone theorem; this is a special case of ESS, and as we'll see in a moment, it's really the same thing.

Theorem 5.4 (Erdős–Stone)
For every s and k, we have

$$ex(n, K_k[s]) = \left(1 - \frac{1}{k - 1} + o(1)\right)\binom{n}{2}.$$

Definition 5.5. The graph $K_k[s]$, called the s-blowup of K_k, is the complete k-partite graph with parts of size s.

So this is just the graph whose vertex set is k parts of size s, where all edges across are present. There's many other names for it, such as the complete k-partite graph. (It is also $T_k(ks)$, though we won't need that.)

Two obvious things: First, $\chi(K_k[s]) = k$. Certainly we can color it with k colors (give each part its own colors). And we can't do it with any fewer, for example because it contains a K_k (and that I can't possibly color with fewer colors).

The reason we point this out is that ES is a special case of ESS (ESS says this holds for any H, and ES says it holds for this specific H). But actually, it's the general case, for the following reason:

Fact 5.6 — If $\chi(H) = k$, then H is a subgraph of $K_k[s]$ for some s.

Proof. By assumption $\chi(H) = k$, so it has some proper k-coloring. Bundle together all the vertices of color 1, color 2, and so on. And then add in all edges between different color classes (I'm trying to make H a subgraph of something, so I can do this). Now I have a complete k-partite graph with parts of some size; call s the largest size, and add more vertices if you need to (that's fine because I'm trying to generate H as a subgraph). $\square$

(This is basically an equivalent definition of the chromatic number, if you want.)

Student Question. *What if you have one color with a lot more vertices?*

Answer. If you have one color with 1000 vertices and the rest with 3, then we just add 997 vertices to each of the other parts. (Subgraph means I can add edges and also add vertices.)

Why is this good? We saw ES is a special case of ESS. But actually it gives us the reverse implication as well — by the easy observation we saw on the first day, this implies

$$\mathrm{ex}(n, H) \leq \mathrm{ex}(n, K_k[s])$$

(whenever something is a subgraph of something else, the extremal numbers satisfy this inequality). And we already proved a lower bound in ESS, so our goal is just to prove the upper bound; and by that inequality, it suffices to prove the upper bound for $K_k[s]$.

Student Question. *Aren't the n's different because we might be adding vertices?*

Answer. We're only adding vertices to H to produce $K_k[s]$, but then H and $K_k[s]$ are fixed graphs with $H \subseteq K_k[s]$, and this inequality holds for every n.

We've almost done nothing so far (this is just translating definitions around), but the point is ES is sufficient to prove ESS.

Remark 5.7. Erdős–Stone proved this in 1946; ESS was proved by Erdős–Simonovits in the 1960s. All that happened in the intervening 20 years is this realization. (The real meat is Erdős–Stone, which is why people frequently call it Erdős–Stone; it just took people a while to realize that this statement gives you the result for all H.)

Student Question. *KST was after this, so how did the original proof of Erdős–Stone go?*

Answer. Differently. It involves induction; the ideas are very pretty, but making it work is technically annoying. There's at least 2 other proofs of ESS Yuval knows; he's presenting one which he thinks is nicest, but it's far from the only one.

Proof of Erdős–Stone. We need to prove an upper boudn on $\mathrm{ex}(n, K_k[s])$; that basically means showing that if I stick any ε in place of the $o(1)$, then the theorem is true for all large n. So what we want is that for any ε and for all sufficiently large n, we have

$$\mathrm{ex}(n, K_k[s]) < \left(1 - \frac{1}{k-1} + \varepsilon\right)\binom{n}{2}.$$

(This is just translating what the meaning of $o(1)$ is — it means I can make this extra error term negligibly small as $n \to \infty$, and if you think about what a limit is, that's exactly this.)

What does it mean to prove this? We want to show that if G is an n-vertex graph (where n is sufficiently large) and

$$e(G) \geq \left(1 - \frac{1}{k-1} + \varepsilon\right)\binom{n}{2},$$

then G contains a copy of $K_k[s]$. (Again, we haven't done anything; we're just restating what the definition of an extremal number is.) (To show a bound on the extremal number, I need to show you that any graph with at least that many edges has a copy of $K_k[s]$.)

What do we apply at this point? We've conveniently set everything up with the same notation; indeed, this is precisely the assumption in supersaturation, so we should apply that. What we find is that by supersaturation, there exists $\delta > 0$ (not depending on n) such that G has at least $\delta\binom{n}{k}$ copies of K_k.

That's pretty good, right? What we're looking for is $K_k[s]$, which is sort of a bunch of copies of K_k sitting all over each other. I don't have that yet, but I have something sort of in the right direction, which is a lot of copies of K_k.

What'll we do next? Our goal is somehow to stitch together the K_k's, and it's not totally clear how to do this. We had this whole weird detour into hypergraphs which has not yet been useful for anything, but we promised it was going to be useful for proving ESS, so let's use it.

The point is that we're in a situation where we have a lot of somethings, and want to conclude there is a something (in Turán you have lots of edges and want to say there's a K_k; in the hypergraph setting we had lots of hyperedges and want to conclude there's something). Here we have a lot of K_k's. That doesn't quite look like what we were doing, but there's a way of making it look like those things: Let $\mathcal{G}$ be the k-graph whose vertex set is $V(G)$, whose hyperedges are the copies of K_k in G. So I'm defining a k-uniform hypergraph, by looking at the copies of K_k in my graph, and calling each of those a hyperedge (this is allowed because my hyperedges are allowed to be whatever collection of k-sets I want). How many edges does this hypergraph have? At least $\delta\binom{n}{k}$.

Now I have a hypergraph, and I know it has a lot of edges (where 'a lot' is some tiny fraction of all possible hyperedges). Now I want to use a theorem I know about hypergraphs. Conveniently, we've only said one theorem about hypergraphs. To recall it:

Theorem 5.8 (Hypergraph KST)
We have $\mathrm{ex}(n, K^{(k)}_{s,s,\dots,s}) = O(n^{1-1/s^{k-1}})$.

Crucially, the point is that this is $o(n^k)$ — this grows slower than n^k. On the other hand, we said that $e(\mathcal{G}) \geq \delta\binom{n}{k}$, which is of the form cn^k for some constant c. So the point is that if n is sufficiently large, then

$$e(\mathcal{G}) > \mathrm{ex}(n, K^{(k)}_{s,\dots,s})$$

(because the left-hand side grows as some constant times n^k, and the right-hand side grows slower — it's like $n^3/10^9$, vs. $10^9 n^{2.999}$; and if n is large enough, the billions don't matter, and the fact that the exponent on the left is bigger means that the left will be bigger).

And the point is that then this hypergraph $\mathcal{G}$ contains a $K^{(k)}_{s,s,\dots,s}$. And why are we done? Now I remember that $\mathcal{G}$ came from a graph G, and on the level of the graph, the graph G contains $K_k[s]$. This is because the picture would look basically the same — the hypergraph picture has k parts of size s where every k-tuple is a hyperedge, i.e., a copy of K_k in the graph. □

Student Question. *How do we make sure that each part doesn't contain edges?*

Answer. We're okay with that — we just want to find a copy as a subgraph, but we're okay with there being more edges. (We have to, because e.g., G could be the complete graph; then it will have a copy of whatever we're looking for, but also a lot of extra edges.)

To summarize, the statement of ESS is an explicit asymptotic formula for the extremal number of any H. The lower bound was pretty easy — just noting that the Turán graph itself with $\chi(H) - 1$ parts has no copy of H (by the definition of the chromatic number), and it has the right number of edges. Now we need to do the upper bound. This means I assume I have $(1-\frac{1}{k-1}+\varepsilon)\binom{n}{2}$ edges. Then by supersaturation, I have many copies of K_k — at least $\delta\binom{n}{k}$. Then by hypergraph KST, the hypergraph of copies of K_k contains a large complete k-partite thing — i.e., I see this picture in the hypergraph, which means I see the same-looking picture in the graph. Finally, if I know how to find one of these things $K_k[s]$ in my graph, then certainly I know how to find any graph H with $\chi(H) = k$ (since it lives inside such a graph $K_k[s]$).

§5.3 Exact extremal numbers

Yuval will now move on to a closely related question:

Question 5.9. What actually is $\operatorname{ex}(n, H)$?

In Turán's theorem, we actually found an exact formula, and the unique extremal example. ESS gives a very nice answer, but only for the asymptotic behavior, not the exact formula. On the first or second day, someone asked, is the exact same formula as Turán's theorem the truth in all cases? We had an example on the homework showing it's not always true — we saw that the extremal number of a pentagon with two extra edges is *strictly* larger than $t_2(n)$ (this has chromatic number 3), i.e., the Turán graph is not the extremal example (you can put more edges inside).

This is an instance of a more general phenomenon:

Definition 5.10. We say H is color-critical if there is some edge e in H such that $\chi(H \setminus e) < \chi(H)$.

In other words, H is color-critical if you can find an edge such that deleting that edge strictly reduces its chromatic number.

Example 5.11

Any odd cycle is color-critical (in fact, if you delete any edge, you get a path, which has chromatic number 2).

Example 5.12

A pentagon with two extra edges is not color-critical (you can check this by casework).

This is relevant because of the following cool fact:

Proposition 5.13

If H is not color-critical, then $\operatorname{ex}(n, H) > t_{\chi(H)-1}(n)$.

So for those graphs that are not color-critical, their extremal number is *not* given by the Turán number, and the Turán graph is not the extremal graph.

Proof. We'll take the Turán graph, and then take one of the parts and stick a single edge inside: Consider $T_{\chi(H)-1}(n)$, and then add one edge to one part. (The Turán graph has maybe 3 blobs and all edges between them; we add a single additional edge in one of the parts.) Certainly the number of edges is strictly greater than $t_{\chi(H)-1}(n)$ (since we started with this and added 1). And the claim is that this is H-free.

The proof is similar to what we just did — imagine you did have a copy of H in here, so you have some way of finding H in this thing. The idea is a copy of H in this graph 'almost' gives a proper coloring of H

with $\chi(H)-1$ colors (I just color the vertices according to which part they live in). This is almost a proper coloring, in the sense that for almost every edge in H, its endpoints receive different colors. There's one counterexample — if my copy of H used this special edge, then there's one edge where the two endpoints received the same color. But there's at most one of these — at most one edge has monochromatic endpoints.

And if I have a coloring which is proper except that there's one bad edge, if I deleted that edge, then I would have a graph with chromatic number $\chi(H)-1$ (because this coloring would be a proper coloring of that thing). So this shows H is color-critical, which is a contradiction.

In other words, this graph is H-free and has strictly more edges than the Turán graph, so $\text{ex}(n, H) > t_{\chi(H)-1}(n)$. □

That's nice and easy. But it turns out the converse is also true, which is kind of amazing.

Theorem 5.14 (Simonovits 1968)
If H is color-critical and n is sufficiently large, then $\text{ex}(n, H) = t_{\chi(H)-1}(n)$, and the unique extremal graph is $T_{\chi(H)-1}(n)$.

So the converse is also true — if your graph is color-critical, then Turán's theorem still holds for it (for any sufficiently large n) — the extremal number is *exactly* the Turán number, and that's the unique extremal construction.

There's one obvious question: What's the deal with 'n sufficiently large' — is that necessary? The answer is yes.

Example 5.15 (Odd cycles)
This theorem says that e.g. $\text{ex}(n, C_5) = \lfloor n^2/4 \rfloor$ for all large n. But this is not true for all n — for example, with $n = 4$, the graph K_4 doesn't have a 5-cycle but has more edges than this.

So you do need 'large n'; for odd cycles we know exactly what n this kicks in for, but for general H we don't (but 'large n' is necessary).

Student Question. *The 'large n' condition is not just trivial, in the sense that you just need enough vertices to possibly have a H?*

Answer. Even for odd cycles, for cycles of length ℓ you need something like 2ℓ; and in general it can be worse.

If there is some time tomorrow, Yuval will tell us a couple of ingredients that go into the proof, at least for the C_5 case. The lecture notes have a more complete proof sketch (still missing one step). There are very cool ideas in this proof, but it's also kind of a pain to get it to work (which is why we're not doing it in full detail).

Question 5.16. What's going on with other graphs?

We have a cool dichotomy where if you are color-critical the Turán graph is the actual extremal graph, but if you're not color-critical then it's not. But what can we say about those?

The next example on our homework was the octahedron graph, which we call O_3. What we showed on our homework (maybe) is that $\text{ex}(n, O_3)$ is actually a lot bigger than $t_2(n)$ (it has chromatic number 3) — in fact

$$\text{ex}(n, O_3) \geq t_2(n) + \Omega(n^{3/2}).$$

(We showed if you're not color-critical you're at least 1 more, but this one is actually a lot more.)

The proof was that you start with the Turán graph, which is a complete bipartite graph. Before, we stuck in a single edge into one of the parts. But now what you stick inside is a C_4-free graph (like the one we constructed in class based on geometry over finite fields). By doing that, you add $cn^{3/2}$ edges (that's how many edges this has). And the thing you had to check on the homework was that this works — if you do this, you don't generate a copy of O_3 in the graph.

Question 5.17. What's the more general phenomenon underlying this (just as 'color-critical' was the underlying phenomenon for the first example)?

Definition 5.18. Let $\chi(H) \geq 3$. A pseudocoloring of H is a function $f : V(H) \to [\chi(H) - 1]$ such that if $uv \in E(H)$, then either $f(u) \neq f(v)$, or $f(u) = f(v) = 1$.

This is not normal terminology; Yuval doesn't think this has a real name, so he made one up. (Remember $[k]$ means the set of integers between 1 and k.)

So if we didn't have the $f(u) = f(v) = 1$ condition, this would just be a proper coloring. Here I'm giving you the power to let adjacent vertices be the same color, but only if that same color is 1. (So there's one special color where you are allowed to have adjacent vertices.)

The other definition, which does have a standard name (unfortunately, because the terminology sucks):

Definition 5.19. We say a bipartite graph B is in the decomposition family of H if for every pseudo-coloring of H, we have $B \subseteq f^{-1}(1)$.

In other words, B is in the decomposition family if no matter how you pseudocolor, in the special color which has edges, you will see a copy of B — so no matter how you do the pseudocoloring business where you allow some edges to be monochromatic in a special color, you're guaranteed to find a copy of B among the edges of this color.

Fact 5.20 — C_4 is in the decomposition family of O_3.

(You can check this with casework.) So no matter how you pseudocolor the octahedron, you're guaranteed to find C_4 in the first color. And that's what underlies our proof working. The generalization of this:

Proposition 5.21

If B is in the decomposition family of H, then $\text{ex}(n, H) \geq t_{\chi(H)-1}(n) + \text{ex}(\lfloor n/(\chi(H) - 1) \rfloor, B)$.

So this says if B is in your decomposition family, then $\text{ex}(n, H)$ is at least the Turán number plus something, where the something is a

The proof is you start with a Turán graph, and inside one of the parts you stick a B-free graph with as many edges as possible. You only need to check that this doesn't create a copy of B; and this definition (of the decomposition family) was basically made so that this is true. (If you solved the exercise, this is exactly what you did.)

This means to actually understand what's going on for general H, you can't get around understanding extremal numbers of bipartite graphs.

For O_3, this is basically the true value — in fact, the extremal graph is to take a complete bipartite graph, put a C_4-extremal thing in one part, and put a matching in the other. There's a conjecture that any construction looks like this in some sense — you take the Turán graph and then do something smart inside the parts based on these decomposition families.

§6 July 15, 2025

Today's the last day of extremal graph theory as such; on Thursday we'll move to Ramsey theory (there will be overlap and similarities, but the topics will feel a bit different).

At the end of yesterday, we discussed that if you want to understand the extremal number of a general graph H, you're basically forced to understand the bipartite case. And we talked a fair amount about the extremal number of bipartite graphs. But we'll now say more.

If you care about bipartite graphs, the most natural one is complete bipartite graphs, and the second most natural is even cycles.

Theorem 6.1 (Erdős, Bondy–Simonovits)
We have $\operatorname{ex}(n, C_{2\ell}) = O(n^{1+1/\ell})$ (where the implicit constant may depend on ℓ).

(Who this is due to depends on who you ask; Erdős didn't write it up, and Bondy–Simonovits did. Most people say it's due to them, but they say it's due to him. . .)

One thing to notice is that for complete bipartite graphs, the upper bound we had was $n^{2-1/s}$; in particular, when s is very large, this is very close to 2. But with even cycles, it's the other way — if you have a very long even cycle, the upper bound is very close to n (instead of n^2). The intuition is a complete bipartite graph is very dense, so it's hard to find these. Conversely, a cycle of length 1000 is a very loose structure, with a few edges very spread out, so it's a lot easier to find it.

One final observation is that when $\ell = 2$, we already proved this — we saw $\operatorname{ex}(n, C_4) = \Theta(n^{3/2})$.

We're not actually going to prove this; we'll prove something slightly weaker.

Theorem 6.2
If G is an n-vertex graph with at least $Cn^{1+1/\ell}$ edges (for some appropriate constant C), then G contains a cycle of length *at most* 2ℓ.

If we didn't have 'at most,' then this would be a restatement of the previous theorem — that if I have a graph with this many edges, it contains a cycle of length 2ℓ. But we'll prove it with the 'at most,' which is slightly weaker.

There's basically only one way of proving this, which we'll see. There are many ways of proving Bondy–Simonovits, but all of them take the proof of this second theorem and then fight for a while to show you're not finding a shorter cycle. (There is no very nice proof of that; there are very nasty proofs and merely nasty proofs, so we're not going to do it.)

Remark 6.3. In the terminology you may have seen on the homework, this second theorem is equivalent to saying $\operatorname{ex}(n, \{C_k \mid k \leq 2\ell\}) \leq Cn^{1+1/\ell}$ (where the extremal number of a family means you want to exclude all things in the family).

To prove this, we'll start with a lemma (also from the homework).

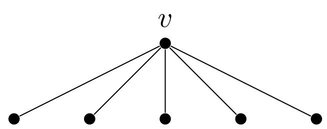
Lemma 6.4
Every n-vertex graph G contains a subgraph H with minimum degree at least $e(G)/n$.

This was on the homework (there was a problem about the extremal number of trees). It's a very useful statement used all over the place in EGT — if I have a graph with lots of edges, I can find a subgraph with high minimum degree. Equivalently, if I know something about the average degree of G, I can find a subgraph with basically the same bound on the minimum degree. This is useful in lots of contexts.

Proof sketch. The idea is you repeatedly delete vertices of degree less than $e(G)/n$. The point is when you do this, you delete at most n vertices (there's only that many to start with), and every time the number of edges you lose is its degree. So the total number of edges you lose is less than $e(G)$, which means you have at least one edge remaining. But the only way you can stop the process is when you have no low-degree vertices remaining. So in the end you have a nonempty subgraph with this minimum degree condition. □

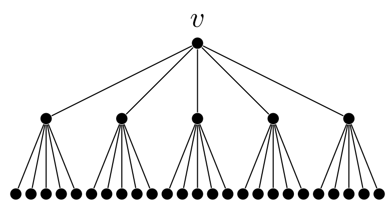
Proof of theorem. The assumption is we have a graph with $Cn^{1+1/\ell}$ edges, and our goal is to find a cycle. We'll start by applying the lemma; by the lemma, there exists some subgraph $H \subseteq G$; let's say it has m vertices (so $m \leq n$) and minimum degree at least $Cn^{1/\ell}$. Let's call this d.

Now comes the smart step. I'm going to fix some v, which is an arbitrary vertex of H. And I'm going to look at its neighbors; I'll call this set $N^1(v)$ (this is just all neighbors of v). Note that $|N^1(v)| \geq d$, because we assumed H had minimum degree at least d.



First, what happens if there is some edge inside this neighborhood? Then we have a triangle, and we're done (our goal was to find a cycle of length at most 2ℓ, so if there is an edge, we're done).

So let's assume there's no such edge. Then we look at all vertices in here (there's at least d). Each has degree at least d; one edge is accounted for, but each has $d-1$ edges coming out (and those edges can't stay inside the set, because then we'd have a triangle, so they must all leave). So we call this set $N^2(v)$, the set of all vertices at distance exactly 2 from v.



We have no idea what the size of this thing is yet. But here's another bad thing that could happen: What if two of these edges out of $N^1(v)$ collided? Then we'd have a 4-cycle, and then we're done. That means we might as well assume each of these edges are going to different places — there's no collisions whatsoever. And then we get

$$|N^2(v)| \geq d(d-1)$$

(we had d vertices, each sends out $d-1$ edges, and we didn't overcount anything). And if we have an edge inside $N^2(v)$, we'd have a 5-cycle, so we're also done.

Now we continue this process until we get to $N^\ell(v)$. The point is that we claim

$$|N^\ell(v)| \geq d(d-1)^{\ell-1}.$$

This is just by induction, basically. What's going on is every time we do one of these steps, we have some number of vertices, and each one of them sends at least $d-1$ edges out. Those edges can't stay inside the same set (or else we'd be done), and they can't collide (then we'd be done), so they must all go to different places outside. We can push this all the way to level ℓ (the worst thing that could happen is if we get a collision at the bottom layer; that gives us two paths of length ℓ, giving a cycle of length 2ℓ). We can't push it past this point, but up to $N^\ell(v)$ we get that there are no collisions.

Student Question. *Can you have a collision between different layers?*

Answer. This can't actually happen, because we're defining the next layer to be all the vertices you get from the previous layer. So anything that looks like that would actually look like an edge inside a layer.

Another way of thinking about this is we're exploring all vertices at distance at most ℓ from v. And if we had any collision in any direction, that would give us two paths of length at most ℓ from v that collide; and that's a cycle of length at most 2ℓ.

This $d-1$ is annoying, but the -1 doesn't matter; this is certainly $\Omega(d^\ell)$. And we said $d \geq Cn^\ell$, so this is at least $\Omega(C^\ell n)$ (where C is whatever constant we pick).

On the other hand, we claim that certainly

$$|N^\ell(v)| \leq m \leq n.$$

This is because this is some set of vertices in H, so it had better not have more vertices than the graph H (and in particular, it also lives in the original graph G).

So this is a contradiction if we pick C appropriately (in fact, C being a tiny bit more than 1 already works). □

Yuval will tell us an equivalent way of thinking about this proof, if we didn't like this one. An equivalent way of saying this is we're going to count walks of length ℓ from v — I just want to count how many ways there are of starting from v, going to a neighbor, then another neighbor, and so on ('walk' just means we're allowed to repeat vertices in this picture, but let's say I'm not allowed to backtrack). The number of walks is at least $d(d-1)^{\ell-1}$ — you have d choices for the first step, then $d-1$ for the second (you're not allowed to go back), then $d-1$ for the next, and so on. So the total number of walks is at least this, which by the same computation is going to be strictly larger than n. But that means by pigeonhole that two of them must end in the same place. So I start at v and have some crazy walk of length ℓ that ends somewhere, and another crazy walk of length ℓ that ends at the same point.

If they don't intersect before that last point, we get a cycle of length exactly 2ℓ. If they do intersect, we get a shorter cycle, which is still fine.

Student Question. *How do we guarantee $C^\ell > 1$?*

Answer. We get to pick C. So we can just pick $C = 1000$, and that's fine. (In fact, the implicit constant in $\Omega(d^\ell)$ doesn't matter, since all we did was change $d-1$ to d. So if you're careful you can just take C close to 1.)

How do you change this to get to the full theorem of Bondy–Simonovits? The kind of argument you want to do is, let's go back to the first step. We said if there was an edge inside the first set, we're done. If we're looking for a cycle of length *exactly* 2ℓ, that's no longer true. But if inside this set I find a *path* of length $2\ell-2$, then I'm done (that really does close a cycle of length 2ℓ). But we know from a previous homework problem that if a graph does not have a path of length $2\ell-2$, then it has really few edges (c times its number of vertices). So there really are very few edges inside $N^1(v)$ (it's not that there are no edges, but there's very few, or else we could find a long path and close a cycle). So it's no longer the case that all of them send $d-1$ edges out, but now most of them send lots of edges out (e.g. $d-10^6$).

So $N^2(v)$ is still pretty big. You also need to show there's not too many collisions. There could be some, but if there were too many, there would be ways of walking back and forth between these two layers to get a cycle of length 2ℓ. So the size of this is maybe not $d(d-1)$, but that divided by a million. And the next time you maybe get $d(d-1)^2$ divided by a trillion. And so on.

(Turning this into a real proof is not very easy; Yuval may have given it two stars on the homework, but it's probably harder than that. . . He also put a five-star problem on the homework; it's easier than that. The only known proof technique uses very advanced methods we don't know, but it's a huge open problem to find a proof not using those methods; Yuval would be very interested, and so would everyone else at the conference. So do try it.)

One more thing, which we mentioned on a previous day: We know that $\text{ex}(n, C_{2\ell}) = \Theta(n^{1+1/\ell})$ — so this upper bound is tight — if ℓ is one of 2, 3, and 5. So we know that for 4-cycles, 6-cycles, and 10-cycles, this upper bound is the truth. (The construction is on the homework; it's another one of these algebraic things, and it's a beautiful elegant proof.) It's a major open problem to determine whether this bound is tight for any other value of ℓ. Notably, for C_8 we don't know the answer. Maybe C_{14} is more accessible (no one knows how to do it, but there's good reason to believe it maybe can be done). This is another very annoying state of affairs; if you're asking what is wrong with 8, do the homework problem and you'll see why that construction doesn't work for C_8, which is a very frustrating state of affairs.

§6.1 Stability

A completely new topic now — the new topic, which will take the rest of the class, is stability.

Stability is a very important general type of result that occurs in many places in EGT. A general EGT kind of question is like Mantel — if a graph doesn't have a triangle, it has at most $\lfloor n^2/4 \rfloor$ edges, and the unique extremal example is the complete bipartite. Stability says, suppose I'm triangle-free and have very close to $n^2/4$ edges. Then I must actually look like the extremal example (the complete bipartite graph).

Let's give a formal statement. (This specific statement, which is very pretty and elegant, is due to Füredi, but stability in general is due to Erdős–Simonovits.)

Theorem 6.5 (Füredi)
Let G be an n-vertex K_r-free graph. Suppose that $e(G) \geq t_{r-1}(n) - s$. Then G can be made $(r-1)$-partite by deleting at most s edges.

My assumption is that my graph has almost as many edges as the Turán graph; it just differs from it by at most s edges.

So basically the assumption is my graph is K_r-free and has 'close' to the maximum possible number of edges. The conclusion is it looks a lot like the extremal graph. The extremal graph is the complete $(r-1)$-partite graph with all parts of the same size. This says you can delete a small number of edges to make it $(r-1)$-partite. On the homework you'll show you can use this statement to prove stronger things (you can add or delete a small number of edges to get the actual Turán graph).

Note that this is a strict generalization of Turán's theorem — Turán's theorem is exactly the $s = 0$ case of this statement (if I have a K_r-free graph with at least $t_{r-1}(n)$ edges, then it must be $(r-1)$-partite — I can make it $(r-1)$-partite by deleting 0 edges — which means it has at most as many edges as the Turán graph).

We'll only prove it for the case $r = 3$; on the homework you'll do it for all r (it's not much harder).

Proof for $r = 3$. We have a graph; we assume it's triangle-free and has nearly as many edges as the complete bipartite graph, and our goal is to make it bipartite. So the first step is to come up with a good partition of the vertex set into two parts, with the hope these will have few edges inside them. If we can do that, we can delete those few edges inside, and then we'll have a complete bipartite graph.

How are we going to do this? Let v be a vertex of maximum degree. Let $B = N(v)$ be its neighborhood, and let $A = V(G) \setminus N(v)$.

So the picture is I take this vertex v of maximum degree, look at its neighborhood and call that B, and then A is everything else (including v itself).

We'll show this is a good choice of the bipartition — we'll show there's few edges inside A and B, so by deleting those edges we'll get a bipartite graph.

First, I claim there's *really* few edges in B. Why? There's no edges in B, because if there were an edge in B, we'd have a triangle (consisting of v and that edge).

It'd be awesome if we could say there's no edges in A. That's not true, but we'll show there's *few* edges in A. The trick for doing this is we'll consider $\sum_{w \in A} \deg(w)$.

First, we claim this is at most $|A|\,|B|$. Why? We have $\deg(w) \leq \deg(v) = |B|$ for all w, because we picked v to be a vertex of maximum degree, so any vertex has degree at most $\deg(v)$, which by definition is $|B|$. So every term in this sum is at most $|B|$, and there are $|A|$ many terms in the sum.

Next, we know what this sum equals — we have

$$\sum_{w \in A} \deg(w) = e(A, B) + 2e(A).$$

Why? When I sum all degrees in A, every edge between A and B is counted exactly once; every edge inside A is counted twice (once for each endpoint).

Now comes a few lines of magic manipulations. First, we claim that $e(G) = e(A, B) + e(A)$. Why? Because there's no edges in B; so edges can only live inside A or between A and B.

Writing this in a slightly different form, we can say

$$e(G) = e(A, B) + 2e(A) - e(A) = \sum_{w \in A} \deg(w) - e(A) \leq |A|\,|B| - e(A)$$

(so far, this is all the bounds we just did). And the real magic is that this is at most $t_2(n) - e(A)$. Why? This is AM–GM (or the thing we saw on the first day) — we know $t_2(n)$ is the number of edges in the balanced complete bipartite graph, and among all complete bipartite graphs, the one with the most edges is the balanced one. So $|A|\,|B|$ (which is the number of edges in the complete bipartite graph between A and B) is at most what it would be if they had the same size.

Rearranging this, we get

$$e(A) \leq t_2(G) - e(G) \leq s$$

(by our assumption on $e(G)$). And why are we done? Our goal was to find a bipartition and at most s edges we could delete so that everything else goes across. We've just shown there are 0 edges in B and at most s in A, so if we delete those we're done. □

Student Question. *Did we have a stronger result in the homework, with* $2/5$*?*

Answer. That was about minimum degree assumptions — if your minimum degree is at least $2/5$, you are bipartite. But that proof really used the minimum degree. You need to be allowed to delete some edges to become bipartite in general. An example to keep in mind, if n is odd (say $n = 2m + 1$): Take two parts of size m, put a complete bipartite graph across, except for a single edge uw. Take this last vertex, and join it to just u and w. This doesn't have a triangle (this last vertex doesn't participate in a triangle, and other than that it's bipartite). This graph is also not bipartite (it has plenty of cycles of length 5). This has nearly as many edges as the Turán graph, and is not bipartite. So this shows you really need the minimum degree assumption in the homework problem (of course, it doesn't violate this theorem, because you can delete one of the edges from v to make it bipartite).

Why do we care about stability? It's very nice in its own right — it's cool these results are true. One way of thinking about it (Yuval has no idea if this analogy is helpful): EGT is about maximizing stuff subject to constraints. Think of this like optimizing in calculus — you have some weird search space and you're trying to maximize something over it. Normal EGT results say I've found the maximizer (e.g., the Turán graph) — that's the unique global maximum. What stability tells you is that the picture really looks like a hump at the maximum; it doesn't look like a weird picture where somewhere else you have another local maximum really close to the global one. That would be an example where you don't have stability — this first point is really the maximum, but we have something very far from it whose value is nearly the same. That would not be stability. And stability is the negation of this. So when you do have stability, the search space looks like a single bump, in the sense that if you're close to the maximum value, you must actually be close in the space of possibilities.

So it's cool when stability works, but it's also really useful. In the last 15 minutes, Yuval will try to sketch the proof of a theorem he stated yesterday:

Theorem 6.6

We have $\mathrm{ex}(n, C_5) = \lfloor n^2/4 \rfloor$ for large n, and the unique extremal graph is $T_2(n)$.

This was a special case of this theorem of Simonovits that if you're color-critical, your extremal number is exactly given by the Turán graph.

This is proved by what's called the stability method, which looks like this:

(1) Prove stability, which says that any extremal graph 'looks like' $T_2(n)$ (in general, you first prove that any extremal graph, in some rough sense, looks like the thing you're trying to show is the unique extremal graph).

(2) Work hard to iron out the imperfections. In our setting, what we'll see is that the extremal graph is *close* to a complete bipartite graph, but maybe not completely — there might be some edges inside the parts, and some missing edges across. But what you show is these imperfections can only hurt you, not help. So you basically show that in a genuinely extremal graph, you actually don't have imperfections. (This is the hard part.)

(3) Profit (once you've ironed out the imperfections, you're done).

The stability lemma we'll need:

Lemma 6.7 (Stability for C_5)

If G is an n-vertex C_5-free graph with $e(G) \geq \lfloor n^2/4 \rfloor$, then we can delete at most $Cn^{5/3}$ edges from G to obtain a bipartite graph.

This stability theorem is the unique nice stability theorem; every other one has other random constants and things tending to 0 as $n \to \infty$, so they become annoying to state. But this one says if I give you a C_5-free graph with at least as many edges as the Turán graph, I can make it bipartite by deleting a small number of edges (where 'small' is $Cn^{5/3}$, but the important thing is it's quadratic.)

Student Question. *Don't we know you have a C_5 if you have more edges than $\lfloor n^2/4 \rfloor$?*

Answer. We'll eventually show that, but we haven't shown it yet. So we start with a potential counterexample, prove it has some nice structure (that's what this statement says), and then massage that structure to show it actually can't exist. So this whole proof is really by contradiction (you assume for contradiction you have a counterexample, apply that lemma to it, and keep working).

We won't fully prove this, but the idea:

Proof idea. Let H be the subgraph of G comprising of all edges on a triangle (equivalently, the union of all triangles in G, viewed as a subgraph). The key observation is that H is $K_{3,3}$-free (for example); the point is if you had a $K_{3,3}$ in H, every one of these edges lives on a triangle, and once you get a triangle, it's not very hard to trace it around and get a C_5 (with the rest of the $K_{3,3}$). And we assumed the graph had no C_5. (The reason we want $K_{3,3}$ is because there are annoying issues where the triangle coincides with one of the other vertices.)

So the point is this subgraph has no $K_{3,3}$. Why is that good? That implies $e(H) \leq cn^{5/3}$ by KST. So now we're in business. We first delete all edges of H from my graph. Now I end up with a graph with no triangles (because I deleted all edges on triangles). And now I have a grpah with no triangles, and at least $n^2/4 - n^{5/3}$ edges (because I started with $n^2/4$ and deleted this many). So now I apply Füredi's theorem to delete another $n^{5/3}$ edges and get a bipartite graph. □

(This is spelled out in the notes in real detail.)

Now comes the sketchy part that's not done in the notes in full detail. We're done with Step (1) — if you have lots of edges and are C_5-free, you look like bipartite. Now comes the hard step (this will be 'no thoughts only vibes'). Our graph looks like two parts A and B, and it has some imperfections. First, we know we can delete a small number of edges to make it bipartite, so there may be some edges inside these parts, but a small number. Additionally, you can easily check — it has a ton of edges, something like $n^2/4 - n^2/10^9$. What that means is because A and B have very few edges, almost all the edges have to be across. So across, we have a ton of edges, but it might not be perfect — there might be some imperfections (missing edges) going across.

Our goal is to now iron out these imperfections. The way we'll do this is: Cheating a bit (this is true, but hard to justify), we'll assume that all vertices in A have at least $0.99\,|B|$ neighbors in B. So this graph across is extremely close to complete (there's just a very small number of imperfections), and we can even massage it to make sure these imperfections are not very clustered at any vertex. So that means every vertex in A has a ton of neighbors in B (every vertex in A is adjacent to almost all of B).

That implies that any two vertices in A have a common neighbor in B. In fact, they have a ton of common neighbors in B, just by pigeonhole (if a_1 is adjacent to 99% of the right and a_2 is too, the number of common neighbors is at least 98% of the right, so there's certainly at least one).

The reason this is good is, I want to iron out the imperfections. So I want to suppose I had an edge inside B. I'll look at its two endpoints; each of these guys (by a symmetric argument) has a ton of neighbors in A, certainly at least one (and I can make sure those two are different neighbors). Now, these two guys in A have at least one common neighbor in B. And that's a 5-cycle.

So once you get this (plus a symmetric statement), this implies there are no edges in B, because otherwise there would be a C_5. By the same argument, there's also no edges in A. So now we're actually done — now I have genuinely a bipartite graph; if it's an extremal graph it shouldn't have any missing edges and the sizes should be as equal as possible, so we learn $G = T_2(n)$.

This is how all stability arguments work. You start by showing your graph is very 'similar' to what you expect to be the extremizer. Then you do some ad hoc arguments to show any imperfections can't exist, so it must actually be equal to the suspected extremizer.

(You'll see more details in the notes, and there's a homework assigmnent.)

§7 July 17, 2025

We'll start Ramsey theory today. It's what Yuval loves (he also loves EGT, but this is his favorite).

The pithiest encapsulation of Ramsey theory is 'complete disorder is impossible.' This doesn't mean anything mathematical, but we'll see many instances. Basically, if you have some sort of structure, you can find within it some order, regardless of how unstructured the global thing may be. It's undergirded by theorems, but the general philosophy is very influential — in combinatorics, but also logic, geometry, functional analysis, CS, ... It's a general phenomenon that you want to understand something and it's super complicated, but you can zoom in on it and get something with a ton of structure.

But it'll take a while before we get there.

§7.1 Fermat's last theorem

Theorem 7.1 (Fermat's last theorem, Wiles)
For any integer $q \geq 3$, there do not exist nontrivial positive integer solutions to $x^q + y^q = z^q$.

Nontrivial just means there's always the trivial solution $x = y = z = 0$. (If we say positive, this is implied, but in any case...)

Hopefully we've heard of this. It's hard to prove, so we will not be proving it. Let's start with something easier.

Why $q = 3$? If we set $q = 2$, we get the Pythagoras solution, which does have nontrivial solutions — $(3, 4, 5)$, $(5, 12, 13)$, and so on.

Let's change this a bit — what about $x^2 + y^2 = 3z^2$? This equation does not have any nontrivial integer solutions! This is great, because it's easier than the FLT one.

Proposition 7.2
There do not exist nontrivial integer solutions to $x^2 + y^2 = 3z^2$.

There's really one main technique for proving something like this — which is to reduce mod something. For this equation, one thing you can do is reduce mod 3. If we had a solution in the integers, certainly it would also remain a solution mod 3. But mod 3, this equation just becomes

$$x^2 + y^2 \equiv 0 \pmod{3}.$$

And now you can just do a tiny case check to see there are no options — mod 3, we have $1^2 = 2^2 = 1 \pmod{3}$. So no matter what values they take mod 3, we'll have $x^2 \equiv y^2 \equiv 1$, so their sum will be 2 mod 3.

There's a tiny step where we cheated — we could have a nontrivial solution over the integers that becomes trivial mod 3. But there's an easy argument that shows that by considering the number of 3's dividing your numbers, you can deal with this.

This is a standard argument in number theory — reducing mod something. There's something called the Hasse principle that says in many situations this strategy is guaranteed to work — if there isn't a solution, you can witness it by reducing mod something.

So for a long time, people tried to prove FLT this way — by reducing mod something.

The theorem Yuval will tell us about (it was first proved by Dickson, but we won't talk about that; it was really proved by Schur 1916, the same one from representation theory):

Theorem 7.3 (Dickson, Schur 1916)
For any $q \geq 3$, there exists $N = N(q) \in \mathbb{N}$ such that for any prime $p > N$, there exist nonzero $x, y, z \in \mathbb{Z}/p\mathbb{Z}$ such that $x^q + y^q \equiv z^q \pmod{p}$.

This says that the approach we just sketched can't possibly work to prove FLT. We know the Fermat equation doesn't have nontrivial integer solutions, but it actually does have nontrivial solutions modulo any sufficiently large prime.

Student Question. *What if there's a prime under $N(q)$?*

Answer. This doesn't say for all primes you don't have nontrivial solutions; you do have to worry about this a bit. But the argument we said about common factors says you always might have some silly finite obstructions.

Dickson proved this with an insane casework proof involving the largest integers Yuval has ever seen in a math paper. Schur gave an extremely beautiful proof, which is what we'll see. Schur was really a great visionary. He was writing this paper at a time combinatorics did not really exist. And he said this theorem, despite being important in NT, follows almost immediately from a lemma which is really just combinatorial. It's amazing that he could see at that time that what appears to be a deep NT fact is actually just combinatorics.

Lemma 7.4 (Schur)

For any $q \geq 3$, there exists $N = N(q)$ such that if $[N] = \{1, \dots, N\}$ is q-colored, there exists a monochromatic solution to $x + y = z$.

In Ramsey theory, we're always talking about coloring things. In contrast with what we were doing with chromatic numbers, our coloring has no structure whatsoever — when we talk about coloring a set (or later the edges of a graph), a coloring is completely arbitrary. So a q-coloring of $[N]$ is just a function $\chi : [N] \to [q]$ (the colors don't mean anything, so we can just call them 1, ..., q — though we'll usually refer to them as red/blue/green).

What does monochromatic mean? Mono means 1, and chromatic means color. So it means they get the same color — i.e., there exists $x, y, z \in [N]$ such that $x + y = z$ and they all receive the same color, so $\chi(x) = \chi(y) = \chi(z)$. Again, what this says is, you give me some number of colors (e.g., 1000). And then as long as I make N large enough (e.g., a gajillion), no matter how you go around coloring them, I'm guaranteed to be able to find three numbers with the same color such that the first two sum to the third.

Yuval isn't going to show this right now, but the lemma actually implies the theorem (on $x^q + y^q \equiv z^q$); we're not doing it because it requires a tiny bit of group theory and elementary number theory, which we're not assumed to know. But we will prove this lemma.

Student Question. *When you say there exists N, you don't mean for all N sufficiently large?*

Answer. Once you get it for a fixed N, you also get it for all larger N, because you can just ignore the extra numbers. So they're equivalent.

Student Question. *Do we need $q \geq 3$?*

Answer. No, this lemma is true for $q \geq 1$. Actually even the original theorem is true for $q \geq 1$, but if $q \leq 2$ we have nontrivial integer solutions, so it's not interesting.

This is the 'complete disorder is impossible' phenomenon. The global thing that's unstructured is this completely arbitrary coloring of $[N]$ — your worst enemy comes and splits the numbers into buckets and you have no idea how they did this, but you're still able to find a solution to $x + y = z$ in the same bucket.

Schur gave a direct proof of this lemma. The way we think about this now is we view it as coming from a graph theory statement. So we'll say a graph theory statement that implies this.

Lemma 7.5

For every $q \geq 1$, there exists $N = N(q) \in \mathbb{N}$ such that in any q-coloring of the edges of K_N, there exists a monochromatic triangle.

So now I'm not coloring integers; I'm assigning the *edges* of the complete graph colors (in a completely arbitrary fashion). Schur didn't write it like this (because he didn't have graphs), but this is how we think about it now.

Proof. We'll prove that $N(q) = 3q!$ — no matter how you color the complete graph on that many vertices, you're guaranteed to find a monochromatic triangle.

The proof will be by induction on q. The base case is when $q = 1$; then the statement says I take the complete graph on 3 vertices and color the edges with one color, and I can find a triangle whose edges all receive the same color. That's a true statement.

For the inductive step, the trick is the following: Here's my universe of size $(3q)!$, where the edges are colored totally arbitrarily. I'm going to fix a single vertex v and look at the edges coming out of it. They come in potentially q different colors; so maybe there's some green ones, some orange ones, and some red ones. But the observation is by pigeonhole, there exists some color (let's call it orange) such that v has at least $\lceil (N-1)/q \rceil$ neighbors in that color. What's going on is that there's $N - 1$ additional vertices, and q different colors touching v, so one of these must be the largest — it must have size at least $(N-1)/q$ (and I get to round up, since this may not be an integer). So maybe there's this big orange neighborhood.

What is this? By definition, that's

$$\left\lceil \frac{3q! - 1}{q} \right\rceil = 3(q-1)!$$

(apart from the -1 it'd be exactly $3(q-1)!$, and the -1 goes away when we round up). So this set (the orange neighborhood) has size at least $3(q-1)!$.

Now we split into two cases. First, if there's an orange edge in there, then we're done — we get an orange triangle.

If not, then you can use the induction hypothesis — I've found $3(q-1)!$ vertices colored by at most $q-1$ colors (since I started with q colors, and orange doesn't appear there anymore). So by induction, I get a monochromatic triangle. □

What does this have to do with the number theory question? At some point, numbers have to come in. The point is this graph theoretic lemma implies Lemma 7.4. This is a general phenomenon in combinatorics — abstract combinatorial objects are abstract, so you can often encode in them things you care about in other areas of math. So we're going to prove Lemma 7.4 by carefully constructing a coloring.

Proof of Schur's lemma. We need to show there's some N; as we've set up the notation, we'll use the same N as in Lemma 7.5. Now what happens is our worst enemy picks a coloring of the integers 1 through N; so we fix some coloring $\chi : [N] \to [q]$ (which I have no control over). My goal is to find a monochromatic solution to $x + y = z$. And the only tool I have at my disposal is Lemma 7.5, so we want to construct a graph coloring somehow.

We'll construct the edges of K_N as follows: We'll identify the vertices of K_N with $[N]$ (so we name the vertices $1, \ldots, N$). And we'll color the edge ab (where $a < b$) by $\chi(b - a)$.

Example 7.6

We color the edge 13 green if $\chi(3 - 1) =$ green. In particular, the edge 24 will receive the same color, because $4 - 2$ is also 2, which is green.

Student Question. *Both the vertices and edge are colored?*

Answer. We're only coloring edges (and that'll be the case for everything we do). You're right that we started with no graph whatsoever, just a coloring of the integers. But we're using that to color the edges of this new graph. The point is that my worst enemy picked the original coloring — I have no control over it. But when I apply Lemma 7.5, I get to choose the coloring; so I can do something intelligent to exploit that.

Now we want to apply Lemma 7.5; so by the lemma, there exists some monochromatic triangle, say with vertices $a < b < c$. Then we'll let $x = b - a$, $y = c - b$, and $z = c - a$. The point is that I know $\chi(x) = \chi(y) = \chi(z)$, and the reason is that in the graph, these edges had the same color, and the edge ab had color $\chi(x)$, the edge bc had color $\chi(y)$, and the edge ac had color $\chi(z)$ (and I assumed it's a monochromatic triangle, so they must have had the same color). And $x + y = z$, because $(b - a) + (c - b) = (c - a)$. □

Student Question. *How does that imply the Fermat's last theorem thing?*

Answer. That's in the notes. You need a bit of group theory to do this, so we're not doing this in class.

Student Question. *Can we generalize it to more complicated equations like $x + y + z = w$?*

Answer. Yes, some things like this will be on the homework. Some equations this is quite easy to generalize; for some it's extremely difficult. There is a good reason. For some it's not true, so you need to be careful. There's a whole theory of this. Schur had a brilliant student called Richard Rado. Schur did only this one theorem on this topic, but he gave to Rado as his PhD thesis to figure out what's going on, and Rado proved the most general theorem about this type of thing.

§7.2 Ramsey's theorem

We're going to generalize this now, but in a different direction. You can generalize this in arithmetic Ramsey theory, where you ask these kinds of questions about arithmetic (solutions to equations). We're going to do graph Ramsey theory instead.

Theorem 7.7 (Ramsey 1929)

For all $k, q \geq 2$, there exists some $N = N(k, q) \in \mathbb{N}$ such that if the edges of K_N are q-colored, there exists a monochromatic K_k.

This is really the birth of Ramsey theory, but we now recognize Schur's theorem as an instance of Ramsey theory (from about a decade earlier).

Ramsey's theorem is the exact same statement as Lemma 7.5, except now I'm not guaranteed a monochromatic triangle, but rather a monochromatic larger thing (larger complete graph).

Ramsey theory often gives results like this; they say that for any whatever parameters you fix, there's some large N such that for any structure on N vertices (here a q-coloring of the edges), no matter how disordered it is, you can zoom in on some order. Lots of Ramsey theory is about qualitative statements like this, but in a lot of Ramsey theory we're concerned about quantitative questions — what is this N?

Definition 7.8. The Ramsey number $r(k; q)$ is the minimum N such that Ramsey's theorem holds.

There's a ton of notations for this; this is the one we'll use. So $r(k; q)$ is the smallest N I could take here to get a true statement. Ramsey's theorem tells us this number is well-defined. We'll care about its values.

One reason is many proofs (like the one we saw) are by some sort of induction, so it's nice to have an actual number.

We'll focus for the rest of this lecture on two colors (that captures most of the difficulty); when $q = 2$, we usually just write $r(k)$.

> **Student Question.** *Did Ramsey himself look at this number?*
>
> **Answer.** Most people at this conference would tell you no, but he actually did…

Ramsey himself proved $r(k) \le k!$. For his application, he didn't care about the actual value. But he remarks in the paper he's certain this bound is very far from the truth and feels it must be much smaller, but he's not able to find a better argument. And a few years later, the next big theorem: Erdős–Szekeres basically rediscovered Ramsey's theorem with a different proof, and they proved that:

Theorem 7.9 (Erdős–Szekeres 1935)
We have $r(k) < 4^k$.

This is a lot smaller (for large k), because it's exponential in k while $k!$ is super-exponential. We'll see this proof, because it's one of the nicest.

Erdős–Szekeres realized that to make the proof work nicely, it's useful to generalize the problem (so that induction works better):

> **Definition 7.10.** The off-diagonal Ramsey number $r(k, \ell)$ is the minimum N such that in any red/blue coloring of the edges of K_N, there exists a monochromatic red K_k or blue K_ℓ.

(Note that before we had a semicolon, while here we have a comma.)

So the definition is just that now I'm not looking for the same thing in red and blue, but a red K_k or blue K_ℓ. Of course, we have $r(k) = r(k, k)$ (since I am looking for the same thing in both colors). And also $r(k, \ell) = r(\ell, k)$, because I can swap the roles of the two colors.

Now we'll state more precisely what E–S proved:

Theorem 7.11 (E–S 1935)
We have $r(k, \ell) \le \binom{k+\ell-2}{k-1}$.

(Note that this is the same as $\binom{k+\ell-2}{\ell-1}$.) .

This implies the bound $r(k) < 4^k$ — if we plug in $k = \ell$, we get $\binom{2k-2}{k-1}$. That's the number of ways of picking something out of a universe of size $2k-2$, which is certainly at most $2^{2k-2} < 4^k$ (in fact, it's a bit better — you can divide by $\sqrt{k}$ or something).

So we'll prove E–S now, which will imply Ramsey's theorem (at least for two colors), since that just says there is some finite upper bound.

Proof of E–S. The proof will be a two-parameter induction. The base case is when $k = 1$ or $\ell = 1$. You maybe don't like that because it's not obvious what a monochromatic K_1 is — it has no edges at all. But the obvious thing to say is that any complete graph has a red K_1 and a blue K_1, because there are no edges. (If you dislike this, you can take the base case to be $k = 2$ or $\ell = 2$, which is also quite easy.)

For the inductive step, we're going to prove the key inequality

$$r(k, \ell) \le r(k-1, \ell) + r(k, \ell-1).$$

This recursion is the reason we wanted to go to the more general off-diagonal case — the inductive proof only finds a smaller clique in *one* of the colors. Once we prove this, we certainly will have proved *some* finite upper bound on Ramsey numbers (if I know by induction that the two terms on the RHS are finite, so is the LHS); and as we'll see, it will also give the specific E–S bound.

So let's prove this recursion. The proof is going to look extremely familiar. We draw our universe of size $N = r(k-1,\ell) + r(k,\ell-1)$. What does it mean to prove the LHS is at most the RHS? That means no matter how I color the edges of K_N, I'm guaranteed to find a monochromatic red K_k or blue K_ℓ. So I make my universe this size, and let my adversary pick an arbitrary coloring in red and blue. And then I fix my favorite vertex v and look at the edges coming out of it; there's some red ones and some blue ones.

Case 1 (v has at least $r(k-1,\ell)$ red neighbors). Then we're going to look at the coloring just among these neighbors; among them, there exists a red K_{k-1} or a blue K_ℓ (because I'm looking at a coloring on at least $r(k-1,\ell)$ vertices, and by the definition of this number, there's a red K_{k-1} or blue K_ℓ).

Case 1a (There's a blue K_ℓ). Then we're done, because that's one of the things I was looking for — my goal was to find a red K_k or blue K_ℓ, so if I find a blue K_ℓ I'm happy.

Case 1b (There's a red K_{k-1}). Now I can use the fact that we're working inside the red neighborhood of v! So if I find $k-1$ vertices in there making a red clique, v is adjacent in red to all of them; and so together with v, we get a red K_k. So we're again done.

Case 2 (v has at least $r(k,\ell-1)$ blue neighbors). The argument is completely symmetric. So if this happens, then again in that neighborhood I either find a red K_k (and I'm done), or a blue $K_{\ell-1}$ (and I can add v to that and be done).

So I'm good in both of these cases. Case 3 is that neither of those happens:

Case 3 (v has at most $r(k-1,\ell)-1$ red neighbors and at most $r(k,\ell-1)-1$ blue neighbors). We claim that you get a contradiction in this case. Why? The total number of neighbors is $N-1$, and by this we get

$$N - 1 \le (r(k-1,\ell)-1) + (r(k\ell-1)-1) = N-2.$$

So I've proved $N-1 \le N-2$, which is not possible; so this case can't happen.

So that proves the recursion. This is already good enough to prove Ramsey's theorem, at least for two colors — we've proved some finite upper bound. For completeness, let's see why it gives the E–S bound.

The point is that we've shown

$$r(k,\ell) \le r(k-1,\ell) + r(k,\ell-1) \le \binom{k-1+\ell-2}{k-2} + \binom{k+\ell-1-2}{k-1}$$

by induction. And now we just apply Pascal's identity; these two binomial coefficients have the same top, and their bottoms differ by 1, so this is equal to what we said it should be, which is $\binom{k+\ell-2}{k-1}$. □

So we've proved Erdős–Szekeres, and therefore also Ramsey's theorem for two colors. The proof is very simple, it's just induction; we did it in all this detail, but the basic idea is extraordinarily simple. I have a huge universe, I look at a vertex, it has either a bunch of red or blue neighbors. In the first case, I apply induction where I look for a clique one smaller in red; in the second I look for a clique one smaller in blue.

And somehow, Ramsey's theorem is a highly nonobvious statement — the fact that we can always find order in chaos is highly nonobvious — and the fact that the proof is easy shouldn't make us forget this.

There's a story that a Hungarian sociologist wanted to understand how cliquishness forms in small children. So he went to elementary schools and asked all the children who they're friends or not with. And he found that in every classroom, there's either 4 children who were pairwise friends, or 4 who were not. His first instinct was that this was amazing — he's found that cliquishness starts at a young age! But he asked some combinatorialists what's going on, and they said congratulations you've discovered Ramsey's theorem. And this has nothign to do with children! If you could design a classroom, it turns out once you have 18 children, you can't avoid 4 being friends or 4 being non-friends. When you phrase this in children it's surprising, but it's just Ramsey's theorem.

(His name was Salai. This is one of the stories that sounds obviously fake, but Yuval is told it's probably real.)

So far we've only proved Ramsey for 2 colors. There's two ways to get it for more colors. One is that you can easily reduce from more colors to 2 colors (this is on the homework), if you don't actually care about the quantitative bound. The other is that the same argument we just did gives that

$$r(k_1, \ldots, k_q) \leq \binom{k_1 + \cdots + k_q - q}{k_1 - 1, k_2 - 1, \ldots, k_q - 1}$$

(we haven't defined this notation, but it's the obvious off-diagonal generalization with more colors — we have q colors, and we're looking for K_{k_1} in the first color, K_{k_2} in the second, and so on). The important thing is that in the diagonal case, this gives $r(k; q) < q^{qk}$ (when $q = 2$ this matches 4^k). On the homework you'll also prove this.

In two minutes, we'll talk about lower bounds. To lower bound a Ramsey number, you want an example. Ramsey numbers are the smallest N so that no matter how you color you'll find what you're looking for. So for a lower bound, you need to give me a coloring that does not have the structure you're looking for.

Lemma 7.12

We have $r(k) > (k-1)^2$.

So that's at least some reasonable lower bound.

Proof. We'll use our good old buddy the Turán graph — we take $k-1$ blobs, each of size $k-1$. We put red inside each blob, and we put blue across.

There can't be a red K_k because it doesn't fit (the red bits have size $k-1$), and there can't be a blue K_k because it's the Turán graph (or by pigeonhole,) □

It's really difficult to come up with a better example — in the 1940s Turán was certain this was the right behavior. Then the 4^k upper bound would be really bad (quadratic vs. exponential). But...

Theorem 7.13 (Erdős 1947)

We have $r(k) > \sqrt{2}^k$.

This is way bigger — and it tells us the true value is exponential. This is one of Yuval's favorite proofs, and we'll see it tomorrow.

Since then, we had these two bounds $\sqrt{2}^k < r(k) < 4^k$, and there was basically no movement for 75 years. There was an incredible breakthrough (we may have seen in Rob Morris's cross-program talk). The original proof gave 3.993^k; the current record is 3.8^k. This was an incredible breakthrough, something lots of people thought we wouldn't live to see in our lifetimes. Yuval won't be teaching it in this course, but there's a survey on his website.

§8 July 18, 2025

§8.1 Lower bounds

Yesterday we defined the Ramsey number $r(k)$, the minimum N such that in any two-coloring of the edges of K_N, there exists a monochromatic copy of K_k. Yesterday we saw a few variants — the off-diagonal Ramsey number (looking for K_k in red or K_ℓ in blue) or the multicolor one (where we had more than two colors). But right now we'll stick with this one. Last class we proved $r(k) < 4^k$ (Erdős–Szekeres). We also saw the pretty easy lower bound $r(k) < (k-1)^2$. The reason we did it quickly is that it's super far from the truth. People thought it might be close to the truth, but it's really not:

Theorem 8.1 (Erdős 1947)
We have $r(k) \geq 2^{k/2}$.

The way you prove lower bounds on Ramsey numbers is give examples — if you give me a coloring on some number of vertices with no monochromatic K_k, you've convinced me the Ramsey number is larger than that number of vertices. It's actually very difficult to come up with a construction that does better than the $(k-1)^2$ one. Erdős's main innovation is that you don't have to — you can prove this bound without giving an explicit construction!

Proof. Let $N = 2^{k/2}$ (this may not be an integer, but we won't care — you fix this by inserting floor and ceiling signs, but you can always fix it that way, so we won't worry about it; if you're writing a paper you should, but if you're just learning about things you shoudln't, because it doesn't matter). Our goal is to find a coloring on N vertices with no monochromatic K_k. (That should prove $r(k) > N$; the reason we're just claiming $\geq$ is because of this stupid rounding issue, but it doesn't matter.)

So we want to color K_N red and blue. We'll do this randomly! So for every edge we flip a coin; if it's heads we color it red, if tails we color it blue.

The point is, Erdős is just saying, let's try random and see what happens (instead of saying I'll color this edge red and that one blue, I just flip a coin).

First, for any given set of k vertices, the probabilty that they are a monochromatic K_k is $2^{1-\binom{k}{2}}$. Why? You have $\binom{k}{2}$ edges, and they all have to come out the same way; and they could all come out red or all blue, which explains the extra 2^1.

Our goal is we hope to end up with no monochromatic K_k's at all. So our goal would ideally be to find the probability there exists *any* monochromatic K_k. But it's super hard to determine this precisely — different sets of k vertices will interact with others in complicated ways. So it's difficult to write down an actual formula. Instead, we'll be super lazy and upper-bound this by

$$\mathbb{P}[\text{exists monochromatic } K_k] \leq \binom{N}{k} 2^{1-\binom{k}{2}}.$$

Here we used the union bound, the most important thing in probability:

Fact 8.2 (Union bound) — We have $\mathbb{P}[A \text{ or } B] \leq \mathbb{P}[A] + \mathbb{P}[B]$.

We only get an inequality because maybe sometimes both happen, and I've overcounted that; but certainly the inequality is true. In this computation, we applied the union bound $\binom{N}{k}$ times. What's the probability there's some monochromatic K_k? That's the probability that this set is a monochromatic K_k or that set is or that set is or so on. There's $\binom{N}{k}$ such sets, so I'm adding $\binom{N}{k}$ things; and for each of those sets, the probability is exactly $2^{1-\binom{k}{2}}$. (This is only an inequality; the actual value is a bit smaller because of this interactions, but we won't need that.)

Our goal is now to estimate this quantity. We'll be a bit lazy about this — we have $\binom{N}{k} \leq N^k/k! \leq N^k$. (This is being quite lazy, but. . .) For the other term, we'll first write something false, that $2^{1-\binom{k}{2}} \leq 2^{-k^2/2}$. That's not true, but it's almost true — $\binom{k}{2} = k^2/2 - k/2$. So this is really $2^{1+k/2-k^2/2}$. This is not less than $2^{-k^2/2}$ because we have positive things in the exponent. But it is true that $2^{-k^2/2}$ is the important term — when k is large, it completely swamps the other stuff. And it is true that

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

(The notes have a formal derivation of this, but the point is that $\binom{N}{k} < N^k$, and $2^{1-\binom{k}{2}}$ isn't smaller than $2^{-k^2/2}$, but they're small enough that the $k!$ you threw away in the first inequality can handle that — $k!$ is super-exponential in k, while these extra terms here are exponential in k.)

And we picked $N = 2^{k/2}$, so this equals

$$(2^{k/2})^k 2^{-k^2/2} = 1.$$

So we've just proven that

$$\mathbb{P}[\text{exists a monochromatic } K_k] < 1.$$

Why is this good news? Our goal is to find a specific coloring (we don't want a random coloring, we want a real one). But we've shown if we pick one at random, the probability it's bad (that it has a monochromatic K_k) is strictly less than 1; equivalently,

$$\mathbb{P}[\text{does not exist a monochromatic } K_k] > 0.$$

If I have any random whatever and the probability it's good is positive, there must be a good whatever (if there were not a good whatever, the probability the random one is good would be 0). So as a consequence, there exists a specific coloring of the edges of K_N without a monochromatic K_k. And that's exactly what we wanted to prove. □

Student Question. *You can't do meaningfully better by not just applying the union bound and being more rigorous with trying to find a better one?*

Answer. No. On the homework you have an exercise to fully optimize this argument and get the best thing, but still there the union bound is the best thing to do. The intuitive reason why the union bound is fine is that because N is huge, most k-sets don't interact with each other at all; so even though we are throwing away something, it's really negligible.

Student Question. *Do we know what the probability actually is?*

Answer. In fact, if you're a little more careful, you'll see that the probability there is a monochromatic K_k tends to 0 very quickly. So if you pick a random coloring on this many vertices, it is overwhelmingly likely to work (for $k = 10$, already 99.999% of colorings work). This brings us very nicely to what we'll say next. . .

A major open problem:

Question 8.3 (Erdős). Find an explicit coloring on 1.0001^k vertices with no monochromatic K_k.

This is nuts. To stress how crazy this is, first Erdős invented this incredible magic trick (this is the origin of the probabilistic method in combinatorics, now extremely important). I have no idea how to construct these things, so I'll pick one at random and show with positive probability it works. That's the magic trick. But in fact, on this number of vertices, the overwhelming majority of colorings do work. So if you pick one at random, the probability you'll fail is less than 1 over the number of atoms in the universe. Nonetheless, we don't know how to describe such a coloring — not just on $2^{k/2}$ vertices, but on anything growing exponentially in k. This is what some people call 'finding hay in a haystack.' The joke is usually it's hard to find a needle in a haystack. Here you have a ton of hay, you know it's almost all hay, and everytime you reach in and grab something a needle comes out. This is very frustrating, but not unique to this problem — there are many situations we have probabilistic ways of constructing things (we know they exist and even that almost all things work), but we can't find one.

There was a major breakthrough recently:

Theorem 8.4 (Li 2023)
There is an explicit construction on $2^{k^{0.0001}}$ vertices.

This is not truly exponential in k, but it's exponential in some power of k, which is the current state of the art. There's an annoying thing, which is that 'explicit' does not mean 'easy to describe.' Li is a computer scientist; explicit means there's a deterministic algorithm that produces the coloring.

Student Question. *Why doesn't something stupid like reading off the digits of π and using that to produce a pseudorandom coloring work?*

Answer. To prove that you'd have to prove something about the digits of π. It's probably true, but we can't prove it.

Yuval was planning to move on after this point, but literally yesterday there was a breakthrough posted on arxiv. The same argument as Erdős's gives some specific lower bound for $r(k, ck)$ for any fixed $c \geq 1$. The point is, suppose I want an off-diagonal Ramsey number of k vs. $2k$, or k vs. $1.1k$. There's an obvious random thing you can do, which is you do the same thing. But to optimize the bound, you shouldn't color red and blue with equal probability. There's some optimization to do, but it's not hard — based on c, there's a specific probability you pick, and that'll give you some bound, which until yesterday was the best bound. But very recently:

Theorem 8.5 (Ma–Shen–Xie)
You can do a little better for any $c > 1$.

So they have an exponential lower bound where the base of the exponent is a tiny better than what you'd get from optimizing this probabilistic thing. Very annoyingly their proof does not give an improvement for $c = 1$; but for any $c > 1$, they're able to get a better lower bound than what was previously known.

It is probabilistic, but not completely independent. Instead they work in a high-dimensional space, randomly place points on a sphere, and color red or blue depending on whether they're close or far on the sphere. Analyzing this is difficult (you have to do lots of high-dimensional geometry), so the paper is 40 pages long with lots of integrals.

§8.2 Hypergraph Ramsey

Now we're going to shift gears a little bit and talk about hypergraphs.

Theorem 8.6 (Ramsey)
For any $k \geq t \geq 2$ and $q \geq 2$, there exists some N such that no matter how we q-color the edges of the complete t-uniform hypergraph on n vertices, there exists a monochromatic $K_k^{(t)}$.

So it's the same statement, just for hypergraphs. Now I'm working with some uniformity t (previously we had $t = 2$), and I'm coloring the edges of the complete t-uniform hypergraph (so if $t = 3$, every triple receives a color). And I'm claiming I can find a monochromatic clique of size k, i.e., k vertices where all t-tuples within them receive the same color.

So let's prove this. We'll only prove it for $q = 2$, from which (as you saw on yesterday's homework) you can easily derive the general case (and it captures most of the difficulty). We'll follow our nose as much as possible from the proof for graphs. We'll also go through the off-diagonal case:

Definition 8.7. We define $r_t(k, \ell)$ as the minimum N such that we can always find a red $K_k^{(t)}$ or blue $K_\ell^{(t)}$.

We'll prove by induction that this number is finite (that's what Ramsey's theorem is saying). The key claim is that:

Claim 8.8 — We have $r_t(k, \ell) \leq r_{t-1}(r_t(k-1, \ell), r_t(k, \ell-1)) + 1$.

Before we had a joint induction on k and ℓ, where we decremented either by one. Now we're still doing that, but also built into that is an induction on t. The key claim is that what I'm trying to prove is finite is upper-bounded by a bunch of stuff that's also finite. As yesterday, we had $r_t(k-1, \ell)$ and $r_t(k, \ell-1)$. Yesterday we just added them. Today we have to do something much worse, which is to take r_{t-1} of those numbers. But we know these two numbers are finite, and r_{t-1} of anything is finite; so this does prove the thing is finite.

The crazy formula is crazy, but where it comes from is the same as yesterday:

Proof. As a picture, let's pretend $t = 3$. I want to prove that in any coloring of the t-uniform hypergraph on N vertices, I can find a red $K_k^{(t)}$ or blue $K_\ell^{(t)}$. As before, we'll look at a fixed vertex v. Before we looked at the edges coming off v; now they'll be t-sets. So if $t = 3$, I have some triangles coming off v in red, and some in blue.

Yesterday we looked at the red neighborhood and the blue neighborhood; now that doesn't make any sense. But what does make sense is that if I ignore v, each of these red triangles containing v looks like a single red edge (in uniformity 2). So rather than looking at the neighborhood, I can look at an auxiliary hypergraph of uniformity one lower.

Let's write that down: The coloring on the t-hyperedges containing v (we're working in uniformity t, so the hyperedges have size t) naturally yields a coloring of the edges of $K_{N-1}^{(t-1)}$, the complete hypergraph on all the vertices other than v in one lower uniformity. In the picture, this means we only look at the triangles containing v. Then we delete v, and now they look like edges. So I've just colored all the edges not containing v. And in general I'd have t-sets containing v, and if I ignore v in all of them I get $(t-1)$-sets.

Now what's the point? What is N? Well, N was some crazy thing plus 1, so $N - 1$ is that crazy thing. S by induction, this coloring contains a red something or a blue something. And the red something is a $(t-1)$-uniform clique of size $r_t(k-1, \ell)$, and the blue something is a $(t-1)$-uniform clique of size $r_t(k, \ell-1)$. So I've found either $r_t(k-1, \ell)$ vertices where all the $(t-1)$-hyperedges are red, or likewise with blue.

And if I remember what that means in the original thing, in the first case, that means all the t-hyperedges containing v and $t-1$ vertices from this set of size $r_t(k-1, \ell)$ are red. So in my picture (where $t = 3$), what this looks like is that I've found a set of size $r_t(k-1,\ell)$ with the property that every time I take two vertices from there as well as v, no matter how I do this, all those triangles are red.

Now what do we do? This is another Ramsey number, so I apply Ramsey again, and in there I get a red $K_{k-1}^{(t)}$ or a blue $K_\ell^{(t)}$. So now I'm looking just inside this set and remembering the original coloring of the t-hyperedges that was there, and in that thing by induction I either find a blue $K_\ell^{(t)}$ (in which case I'm done because that was one of the things I was looking for), or I find a red $K_{k-1}^{(t)}$, in which case I'm again done because I can add v to that and get a red $K_k^{(t)}$.

And the other case is symmetric (we swap the roles of red/blue and k/ℓ, and everything is the same), and that concludes the proof. □

Student Question. *When we got a coloring on $K_{N-1}^{(t-1)}$, what do we do with the edges that don't contain v?*

Answer. We don't — we just forget about them. In the picture with triangles, all the triangles were colored, but now I'm only trying to color single edges. So the way I color any single edge is I add v back to it and look at the color of that triangle in the original.

Student Question. *Can you do induction on the 1-uniform case and recover the proof from yesterday?*

Answer. Yes — this looks very similar to what we did yesterday, and in fact you can combine them (starting the induction at 1).

Let's spend five minutes talking about big numbers. If you think about what this recursion gives, in the 2-uniform case it gave an exponential bound — every time we decremented k or ℓ and added those, so roughly at each step of the recursion, we were doubling our bound.

If you think even about $t=3$, one step of this recursion tells you

$$r_3(k,k) \le r_2(r_3(k-1,k), r_3(k,k-1)) + 1.$$

(The $+1$ won't matter, as we'll soon see.) Plugging in our 2-uniform bound, this is at most $4^{r_3(k-1,k)}$. (You can use 3.8, but that also won't matter.) So the recursion means I can decrease one of the numbers, at the expense of an exponential. That tells you $r_3(k) < 4^{4^{4^{\cdot^{\cdot^{\cdot}}}}}$, where the height of this tower is $2k$. That's very bad. And that's just 3-uniform. In 4-uniform, each step of the recursion is going to cost you one of these things. So $r_4(k) < 2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}$ where the height of the tower is itself $2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}$ where so on, where the number of times I do this is $2k$. (It doesn't matter whether we write 4 vs. 2.)

You can ask whether these terrible bounds are necessary. Your first reaction might be yes — in the 2-uniform case exponential was the truth, even though our instinct was quadratic. But it's actually not necessary. We won't see it, but there's a completely different proof by Erdős–Rado (which we won't see), which does the induction in a different order; and it basically says

$$r_t(k) \le 2^{c r_{t-1}(k)}.$$

(There's some other junk that doesn't matter.)

Here, every time we increased uniformity by 1, we had to crazily iterate the bound from the previous uniformity, which gave us something terrible. Here, every time you increase the uniformity by 1, you only have to do one exponential, which is way better. This gives $r_3(k) \le 2^{2^{ck}}$ and $r_4(k) \le 2^{2^{2^{ck}}}$, which is way more reasonable.

Now you can ask whether *these* bounds are tight. The answer is that no one really knows. Here's something that is true. There's another amazing theorem due to Erdős–Hajnal–Rado which basically says

$$r_t(k) \ge 2^{c r_{t-1}(k)}.$$

So they get a completely matching lower bound, which is awesome — we can also go up a single exponential at every uniformity. But unfortunately this only works when $t \geq 4$. And for 3-uniform, the best-known bound is just $r_3(k) \geq 2^{ck^2}$ (which follows just from a random argument).

So we have this annoying situation where for 3-uniform we have a single-exponential lower bound and double-exponential upper bound. From there, we can step up — at uniformity t we have an upper bound that's a tower of height t, and a lower bound of height $t-1$. This persists all the way up; but if you closed this bound at any uniformity, you'd solve it for all uniformities, because of this stepping up.

This is a super open problem. Most people believe $r_3(k)$ should be double-exponential. Why? We do know that if you let me use four colors, then that is true — Hajnal had an amazing argument that got a double-exponential bound

$$r_3(k; 4) \geq 2^{2^{ck}}.$$

And all the stepping up and down business does work for any number of colors. So for four colors and above, we do know the right tower heights; it's just two and three colors for which we don't.

§8.3 Points in convex position

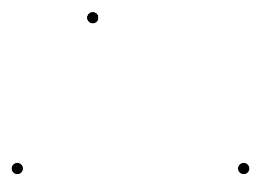
Why did we go to hypergraphs? It's a cool story in its own right, but there's a beautiful application (which was sketched in Rob Morris's talk, but we'll do it in more detail).

> **Fact 8.9** (Klein) — Among any five points in the plane, no three collinear, there exist four in convex position.

Convex position means they form the vertices of a convex polygon.

Proof. We have five points, so let's look at their convex hull (the polygon they define). If the polygon they define has 4 or 5 vertices, we're certainly done (then 4 of those points will automatically be 4 points in convex position). So the only remaining case is when it's a triangle with two points in the middle.

In this case, we consider the line through those two inner points. By the pigeonhole principle, two of the points of the triangle are on the same side of this line; and those two points, together with the two inner points, form a convex quadrilateral. □



Esther Klein is also the same person who came up with the C_4-free construction we saw a week ago. his was when she was in college, and so was Erdős and Szekeres; they'd meet at some big statue in Budapest and walk through the woods and talk math. She asked the following natural generalization:

> **Question 8.10** (Klein). Is it the case that for every k there exists N such that among any N points in the plane with no three collinear, there exist k in convex position?

For example, what if instead of finding four points in convex position, I want ten? Is it the case that if I start with enough, I can find it?

Theorem 8.11 (Erdős–Szekeres)
The answer is yes.

In fact, the paper they gave the E–S bound on Ramsey numbers was motivated by this question. This is a Ramsey-theoretic statement — we have some totally chaotic thing and can find within it order (k points in convex position). And the proof uses hypergraph Ramsey.

They gave two proofs, but we'll only see one.

Proof. We claim that $N = r_4(5, k)$ (the 4-uniform Ramsey number of 5 vs. k) suffices.

To show this works, we fix N points in the plane, with no three collinear. And we want to define some coloring of a hypergraph. Here's how we do it: We color a 4-tuple *abcd red* if they're not in convex position, and *blue* if they are in convex position. (My goal is to define a coloring of a 4-uniform hypergraph, so I need to tell you the color of any 4-tuple; and my rule is that I color them red if the four points are not in convex position, and blue if they are.)

What do I do now? Well, we've sort of set it up so that there's only one thing we could do. I defined N to be some Ramsey number and gave a coloring, so I certainly want to apply the definition of a Ramsey number. So by definition, there exists a red $K_5^{(4)}$ or a blue $K_k^{(4)}$.

We claim the first case, a red $K_5^{(4)}$, can't happen. Why? Then you get a counterexample to Klein's observation! This would mean 5 points where all 4-tuples are non-convex. And we just proved that that's impossible.

So the other one must happen — we do have a blue $K_k^{(4)}$. So now we just need to argue that this finishes the job. Clearly we want to look at these k points and hope they'll all be in convex position. So the claim is that all k of them are in convex position. Why? Suppose not. If not, that means some of them are on the outside (the convex hull), but at least one of them is somewhere on the inside. And the cool trick is now, let's draw the polygon (using the convex hull) and triangulate it arbitrarily. This point that was on the inside is in one of the triangles! And that means it, together with the vertices of the triangle, are four points not in convex position. So they're a red edge in my coloring. But I said this was supposed to be a blue $K_k^{(4)}$, so that's not possible. □

Student Question. *What if that inside point is on the edge of the triangle?*

Answer. We assumed there are no three collinear. (That's important — this is false if you allow collinear points, which is an exercise on the homework.)

How far is this off the true value?

Definition 8.12. We define the Klein number $\mathrm{Kl}(k)$ to be the smallest N for which this works.

For example, we know $\mathrm{Kl}(4) = 5$. On the homework, you can work out (with a lot of casework) that $\mathrm{Kl}(5) = 9$. It's known (with a lot of computer check) that $\mathrm{Kl}(6) = 17$. What's the pattern? These are not Ramsey numbers; they're always one more than a power of 2.

Conjecture 8.13 (Erdős–Szekeres) — For all k, we have $\mathrm{Kl}(k) = 2^{k-2} + 1$.

It's known in only these cases ($k \leq 6$). This conjecture they made in 1935. Thirty years later, they proved:

Theorem 8.14
We have $\mathrm{Kl}(k) \geq 2^{k-2} + 1$.

So they found a construction on 2^{k-2} points without k points in convex position.

The upper bound Yuval just showed us shows that

$$\mathrm{Kl}(k) \le r_4(5, k) \le 2^{2^{ck}}$$

(we showed r_4 is at most triple-exponential, but it turns out the off-diagonal is at most double).

There's another proof in the notes that shows $\mathrm{Kl}(k) \le r_3(k, k)$ (a diagonal 3-uniform Ramsey number). Unfortunately, this is also of the same order. The story is a mathematician was in an introductory class in combinatorics, the professor presented this proof, the mathematician wasn't at lecture that day; and on the exam they had to prove $\mathrm{Kl}(k)$ is finite. He hadn't been at lecture that day, and on the exam he invented a different proof, the one giving this. Unfortunately, it's an equally bad bound.

In the E–S, they gave a different (more geometric) proof that shows

$$\mathrm{Kl}(k) \le \binom{2k-4}{k-2} < 4^k.$$

This is way better — the conjecture is exponential in k, and certainly this is better than double-exponential (the previous one), but the base of the exponent doesn't match. This was a major open problem for a gajillion years. Ten years ago, Suk showed that $\mathrm{Kl}(k) \le 2^{k+o(k)}$. The exact formula is open, but we do now know the right base is 2.

The other story Yuval is required to tell us (it's. alaw of Ramsey theory) is that when Klein asked this question, Szekeres had a big crush on her and really wanted to solve this question; and after they proved this theorem, he and Klein actually got married. They moved to Australia (they were Jewish and the war was starting) and lived happily ever after. So Erdős always called this the happy ending problem.

§9 July 21, 2025

Today we'll start a new topic, which is graph Ramsey theory; this is maybe poorly named because you might think much of the Ramsey theory we've been doing is about graphs, but we'll see the reason.

§9.1 Graph Ramsey numbers

Definition 9.1. Let $H_1, \ldots, H_q$ be graphs. The Ramsey number $r(H_1, \ldots, H_q)$ is the minimum N such that every q-coloring of the edges of K_N contains a monochromatic copy of H_i in color i for some $i \in [q]$.

This is a mouthful, but very similar to what we've been doing so far. Before we had Ramsey numbers where we put integers $k_1, \ldots, k_q$; and these were the smallest N such that no matter how you color, you're guaranteed a monochromatic clique of the right size. This is the same thing, but instead of looking for a clique, we're looking for any graph we like (e.g., maybe we want a red Petersen graph or blue triangle or green K_{1007}).

Notation 9.2. If $H_1 = \cdots = H_q = H$, we write this as $r(H; q)$. If $q = 2$, we just write it as $r(H)$.

First, why is this well-defined? It's bounded above by the Ramsey number for cliques on those sizes:

Fact 9.3 — If $H_i \subseteq H_i'$ for all i, then $r(H_1, \ldots, H_q) \le r(H_1', \ldots, H_q')$.

This is because the right-hand side says that if I color that number of vertices, I'm guaranteed to find some monochromatic copy of H_i'; and once I have that, because H_i is a subgraph of that, I also have a monochromatic copy of H_i.

In particular, if H_i has k_i vertices, then

$$r(H_1, \ldots, H_q) \leq r(k_1, \ldots, k_q)$$

(where the right-hand sides is the notation we had before, where we're looking for a monochromatic clique in that color); the point is that $H_i \subseteq K_{k_i}$. And we know the RHS is finite, so the LHS is too.

Let's see some examples. We'll mostly be working with the two-color diagonal case $r(H)$. As a consequence of this observation:

Corollary 9.4

If H has n vertices, then $r(H) \leq r(K_n) \leq 4^n$.

So we get an exponential upper bound on $r(H)$. What about a lower bound? We have

$$r(H) \geq n,$$

since if you have fewer than n vertices, you can't possibly have H. More explicitly, what does it mean to prove $r(H) \geq n$? I need to give you a coloring on $n-1$ vertices with no monochromatic H. But any coloring works, since I can't fit a copy of H on anything with fewer than n vertices.

As a reminder, we know that for $H = K_n$, $r(K_n)$ really is exponential in n — we proved Erdős's lower bound $r(K_n) \geq 2^{n/2}$. So the base of the exponent is a huge open problem, but at least the correct behavior is exponential.

This makes the lower bound seem like trash (it's just linear). But for some graphs, this is an equality!

Example 9.5

If H has zero edges and n vertices (i.e., H is the empty graph), then $r(H) = n$.

Proof. We already proved $r(H) \geq n$. For the upper bound, I need to convince you every coloring of n vertices has a monochromatic copy of the graph with no edges; and it does, since there are no edges. □

In fact, you can boost this:

Example 9.6

If H has zero or one edges and n vertices, then $r(H) = n$.

This is kind of stupid, but it shows we can't hope to do better than $r(H) \geq n$ in general — certainly some graphs have $r(H) = n$. But the intuition you should have is that 'most' graphs have behavior more like exponential.

That's kind of right and kind of wrong. Let's look at some slightly more interesting examples.

§9.2 Ramsey numbers of trees

Theorem 9.7 (Erdős–Graham)
If T is an n-vertex tree, then $r(T) \leq 4n - 3$.

The specific bound of $4n - 3$ isn't super important (you can do a bit better); the important thing is this really is linear in n (so for trees, the answer really is linear and not exponential).

To prove this, we'll need two lemmas. One we already saw:

Lemma 9.8
If G is an N-vertex graph, it has a subgraph with minimum degree at least $e(G)/N$.

We saw this when proving extremal numbers for even cycles; it was also on the homework. (So we're not going to prove this.)

Lemma 9.9
Let T be an n-vertex tree. If G is a graph and has minimum degree at least $n - 1$, then $T \subseteq G$.

(This was also on the homework, but we'll prove it.) This says any graph of minimum degree at least $n - 1$ contains every n-vertex tree as a subgraph.

Proof. We use induction on n. The base case $n = 1$ is easy — it says if I have a 1-vertex tree then it's a subgraph of any graph with minimum degree 0, which is true (a 1-vertex tree is a single vertex, so it's a subgraph of anything).

For the induction step, we'll just draw a picture: Imagine we have our n-vertex tree, and let's pick our favorite leaf in this tree and forget about it for a moment. Now I have an $(n - 1)$-vertex tree. And G has minimum degree at least $n - 1 \geq n - 2$, so by induction, I can find that $(n - 1)$-vertex tree inside it. And now I look at the vertex where the leaf is attached. And only now I'm going to use my minimum degree assumption — this vertex in G has degree at least $n - 1$. And I've potentially already used up $n - 2$ vertices (I'm not allowed to touch all other vertices of the tree; and the tree has $n - 1$ vertices, one of which is this one). So it has at least one neighbor I haven't touched yet, and that gives what we want (a way to place this leaf). $\square$

Proof of Theorem 9.7. We need to prove an upper bound on $r(T)$. So in other words, let $N = 4n - 3$, and consider a 2-coloring (in red and blue) of the edges of K_N. I'm going to do something very lazy and say that without loss of generality, at least half the edges are red (I've colored the edges in two colors, so one hits at least half the edges; let's say it's red). Let G be the graph formed by the red edges. (My goal is to show one of the two colors contains a monochromatic T, and I'm saying let's just be super lazy and look at the color with more edges and try to find the tree in there; and the point is that's going to work.)

By Lemma 9.8, there is some subgraph $G' \subseteq G$ with minimum degree at least $e(G)/N$. So let's think about what $e(G)$ is — by assumption it's at least $\frac{1}{2}\binom{N}{2}$ (I colored the complete graph, and I'm saying half the edges are red), which is

$$\frac{1}{2} \cdot \frac{(4n-3)(4n-4)}{2} = (4n-3)(n-1) = N(n-1).$$

So G had $N(n - 1)$ edges, which means the minimum degree of G' is at least $e(G)/N = n - 1$.

Now I use Lemma 9.9, which implies T is a subgraph of G' (which is a subgraph of G).

And now I'm done — G was all the red edges, and I've found T as a subgraph of that, i.e., I've found a red copy of T. $\square$

§9.3 Complete bipartite graphs

Let's do another example. We've seen complete graphs have Ramsey numbers exponential in their number of vertices; trees have Ramsey numbers *linear* in their number of vertices. Let's see another example that interpolates between these.

Theorem 9.10 (Chvatál)
For all $s \le t$, we have $r(K_{s,t}) \le 2^{s+1}t$.

So now I'm looking at complete bipartite graphs, and I get a bound that's exponential in s and linear in t. That's interesting because it interpolates between the regimes we've seen. For example, if $s = t = n$ we get $r(K_{n,n}) \le 2^{n+1}n$, which is exponential in the number of vertices (which is $2n$). But despite being exponential, it's much better than the naive thing you'd get from plugging in $r(K_{n,n}) \le r(K_{2n}) \le 16^n$. So despite being exponential, it's still much better than what that gives us.

On the other hand, if s is fixed (e.g., $s = 17$) and t is large, then this says $r(K_{s,t}) \le C_s t \le C_s(s+t)$. So the point is that if s is fixed but t is large, then the behavior is again *linear* in the number of vertices.

So for these complete bipartite graphs — at least, this upper bound — if we're highly unbalanced then it's linear, while if we're doing balanced we get an exponential upper bound (and if $s = \sqrt{t}$, for example, you get something kind of in the middle).

Student Question. *Do we know how good this bound is?*

Answer. We'll talk about this in a bit. In the linear regime (with s is fixed), you can't really do better than linear, since there's always a linear lower bound. And later we'll see that for $r(K_{n,n})$, the truth really is exponential, though we don't know the base.

Proof. This proof is going to look a lot like the proof of KST. Let $N = 2^{s+1}t$, and fix a 2-coloring of the edges of K_N. Our goal is to find a monochromatic $K_{s,t}$ in there. The reason this is like KST is it's a different setup (I had a graph with many edges), but there my goal was also to find a $K_{s,t}$. (In fact, you could get a slightly weaker result by just invoking KST, but to get this exact bound we need to rerun the proof.)

So let S be the number of monochromatic $K_{1,s}$ in the coloring. (In the proof of KST we counted stars $K_{1,s}$; now we also want to do that, but since we're working with a coloring we might as well count all monochromatic ones.) First, we have

$$S = \sum_{v \in V(K_n)} \left(\binom{\deg_R(v)}{s} + \binom{\deg_B(v)}{s} \right)$$

(where $\deg_R(v)$ and $\deg_B(v)$ are the red and blue degrees of v). This is because we're counting these stars by first thinking about where the central vertex goes — we sum over all choices v and ask how many times it's the center of a monochromatic $K_{1,s}$. That $K_{1,s}$ can be either red, in which case we have $\binom{\deg_R(v)}{s}$ choices (we need to choose s red neighbors), or blue, which is similar.

Let's simplify this a bit — note that $\deg_R(v) + \deg_B(v) = N - 1$. So we're doing $\binom{x}{s} + \binom{N-1-x}{s}$; if you think about how to minimize this, it's smallest when $x = N - 1 - s$ (this is because the plot of $\binom{x}{s}$ is convex; this also follows from Jensen's inequality, applied with just two terms). (We're ignoring the sum over v, and just applying it on these two terms.) So by Jensen,

$$S \ge \sum_{v \in V(K_n)} 2\binom{(N-1)/2}{s}$$

(this is what would happen if $\deg_R(v)$ and $\deg_B(v)$ were equal). Now we can stop forgetting about the sum — we're just summing the same quantity N times, so

$$S \geq 2N\binom{(N-1)/2}{s}.$$

So I now have a lower bound on this quantity S.

Now we're going to prove an upper bound. Just as in KST, the lower bound came from summing over the vertex in the middle and using Jensen; the upper bound comes from assuming we don't have a $K_{s,t}$ and summing over the outer vertices.

So assume for contradiction that there does not exist a monochromatic $K_{s,t}$. That implies that for any s vertices, they are the outer vertices of strictly fewer than t red $K_{1,s}$, and strictly fewer than t blue $K_{1,s}$ (for the same reason as in the KST theorem — if I had s vertices that were the outer vertices of t red $K_{1,s}$, that would be a red $K_{s,t}$, and similarly in blue). That implies

$$S < \sum_{u_1,\dots,u_s}(t+t) = 2t\binom{N}{s}$$

(where we're summing over distinct vertices — I'm summing over all choices for the outer vertices, and then given where I've placed those, they're the outer vertices of strictly fewer than t red $K_{1,s}$ and strictly fewer than t blue $K_{1,s}$).

So now I have a lower and upper bound on S, so we just compare them and try to get a contradiction. Putting these together, we have

$$2N\binom{(N-1)/2}{s} \leq S < 2t\binom{N}{s}.$$

First, the 2's cancel, which is nice. We again have a completely formal derivation in the notes, but we'll just vibe this one out (the point of doing this in class is if you want to learn how to think about the asymptotics, it's much better to learn to think about how big they should be than to do the formal derivations of how big they literally are). The point is that $\binom{x}{s} \approx \frac{x^s}{s!}$. So this is roughly the same as

$$N\frac{((N-1)/2)^s}{s!} < t\cdot\frac{N^s}{s!}.$$

(Again, these do not literally imply each other, but this is how you should think about it, and once you do that you can check that it actually works.) Now more things cancel — we can forget about the $s!$'s. Also let's continue vibing and ignore the -1 (N is large, so it doesn't really matter). Now things are looking really good — we get

$$\frac{N^{s+1}}{2^s} < t\cdot N^s.$$

So a lot cancels, and I get $N < 2^s t$. That's a contradiction, because I said $N = 2^{s+1}t$, which is larger. And the point is because I gave myself this extra room with a factor of 2, the fact that a bunch of stuff I said is not literally true is fine (all that stuff can be absorbed into the extra factor of 2). So we really do get a contradiction (in vibes but also in truth). □

Student Question. *Can you optimize that factor of 2?*

Answer. Yes, really the truth is $2^s t$ plus a lower-order term (we're just being lazy and saying the lower-order term is at most the main term).

§9.4 Ramsey numbers vs. average degree

We've seen a bunch of examples. We saw sometimes the Ramsey number is exponential, sometimes it's linear, and sometimes it's in the middle. Does anyone have a sense of what's the reason? What's different between K_n and trees that makes it so that $r(K_n)$ has large Ramsey number, and trees have small Ramsey number?

The difference appears to be how many edges these things have — K_n has the most possible edges, trees have very few; and complete bipartite edges nicely interpolate (in $K_{n,n}$ we have half the possible edges, which is a lot; but in $K_{7,t}$ we have only $7t$ edges out of t^2 possible ones). So a natural guess is that edge density has something to do with it.

It's a bit easier to think about it in terms of average degree (which is roughly the same, since the average degree is $2e(G)/n$). So a natural guess is that you have high average degree if and only if you have big Ramsey number. Let's explore this idea. It turns out to be pretty close to right. Here's one version which is true:

Proposition 9.11

If H has average degree d, then $r(H) > 2^{(d-1)/2}$.

This tells you your Ramsey number is at least exponential in your average degree. In particular, it answers the question from earlier about how good the $K_{s,t}$ bounds were — it's giving you an upper bound that's exponential in the average degree (for $K_{s,t}$, your Ramsey number is basically s). So that gave an upper bound that's basically exponential in the average degree, and this gives a lower bound (the bases don't match, but at least both are exponential). It's consistent with trees (which have average degree 2, so this gives nothing — we know linear is the truth, so we shouldn't expect to get something good).

Student Question. *What about a cycle?*

Answer. A cycle also has average degree 2, so this gives a lower bound of 2, which is not good. The truth is some constant times its length (and it turns out whether it's an even or odd cycle makes a difference — for even cycles the constant is like $\frac{4}{3}$, and for odd cycles it's 2).

Student Question. *Are there things in the middle, which are e.g. quadratic?*

Answer. Yes, and we'll get to that soon.

This also recovers Erdős's lower bound from Friday (up to the -1) — K_n has average degree $n-1$. We said at that time that we don't have any idea how to prove that except by a random coloring, so naturally the proof of this will also be a random coloring.

Proof. Let H have n vertices, so $e(H) = dn/2$. Now let $N = 2^{(d-1)/2}$ (again if you don't like that that's not an integer, stop...). We'll randomly color the edges of K_N. We claim that

$$\mathbb{P}[\text{exists monochromatic } H] \leq n!\binom{N}{n}2^{1-dn/2}.$$

Let's walk through this slowly; it's very similar to what we did on Friday. First, for any fixed set of n vertices and any fixed assignment of how they might be a copy of H, I have $dn/2$ edges of H, and they all need to agree; the probability that happens is $2 \cdot 2^{-dn/2}$ (where the extra factor of 2 is because they could all be red or all be blue). So that's for a single copy of H.

Then $\binom{N}{n}$ is because I have N vertices and need to pick n out of them for a copy of H. Finally, $n!$ wasn't present when we did the computation on Friday. The difference is cliques are highly symmetric (all the vertices look the same). If H is arbitrary, in addition to fixing the vertices it lives on, I also need to look at all ways H could live on them. There's at most $n!$ ways of doing this (I can label the vertices of H from 1 through n and say this one's the first, that one's the second, and so on). And the inequality is the union bound. (I could be overcounting — maybe H has lots of symmetries so $n!\binom{N}{n}$ is an overcount, and the union bound also overcounts; but that's fine, since we always get an upper bound).

Now for once we'll do a true computation and not cheat — we have

$$n!\binom{N}{n}2^{1-dn/2} < n!\,\frac{N^n}{n!}2^{1-dn/2} = N^n 2^{1-dn/2} = \left(2^{(d-1)/2}\right)^n 2^{1-dn/2} \leq 1$$

(the main term is the $dn/2$ in the exponent, which we get with a positive sign in the first thing and negative sign in the second; we have an extra $+1$ in the second term but $-n/2$ in the first). (We can assume H has at least two vertices; otherwise the theorem is just true.)

And why are we done? We picked a random coloring and showed that the probability it has a monochromatic H is strictly less than 1, so the probability it doesn't is strictly larger than 0; therefore there must exist such a coloring (even though we have no idea what it looks like). □

Student Question. *When we did it before, and we didn't have the $n!$, you pointed out these events are almost independent so it's a good bound. But here, is the $n!$ necessary?*

Answer. If H is totally unstructured, you probably need the $n!$. But luckily it doesn't matter — the only difference is this -1 in the exponent, which is a constant factor up front (so we won't be bothered by things like this).

Student Question. *This probability is super small, right?*

Answer. Yes, if you take N to be even 1 or 2 bigger than this, this probability is unbelievably close to 0. So again it's the finding hay in a haystack situation where we know almost all colorings are good, but it's very hard to find one.

Our goal is to try to understand when graphs have very large vs. small Ramsey number. And we seem to be in business — our guess was that it has to do with average degree, and we've validated at least one direction (if you have large average degree, you have large Ramsey number). Now we'd like to do the other direction. But that's not true!

Proposition 9.12

For every n, there exists H on n vertices with average degree at most 1 and with $r(H) \geq 2^{\sqrt{n}/2}$.

That's really bad — it has really tiny average degree (less than a tree), and its Ramsey number is really big (not exponential in n, but exponential in $\sqrt{n}$). That's bad — it tells us one direction of the implication is true, but the other is super false.

This is maybe a surprising result, but the proof is one line.

Proof. Let H be a complete graph on $\sqrt{n}$ vertices, plus $n-\sqrt{n}$ isolated vertices (we're not going to worry that $\sqrt{n}$ may not be an integer). This has $\binom{\sqrt{n}2}{\leq}n/2$ edges, so its average degree is at most 1. And

$$r(H) \geq r(K_{\sqrt{n}}) \geq 2^{\sqrt{n}/2}.$$ □

§9.5 Degeneracy vs. large Ramsey numbers

Student Question. *Can we fix this by requiring connectedness?*

Answer. No — it basically doesn't matter that these vertices are isolated, as opposed to e.g. being joined to the clique as a star. You could have a little clique of size $\sqrt{n}$, with everything else joined to it. Then the average degree would be 2 instead of 1, but we'd still get this bound.

So connectedness won't help us, but we can still try to fix this.

The intuition is that in this example, why did the average degree fail us? We made the average degree artificially small by hiding a super dense graph in there and adding junk. So we shouldn't just look at the graph as a whole, but all subgraphs, to make sure we haven't hidden something denser inside. There are lots of ways of formalizing this; we'll give the most useful.

Definition 9.13. The degeneracy of H is the maximum over all subgraphs $F \subseteq H$ of the minimum degree of F.

So I run over all subgraphs and look at the minimum degree, and I take the maximum of that, and that's the degeneracy. In particular, if I have high minimum degree I have high degeneracy (I can just take $F = H$). But in an example like the above one, this will let me isolate the dense spot and say that has high minimum degree, so the graph as a whole has high degeneracy.

Corollary 9.14

If H has degeneracy d, then $r(H) > 2^{(d-1)/2}$.

Proof. Let F be a subgraph with minimum degree at least d (that exists by the definition of degeneracy). Then the average degree of F is certainly also at least d. So by our earlier theorem, $r(F) \geq 2^{(d-1)/2}$; and $r(H) \geq r(F) \geq 2^{(d-1)/2}$ (since F is a subgraph of H). □

Student Question. *Would it be substantially better to look at the average degree of a subgraph instead of minimum?*

Answer. It turns out to be the same up to a factor of 2, so it doesn't really matter (at the level of roughness we're doing things at).

So we want to correct our guess to

$$\text{high degeneracy} \iff \text{big Ramsey number}.$$

We've already proved one direction. When we tried with average degree, the reverse implication was not true. But here it turns out that it is!

Conjecture 9.15 (Burr-Erdős) — Bounded degeneracy graphs have *linear* Ramsey number — i.e., for every d, there exists some C such that if H is an n-vertex d-degenerate graph, then $r(H) \leq Cn$.

Definition 9.16. We say H is d-degenerate if it has degeneracy at most d.

This is in some sense the reverse implication — if you have low degeneracy (specifically, bounded), then your Ramsey number is linear, which is as low as it could be.

This is true! It was a major breakthrough due to Lee about ten years ago.

Theorem 9.17 (Lee 2017)
The Burr–Erdős conjecture is true.

This was an enormous breakthrough; Lee built on a ton of existing techniques. Tomorrow we'll talk about one of the most important pieces of partial progress towards the conjecture from before.

There's two ways of phrasing the question. One is, how big is C as a function of d? The other is, what if I don't care about graphs of bounded degeneracy, but instead tell you the degeneracy is at most $\sqrt{n}$, or $(\log n)^{100}$? Unfortunately we don't know, but it's widely believed:

Conjecture 9.18 (Conlon–Fox–Sudakov) — If H has n vertices and degeneracy d, then

$$r(H) \le 2^{\Theta(d+\log n)}.$$

What does this mean? If d is super large (a lot bigger than $\log n$), then you might as well forget about the $\log n$ term, and it basically says your Ramsey number is exponential in the degeneracy (we proved a lower bound of that type; the conjecture is that there's also an upper bound). When $d \ll \log n$, this is $2^{C \log n}$, which is some polynomial in n.

But what this means is if I'm hiding some graph and you want a guess for its Ramsey number, and you have only two questions (and you're not allowed to directly ask its Ramsey number), you should ask what are its number of vertices and degeneracy; the rest of its structure shouldn't matter.

This is still a conjecture, but it's quite close — we have it with some extra factor of $\log d$ in the exponent. It's still quite difficult, but it's almost a theorem.

§10 July 22, 2025

Yesterday Yuval told us about the famous Burr-Erdős conjecture, now a theorem, that graphs of bounded degeneracy have linear Ramsey number:

Theorem 10.1 (Lee 2017)
For every d, there exists C such that if H is an n-vertex graph of degeneracy at most d, then $r(H) \le Cn$.

The point is, what does linear mean? Every n-vertex graph has Ramsey number at least n. So this says if you're sparse in the sense of having bounded degeneracy, then your Ramsey number is almost as small as it possibly could be (i.e., it's linear). For example, trees have degeneracy 1, so the $d = 1$ case is something we did in class; but it becomes much harder when d is larger.

§10.1 Ramsey numbers of bounded-degree graphs

We won't prove this, but we'll talk about important partial progress which for a while was the best progress in this direction, which is the same thing but for bounded *degree* instead of degeneracy.

Theorem 10.2 (Chvatál–Rödl–Szemerédi–Trotter)
For every Δ, there exists C such that if H is an n-vertex graph with maximum degree at most Δ, then $r(H) \le Cn$.

Why is this a weaker result? It's easy to see that if you have bounded degree, then you have bounded degeneracy (if you have degree at most Δ, all subgraphs have maximum degree at most Δ, and therefore also minimum degree at most Δ). So this is a stronger assumption, and therefore a weaker result. An example showing it's a strictly stronger assumption is a star — stars (like all trees) have degeneracy 1, but huge maximum degree.

Nonetheless, this is very important partial progress, and for a long time it was the best result in this direction.

This result is also very important for the proof technique they introduced — this was the first application of Szemerédi's regularity lemma in Ramsey theory (not the first application overall, but the first in Ramsey theory, and one of the first overall). There's no reason we should have heard of SRL, but it's arguably the most important development in 20th century graph theory. It's hugely important, but we won't say much about it in this course. It's extremely powerful and it took some time for people to realize how to use it. This was one of the major applications — it really made clear just how powerful the regularity approach is (once you have those tools it's not hard).

Unfortunately there are some downsides of SRL, and one is that it comes with absolutely abysmal bounds. This theorem says for every Δ there exists C; and that proof gave $C \leq 2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}$, where the height of the tower is something like $2^{100\Delta}$. We now know that proofs using the regularity lemma *necessarily* invoke bounds this bad.

On one hand, who cares — the theorem is very elegant as it is (bounded-degree graphs have linear Ramsey number). But it's natural to ask, is this really necessary? This motivated people to find other proofs. Now there are at least three genuinely different proofs. One is the regularity proof. This was also improved a lot — someone found a variant of the regularity proof where C is only $2^{2^{\Delta}}$, which is way better.

In another direction, there was a development of something called the dependent random choice technique, developed in the early 2000s (in this context, especially by Kostochka–Rödl and Kostochka–Sudakov). This gives you a new proof with better bounds. But DRC is very flexible, and that was basically the main ingredient in Lee's proof of Burr–Erdős — Lee was able to develop it further and push some aspects to get the full conjecture. So this was an important development not just because it got a better constant, but because it led to a full proof of the conjecture.

The third proof, which we'll talk about, is called the greedy embedding technique. In some sense it goes back all the way to work of Erdős–Hajnal from the 1970s, but in this context it's due to Graham–Rödl–Ruciński from 2000. One thing their approach did is give a really good bound — they proved $C \leq 2^{O(\Delta \log^2 \Delta)}$, which is a much better bound than the crazy one from SRL. They also proved a lower bound $C \geq 2^{c\Delta}$. So their technique is really powerful and gets you essentially the optimal dependence on Δ.

We won't talk about the actual dependence very much, but this theorem is important enough that we'll say something about the proof. Greedy embedding is the one easiest to explain the ideas of, but it's not that easy. So Yuval will give us a proof sketch using this greedy embedding technique. The word 'sketch' is really doing a lot of work here — this will not be anything like full details. Also this is almost certainly the hardest topic in the course, but hopefully we'll get a picture of what's going on.

One more thing about dependent random choice: If you did the exercise from sometime in the first week about extremal numbers of bipartite graphs with bounded degree on one side, if you want to learn what DRC is, go back and do that exercise — that's the simplest application of the DRC technique.

§10.2 Overview of greedy embedding

Now we'll give a proof sketch of this theorem using greedy embedding.

We fix our favorite H with n vertices and maximum degree at most Δ. Let C be some huge constant (in principle you have to eventually pick it, but we won't be doing that). Our goal is to prove an upper bound on $r(H)$, so let $N = Cn$ and fix a 2-coloring of the edges of K_N. Our goal is to find a monochromatic copy of H.

The genius of the greedy embedding technique is we're going to be super stubborn and try as hard as possible to find a red copy of H:

> **Goal 10.3.** Try super hard to find a red copy of H.

And we're just going to really blindly try to do this. Obviously we can't guarantee this will work — for example, maybe our adversary colored everything in blue and then we have no hope. But we don't care; we'll just try really hard to find something in red.

If we succeed, then of course we're happy. But we might fail — there's no reason to expect to succeed. And somehow, the genius (really going back to this work of Erdős–Hajnal) is that if we fail, we learn something about the blue graph. (Yuval will tell us what this means soon.)

(Somehow he feels he could've come up with the technique of 'try really hard to embed in red'; the real insight is understanding what it means to fail and why that helps you.)

So let's do it.

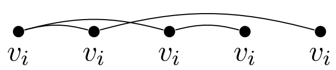
§10.3 An embedding algorithm

Let the vertices of H be $v_1, \ldots, v_n$ (we're just arbitrarily labelling them). We'll also define some sets — let $V_1 = \cdots = V_n = V(K_N)$. So far, we're just setting all these sets to $V(K_N)$. The point is we'll have an algorithm where at each step, I attempt to embed the next vertex in the list into my big K_N; and I'm trying to do this so that one by one I'm building a red copy of H. At the beginning I can put my vertices anywhere; but after I've made some choices, the places I can put future vertices becomes constrained. So these things V_i are going to be the 'candidate sets' for the vertices v_i — each index comes with its own candidate set V_i, which is where we're going to try to embed it. (So we'll constantly update the sets $V_1, \ldots, V_n$ as we proceed in the algorithm.)

Let's try to draw a picture of what happens. These sets $V_1, \ldots, V_n$ are not disjoint (we defined them to start out as literally the same set, so they can't be); but it makes more sense to draw them as though they were (so this is not a completely realistic picture). We have our sets $V_1, \ldots, V_n$ (drawn as n circles).

It's called the greedy embedding procedure, so I'm going to greedily try to embed the vertices one by one, starting with v_1. It has a candidate set V_1, so what if we try putting v_1 as some vertex in it?

What happens? My goal is to super stubbornly try to find a copy of H. Let's imagine H has only five vertices (with edges v_1v_2, v_1v_3, v_2v_5, v_3v_4). What happens? I'm asking myself, is this a good place to embed the vertex v_1? What does that mean? Well, if I choose to put v_1 there, this puts constraints on what I can do in the future. Concretely, it means v_2 can only ever go to the red neighborhood of this guy (I'm trying to find a red H, and $v_1 \sim v_2$). So if I put v_1 there, v_2 definitely has to live in its red neighborhood in V_2; and similarly v_3 has to live in its red neighborhood in V_3. But regardless of where I put v_1, that doesn't put any constraints on what I do with v_4 or v_5 (they're not adjacent to v_1).



So if I choose to put v_i somewhere, I must update the candidate sets v_j for all $j > i$ such that v_iv_j is an edge of H, to the red neighborhood. That's just saying that if I put v_1 over there, I must update V_2 and V_3 — I shrink them to be the red neighborhood of this vertex v_1.

Student Question. *Don't v_4 and v_5 need to be in the red neighborhoods of red neighborhoods?*

Answer. I've only placed v_1. It's possible this could rule out some vertices in V_4 and V_5, but we won't worry about that right now — we'll only look one step in the future, and when it comes time to pick v_2, *then* V_5 will have to shrink. But I'm only keeping track of candidate sets consistent with the choices I've made so far, and not thinking about the future.

We'll also fix some small constant $\varepsilon > 0$. What's the point of this? I need to decide, is this a good or bad place to put v_1? And I'll call it a good choice if these candidate sets don't shrink by too much. It's a bad choice if they shrink by too much. For example, if v_1 has no red neighbors in V_2, this would definitely be a bad choice (I'd get stuck). But if v_1 had only three red neighbors in V_2, that would also be bad (I want to maintain that at every step I have *many* choices).

Definition 10.4. We call a vertex $w \in V_j$ prolific if w has at least $\varepsilon |V_j|$ neighbors in V_j for all $j > i$ such that $v_i v_j \in E(H)$.

(If you're doing a formal proof you have to decide at some point what ε is, but we will not be doing that.)

One important thing is that as the process evolves, all these sets are shrinking. So vertices might have previously been prolific that stop being prolific. Conversely, a vertex that might have been non-prolific could become prolific (maybe it only had a tiny fraction of neighbors in the original universe, but now the universe has shrunk and that becomes a bigger fraction).

Now let's describe the algorithm.

Algorithm 10.5

We embed vertices one by one in order (starting with v_1, then v_2, then v_3, and so on). At step i:

- (Good case) If there is a prolific $w \in V_i$, embed v_i to w. (If there are many choices, pick one arbitrarily.) Update the sets V_j for all $j > i$ such that $v_i v_j \in E(H)$, to the red neighborhood.
- (Bad case) If there is no prolific w, output FAIL and stop.

So if at step i there's at least one prolific vertex, I pick one arbitrarily and say that's where I'm embedding v_i. That's the good situation — that I find a prolific vertex (if there's many, I'm even happier, but I just pick one.) And I update my V_js according to the rule — if I embed v_1 in some place, then its red neighborhood in V_2 becomes the new V_2, and the same with V_3 (and V_4 and V_5 stay unchanged.) That's the good step.

I keep running this as long as we can; and if I get to a point I can't do it, we just say we fail. (You could try to do something smarter like backtracking, but we won't; we'll just say I failed.)

Claim 10.6 — If we don't fail, we succeed (i.e., we find a red copy of H).

The point is that I set this up to stubbornly find a red copy of H; so if the algorithm doesn't fail, it works and really does find a red copy.

We said that if we succeed we're happy; if we fail, we're supposed to learn something. So what do we learn? Suppose we fail at step i. We have our set V_i, and we'll only draw the sets relevant to V_i, which are $V_{j_1}, \ldots, V_{j_\Delta}$ (where $v_{j_1}, \ldots, v_{j_\Delta}$ are the neighbors of v_i in H; here is one place I'm using the bounded maximum degree assumption to say v_i has at most Δ neighbors). (As we've drawn it here it has exactly Δ, but that doesn't matter; this will also be fine if it has fewer.)

And what do I know? I know I fail at this step, meaning that no matter what vertex $w \in V_i$ I try there, it's not prolific (since if there were some prolific vertex, I'd have not failed). And that means in one of these sets V_{j_k}, it has too few red neighbors — maybe it has way less than $\varepsilon |V_{j_1}|$ neighbors in V_{j_1}. Maybe another one also has few red neighbors in V_{j_1}, and another one has few red neighbors in V_{j_2}. But certainly I know that for every vertex in V_i, for at least *one* of these sets it has few red neighbors.

The point of this is that by pigeonhole, there exists some j such that at least $\frac{1}{\Delta}|V_i|$ vertices have fewer than $\varepsilon|V_j|$ red neighbors in V_j. So what I'm pigeonholing is that each of these vertices in V_i is not prolific for a reason, and that reason is one of these Δ other sets. I have at most Δ possible reasons, so at least a $1/\Delta$-fraction of my vertices must give the same reason. That says there's some j such that at least a $1/\Delta$-fraction of my vertices give the same reason V_j, meaning that they have few neighbors in that specific set.

Now, why is this good? My goal when I failed was to learn something about blue. And what I learned is — let's say this top portion of V_i is my $1/\Delta$-fraction of vertices which correspond to V_j. So every vertex in this top portion of V_i has few red neighbors in this second set V_j. In other words, I've found a pair (V_i', V_j) with red edge density strictly less than ε. What we mean by that is, this top portion is my V_i'. If I look between V_i' and V_j, at most an ε-fraction of the edges are red (since each $w \in V_i'$ sends at most an ε-fraction of red edges to V_j).

And here's where I learn something about blue: That's the same as saying the blue edge density is larger than $1 - \varepsilon$ (because every edge is red or blue, so if I have few red then I must have many blue).

There's one final observation: Throughout the process, each set shrinks by a factor of no more than ε^Δ. And the reason for this is, when does a set shrink? The only time a set shrinks is when I embed one of its neighbors, and then it shrunk by a factor of ε. And every vertex had at most Δ neighbors, so that happens at most Δ points. So every step, I shrink by a factor of ε; and I do that at most Δ times.

> **Student Question.** *Don't you need to delete the vertices you embed (from future candidate sets)?*
>
> **Answer.** Yes, we're kind of cheating here by drawing them as disjoint, and for a full proof you need to be careful with this. But it works out fine.

Here's the point of this:

> **Lemma 10.7**
>
> If the algorithm fails, then we find sets A_1 and A_2 where $|A_1|, |A_2| \geq \frac{1}{\Delta}\varepsilon^\Delta N$, such that the blue density between them is at least $1 - \varepsilon$.

This is just reformulating everything we've said so far — I've just told you I find a pair of sets with large blue density, and these sets are pretty large (because at every step I don't shrink too much). (We pay the $1/\Delta$ because of the pigeonholing thing.)

This is really the key point — when we fail, we learn something. And what we get when we fail is two sets that are still quite large, such that almost all edges between them are blue. (Not literally all edges, but the vast majority.)

This is the key observation. We're not done yet. But now we have a completely generic lemma that says every time the algorithm fails, something good happens. What we'll do is repeat this lemma inside each part — now I restrict my universe to just this set, and start from scratch to run this algorithm. If it succeeds, I'm happy. If I fail, then within the left part I find two large sets that have very high density in blue. And I do the same in the right part — either I find a copy of H in red and I'm happy, or I find two large sets that are almost complete in blue.

And I keep iterating this. The point is after I iterate this enough times, what I end up with is a picture that looks like a bunch of sets that are all fairly large, and where between every pair of them, I have a ton of blue edges — almost everywhere I have a ton of blue edges (that's exactly what I get by iterating this picture over and over again).

And now Yuval will spare us the details of the final step, but once we get that picture, we're going to greedily embed, but now using blue. So now we're finally going to give up our stubbornness of always trying to embed in red, and we'll try to embed in blue. And the point is because we have so many blue edges, we're guaranteed to succeed. So that's a computation you have to do, which we won't do here. But the point is you can choose ε small enough that when you get to this stage, you simply cannot fail.

And that completes the proof — the point is you try to embed in red as much as you can and get this one structure, and iterate it until you get this picture (any time you succeed you're happy, and otherwise you keep failing and find this picture, and then embed in blue).

> **Student Question.** *How many of these groups do you need?*
>
> **Answer.** You want $\Delta + 1$ blobs. The reason is a graph with maximum degree Δ has chromatic number at most $\Delta + 1$, and you use the proper coloring to embed into these blobs.

> **Student Question.** ...
>
> **Answer.** We said 99% of the edges between the big sets are blue, but maybe one of the smaller sets inside it is a millionth percent of that, so maybe it doesn't actually touch any blue edges. The way you fix this is every time you do this, you clean to go from many blue edges to high blue minimum degree. You do that every single time to ensure you never run into this issue.

(In the end ε ends up being something like $1/\Delta^2$; this is not obvious, and comes from all the computations we're omitting.)

§10.4 Multiple colors

One final remark about this topic: The actual CRST theorem, and Burr–Erdős conjecture (now Lee's theorem), are true for any number of colors (bounded-degree graphs have linear Ramsey number even with 17 colors). But this proof only works for 2 colors, because we used in this proof the really key observation that in a 2-coloring, if an edge is not red it must be blue. Once you have more than two colors, you lose that entirely. It is a major open problem to get this to work for three colors. And this would be a huge advance — somehow this entire technique is predicated on the fact that if it's not red it must be blue, and we don't know how to get past that, even though it sounds silly.

> **Student Question.** *What happens if you try a trick like splitting the colors into groups to turn two colors into one?*
>
> **Answer.** No one has gotten it to work. The problem is you really want a linear bound; you could imagine something like that, but every time you merge you might need to pay another factor of n or something; and then you might get a polynomial bound but not a linear bound.

§10.5 Monotone subsequences

Everyone who fell asleep should wake up, because we're going to do something much more fun (?) (Yuval loves this, but understands it's a bit brutal — he debated whether to present it, asked people, and ignored their opinions and decided to present it). He thinks the result is important and the technique is very pretty so we should get to see it, but it's also difficult. So now we'll do something completely different. Here's an amazing fact:

> **Fact 10.8** — Eight of you walked in today sorted by height.

Yuval wasn't keeping track of the order we came in, but he guarantees that at some point eight of us came in (possibly with others in between) sorted (either increasing or decreasing) by height. And of course there's nothing special about height; it's also true for age, and everything else. The thing underlying this is another result of Erdős–Szekerés from the same 1935 paper:

Theorem 10.9 (Erdős–Szekeres 1935)
Let $a_1, \ldots, a_N$ be distinct real numbers. If $N \geq (k-1)^2 + 1$, then there is a monotone subsequence of length k, i.e., there are indices $i_1 < \cdots < i_k$ such that

$$a_{i_1} < a_{i_2} < \cdots < a_{i_k} \quad \text{or} \quad a_{i_1} > a_{i_2} > \cdots > a_{i_k}.$$

(Yuval is pretty sure there's more than 50 people in this room; he's assuming for simplicity that our heights are all distinct, which is probably true if you measure accurately enough.)

This is a Ramsey-type question — we have some object (in this case, a sequence of real numbers), and I find order or structure within it. So you walked in in a completely crazy order, but I'm allowed to find a subsequence which is highly structured.

As some philosophy, this is something called a canonical Ramsey theorem. Canonical Ramsey theory is one of Yuval's favorite kinds. A thing that happens a lot of time in math is that you study things, and any type of thing you study almost always has a notion of something. If you study groups they have subgroups, rings have subrings/ideals, vector spaces have subspaces, and so on. And you care about the subobject relation.

There exist special objects all of whose subobjects look like the big object. Examples you might have seen in group theory: If you have any cyclic group, all its subgroups are cyclic. Or if you look at $\mathbb{Z}$, all its subgroups are isomorphic to $\mathbb{Z}$. A harder theorem is in any free group, all its subgroups are also free group. So these are objects where when you restrict to subobjects, you get either the same thing or a thing of the same type. (This is very vague.) These special objects are called *canonical*. (So the objects all of whose subobjects look the same as themselves are the ones called canonical.)

The point of canonical Ramsey theory is it has two steps:

(1) Identify the canonical objects.

(2) Prove that any object has a canonical subobject.

For example, in group theory, cyclic groups are canonical; and it's quite easy to prove every group contains a cyclic subgroup. But that's generally what this is about — you have these special objects you really like, and prove that every object has one of these.

This is useful — for example in Banach space theory you have a big space you have no hope of understanding, and in it you find a special one that's canonical; and then you can work in that space and do nice things in there.

(You can view Ramsey theory as this with colored complete graphs. The colored complete graphs all of whose subgraphs look the same are monochromatic cliques; and Ramsey's theorem tells you any colored clique has a monochromatic subclique.)

The reason to bring this up now is on Thursday we'll prove the canonical Ramsey theorem, which is another theorem of this type.

This theorem is another instance of this: The objects are sequences of real numbers, and subobjects are subsequences. The only things where all subsequences look the same are monotone — if I start with a monotone thing, all its subsequences are monotone (while if it's not monotone, some subsequences go up and some go down).

Student Question. *Could you say forests are also an example of this?*

Answer. Yes (all forests, all their subgraphs are also forests). This is so vague that you can come up with many examples; some are interesting and some are not (e.g., you can say any subgraph of anything is still a graph; this one is more interesting in that).

We'll see a proof due to Seidenberg, which is one of the most beautiful proofs in math.

Proof. Let $\delta(m)$ be the length of the longest decreasing subsequence ending at a_m, and let $\iota(m)$ be the length of the longest *increasing* subsequence ending at a_m. So I take my sequence, and at every element I write down the length of the longest increasing and decreasing sequences ending at them (δ and ι stand for 'decreasing' and 'increasing'). The key observation:

Fact 10.10 — If $\ell \neq m$, then $(\delta(\ell), \iota(\ell)) \neq (\delta(m), \iota(m))$.

So for any two elements of my sequence, this pair (δ, ι) must take on different values.

Proof. Suppose $\ell < m$. Then there's two cases: If $a_\ell < a_m$, what that means is I have my two points a_ℓ and then a_m, and this thing goes up (a_m is higher than a_ℓ). Then $\iota(m) > \iota(\ell)$. And the reason is you can take the longest increasing sequence ending at a_ℓ; I can always add a_m to the end of that sequence, and get a strictly longer increasing sequence. So that means $\iota(m) \geq \iota(\ell) + 1$; certainly they cannot be equal.

And the other case is symmetric — if $a_\ell > a_m$, then $\delta(m) > \delta(\ell)$, by a symmetric picture (where it goes down). □

Now, the key observation is that this thing where I record for every value both δ and ι — that's an injection. So $m \mapsto (\delta(m), \iota(m))$ is an injection (that's what I just proved — if I have two different points, they have different values of this pair). But now I have at least $(k-1)^2 + 1$ values, so this injection cannot be valued on $[k-1]^2$ — it can't be the case that δ and ι are always at most $k-1$, because if they were, I'd have an injection from a set of size $(k-1)^2 + 1$ to a set of size $(k-1)^2$, which is not possible. So there must be some point where $\delta(m)$ or $\iota(m)$ is at least k, which gives an increasing or decreasing subsequence of length at least k. □