1 What is extremal combinatorics?

We begin with a magic trick. I ask the audience to give me ten integers between 1 and 100, and then I find two disjoint subsets of this set of integers which have the same sum. The following theorem guarantees that I can always do this. Here, and throughout the course, we'll denote by $[\![n]\!]$ the range $\{1, 2, \ldots, n\}$.

Theorem 1.1. *Let $a_1, \ldots, a_{10}$ be integers in $[\![100]\!]$. Then there exist two disjoint, non-empty sets $S, T \subseteq [\![10]\!]$ such that*

$$\sum_{i \in S} a_i = \sum_{j \in T} a_j.$$

Proof. For every non-empty set $S \subseteq [\![10]\!]$, let us denote by a_S the number $a_S := \sum_{i \in S} a_i$. Note that there are $2^{10} - 1 = 1023$ non-empty subsets of $[\![10]\!]$, so we have defined 1023 numbers. Moreover, for every set S, we have that

$$a_S = \sum_{i \in S} a_i \leqslant |S| \cdot \max\{a_1, \ldots, a_{10}\} \leqslant 10 \cdot 100 = 1000.$$

That is, each of the numbers a_S lies in the interval $[1, 1000]$. As we have defined 1023 numbers, by the pigeonhole principle, two of them must be equal. That is, there must exist $S \neq T$ such that $a_S = a_T$.

We are almost done, except that S and T need not be disjoint, but this can be easily remedied: we define $S' := S \setminus T = S \setminus (S \cap T)$, and similarly $T' := T \setminus S$. Then S and T are disjoint. Moreover, we have that

$$a_{S'} = \sum_{i \in S'} a_i = \sum_{i \in S} a_i - \sum_{i \in S \cap T} a_i = a_S - a_{S \cap T},$$

and similarly $a_{T'} = a_T - a_{S \cap T}$, so $a_{S'} = a_{T'}$. Finally, these two sets are non-empty: this is because each of S, T was non-empty, and we cannot have $S \subseteq T$ or $T \subseteq S$, as this would contradict that they have the same sum. Therefore even when we remove $S \cap T$ from each of them, they remain non-empty. □

Note that the proof above shows that in fact, I could have allowed the audience to pick ten numbers in the range $[\![102]\!]$: in that case, each sum a_S would be at most 1020, and I could still apply the pigeonhole principle. This leads to a natural generalization of this question. Suppose I want to let my audience pick n numbers from some range $[\![M]\!]$, while still ensuring that no matter what they pick, I can find two disjoint subsets with equal sum. How large can I take M while ensuring that this is possible? Formally, we can define

$$\text{magic}(n) := \max\Bigg\{M \in \mathbb{N} \mid \text{for all } a_1, \ldots, a_n \in [\![M]\!], \\ \text{there are disjoint } S, T \subseteq [\![n]\!] \text{ with } \sum_{i \in S} a_i = \sum_{j \in T} a_j\Bigg\}.$$

Note that proving a *lower bound* on magic(n) boils down to proving a variant of Theorem 1.1. Indeed, to prove a lower bound on magic(n), it suffices to exhibit some M that works, which implies that magic(n) $\geqslant M$. Thus, for example, Theorem 1.1 states that magic(10) $\geqslant 100$, and as pointed out above, the proof actually gives magic(10) $\geqslant 102$. By following the proof of Theorem 1.1, one can more generally show that

$$\text{magic}(n) \geqslant \frac{2^n - 1}{n}.$$

What about proving *upper bounds* on magic(n)? This, in turn, boils down to finding a clever set of integers $a_1, \ldots, a_n$ such that all their subset sums are distinct. If we can find such integers $a_1, \ldots, a_n \in [\![M]\!]$, we have proved that M does not work for the magic trick, i.e. that magic(n) $< M$. A natural idea for how to pick such numbers $a_1, \ldots, a_n$ is to let them be the powers of 2, i.e. to set $a_i = 2^{i-1}$. Then it is not hard to check that all subset sums are distinct, which implies that

$$\text{magic}(n) < 2^{n-1}.$$

Our lower and upper bounds are not too far apart, but they differ by roughly a factor of n. And although obtaining these nearly-matching bounds was quite easy, closing this gap is a major open problem! Erdős offered \$500 for a proof or disproof of the following conjecture.

Conjecture 1.2 (Erdős's distinct sums conjecture)**.** *We have that* magic(n) $\geqslant \Omega(2^n)$. *That is, the upper bound is best possible up to a constant factor.*

Here, we recall the asymptotic notation that we'll use a lot throughout the course: the statement $f(n) \geqslant \Omega(g(n))$ means that for all sufficiently large n, we have that $f(n) \geqslant c \cdot g(n)$, where $c > 0$ is some constant independent of n.

The best known bounds on this problem are as follows. First, for the upper bound, Bohman found a construction of a set of integers beating the powers of two, which implies that

$$\text{magic}(n) \leqslant 0.22002 \cdot 2^n$$

for all sufficiently large n. In the other direction, a beautiful probabilistic[1] argument of Erdős and Moser proves that

$$\text{magic}(n) \geqslant \Omega\left(\frac{2^n}{\sqrt{n}}\right),$$

which improves the argument of Theorem 1.1 by roughly a factor of $\sqrt{n}$. The best known constant factor was obtained only recently by Dubroff, Fox, and Xu, who proved that

$$\text{magic}(n) \geqslant \left(\sqrt{\frac{2}{\pi}} - o(1)\right)\frac{2^n}{\sqrt{n}},$$

[1]You may wonder what probability has to do with any of this. Somewhat amazingly, even for problems like this that have nothing do with probability, many of our most powerful techniques are probabilistic. This introduction of random tools into non-random problems is called the *probabilistic method*, and we will see several examples of it throughout the course.

where $o(1)$ represents a quantity that tends to 0 as n tends to infinity.

This simple problem is an instance of extremal combinatorics, which is the topic of this course. In this context, "extremal" means "maximum or minimum": extremal questions are questions of the form "how large or small can an object be, subject to certain constraints?" For example, magic(n) is an extremal function: it asks how large M can be so long as $[\![M]\!]$ contains no set of n integers with distinct subset sums. Equivalently, we can view this function in the "opposite" perspective, and ask how many elements we can fit into $[\![M]\!]$ without creating two equal subset sums. The "combinatorics" in "extremal combinatorics" means that we are working with discrete structures, such as, in this case, finite sets of integers. For most of this course we will be focusing on extremal graph theory, where the objects of study are graphs (and their variants).

As in this example, whenever we are confronted with an extremal function, we would like to prove upper and lower bounds on it that are as close as possible. Generally speaking, one bound is proved by constructions of large or small objects satisfying the requisite property (like the powers of 2 above), and the other bound is proved by showing that *any* object of appropriate size must not have the requisite property (like the proof of Theorem 1.1). One of the things that I love about extremal combinatorics is that there are many natural and easy-to-state questions (such as the one discussed above), for which we can easily prove some upper and lower bounds, but for which it currently seems extremely difficult (or downright hopeless) to pin down the answer. That said, there keep being remarkable breakthroughs on some previously intractable problems, usually via the introduction of remarkable new ideas. In this course, I will try to give you a flavor of what extremal combinatorics is all about, some of the beautiful ideas that are used in the field, what some of the major open problems are, and hint at some of the recent breakthroughs on these and other questions.

2 What is extremal graph theory?

A simple and well-known fact in graph theory is that every n-vertex tree has $n-1$ edges. This immediately implies that if an n-vertex graph G has no cycles, then G has at most $n-1$ edges. Another well-known result in graph theory, following quickly from Euler's formula, is that an n-vertex planar graph G with $n \geqslant 3$ has at most $3n-6$ edges.

This class is about extremal graph theory, the study of results of this type. How many edges can an n-vertex graph have, given that it satisfies some natural constraint? Our major goal, for the next few lectures, is to prove the Erdős–Stone–Simonovits theorem, sometimes called the Fundamental Theorem of Extremal Graph Theory, which answers this question more or less completely for a very wide range of constraints.

The question that will occupy us for some time is what happens when the constraint is excluding a single "forbidden subgraph".

Definition 2.1. Let H and G be graphs. We say that G is *H-free* if H is not a subgraph of G (or, more formally, if G has no subgraph isomorphic to H). We will often also say that G has *no copy of H*.

The basic question we will be attempting to answer is "how many edges can an n-vertex H-free graph have?". Because we will be using this notion over and over again, it's best to just give it a name. We use $e(G)$ to denote the number of edges of a graph G.

Definition 2.2. The *extremal number* of H is defined as

$$\mathrm{ex}(n, H) = \max\{e(G) \mid G \text{ is an } n\text{-vertex } H\text{-free graph}\}.$$

In other words, $\mathrm{ex}(n, H)$ is simply the most number of edges that an H-free graph on n vertices can have. Note that this quantity is well-defined, since there are only finitely many n-vertex graphs.

In this class, we will attempt to understand how the function $\mathrm{ex}(n, H)$ behaves when H is some fixed graph, and when n tends to infinity. Additionally, we will often try to understand which graphs G are the maximizers in the definition of $\mathrm{ex}(n, H)$; that is, which graphs G have the most edges among all n-vertex H-free graphs.

Before getting into specific examples, let's briefly think about what it means to prove upper and lower bounds on $\mathrm{ex}(n, H)$. Since $\mathrm{ex}(n, H)$ is defined as the maximum of something, to prove a *lower* bound on $\mathrm{ex}(n, H)$, it suffices to exhibit an n-vertex graph G with no copy of H; such a G gives us the lower bound $\mathrm{ex}(n, H) \geqslant e(G)$. On the other hand, to prove an *upper* bound on $\mathrm{ex}(n, H)$, we need to prove that *every* n-vertex graph G with m edges has a copy of H; this yields the upper bound $\mathrm{ex}(n, H) < m$.

3 Forbidden cliques: Mantel's and Turán's theorems

The earliest result in extremal graph theory is due to Mantel, from more than 100 years ago. Mantel studied (though not in this language) the extremal number of the triangle graph, K_3. Let's begin by coming up with a lower bound on $\mathrm{ex}(n, K_3)$.

After playing around with it a bit, it's pretty natural to come up with the following construction. Let $G = K_{a,b}$ be a complete bipartite graph, where $a + b = n$. Then G is certainly triangle-free, since K_3 is not bipartite. Moreover, the number of edges in G is simply ab. So we find that

$$\mathrm{ex}(n, K_3) \geqslant ab \qquad \text{for all integers } a, b \text{ with } a + b = n.$$

Since we want as good a lower bound as possible, we want to pick a, b so that ab is maximized, subject to the constraint that $a + b = n$. Using the AM-GM inequality, we see that

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

Moreover, equality holds if and only if $a = b = n/2$. If n is odd, then we can't have $a = b = n/2$ if a and b are both integers; the product ab is maximized when $a = \lfloor n/2 \rfloor, b = \lceil n/2 \rceil$. But in any case, we find that

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

with the example of a K_3-free n-vertex graph with $\lfloor n^2/4 \rfloor$ edges given by the complete bipartite graph $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.

Mantel's theorem says that this is the best we can do.

Theorem 3.1 (Mantel 1907). $\mathrm{ex}(n, K_3) = \lfloor n^2/4 \rfloor$. *Moreover, the unique n-vertex triangle-free graph with $\lfloor n^2/4 \rfloor$ edges is $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

We won't prove this right now. Instead, we'll first generalize Mantel's theorem, and then prove the generalization.

Almost 40 years after Mantel, Turán started thinking about similar questions, and it is thanks to his work that the field of extremal graph theory exists at all. Turán was studying what happens when, rather than excluding a triangle, we exclude some larger complete graph (also known as a *clique*). Namely, he was studying $\mathrm{ex}(n, K_r)$ for $r \geqslant 3$.

Again, there is a natural type of example we can come up with to lower-bound $\mathrm{ex}(n, K_r)$. Namely, let G be a complete $(r-1)$-partite graph on n vertices, namely a graph obtained by splitting the n vertices into $r-1$ parts, then putting all edges between pairs of vertices in different parts and no edges within a part. Then G certainly will not have a copy of K_r: by the pigeonhole principle, if we take any r vertices in G, two of them must lie in the same part, and thus there cannot be an edge between them. Moreover, another simple application of the AM-GM inequality (or Jensen's inequality) shows that the way to do this in order to maximize the number of edges of G is to make all the parts have as equal sizes as possible, namely to make each part have size either $\lfloor n/(r-1) \rfloor$ or $\lceil n/(r-1) \rceil$. This motivates the following definition.

Definition 3.2. The *Turán graph* $T_{r-1}(n)$ is the n-vertex complete $(r-1)$-partite graph with all parts of size either $\lfloor n/(r-1) \rfloor$ or $\lceil n/(r-1) \rceil$. We denote its number of edges by

$$t_{r-1}(n) \coloneqq e(T_{r-1}(n)).$$

Remark. In case n is divisible by $r-1$, then every part of the Turán graph $T_{r-1}(n)$ has exactly $n/(r-1)$ vertices in each part, so

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

In case n is not divisible by $r-1$, the formula is a little messier, involving the remainder of n when divided by $r-1$. However, we still have that for all fixed r and $n \to \infty$,

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

where $o(1)$ represents a quantity that tends to 0 as n tends to infinity. In other words, if we fix $\varepsilon > 0$, then for any sufficiently large n, we have that

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

One other useful observation is that for any n, if we delete one vertex from each of the $r-1$ parts of $T_{r-1}(n)$, we obtain a copy of $T_{r-1}(n-r+1)$. Moreover, each non-deleted vertex is adjacent to exactly $r-2$ deleted vertices. So we delete $(r-2)(n-r+1)+\binom{r-1}{2}$ edges to obtain $T_{r-1}(n-r+1)$ from $T_{r-1}(n)$. This shows that

$$t_{r-1}(n) = t_{r-1}(n-r+1) + (r-2)(n-r+1) + \binom{r-1}{2}. \tag{1}$$

Note that $T_2(n) = K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$, so Mantel's theorem can be rephrased as saying that $\mathrm{ex}(n, K_3) = t_2(n)$ and that $T_2(n)$ is the unique n-vertex K_3-free graph with $t_2(n)$ edges. Turán's theorem generalizes this to $\mathrm{ex}(n, K_r)$ for all $r \geqslant 3$.

Theorem 3.3 (Turán 1941)**.** *For every $r \geqslant 3$, we have $\mathrm{ex}(n, K_r) = t_{r-1}(n)$. Moreover, the unique n-vertex K_r-free graph with $t_{r-1}(n)$ edges is $T_{r-1}(n)$.*

Proof. We proceed by induction, with steps of size $r-1$. So we need $r-1$ base cases, corresponding to $n = 1, 2, \ldots, r-1$. But the theorem holds for such n, because for such n, any n-vertex graph has no K_r subgraph. So $\mathrm{ex}(n, K_r) = \binom{n}{2}$ for $1 \leqslant n \leqslant r-1$. Moreover, $T_{r-1}(n)$ is exactly K_n in these cases. This proves the base cases of the induction.

Now let $n > r-1$, and assume the theorem is true for $n-r+1$. Let G be an n-vertex graph with no copy of K_r and as many edges as possible. G must contain a copy of K_{r-1}, for otherwise we could add an edge and get a K_r-free graph with strictly more edges. Let K be some such K_{r-1} subgraph, and let $F \subseteq G$ be the subgraph obtained by deleting K. We know that $e(K) = \binom{r-1}{2}$. By induction, we know that

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

Finally, each vertex of F cannot be adjacent to every vertex of K, for otherwise we would get a K_r. So the number of edges between F and K is at most $(r-2)(n-r+1)$. So

$$e(G) \leqslant \binom{r-1}{2} + t_{r-1}(n-r+1) + (r-2)(n-r+1) = t_{r-1}(n),$$

by (1).

If $e(G) = t_{r-1}(n)$, then every inequality above must be an equality. In particular, the induction hypothesis implies that $F \cong T_{r-1}(n-r+1)$. Moreover, each vertex in F must be adjacent to exactly $r-2$ vertices in K, since we assume we have equality in the number of edges. Moreover, given two adjacent vertices in F, they cannot be non-adjacent to the same vertex of K, for otherwise we could take the remaining $r-2$ vertices and these two to get a K_r. So this implies that each part of F is associated to exactly one missed vertex. So by adding this missed vertex to its part, we see that $G \cong T_{r-1}(n)$. □

On the homework over the next few days, you will see many different proofs of Turán's theorem. It is one of those amazing mathematical theorems with dozens of different, and differently informative, proofs. It is also extremely useful, as you'll see on the homework!

Before moving on, let me just mention one convenient way to think about Turán's theorem is as follows. Note that

$$\binom{n}{2} = \frac{n^2}{2} - \frac{n}{2} = (1 + o(1))\frac{n^2}{2}.$$

This shows that

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

Note that an n-vertex graph can have anywhere between 0 and $\binom{n}{2}$ edges. So Turán's theorem implies that a K_r-free n-vertex graph can have at most, asymptotically, a $1-1/(r-1)$ fraction of all possible edges.

4 Beyond Turán's theorem

Turán's theorem is great, and tells us exactly what $\text{ex}(n, K_r)$ is for any r. But we started this class by asking about $\text{ex}(n, H)$ for general H; what can we say about that? In general, we'd probably expect this problem to be really hard, and the answer should depend in complicated ways on the fixed graph H.

But it turns out that's not the case! Kind of amazingly, the answer depends, essentially, on a single parameter of the graph H—its chromatic number.

Theorem 4.1 (Erdős–Stone(–Simonovits) 1946 (1966))**.** *For any graph H,*

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

Remark.

- This is sometimes called the Fundamental Theorem of Extremal Graph Theory, for hopefully obvious reasons: it more or less completely resolves the main question that we started with.
- The history (and naming) of this theorem is a bit confusing. Erdős and Stone proved a special case of it in 1946. In 1966, Erdős and Simonovits realized that the special case actually implies (with a one-line implication) the general case, which had not been really studied before. We will soon see the special case, and how it implies the general case.
- Notice that if H is bipartite (i.e. if $\chi(H) = 2$), then $1 - 1/(\chi(H) - 1) = 0$. So the theorem simply says that if H is bipartite, then

$$\text{ex}(n, H) = o(1) \cdot \binom{n}{2},$$

which we usually write as $\mathrm{ex}(n, H) = o(n^2)$. In other words, if G is an n-vertex graph containing no copy of some fixed bipartite graph H, then G must have *very few* edges—its number of edges grows sub-quadratically in n. Said differently, the fraction of all possible edges that we can put in such a graph is a vanishingly small fraction; the fraction tends to 0 as $n \to \infty$.

Already this statement is far from obvious, and we'll soon prove it. In fact, as we'll see, proving the statement for bipartite H implies, in a certain sense, the full Erdős–Stone–Simonovits theorem.

The next few lectures will be spent on proving the Erdős–Stone–Simonovits theorem. To do so, we'll prove upper and lower bounds on $\mathrm{ex}(n, H)$ of the form $(1-1/(\chi(H)-1)+o(1))\binom{n}{2}$. In fact, we can easily dispense with the lower bound.

Proposition 4.2. *For any fixed graph H and integer n,*

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

Proof. We claim that the Turán graph $T_{\chi(H)-1}(n)$ has no copy of H. Indeed, suppose we had some vertices in $T_{\chi(H)-1}(n)$ that defined a copy of H. Give the parts of $T_{\chi(H)-1}(n)$ names, say $V_1, \ldots, V_{\chi(H)-1}$. Then note that any two vertices of H that lie in the same part V_i cannot be adjacent in H, since $T_{\chi(H)-1}(n)$ has no edges inside any part V_i. Said differently, if we assign to any vertex v of H the number i so that $v \in V_i$, then two adjacent vertices are assigned different numbers. In other words, this yields a proper coloring of H with $\chi(H) - 1$ colors. But this contradicts the definition of the chromatic number. □

5 Extremal numbers of bipartite graphs

5.1 Upper bounds

Let H be a bipartite graph. Recall that the Erdős–Stone–Simonovits theorem implies that in this case, $\mathrm{ex}(n, H) = o(n^2)$, or equivalently that

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

This is pretty surprising! For example, the complete bipartite graph $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ has $\lfloor n^2/4\rfloor$ edges and no copy of any odd cycle $C_{2\ell+1}$. Thus, $\mathrm{ex}(n, C_{2\ell+1}) \geqslant \lfloor n^2/4\rfloor$ for all ℓ. But the four-cycle (or any other even cycle) is bipartite, so the Erdős–Stone–Simonovits theorem implies that $\mathrm{ex}(n, C_4) = o(n^2)$. What's up with that?

We will shortly prove that in fact, $\mathrm{ex}(n, C_4) \leqslant O(n^{3/2})$. In case you haven't seen it before, the big-O notation means that $\mathrm{ex}(n, C_4) \leqslant Cn^{3/2}$ for some absolute constant C, which we won't specify. In other words, we will shortly prove that if G is an n-vertex graph with at least $Cn^{3/2}$ edges, then G has a copy of C_4, assuming C is an appropriately chosen constant.

As a warm-up, we will begin with an easier special case of this result, which is the case when G is d-regular (i.e. every vertex in G has degree d). Recall that in any graph, the sum of the degrees of all the vertices equals twice the number of edges, so if G is d-regular then it has $dn/2$ edges. Thus, if G has $Cn^{3/2}$ edges and is d-regular, then $d = 2C\sqrt{n}$.

Proposition 5.1. *Let G be a d-regular n-vertex graph. If $d \geqslant 2\sqrt{n}$, then G contains a copy of C_4.*

Proof. Suppose for contradiction that G is C_4-free. We count the number of copies of $K_{1,2}$ in G, where $K_{1,2} =$ ●—●—● consists of one central vertex adjacent to two outer vertices. On the one hand, if we sum over all possibilities for the central vertex, we see that

$$\#(K_{1,2} \text{ in } G) = \sum_{v \in V(G)} \#(K_{1,2} \text{ with central vertex } v) = \sum_{v \in V(G)} \binom{\deg(v)}{2} = n\binom{d}{2}.$$

On the other hand, suppose we fix some $u, w \in V(G)$. We claim that they can be the outer vertices of at most one copy of $K_{1,2}$. Indeed, if not, then we would have two $K_{1,2}$s agreeing on the outer vertices, which yields a copy of C_4, a contradiction. So we conclude that

$$\#(K_{1,2} \text{ in } G) = \sum_{\substack{u,w \in V(G) \\ \text{distinct}}} \#(K_{1,2} \text{ with outer vertices } u, w) \leqslant \sum_{\substack{u,w \in V(G) \\ \text{distinct}}} 1 = \binom{n}{2}.$$

Rearranging, we see that

$$n\binom{d}{2} \leqslant \binom{n}{2} \qquad \iff \qquad \binom{d}{2} \leqslant \frac{n-1}{2} \qquad \iff \qquad d(d-1) \leqslant n-1.$$

But if $d \geqslant 2\sqrt{n}$ and $n \geqslant 0$, then this is a contradiction. □

To prove the real result, we will need one extraordinarily useful analytic tool, called *Jensen's inequality.* We will actually only need the following special case. For a real number x and a positive integer r, we extend the definition of the binomial coefficient as

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

Lemma 5.2 ((Consequence of) Jensen's inequality)**.** *Let $r \geqslant 1$ be a positive integer, and let $x_1, \ldots, x_n$ be non-negative integers. Suppose that $\frac{1}{n}\sum_{i=1}^n x_i \geqslant r$. Then*

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

The point of this is that if we add up terms of the form $\binom{x_i}{r}$, we can only decrease the sum if we replace each x_i by the *average* of all the x_i. One says that the function $x \mapsto \binom{x}{r}$ is *convex*: the sum of its values is minimized when all the variables are equal (to their average).

We won't prove Jensen's inequality in class, but its proof is on the homework if you're interested. Once we have Jensen's inequality, we can easily prove the full result that $\mathrm{ex}(n, H) \leqslant O(n^{3/2})$. In fact, we will prove the following much more general result.

Theorem 5.3 (Kővári–Sós–Turán 1954). *For positive integers $s \leqslant t$, we have*

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

Here, the implicit constant may depend on s and t (which we think of as fixed).

Proof. We proceed much as in the proof of Proposition 5.1. Let G be an n-vertex graph with at least $Cn^{2-1/s}$ edges, where C is some large constant we will pick later. Let d be the average degree in G, so that $d = \frac{2}{n}e(G) \geqslant 2Cn^{1-1/s}$. Suppose for contradiction that G is $K_{s,t}$-free. We count the number of copies of $K_{1,s}$ in G in two ways. First, by summing over the options for the central vertex, we have that

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

using Lemma 5.2, as well as the fact that $d \geqslant s$ by picking C sufficiently large. On the other hand, by counting over the s outer vertices of $K_{1,s}$, we have that every $u_1, \ldots, u_s \in V(G)$ can be the outer vertices of at most $t-1$ copies of $K_{1,s}$. So

$$\#(K_{1,s} \text{ in } G) \leqslant \sum_{\substack{u_1,\ldots,u_s \in V(G) \\ \text{distinct}}} (t-1) = (t-1)\binom{n}{s}.$$

Combining these, we see that

$$(t-1)\binom{n}{s} \geqslant n\binom{d}{s} \qquad \iff \qquad (t-1)(n-1)(n-2)\cdots(n-s+1) \geqslant d(d-1)\cdots(d-s+1).$$

Now, if n is very large (which is the regime we care about anyway), all this subtracting stuff doesn't matter. So this is roughly equivalent to

$$(t-1)n^{s-1} \geqslant d^s \qquad \iff \qquad d \leqslant (t-1)^{1/s}n^{1-1/s}.$$

If C is sufficiently large, then this is a contradiction. Moreover, if C is sufficiently large, then the slighty sketchy step above where we dropped the subtractions is also OK, and we get the desired contradiction. □

Note that $C_4 = K_{2,2}$, so in the case $s = t = 2$, we indeed get the claimed bound of $\mathrm{ex}(n, C_4) \leqslant O(n^{3/2})$.

There are a number of important consequences of the Kővári–Sós–Turán theorem. The first is that it immediately gives us a bound on $\mathrm{ex}(n, H)$ for all bipartite H. Indeed, note that if H_1 is a subgraph of H_2, then

$$\mathrm{ex}(n, H_1) \leqslant \mathrm{ex}(n, H_2)$$

for all n, as any H_1-free graph is also H_2-free. Now, if H is a bipartite graph, then H is a subgraph of $K_{s,t}$ for some $s \leqslant t$. So

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

In particular, this proves that $\mathrm{ex}(n,H) = o(n^2) = o(1) \cdot \binom{n}{2}$ for bipartite H. Indeed,

$$\lim_{n\to\infty} \frac{\mathrm{ex}(n,H)}{n^2} \leqslant \lim_{n\to\infty} \frac{O(n^{2-1/s})}{n^2} = \lim_{n\to\infty} O(n^{-1/s}) = 0.$$

Recall that this was a consequence of the Erdős–Stone–Simonovits theorem.

5.2 Lower bounds

How good are the upper bounds we proved? Let's begin with the one we started with, $\mathrm{ex}(n,C_4) \leqslant O(n^{3/2})$. Can we construct an n-vertex C_4-free graph with roughly that many edges?

As it turns out, we can! The following construction was originally due to Eszter Klein in 1938 (as reported in a paper of Erdős). Note that this is before Turán's theorem, so before the birth of extremal graph theory! As such, no one really appreciated what this construction was or meant, and it was later rediscovered by Erdős, Rényi, and Sós (and independently Brown). These days, it is often called the "Erdős–Rényi" construction, which I find a little odd, both because they weren't the first to discover it, and because there are many other things named after Erdős and Rényi.

Theorem 5.4 (Klein 1938)**.** *For every $n \geqslant 1$, there is an n-vertex C_4-free graph with at least $n^{3/2}/64$ edges.*

Proof. First, suppose that $n = 2p^2$ for some prime p; we will later get rid of this assumption. Consider the integers mod p, which form a field that we denote $\mathbb{F}_p$. (If you don't know what the word "field" means, just believe me that among the integers mod p, we can use addition, multiplication, and division and have them work basically the same way they do in $\mathbb{R}$.)

Let $\mathbb{F}_p^2$ denote the two-dimensional plane over $\mathbb{F}_p$, i.e. the set of points (x,y) with $x,y \in \mathbb{F}_p$. For $m,b \in \mathbb{F}_p$, let $\ell_{m,b}$ denote the line $y = mx + b$ in $\mathbb{F}_p^2$. In other words, $\ell_{m,b}$ is the set of points $(x,y) \in \mathbb{F}_p^2$ satisfying $y = mx + b$.

We define a bipartite graph G with parts P, L, where $P = \mathbb{F}_p^2$ and $L = \{\ell_{m,b} : m,b \in \mathbb{F}_p\}$. The edges of G are given by *incidence*: we connect $(x,y) \in P$ to $\ell_{m,b} \in L$ if and only if (x,y) lies on the line $\ell_{m,b}$, i.e. if and only if $y = mx + b$.

Note that $|P| = |L| = p^2$, so G has $n = 2p^2$ vertices. Moreover, every line $\ell_{m,b} \in L$ has exactly p points on it, so every vertex in L has degree p in G. Therefore, $e(G) = p|L| = p^3 = (n/2)^{3/2}$.

Finally, we claim that G is C_4-free. To see this, note that G is bipartite, so the only way we could have a copy of C_4 in G is to have distinct $p_1, p_2 \in P$ and distinct $\ell_1 \ell_2 \in L$ so that $p_1 \ell_1 p_2 \ell_2$ forms a 4-cycle. But this means that p_1 lies on the lines ℓ_1, ℓ_2, and that p_2 also lies on both these lines. So we have two lines which intersect at two distinct points!

Using our intuition from $\mathbb{R}$, we expect this to be impossible, and it's impossible over $\mathbb{F}_p$ as well. Formally, let $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$, and $\ell_1 = \ell_{m_1,b_1}, \ell_2 = \ell_{m_2,b_2}$. Then we have the equations

$$\begin{aligned} y_1 &= m_1x_1 + b_1 & \qquad y_2 &= m_1x_2 + b_1 \\ y_1 &= m_2x_1 + b_2 & \qquad y_2 &= m_2x_2 + b_2 \end{aligned}$$

Rearranging the first column, we see that $m_1x_1 + b_1 = m_2x_1 + b_2$, or equivalently that $(m_2 - m_1)x_1 = b_1 - b_2$. If $m_1 = m_2$ then this implies that $b_1 = b_2$, contradicting that ℓ_1, ℓ_2 are distinct. So we have that $m_1 \neq m_2$, so $x_1 = (b_1 - b_2)/(m_2 - m_1)$. But from the second column of equations, we conclude that $x_2 = (b_1 - b_2)/(m_2 - m_1)$ as well, so $x_1 = x_2$. But if we plug this into any of the equations, we conclude that $y_1 = y_2$, and thus that $p_1 = p_2$, a contradiction. So G is C_4-free.

The only remaining thing is to deal with the fact that n need not equal twice the square of a prime. So let n be arbitrary. There is an important result in number theory, called Bertrand's postulate, which says that there is always a prime between m and $2m$ for all positive integers m. Let $m = \lfloor \sqrt{n}/4 \rfloor$, and let p be a prime between m and $2m$, so that $n/8 \leqslant 2p^2 \leqslant n$. Using the construction above, we obtain a C_4-free graph G on $2p^2$ vertices with p^3 edges. We add to this graph $n - 2p^2$ isolated vertices, and we obtain a new C_4-free n-vertex graph with $p^3 \geqslant (n/16)^{3/2} = n^{3/2}/64$ edges. □

Using these finite fields and finite geometries might seem like a neat trick, but it turns out that it's essentially the only thing one can do. Indeed, all constructions we know of for C_4-free graphs with many edges use such techniques. Moreover, there is a powerful result of Füredi, which says that for those n for which such a construction (appropriately defined) exists, the *unique* C_4-free n-vertex graph with the most edges comes from such a construction.

So we conclude that $\text{ex}(n, C_4) = \Theta(n^{3/2})$, where the big-$\Theta$ means that we have upper and lower bounds that agree up to a constant factor. Since $\text{ex}(n, K_{2,t}) \geqslant \text{ex}(n, C_4)$ for all $t \geqslant 2$, we conclude that $\text{ex}(n, K_{2,t}) = \Theta(n^{3/2})$ for all $t \geqslant 2$.

What about $\text{ex}(n, K_{3,3})$? We proved in Theorem 5.3 that $\text{ex}(n, K_{3,3}) \leqslant O(n^{5/3})$. As it turns out, this is also tight.

Theorem 5.5 (Brown 1966)**.** *For every n, there exists an n-vertex $K_{3,3}$-free graph G with $n^{5/3}/100$ edges.*

Proof sketch. I won't present the proof in detail, but will explain the big idea. Suppose that $n = p^3$. Construct a graph G with vertex set $\mathbb{F}_p^3$, where we connect two vertices (x, y, z) and (x', y', z') by an edge if and only if

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

In other words, the neighborhood of any vertex looks like a "unit sphere" centered at that vertex, except that "spheres" don't really exist over $\mathbb{F}_p$.

Nonetheless, if we were working in $\mathbb{R}^3$, then we'd expect that any three unit spheres can intersect in at most two points: two unit spheres can intersect in a circle, and that circle can intersect a thid unit sphere in only two points. So we'd expect that G is $K_{3,3}$-free, since any three vertices have at most two common neighbors.

Since we expect a sphere to be "two-dimensional", we should expect every unit sphere to have roughly p^2 points on it, and this turns out to be true. So G has $n = p^3$ vertices, and every vertex has degree around p^2, so we expect $e(G) \approx p^5 = n^{5/3}$.

All of this intuition can be made precise, some of it with some annoyance. For example, it turns out that this only really works if $p \equiv 3 \pmod 4$. But the high-level idea is correct. □

So we see that the Kővári–Sós–Turán theorem is best possible (up to the constant factor) for $s = 2$ and $s = 3$. The case of $s = 1$ is much easier, but it's also best possible there, as you'll see on the homework. So it is natural to conjecture, as many have done, that the Kővári–Sós–Turán theorem is tight in general.

Conjecture 5.6 (Many people)**.** *For all $s \leqslant t$,*

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

Moreover, based on what I've told you so far, it is natural to expect that not only is this conjecture proved, but that the constructions look kind of the same as above. You work with the s-dimensional space $\mathbb{F}_p^s$ over the field $\mathbb{F}_p$, and use some kind of cleverly chosen polynomial or set of polynomial equations to define the adjacency condition. However, despite many people having this same idea, Conjecture 5.6 remains unproved. Moreover, many experts in the field now even question whether it is true.

Nonetheless, some other things are known about $\mathrm{ex}(n, K_{s,t})$. Namely, it is known that $\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s})$ if t is sufficiently large compared to s. The first result of this type is due to Kollár, Rónyai, and Szabó in 1996, who proved that

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

To do this, they constructed a $K_{s,t}$-free graph, again using the space $\mathbb{F}_p^s$, called the *norm graph*. Their construction was later modified by Alon, Rónyai, and Szabó in 1999, who defined a similar graph called the *projective norm graph* (again over $\mathbb{F}_p^s$), which implies that

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

So, for example, we know that $\mathrm{ex}(n, K_{4,7}) = \Theta(n^{7/4})$, but have no such lower bound for $\mathrm{ex}(n, K_{4,4})$.

For about 20 years, the Alon–Rónayi–Szabó result was the best known. But very recently, Bukh proved the following theorem.

Theorem 5.7 (Bukh 2021)**.** *Suppose $s \geqslant 2$ and $t \geqslant 9^s \cdot s^{4s^{2/3}}$ are integers. Then*

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

The key point is that for large s, the previous bound on t, namely $(s-1)!$, grew super-exponentially in s. But Bukh's bound, for large s, grows merely exponentially in s. The key to Bukh's construction is again to work over $\mathbb{F}_p^s$, but not to pick a *clever* polynomial. Instead, he picks a *random* polynomial, and then uses arguments from probability, combinatorics, and algebraic geometry to prove that the resulting graph is $K_{s,t}$-free with positive probability.

While similar algebraic constructions exist for certain specific bipartite H, there is essentially only one general-purpose lower bound that is known. In general, algebraic techniques like the ones described above are the best techniques we have for constructing lower bounds, but they often rely on specific structures that we can exploit. The following bound holds for any bipartite graph.

Given a graph H, we define its *2-density* to be

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

Theorem 5.8. *For any bipartite H, we have*

$$\mathrm{ex}(n,H) \geqslant \Omega(n^{2-1/m_2(H)}).$$

The proof of Theorem 5.8 uses the probabilistic method, and I won't cover it in class. But at a high level, the idea is to pick a *random* graph G with n vertices and roughly $n^{2-1/m_2(H)}$ edges. One can then show that with positive probability, the number of copies of H in G is less than half the number of edges of G. By deleting a single edge from each copy of H, we obtain a graph with half as many edges—so still $\Omega(n^{2-1/m_2(H)})$—and no copy of H.

6 Extremal numbers of hypergraphs

It's time for everything to get more *hyper*.

If we go back to bare basics, a graph is a collection V of vertices, plus a collection E of edges, which are simply unordered pairs of vertices. Why restrict ourselves to pairs?

Definition 6.1. A *k-uniform hypergraph* (sometimes called an k-graph for short) consists of a finite collection V of *vertices*, as well as a collection E of *k-uniform hyperedges*, which are simply subsets of V of size k.

As with graphs, we say that one k-graph $\mathcal{H}$ is a *subhypergraph* (or simply *subgraph*) of another k-graph $\mathcal{G}$ if we can obtain $\mathcal{H}$ from $\mathcal{G}$ by deleting some vertices and edges. We say that $\mathcal{G}$ is *$\mathcal{H}$-free* if $\mathcal{G}$ does not contain $\mathcal{H}$ as a subgraph (and we also say that $\mathcal{G}$ has no copy of $\mathcal{H}$).

As with graphs, we define the *extremal number* of $\mathcal{H}$ as

$$\mathrm{ex}(n,\mathcal{H}) = \max\{e(\mathcal{G}) : \mathcal{G} \text{ is an } n\text{-vertex } \mathcal{H}\text{-free } k\text{-graph}\}.$$

In contrast to graphs (the case $k=2$), we know extraordinarily little about $\mathrm{ex}(n,\mathcal{H})$ for k-graphs $\mathcal{H}$ with $k \geqslant 3$. For example, even the hypergraph analogue of Mantel's theorem is a famous open problem. To explain this formally, we make the following definition.

Definition 6.2. The *complete k-graph* on r vertices, denoted $K_r^{(k)}$, is the k-graph with r vertices whose edge set consists of all subsets of size k.

Then the amazing fact is that for any $k > r \geqslant 3$, we do not know the value of $\text{ex}(n, K_r^{(k)})$. For literally no pair of (r, k)! This problem was proposed by Turán already in 1941, and he made the following conjecture, which is a natural analogue of Mantel's theorem.

Conjecture 6.3 (Turán 1941)**.**

$$\text{ex}(n, K_4^{(3)}) = \left(\frac{5}{9} + o(1)\right)\binom{n}{3}.$$

The reason for 5/9 is a specific construction of an n-vertex $K_4^{(3)}$-free 3-graph, which Turán came up with, and which was the best he could come up with. You'll see Turán's construction on the homework.

Erdős offered \$500 for the resolution of Conjecture 6.3, and \$1000 for a general formula for $\text{ex}(n, K_r^{(k)})$. So far, very little progress has been made on these questions. The best known bound for $\text{ex}(n, K_4^{(3)})$ is due to Razborov, who proved that

$$\text{ex}(n, K_4^{(3)}) \leqslant (0.561666 + o(1))\binom{n}{3}.$$

Note that $5/9 = 0.555\ldots$, so this is pretty close to Turán's conjecture. Unfortunately, Razborov's technique is unlikely to yield the full resolution of Conjecture 6.3, because his technique uses a computer to do complicated computations to what is essentially a "finite approximation" to the problem.

In general, the best known lower bound is due to de Caen, who proved that

$$\text{ex}(n, K_r^{(k)}) \leqslant \left(1 - \frac{1}{\binom{r-1}{k-1}} + o(1)\right)\binom{n}{k}.$$

The best known general lower bound, due to Sidorenko, is

$$\text{ex}(n, K_r^{(k)}) \geqslant \left(1 - \left(\frac{k-1}{r-1}\right)^{k-1} + o(1)\right)\binom{n}{k}.$$

In the case of $k = 3$, this says that

$$\text{ex}(n, K_r^{(3)}) \geqslant \left(1 - \left(\frac{2}{r-1}\right)^{2} + o(1)\right)\binom{n}{3},$$

and this was conjectured to be optimal by Turán. You'll see the construction in the homework.

Despite not knowing the hypergraph analogues of Mantel's or Turán's theorems, the hypergraph analogue of the Kővári–Sós–Turán theorem *is* known, and is due to Erdős (1965). To state this, we need to define the hypergraph analogue of a bipartite graph.

Definition 6.4. A k-graph $\mathcal{H}$ is called *k-partite* if its vertex set can be split into k parts, so that every hyperedge of $\mathcal{H}$ contains exactly one vertex from each part.

The *complete k-partite k-graph with parts of sizes $s_1, \dots, s_k$*, denoted $K^{(k)}_{s_1,\dots,s_k}$, is the k-graph with parts of sizes $s_1, \dots, s_k$, containing every edge with exactly one vertex from each part.

Note that in case $k = 2$, this simply recovers the definition of a bipartite graph and a complete bipartite graph. Because of this, the following result generalizes the Kővári–Sós–Turán theorem.

Theorem 6.5 (Erdős 1965)**.** *Let $s_1 \leqslant \cdots \leqslant s_k$ be positive integers. Then*

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

The most important thing here is that any n-vertex k-graph has at most $\binom{n}{k} = \Theta(n^k)$ hyperedges. So this upper bound has a smaller exponent on n. This implies that if $\mathcal{H}$ is *any* k-partite k-graph, then

$$\mathrm{ex}(n, \mathcal{H}) = o(1) \cdot \binom{n}{k}.$$

The upper bound in Theorem 5.3 is still the best upper bound we have on extremal numbers of k-partite k-graphs. Moreover, as in the case of graphs, it is known that the bound in Theorem 6.5 is best possible if s_k is sufficiently large with respect to $s_1, \dots, s_{k-1}$.

The proof of Theorem 6.5 very similar to that of Theorem 5.3, except that we combine it with an induction on k. To keep the notation from getting too crazy, we will only prove it in the case $k = 3$, which we will derive from the case $k = 2$, i.e. the Kővári–Sós–Turán theorem. Also, we will only prove it in the case $s_1 = s_2 = s_3 = s$, i.e. we will prove that

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

Hopefully you'll believe me (or convince yourself that it's true if you don't!) that the general result follows from the same technique, just with more bookkeeping.

Proof of (2). Let $\mathcal{G}$ be an n-vertex 3-graph with at least Cn^{3-1/s^2} hyperedges, for some constant $C > 0$ we will pick later. Suppose for contradiction that $\mathcal{G}$ is $K^{(3)}_{s,s,s}$-free. Let X be the number of copies of $K^{(3)}_{1,1,s}$ in $\mathcal{G}$. We bound X in two ways.

First, for a pair of distinct vertices v, w, let $\mathrm{codeg}(v, w)$ denote the number of hyperedges containing both v and w. Then we first claim that

$$\sum_{\substack{v,w \in V(\mathcal{G}) \\ \text{distinct}}} \mathrm{codeg}(v, w) = 3e(\mathcal{G}).$$

This is true for the same reason that the sum of the degrees in a graph equals twice the number of edges. Namely, every hyperedge of G appears exactly three times in the sum on the left-hand side.

Using this, we see that by Jensen's inequality,

$$X = \sum_{\substack{v,w \in V(\mathcal{G}) \\ \text{distinct}}} \binom{\operatorname{codeg}(v,w)}{s} \geqslant \binom{n}{2}\binom{\frac{1}{\binom{n}{2}}\sum_{v,w}\operatorname{codeg}(v,w)}{s} = \binom{n}{2}\binom{3e(\mathcal{G})/\binom{n}{2}}{s}.$$

Note too that since $e(\mathcal{G}) \geqslant Cn^{3-1/s^2}$, we have that $3e(\mathcal{G})/\binom{n}{2} \geqslant \Omega(n^{1-1/s^2})$. This implies that

$$X \geqslant \binom{n}{2}\binom{3e(\mathcal{G})/\binom{n}{2}}{s} \geqslant cn^2 \cdot (Cn^{1-1/s^2})^s = cC^s n^{2+s-1/s}$$

for some absolute constant $c > 0$, depending only on s.

On the other hand, we may upper-bound X by counting over s-sets of vertices which can be the "outer" vertices of the $K_{1,1,s}^{(3)}$. Namely, fix distinct $u_1, \dots, u_s \in V(\mathcal{G})$. We define a new graph (note: not a hypergraph, a *graph*) $G(u_1, \dots, u_s)$ as follows. The vertex set of $G(u_1, \dots, u_s)$ is $V(\mathcal{G}) \setminus \{u_1, \dots, u_s\}$. Moreover, given $v, w \in V(\mathcal{G}) \setminus \{u_1, \dots, u_s\}$, we make vw an edge of $G(u_1, \dots, u_s)$ if and only if $\{v, w, u_1, \dots, u_s\}$ form a copy of $K_{1,1,s}^{(3)}$.

Now, for every choice of $u_1, \dots, u_s$, we claim that $G(u_1, \dots, u_s)$ is a $K_{s,s}$-free graph. Indeed, if we had a copy of $K_{s,s}$ in $G(u_1, \dots, u_s)$, then we would find a copy of $K_{s,s,s}^{(3)}$ in $\mathcal{G}$, which is a contradiction. So by the Kővári–Sós–Turán theorem, we know that

$$e(G(u_1, \dots, u_s)) \leqslant O(n^{2-1/s})$$

for every choice of distinct $u_1, \dots, u_s \in V(\mathcal{G})$.

We can use this to upper-bound X, as follows. Note that $e(G(u_1, \dots, u_s))$ is precisely the number of copies of $K_{1,1,s}^{(3)}$ that have $u_1, \dots, u_s$ as the outer vertices. This implies that

$$X = \sum_{\substack{u_1, \dots, u_s \in V(G) \\ \text{distinct}}} e(G(u_1, \dots, u_s)) \leqslant \sum_{\substack{u_1, \dots, u_s \in V(G) \\ \text{distinct}}} O(n^{2-1/s}) = \binom{n}{s} \cdot O(n^{2-1/s}) = O(n^{2+s-1/s}).$$

Combining our upper and lower bounds on X, we see that

$$cC^s n^{2+s-1/s} \leqslant O(n^{2+s-1/s}),$$

where both c and the implicit constant in the big-O depend only on s. Thus, if we pick C sufficiently large, this is a contradiction, and we conclude that $\mathcal{G}$ has a copy of $K_{s,s,s}^{(3)}$. □

7 Supersaturation

In this section, we discuss a special case of a very important phenomenon in extremal combinatorics, known as *supersaturation*. Roughly speaking, extremal combinatorics proves results of the type "if some discrete structure is sufficiently 'large', then it contains at least one copy of some other structure". The example we've seen of this is Turán's theorem: if a graph (discrete structure) has sufficiently many edges (is large) then it contains a K_k subgraph (a copy of some other structure). Supersaturation, in general, boosts this to a statement of the type "if the discrete structure is just a bit larger, then it contains *very many* copies of the other structure". Specifically, we'll prove the following supersaturation version of Turán's theorem. It was first explicitly stated by Erdős and Simonovits in 1983, but it can implicitly be found in earlier works, e.g. of Erdős from 1971.

Theorem 7.1. *For every integer $k \geqslant 3$ and every real number $\varepsilon > 0$, there exists some $\delta > 0$ so that the following holds for all sufficiently large n. If G is an n-vertex graph with*

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

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

Note that G has *at most* $\binom{n}{k}$ copies of K_k, so this theorem is pretty remarkable: it says that once we have just barely more edges than the Turán graph, we have not only one copy of K_k, but a constant proportion of *all possible* copies of K_k. To prove this theorem, we need the following useful lemma, which is stated in greater generality than we need.

For a graph G and a subset $M \subseteq V(G)$, we denote by $e(M)$ the number of edges entirely contained in M, or equivalently the number of edges in the induced subgraph $G[M]$.

Lemma 7.2. *Let $0 < \alpha < \beta < 1$ be real numbers, let $m \geqslant 2$ be an integer, and let G be an n-vertex graph with $n \geqslant m$. Assume that $e(G) \geqslant \beta\binom{n}{2}$. Then the number of sets $M \subseteq V(G)$ with $|M| = m$ and $e(M) \geqslant \alpha\binom{m}{2}$ is at least $(\beta - \alpha)\binom{n}{m}$.*

Proof. The key identity which underlies this proof is

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

This has a simple bijective proof. On the right-hand side, every edge uv is counted a number of times, and that number of times is simply the number of m-sets M which contain both u and v. But the number of such m-sets is exactly $\binom{n-2}{m-2}$, yielding the formula.

Now, let $\mathcal{M}_0$ denote the set of M with $e(M) < \alpha\binom{m}{2}$, and let $\mathcal{M}_1$ denote the set of M with $e(M) \geqslant \alpha\binom{m}{2}$. So our goal is to prove a lower bound on $|\mathcal{M}_1|$. Continuing the identity above, we can write

$$\begin{aligned}\binom{n-2}{m-2}e(G) &= \sum_{M \in \mathcal{M}_0} e(M) + \sum_{M \in \mathcal{M}_1} e(M)\\ &\leqslant \sum_{M \in \mathcal{M}_0} \alpha\binom{m}{2} + \sum_{M \in \mathcal{M}_1}\binom{m}{2}\\ &= \binom{m}{2}\left(\alpha|\mathcal{M}_0| + |\mathcal{M}_1|\right)\end{aligned}$$

since every m-set in $\mathcal{M}_0$ has at most $\alpha\binom{m}{2}$ edges, and every m-set in $\mathcal{M}_1$ has at most $\binom{m}{2}$ edges.

Note that $|\mathcal{M}_0| + |\mathcal{M}_1| = \binom{n}{m}$. Let $x = |\mathcal{M}_1|/\binom{n}{m}$, so that $1 - x = |\mathcal{M}_0|/\binom{n}{m}$. Dividing by $\binom{n}{m}\binom{m}{2}$, the above inequality yields

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

Now, we recall that $e(G) \geqslant \beta\binom{n}{2}$, so

$$\frac{\binom{n-2}{m-2}}{\binom{n}{m}\binom{m}{2}} e(G) \geqslant \frac{\binom{n-2}{m-2}\binom{n}{2}}{\binom{n}{m}\binom{m}{2}} \beta.$$

The final step is another magic identity, which is that $\binom{n-2}{m-2}\binom{n}{2} = \binom{n}{m}\binom{m}{2}$; in other words, the complicated fraction above is simply equal to 1. Indeed, both sides of this identity count the same object, which is the number of ways of picking an m-set out of n objects, and then picking 2 objects from the m-set.

Combining all these inequalities, we find that

$$\beta \leqslant \alpha + (1-\alpha)x \qquad \Longleftrightarrow \qquad x \geqslant \frac{\beta - \alpha}{1-\alpha},$$

which implies that

$$|\mathcal{M}_1| = x\binom{n}{m} \geqslant \frac{\beta-\alpha}{1-\alpha}\binom{n}{m} \geqslant (\beta - \alpha)\binom{n}{m},$$

as claimed. □

With this lemma, we are ready to prove the supersaturation theorem, Theorem 7.1.

Proof of Theorem 7.1. Fix $\varepsilon > 0$. Recall that $t_{k-1}(m) = (1 - \frac{1}{k-1} + o(1))\binom{m}{2}$, where the $o(1)$ term tends to 0 as $m \to \infty$. This implies that there is some fixed m, depending only on ε, so that

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

Let this m be fixed, and let $n \geqslant m$. Suppose that G is an n-vertex graph with at least $(1 - \frac{1}{k-1} + \varepsilon)\binom{n}{2}$ edges. We apply Lemma 7.2 with $\beta = 1 - \frac{1}{k-1} + \varepsilon$ and $\alpha = 1 - \frac{1}{k-1} + \frac{\varepsilon}{2}$. Then Lemma 7.2 tells us that the number of m-sets $M \subseteq V(G)$ with $e(M) \geqslant (1 - \frac{1}{k-1} + \frac{\varepsilon}{2})$ is at least $\frac{\varepsilon}{2}\binom{n}{m}$.

Every such m-set M has strictly more than $t_{k-1}(m)$ edges, so Turán's theorem implies that such an M contains a copy of K_k. In other words, we've found at least $\frac{\varepsilon}{2}\binom{n}{m}$ copies of K_k, except that we might have over-counted: each copy of K_k can be counted up to $\binom{n-k}{m-k}$ times, since the k vertices of the K_k can appear in $\binom{n-k}{m-k}$ different m-sets M.

So in total, the number of K_k in G is at least

$$\frac{\frac{\varepsilon}{2}\binom{n}{m}}{\binom{n-k}{m-k}} = \frac{\varepsilon}{2} \cdot \frac{\binom{n}{m}}{\binom{n-k}{m-k}} = \frac{\varepsilon}{2} \cdot \frac{\binom{n}{k}}{\binom{m}{k}} = \frac{\varepsilon}{2\binom{m}{k}}\binom{n}{k},$$

where the middle equality uses the same magic identity as in the proof of Lemma 7.2, namely that $\binom{n}{m}\binom{m}{k} = \binom{n}{k}\binom{n-k}{m-k}$.

To conclude, we recall that m depends solely on ε and k. Therefore, if we define $\delta = \varepsilon/(2\binom{m}{k})$, then this will only depend on ε and k, and that yields the desired result. □

8 Proof of the Erdős–Stone–Simonovits theorem

We are finally ready to prove the Erdős–Stone–Simonovits theorem. We begin by observing a simple reduction, due to Erdős and Simonovits, which says that to prove the bound on $\mathrm{ex}(n, H)$ for all H, it suffices to prove it for a very special class of H. Let $K_k[s]$ denote the complete k-partite graph with parts of size s. (Note that this is the same graph as the Turán graph $T_k(ks)$.)

Proposition 8.1. *Suppose that for all positive integers k, s, we have that*

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

Then

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

for every graph H.

Proof. We already proved the lower bound in the Erdős–Stone–Simonovits theorem, namely that

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

So it only suffices to prove the upper bound. Now, the key claim is that if H has chromatic number k, then H is a subgraph of $K_k[s]$ for some positive integer s. Indeed, if H has chromatic number k, then we may split the vertices of H into k color classes, with the property that no edge of H goes between two vertices in the same color class. If s is the maximum size of one of the color classes, this precisely means that H is a subgraph of $K_k[s]$. But in that case, we see that

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

by assumption. □

So it suffices to prove what is often called the Erdős–Stone theorem, namely the statement that $\mathrm{ex}(n, K_k[s]) \leqslant (1 - \frac{1}{k-1} + o(1))\binom{n}{2}$ for every k, s. This is what we now do.

Proof of the Erdős–Stone theorem. Fix some $\varepsilon > 0$. Our goal is to prove that if n is sufficiently large in terms of ε, k, and s, and if G is an n-vertex graph with

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

edges, then G contains a copy of $K_k[s]$.

By the supersaturation theorem, Theorem 7.1, we know that G has at least $\delta\binom{n}{k}$ copies of K_k, where $\delta > 0$ depends only on ε and k. We define a k-uniform hypergraph $\mathcal{G}$ whose vertex set is $V(G)$, and we make a k-tuple of vertices a hyperedge of $\mathcal{G}$ if and only if the k-tuple defines a copy of K_k in G. Then we have that

$$e(\mathcal{G}) = \#(\text{copies of } K_k \text{ in } G) \geqslant \delta\binom{n}{k}.$$

Recall that by Theorem 6.5, we have that

$$\mathrm{ex}(n, K^{(k)}_{s,s,\dots,s}) \leqslant Cn^{k-1/s^{k-1}}$$

for some fixed constant $C > 0$. Now, if δ is fixed (which it is, since it only depends on ε and k), and if n is sufficiently large, then

$$\delta\binom{n}{k} > Cn^{k-1/s^{k-1}}. \tag{3}$$

This is because, as we've discussed previously, $\binom{n}{k}$ grows as $\Theta(n^k)$, and on the right-hand side we have a smaller power of n. So as long as n is sufficiently large in terms of the other parameters, we have that (3) holds.

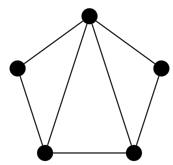
Thus, for sufficiently large n, we have that $e(\mathcal{G}) > \mathrm{ex}(n, K^{(k)}_{s,s,\dots,s})$, which implies that $\mathcal{G}$ contains a copy of $K^{(k)}_{s,s,\dots,s}$. In other words, inside $V(G)$, we can find k sets of s vertices each, with the property that whenever we pick one vertex from each part, they yield a copy of K_k in G. But that precisely means we have found a copy of $K_k[s]$ in G, as claimed. □

9 Beyond the Erdős–Stone–Simonovits theorem

The Erdős–Stone–Simonovits theorem gives a very satisfactory asymptotic answer to the question of how large $\mathrm{ex}(n, H)$ is for any non-bipartite H. However, we could still ask about more precise information. For example, when H is a complete graph, Turán's theorem gives us the *exact* value of $\mathrm{ex}(n, H)$ for all n, as well as a description of the unique extremal graph. Can we get something like this for more general graphs H?

Unfortunately, the answer is "no" in general. And the reason, somewhat surprisingly, is again that we don't understand bipartite graphs that well! In fact, the extremal theory of bipartite graphs is crucial to understanding the extremal theory of general graphs.

To understand this connection, let's begin with a simple example that you already saw on the homework. Let H_0 be the following graph:



It is not hard to verify that $\chi(H_0) = 3$, so the Erdős–Stone–Simonovits theorem implies that $\mathrm{ex}(n, H_0) \leqslant t_2(n) + o(n^2)$. However, unlike the case of triangles, where we know that $\mathrm{ex}(n, K_3) = t_2(n)$ exactly, for this graph H_0 we do not have an equality. Indeed, let us begin with the Turán graph $T_2(n)$, and call its two parts A and B. We now add a perfect matching inside A, and let G be the resulting graph. Then $e(G) = t_2(n) + \lfloor n/4 \rfloor$, and we claim that G is H_0-free. This essentially boils down to a case check: we need to show that no matter how we try to embed the five vertices of H_0 into the two parts of G, we will fail. More precisely, no matter how we assign the letters A and B to the vertices of H_0, either we will label two adjacent vertices by B, or we will label three vertices in a path by the letter A. In either case, we see that this would not be a valid embedding of H_0 into G, as B has no edges in G, and A has no two-edge paths.

How can we generalize this simple example? One first natural thing to try is to more generally understand for which graphs H, we do not have the equality $\mathrm{ex}(n, H) = t_{\chi(H)-1}(n)$, that is, for which graphs H is the Turán graph not extremal. Thinking about the example above, one can come up with the following definition.

Definition 9.1. A graph H is called *color-critical* if there is an edge $e \in E(H)$ for which

$$\chi(H \setminus e) < \chi(H).$$

That is, H is color-critical if we can decrease its chromatic number by deleting a single edge.

The example of H_0 discussed above readily extends to the following simple fact.

Proposition 9.2. *Let H be a graph with $\chi(H) \geqslant 3$. If H is not color-critical, then*

$$\mathrm{ex}(n, H) > t_{\chi(H)-1}(n)$$

for all n.

Proof. Let G be obtained from the Turán graph $T_{\chi(H)-1}(n)$ by adding a single edge in one of the parts (say the first part, for concreteness). We claim that G is H-free. Note that this suffices to prove the proposition, as it implies that

$$\mathrm{ex}(n, H) \geqslant e(G) = t_{\chi(H)-1}(n) + 1 > t_{\chi(H)-1}(n).$$

So suppose for contradiction that there were some copy of H in G. We use this to define a function $f : V(H) \to [\![\chi(H) - 1]\!]$ by mapping each vertex of H to the label of the part of G containing that vertex. Note that f cannot be a proper coloring of H, as this would contradict the definition of the chromatic number. Nonetheless, f is "almost" a proper coloring: at most one pair of adjacent vertices receive the same value under f, and this value must be 1. Indeed, since G contains no edges inside any part except the first one, the only way we can get an edge both of whose endpoints have the same value is if $f(u) = f(v) = 1$, and the copy of H in G uses the edge we inserted as the edge between u and v.

But this exactly shows that f *is* a proper $(\chi(H) - 1)$-coloring of $H \setminus e$, where $e = uv$. That is, f witnesses that

$$\chi(H \setminus e) \leqslant \chi(H) - 1,$$

contradicting our assumption that H is not color-critical. □

A rather amazing theorem of Simonovits shows that this simple necessary condition is actually sufficient!

Theorem 9.3 (Simonovits 1968). *Let H be a color-critical graph with $\chi(H) \geqslant 3$. Then for all sufficiently large n,*

$$\mathrm{ex}(n, H) = t_{\chi(H)-1}(n).$$

Morever, $T_{\chi(H)-1}(n)$ is the unique extremal graph.

For example, all complete graphs are color-critical, so this recovers Turán's theorem (at least for sufficiently large n). But it does more; for example, one can check that every odd cycle is color-critical, so Simonovits' theorem implies that

$$\mathrm{ex}(n, C_{2k+1}) = \left\lfloor \frac{n^2}{4} \right\rfloor$$

for all sufficiently large n. Note that the requirement that n be sufficiently large is necessary, since, for example, K_4 is C_5-free, hence

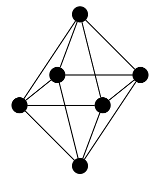
$$\mathrm{ex}(4, C_5) = \binom{4}{2} = 6 > \left\lfloor \frac{4^2}{4} \right\rfloor.$$

We will shortly see a proof of Theorem 9.3 (at least in the case $H = C_5$), but before we do, let us try to learn a bit more about $\mathrm{ex}(n, H)$ when H is not color-critical. One can do better than the argument we used in Proposition 9.2, as follows.

Definition 9.4. Let H be a graph with $\chi(H) \geqslant 3$. A *pseudo-coloring* of H is a function $f : V(H) \to [\![\chi(H) - 1]\!]$ with the property that for all $uv \in E(H)$, either $f(u) \neq f(v)$ or $f(u) = f(v) = 1$. That is, adjacent vertices receive different colors, or they both receive color 1.

Now, let B be a bipartite graph. We say that B is *in the decomposition family* of H if for every pseudo-coloring $f : V(H) \to [\![\chi(H) - 1]\!]$, there is a copy of B among the vertices colored 1 by f.

As an example, consider the octahedron graph O_3:



It is not hard to see that $\chi(O_3) = 3$, and with a little casework, one can verify that C_4 is in the decomposition family of O_3.

By adapting the proof of Proposition 9.2, we obtain the following.

Proposition 9.5. *Let H be a graph with $\chi(H) \geqslant 3$, and let B be a bipartite graph in the decomposition family of H. We have that*

$$\mathrm{ex}(n, H) \geqslant t_{\chi(H)-1}(n) + \mathrm{ex}\left(\left\lceil \frac{n}{\chi(H)-1} \right\rceil, B\right).$$

Proof sketch. Let G be obtained from $T_{\chi(H)-1}$ by inserting an extremal B-free graph into the largest part. By the definition of the extremal function of B, we have that

$$e(G) = t_{\chi(H)-1}(n) + \mathrm{ex}\left(\left\lceil \frac{n}{\chi(H)-1} \right\rceil, B\right).$$

Moreover, G is H-free, for essentially the same reason as in the proof of Proposition 9.2: if there were a copy of H in G, then we would obtain a pseudo-coloring of H in which the vertices of color 1 are B-free, contradicting the assumption that B is in the decomposition family of H. □

Concretely, this argument demonstrates that

$$\mathrm{ex}(n, O_3) \geqslant \left\lfloor \frac{n^2}{4} \right\rfloor + \Omega(n^{3/2}),$$

by our known lower bound for the extremal number of C_4. In fact, a famous theorem of Erdős and Simonovits shows that this is tight, in a very strong sense: every extremal O_3-free graph is obtained from a complete bipartite graph by putting an extremal C_4-free graph in one part, and a perfect matching in the other part. It is conjectured that this should happen more generally, namely that the extremal number of H should be obtained by inserting certain graphs into the parts of $T_{\chi(H)-1}(n)$, and ensuring that these inserted graphs avoid all members of the decomposition family of H. This conjecture remains open in general (and is probably quite difficult), but is known to hold in certain special cases. And in any case, it demonstrates why determining the exact behavior of $\mathrm{ex}(n, H)$ really requires one to understand extremal numbers of bipartite graphs, as the decomposition family of H is what ends up mattering.

10 Extremal numbers of general bipartite graphs

As such, it is a major research direction to study the extremal numbers of bipartite graphs. We have already discussed what we know about complete bipartite graphs, but we now turn our attention to $\mathrm{ex}(n, H)$ for general bipartite H.

As it turns out, we know a lot about this question, but there's much more that we don't know. We already saw a general-purpose upper bound, namely $\mathrm{ex}(n, H) \leqslant \mathrm{ex}(n, K_{s,t}) \leqslant O(n^{2-1/s})$ if H is a subgraph of $K_{s,t}$. But for many specific bipartite graphs, much better upper bounds are known, using arguments specific to the graph at hand. For example, the following is known about the extremal numbers of even cycles.

Theorem 10.1 (Erdős (unpublished), Bondy–Simonovits 1974)**.** *For every $\ell \geqslant 2$, we have*

$$\mathrm{ex}(n, C_{2\ell}) \leqslant O\left(n^{1+1/\ell}\right).$$

We won't quite prove this, but we'll prove a slightly weaker statement that is interesting in its own right; by being a bit more careful with essentially the same proof, one can prove Theorem 10.1. Recall from the homework that if $\mathcal{F}$ is a collection of graphs, then we say that G is $\mathcal{F}$-free if G contains no copy of any $H \in \mathcal{F}$, and we write

$$\mathrm{ex}(n, \mathcal{F}) = \max\{e(G) : G \text{ is an } n\text{-vertex } \mathcal{F}\text{-free graph}\}.$$

Theorem 10.2. *For every $\ell \geqslant 2$, if G is an n-vertex graph containing no cycle of length at most 2ℓ, then*

$$e(G) \leqslant O(n^{1+1/\ell}).$$

In other words,

$$\mathrm{ex}(n, \{C_3, C_4, \dots, C_{2\ell}\}) \leqslant O(n^{1+1/\ell}).$$

In the course of the proof, we will need the following extremely useful lemma which you already encountered on the homework.

Lemma 10.3. *If G is an n-vertex graph, then it has a subgraph G' with minimum degree at least $e(G)/n$.*

Proof. We repeatedly delete from G any vertex of degree strictly less than $e(G)/n$, until we get stuck. Note that we delete at most n vertices, and each time we delete a vertex we delete strictly fewer than $e(G)/n$ edges, hence the total number of edges we delete is strictly less than $e(G)$. That is, when we stop the process, we necessarily have at least one edge left. In particular, when we stop the process, we have some non-empty subgraph G' of G. Since we stopped the process at G', every vertex in G' must have degree at least $e(G)/n$, as claimed. □

We are now ready to prove Theorem 10.2.

Proof of Theorem 10.2. Let G be an n-vertex graph with at least $Cn^{1+1/\ell}$ edges, for some constant C that we'll pick later. Our goal is to show that G must contain a cycle of length at most 2ℓ. By Lemma 10.3, there is a subgraph G' of G with minimum degree at least $e(G)/n \geqslant Cn^{1/\ell}$. Let $m \leqslant n$ be the number of vertices of G', and let $d \geqslant Cn^{1/\ell}$ be the minimum degree of G'.

We now fix some vertex $v \in V(G)$. Let $N^1(v)$ denote the set of neighbors of v, so that $|N^1(v)| \geqslant d$. If there were some edge inside the set $N^1(v)$, it would form a triangle together with v; we would thus find a cycle of length 3 in G', and hence also in G, and we'd be done. We may therefore assume that there are no edges in $N^1(v)$. Since every vertex in $N^1(v)$ has degree at least d, it must have at least $d-1$ neighbors outside of $\{v\} \cup N^1(v)$.

Moreover, suppose that we had $x, y \in N^1(v)$ which had some common neighbor $z \neq v$. Then the vertices v, x, z, y would form a cycle of length 4 in G', and we'd again be done. Therefore, we may assume that all neighbors (apart from v) of vertices in $N^1(v)$ are distinct. In other words, if we define $N^2(v)$ to be the set of vertices at distance 2 from v—that is, the set of neighbors of vertices in $N^1(v)$—then we have that

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

since there are at least d vertices in $N^1(v)$, each of which contributes at least $d-1$ vertices in $N^2(v)$, and we cannot have over-counted.

We now continue in this way, defining $N^3(v), \dots, N^\ell(v)$. At every step of the process, we cannot have any any edges inside any of these sets, nor any collisions: no two vertices in $N^i(v)$ can have a common neighbor in $N^{i+1}(v)$, as this would yield a cycle of length at most 2ℓ in G'. This implies that the size of these sets grows by at least a factor of $d-1$ at every step, that is,

$$|N^i(v)| \geqslant d(d-1)^{i-1}$$

for every $i \leqslant \ell$. In particular,

$$|N^\ell(v)| \geqslant d(d-1)^{\ell-1} \geqslant \Omega(d^\ell) \geqslant \Omega(C^\ell n).$$

However, we also trivially have the bound $|N^\ell(v)| \leqslant m \leqslant n$, since $N^\ell(v)$ is a subset of $V(G')$. This is a contradiction if we pick C sufficiently large. □

As with complete bipartite graphs, it is again widely conjectured that the bound in Theorems 10.1 and 10.2 are tight, but it is only known to be tight in case $\ell \in \{2, 3, 5\}$. This is pretty remarkable: we know that

$$\mathrm{ex}(n, C_4) = \Theta(n^{3/2}) \qquad \mathrm{ex}(n, C_6) = \Theta(n^{4/3}) \qquad \mathrm{ex}(n, C_{10}) = \Theta(n^{6/5})$$

but we have no idea what the value of $\mathrm{ex}(n, C_8)$ is! You will see the construction providing this lower bound on the homework.

To end this section, let me just mention two remarkable conjectures of Erdős and Simonovits, which roughly say that the behavior of $\mathrm{ex}(n, H)$ for general bipartite H is very complicated.

Conjecture 10.4 (Erdős–Simonovits rational exponents conjecture)**.** *For every bipartite H, there exists some rational number $\alpha \in [1, 2)$ so that*

$$\mathrm{ex}(n, H) = \Theta(n^\alpha).$$

Moreover, the converse holds: for every rational $\alpha \in [1, 2)$, there exists some bipartite graph H so that

$$\mathrm{ex}(n, H) = \Theta(n^\alpha)$$

The first part of this conjecture is doubted by some experts, though no one has any idea how to prove or disprove it. However, the second part of the conjecture—that there exists a graph for any rational α—is widely believed to be true, and we are in fact fairly close to proving it. Every few months, a new paper appears finding a new infinite set of rational numbers that are now known to be "achievable", i.e. to be the exponent of $\mathrm{ex}(n, H)$ for some bipartite H.

Moreover, a slight weakening of the second part of the conjecture was recently proved by Bukh and Conlon.

Theorem 10.5 (Bukh–Conlon 2018)**.** *For every rational $\alpha \in [1, 2)$, there exists some finite collection $\mathcal{F}$ of bipartite graphs for which*

$$\mathrm{ex}(n, \mathcal{F}) = \Theta(n^\alpha).$$

11 Stability

Many problems in extremal combinatorics exhibit a phenomenon called *stability*, first discovered by Erdős and Simonovits. Stability is a stronger statement than the determination of an extremal structure: it states that if some structure has *close to* the maximum size given some constraint, then it must be *close to*, in a structural sense, the extremal construction.

In extremal graph theory, the stability form of Turán's theorem says something like the following: if G is an n-vertex K_r-free graph such that $e(G)$ is not much less than $t_{r-1}(n)$, then one can add or delete a small number of edges to G to obtain the Turán graph $T_{r-1}(n)$. Sometimes stating such results formally is a bit cumbersome, but in a moment we'll see a very clean statement along these lines.

In addition to being interesting in their own right, stability results are extremely useful for many applications. Indeed, we will shortly see a proof of (a special case of) Theorem 9.3 using stability; this proof gives a flavor of how stability arguments are often used in extremal graph theory.

But first, here is a formal statement of the stability version of Turán's theorem.

Theorem 11.1 (Füredi 2015)**.** *Let G be an n-vertex K_r-free graph, and suppose that*

$$e(G) \geqslant t_{r-1}(n) - s,$$

for some integer $s \geqslant 0$. Then we may delete at most s edges from G to obtain an $(r-1)$-partite graph.

Note that the case $s = 0$ recovers Turán's theorem. In most applications, however, one usually takes s to be something like εn^2 for a small constant ε. We will only prove Theorem 11.1 in the special case $r = 3$, and you will prove the general case on the homework.

Proof of Theorem 11.1 for $r = 3$. Let v be a vertex of maximum degree in G. Let $B = N(v)$ be the neighborhood of v, and let $A = V(G) \setminus B$. Since v had maximum degree, we have that $\deg(w) \leqslant |B|$ for every vertex w. In particular, applying this to all vertices in A, we find that

$$\sum_{w\in A} \deg(w) \leqslant |A||B|.$$

Moreover, since $A \cup B$ is a partition of $V(G)$ into two parts, we have that $|A||B| \leqslant \lfloor n^2/4 \rfloor = t_2(n)$. On the other hand, the sum $\sum_{w\in A} \deg(w)$ counts every edge between A and B exactly once, and every edge inside A exactly twice, once for each endpoint. Therefore,

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

Finally, we note that B is an independent set, since any edge inside B would form a triangle together with v, and we assumed that G is triangle-free. As a consequence, every edge in G lies either in A or between A and B. Putting this all together, we find that

$$\begin{aligned} e(G) &= e(A, B) + e(A) \\ &= (e(A, B) + 2e(A)) - e(A) \\ &= \left(\sum_{w\in A} \deg(w)\right) - e(A) \\ &\leqslant |A||B| - e(A) \\ &\leqslant t_2(n) - e(A). \end{aligned}$$

Rearranging, this says that

$$e(A) \leqslant t_2(n) - e(G) \leqslant s,$$

where the final inequality is our assumption on G. In other words, we have found a partition of $V(G)$ into two parts A, B, such that B is an independent set and A contains at most s edges. By deleting these s edges, we obtain a bipartite graph, completing the proof. □

While Theorem 11.1 only gives stability for Turán's theorem (i.e. gives an approximate structure for K_r-free graphs with many edges), one can deduce from it a stability version of the Erdős–Stone–Simonovits theorem, i.e. stability for H-free graphs for any non-bipartite H. You'll prove this on the homework. In the meantime, here is one such statement for $H = C_5$.

Proposition 11.2. *There is an absolute constant C such that the following holds. If G is an n-vertex C_5-free graph with $e(G) \geqslant \lfloor n^2/4 \rfloor$, then we may delete at most $Cn^{5/3}$ edges from G to obtain a bipartite graph.*

Proof. Let F be the subgraph of G consisting of all edges of G which lie on a triangle. We claim that F is $K_{3,3}$-free. Indeed, suppose that there were a $K_{3,3}$ in F, say with vertices $a_1, a_2, a_3, b_1, b_2, b_3$, where each a is adjacent to each b in F. As every edge in F lies on a triangle in G, there is some $c \in V(G)$ such that a_1, b_1, c form a triangle. If $c \notin \{a_2, b_2\}$, then we obtain a C_5 in G, namely $c \sim a_1 \sim b_2 \sim a_2 \sim b_1 \sim c$. Similarly, if $c \notin \{a_3, b_3\}$, we obtain a C_5 for the same reason, interchanging the roles of 2 and 3. As c cannot lie in both $\{a_2, b_2\}$ and $\{a_3, b_3\}$, we have proved that F is $K_{3,3}$-free. Consequently, $e(F) \leqslant \mathrm{ex}(n, K_{3,3}) \leqslant cn^{5/3}$, for some absolute constant c, by Theorem 5.3.

That is, at most $cn^{5/3}$ edges of G lie on a triangle. By deleting these edges, we obtain a triangle-free subgraph G' of G, with $e(G') \geqslant e(G) - cn^{5/3} \geqslant t_2(n) - cn^{5/3}$. Now applying Theorem 11.1 to G' with $s = cn^{5/3}$, we conclude that we can delete at most $cn^{5/3}$ additional edges from G' to obtain a bipartite graph. Putting this together, we find that G can be made bipartite by deleting at most $2cn^{5/3}$ edges, proving the result by setting $C = 2c$. □

We now turn to an application of the stability method, namely the proof of Theorem 9.3 in the case $H = C_5$. In order to keep things relatively simple, we will also assume a minimum degree condition, which can be removed with a fairly straightforward, but tedious, argument (see the homework). Here is the precise statement we will prove.

Proposition 11.3. *There is some integer n_0 such that for all $n \geqslant n_0$, the following holds. If G is an n-vertex C_5-free graph with minimum degree at least $n/3$, then $e(G) \leqslant \lfloor n^2/4 \rfloor$, with equality if and only if $G = T_2(n)$.*

Proof. We pick n_0 so that for all $n \geqslant n_0$, we have $Cn^{5/3} \leqslant n/1000$, where C is the constant from Proposition 11.2. By Proposition 11.2, we know that we can make G bipartite by deleting at most $Cn^{5/3}$ edges. That is, there exists a partition $V(G) = A \cup B$ such that $e(A) + e(B) \leqslant Cn^{5/3}$. We pick an optimal such partition, that is, a partition that minimizes $e(A) + e(B)$. We note that every vertex $a \in A$ has at least as many neighbors in B as in A. For if this were not the case, we could move a to B and strictly decrease $e(A) + e(B)$, contradicting our choice of the optimal partition. Similarly, every vertex $b \in B$ has at least as many neighbors in A as in B.

First, we note that we may assume $\frac{2}{5}n < |A|, |B| < \frac{3}{5}n$. For if this is not the case, say if $|A| < \frac{2}{5}n$, then $|A||B| \leqslant (\frac{2}{5}n)(\frac{3}{5}n) = \frac{6}{25}n^2$. But then

$$e(G) = e(A) + e(B) + e(A,B) \leqslant e(A) + e(B) + |A||B| \leqslant \left(\frac{1}{1000} + \frac{6}{25}\right) n^2 < \left\lfloor \frac{n^2}{4} \right\rfloor,$$

which is what we wanted to prove.

Next, suppose that some vertex $a \in A$ has at least $n/30$ neighbors in A. By the observation above, it also has at least $n/30$ neighbors in B. Let A', B' be its set of neighbors in A, B, respectively. Then the subgraph induced by A' and B' cannot contain a four-vertex path P_4, as such a path would yield a C_5 together with a. By an exercise on the homework,

$$e(A', B') \leqslant \mathrm{ex}(n, P_4) \leqslant n.$$

Note that in a complete bipartite graph, A' and B' would have $|A'||B'| \geqslant n^2/900$ edges between them, so there are a ton of "missing edges". More precisely, we see that

$$e(A, B) \leqslant |A||B| - |A'||B'| + e(A', B') \leqslant \frac{n^2}{4} - \frac{n^2}{900} + n.$$

This again implies that

$$e(G) = e(A) + e(B) + e(A, B) < \left(\frac{1}{1000} + \frac{1}{4} - \frac{1}{900}\right) n^2 + n < \left\lfloor \frac{n^2}{4} \right\rfloor,$$

as claimed, where we use the fact that $n \geqslant n_0$ is sufficiently large in the final inequality.

So we may assume that every vertex in A has fewer than $n/30$ neighbors in A. By the minimum degree assumption, it has at least $n/3$ neighbors in total, hence at least $3n/10$ neighbors in B. This implies that any two vertices in A must have a common neighbor in B; indeed, if they did not, then their neighborhoods would be two disjoint subsets of B each of size at least $3n/10$, which is impossible since $|B| < \frac{3}{5}n = 2 \cdot (\frac{3n}{10})$.

We now claim that B is an independent set. Indeed, suppose that there were some edge bb', for $b, b' \in B$. Pick neighbors a, a' of b, b'. By the argument above, a and a' have a common neighbor in B, say b''. But then $b \sim a \sim b'' \sim a' \sim b' \sim b$ is a C_5, a contradiction. The exact same argument, but interchanging the roles of A and B, shows that A is an independent set.

Thus, we have boosted the stability result we started with to an exact structural description: G *is* bipartite, with parts A and B. But now, the only way for G to have the maximum possible number of edges is if $G = T_2(n)$, proving the claim. □

12 Ramsey theory

Ramsey theory is the study of structure and of disorder. The main message of Ramsey theory is that *complete disorder is impossible*—any sufficiently large system, no matter how disordered, must contain within it some highly structured component. This general, highly unintuitive, philosophy manifests itself in topics as diverse as computer science, number theory, geometry, functional analysis, and, of course, graph theory, which is the topic we will mostly be focused on.

However, as Ramsey theory has connections to so many other areas of mathematics and beyond, we will also frequently pause to see how the results we have proved connect to these other fields. This is, in fact, how we begin the course, with perhaps the first-ever Ramsey-theoretic result, published by Issai Schur while Frank Ramsey was only fourteen years old.

12.1 Ramsey theory before Ramsey

Like many other people, Schur was interested in Fermat's last theorem, the statement that the equation $x^q + y^q = z^q$ has no non-trivial integer solutions x, y, z for any fixed $q \geqslant 3$, where a solution is *trivial* if $0 \in \{x, y, z\}$ and *non-trivial* otherwise.

Proving Fermat's last theorem is (very) hard, so let's start with something simpler. There are, of course, non-trivial integer solutions to the Pythagoras equation $x^2 + y^2 = z^2$. What if we change the equation slightly, to, say, $x^2 + y^2 = 3z^2$? After playing around with it for a bit, you might be tempted to conjecture that now, there are no non-trivial integer solutions.

This conjecture is indeed true, and there is a standard technique in number theory for proving such results. Namely, if there *were* some non-trivial solution $x, y, z \in \mathbb{Z}$ to the equation $x^2 + y^2 = 3z^2$, then there would also be a non-trivial[2] solution to the same equation modulo 3, namely the equation $x^2+y^2 \equiv 0 \pmod 3$. However, we know that that $1^2 \equiv 2^2 \equiv 1 \pmod 3$, and we can conclude that there *do not* exist non-trivial solutions modulo 3.

A similar argument can be used to prove that many other polynomial equations have no non-trivial integer solutions, and a general phenomenon called the *Hasse principle* very roughly says that in many instances, such a technique is guaranteed to work. So it is natural to wonder whether Fermat's last theorem can also be proved in this way. This is the question that motivated Schur[3], who proved that this technique *cannot* work for Fermat's last theorem.

Theorem 12.1 (Schur). *For any integer $q \geqslant 3$, there exists an integer $N = N(q)$ such that the following holds for any prime $p > N$. There exist non-zero $x, y, z \in \mathbb{Z}/p$ with*

$$x^q + y^q \equiv z^q \pmod p.$$

As Schur himself realized, despite proving an important and impressive result in number theory, his proof used almost no number theory! He wrote "daß [Theorem 12.1] sich fast unmittelbar aus einem sehr einfachen Hilfssatz ergibt, der mehr der Kombinatorik als der Zahlentheorie angehört."[4] This Hilfssatz is the following.

Theorem 12.2 (Schur). *For any positive integer q, there exists an integer $N = N(q)$ such that the following holds. If $[\![N]\!]$ is colored in q colors, then there exist $x, y, z \in [\![N]\!]$, all receiving the same color, such that $x + y = z$.*

Recall that we use the notation $[\![N]\!] := \{1, \ldots, N\}$. We also now start to use the terminology of *coloring*. By a coloring of $[\![N]\!]$ with q colors, we just mean a partition of $[\![N]\!]$ into q sets $A_1, \ldots, A_q$, where we think of the elements of A_1 as receiving a first color, the elements of A_2 as receiving some second, distinct, color, and so on. We will also frequently use the shorthand *monochromatic* for "receiving the same color", so the conclusion of Theorem 12.2 could also be stated as the existence of a monochromatic solution to $x + y = z$.

[2]One has to be a bit careful here, as a non-trivial solution over $\mathbb{Z}$ may become trivial in $\mathbb{Z}/3$. However, it is not hard to get around this issue, as one can argue that a *minimal* non-trivial solution over $\mathbb{Z}$ cannot have all three of x, y, z divisible by 3.

[3]In fact, the same question had motivated Dickson a few years earlier, and he was the first to prove Theorem 12.1. However, his technique used very messy casework and does not at all connect to Ramsey theory, so we won't discuss it any further.

[4]"that [Theorem 12.1] follows almost immediately from a very simple lemma, which belongs more to combinatorics than to number theory."

As Schur wrote, the derivation of Theorem 12.1 from Theorem 12.2 is almost immediate, but as it requires a few ideas from number theory and group theory, we will defer it for the moment. Let us first see how to prove Theorem 12.2. Schur proved Theorem 12.2 directly, but the modern, Ramsey-theoretic, perspective is to reduce Theorem 12.2 to an even more combinatorial lemma, which we now state.

Lemma 12.3. *For any positive integer q, there exists an integer $N = N(q)$ such that the following holds. If the edges of the complete graph K_N are q-colored, then there exists a monochromatic triangle.*

Proof. We will actually prove something stronger, namely an explicit upper bound on $N(q)$; we will show that $N(q) = 3q!$ satisfies the desired condition. We proceed by induction on q.

The base case $q = 1$ is immediate. We are claiming that any 1-coloring of the edges of K_N, where $N = 3 \cdot 1! = 3$, contains a monochromatic triangle. But as there is only one color, and the complete graph we are "coloring" is itself a triangle, this is certainly true.

For the inductive step, suppose the result is true for $q-1$, i.e. that any $(q-1)$-coloring of $E(K_{3(q-1)!})$ contains a monochromatic triangle. Fix a q-coloring of $E(K_N)$, where $N = 3q!$, and let v be any vertex of K_N. v is incident to $N-1$ edges, each of which receives one of q colors. Therefore, by the pigeonhole principle, there is some color, say red, which appears on at least

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

edges incident to v. Let R denote the set of endpoints of these red edges, and consider the coloring restricted to R. If there is any red edge appearing in R, then it forms a red triangle together with v, and we are done. If not, then R is a set of at least $3(q-1)!$ vertices that are colored by at most $q-1$ colors, and we can find a monochromatic triangle in R by the inductive hypothesis. In either case we are done. □

With Lemma 12.3 in hand, the proof of Theorem 12.2 is almost immediate. All we need to do is to translate the number-theoretic coloring into a graph-theoretic coloring.

Proof of Theorem 12.2. Let $N(q) = 3q!$ be chosen so that Lemma 12.3 holds. We are given a q-coloring χ of $[\![N]\!]$, which we convert to a q-coloring $\hat{\chi}$ of $E(K_N)$ as follows. Identify the vertices of K_N with $[\![N]\!]$, and then color an edge ab, where $1 \leqslant a < b \leqslant N$, according to the color of $b - a \in [\![N]\!]$ in χ.

As $\hat{\chi}$ is a q-coloring of $E(K_N)$, by Lemma 12.3, there is a monochromatic triangle in $\hat{\chi}$. Let the vertices of this triangle be a, b, c, where $a < b < c$. Let $x = b - a, y = c - b$, and $z = c - a$, and note that these satisfy $x + y = z$. Finally, note that they all receive the same color under χ, since $\chi(x) = \hat{\chi}(ab), \chi(y) = \hat{\chi}(bc)$, and $\chi(z) = \hat{\chi}(ac)$, and we assumed that a, b, c is a monochromatic triangle under $\hat{\chi}$. □

This completes the combinatorial part of Schur's work. For completeness, let's see how to derive Theorem 12.1 from Theorem 12.2. As this topic is somewhat outside the main narrative of the class, it will not be covered in lecture.

Deduction of Theorem 12.1 from Theorem 12.2

Proof of Theorem 12.1. Let $N = N(q)$ be as in Theorem 12.2, and fix a prime $p > N$. We recall the well-known fact that the set $\Gamma \coloneqq \{x^q : 0 \neq x \in \mathbb{Z}/p\}$ forms a subgroup of the multiplicative group $(\mathbb{Z}/p)^\times$, and the index of this subgroup is at most† q. Therefore, there are at most q cosets of Γ which partition the non-zero elements of $\mathbb{Z}/p$. By identifying the non-zero elements of $\mathbb{Z}/p$ with $[\![p-1]\!] \supseteq [\![N]\!]$, we obtain a q-coloring of $[\![N]\!]$ according to these cosets.

Now, by Theorem 12.2, there must exist monochromatic $a, b, c \in [\![N]\!]$ such that $a + b = c$. As these three numbers receive the same color, they must lie in some single coset $\alpha\Gamma$ of Γ, for some $\alpha \in (\mathbb{Z}/p)^\times$. By the definition of Γ, this means that we can write

$$a \equiv \alpha x^q \pmod p, \qquad b \equiv \alpha y^q \pmod p, \qquad c \equiv \alpha z^q \pmod p,$$

for some non-zero $x, y, z \in \mathbb{Z}/p$. The equation $a + b = c$ remains true when we reduce it mod p, so we conclude that

$$\alpha x^q + \alpha y^q \equiv \alpha z^q \pmod p.$$

As α is invertible in $\mathbb{Z}/p$, and as $x, y, z \neq 0$, we obtained the desired non-trivial solution $x^q + y^q \equiv z^q \pmod p$. □

†More precisely, the index is exactly $\gcd(q, p-1)$.

13 Classical Ramsey numbers

13.1 Ramsey's theorem and upper bounds on Ramsey numbers

While Schur's theorem can be seen as an early example of Ramsey theory, the theory did not really get going until Frank Ramsey's pioneering work in 1929. Ramsey's theorem, as it is now called, is a generalization of Lemma 12.3 from triangles to arbitrary cliques.

Theorem 13.1 (Ramsey). *For all positive integers k, q, there exists an integer $N = N(k, q)$ such that the following holds. If the edges of the complete graph K_N are q-colored, then there exists a monochromatic K_k, that is, k vertices such that all the $\binom{k}{2}$ edges between them receive the same color.*

Given this theorem, which we will shortly prove, we can make a definition that will be central for much of the rest of the course.

Definition 13.2. Given positive integers k, q, the *q-color Ramsey number of K_k*, denoted $r(k; q)$, is the least N such that the conclusion of Theorem 13.1 is true. That is, $r(k; q)$ is the minimum integer N such that every q-coloring of $E(K_N)$ contains a monochromatic K_k.

In case $q = 2$, we usually abbreviate $r(k; 2)$ as simply $r(k)$, and usually refer to the 2-color Ramsey number as simply the *Ramsey number*.

In this language, Theorem 13.1 can equivalently be stated as saying that $r(k; q) < \infty$ for all k, q. In fact, for much of this course, we will be interested not just in the fact that such Ramsey numbers are finite, but in quantitative estimates on how large they are.

For now, let's focus on the case $q = 2$. Ramsey's original proof of Theorem 13.1 showed that $r(k) \leqslant k!$ for all k. But a few years later, a different proof was found by Erdős and Szekeres, in another foundational paper of the field. In order to present their proof, we need to define a slightly more general notion of Ramsey number.

Definition 13.3. Given positive integers k, ℓ, we denote by $r(k, \ell)$ the *off-diagonal Ramsey number*, defined to be the least N such that every 2-coloring of $E(K_N)$ with colors red and blue contains a red K_k or a blue K_ℓ.

Note that $r(k, \ell) = r(\ell, k)$ as the colors play symmetric roles, and that $r(k) = r(k, k)$.

Theorem 13.4 (Erdős–Szekeres)**.** *For all positive integers k, ℓ, we have*

$$r(k, \ell) \leqslant \binom{k+\ell-2}{k-1}.$$

In particular, we have

$$r(k) \leqslant \binom{2k-2}{k-1} < 4^k.$$

Proof. We proceed by induction on $k + \ell$, with the base case[5] $k = 1$ or $\ell = 1$ being trivial. For the inductive step, the key claim is that the following inequality holds:

$$r(k, \ell) \leqslant r(k-1, \ell) + r(k, \ell-1). \tag{4}$$

To prove (4), fix a red/blue coloring of $E(K_N)$, where $N = r(k-1, \ell)+r(k, \ell-1)$, and fix some vertex $v \in V(K_N)$. Suppose for the moment that v is incident to at least $r(k-1, \ell)$ red edges, and let R denote the set of endpoints of these red edges. By definition, as $|R| \geqslant r(k-1, \ell)$, we know that R contains a red K_{k-1} or a blue K_ℓ. In the latter case we have found a blue K_ℓ (so we are done), and in the former case we can add v to this red K_{k-1} to obtain a red K_k (and we are again done).

So we may assume that v is incident to fewer than $r(k-1, \ell)$ red edges. By the exact same argument, just interchanging the roles of the colors, we may assume that v is incident to fewer than $r(k, \ell-1)$ blue edges. But then the total number of edges incident to v is at most

$$(r(k-1, \ell) - 1) + (r(k, \ell-1) - 1) = N - 2,$$

which is impossible, as v is adjacent to all $N - 1$ other vertices. This is a contradiction, proving (4).

[5] If you don't like starting the induction with $k = 1$—what does a monochromatic K_1 mean, exactly?—you should convince yourself that the base case $k = 2$ or $\ell = 2$ also works.

We can now complete the induction. By (4) and the inductive hypothesis, we find that

$$r(k,\ell) \leqslant r(k-1,\ell)+r(k,\ell-1) \leqslant \binom{(k-1)+\ell-2}{(k-1)-1}+\binom{k+(\ell-1)-2}{k-1}=\binom{k+\ell-2}{k-1},$$

where the final equality is Pascal's identity for binomial coefficients. □

A similar argument works when the number of colors is more than 2. If we denote by $r(k_1,\ldots,k_q)$ the *off-diagonal multicolor Ramsey number* (defined in the natural way), we obtain the following generalization of Theorem 13.4, which you will prove on the homework.

Theorem 13.5. *For all positive integers q and $k_1,\ldots,k_q$, we have*

$$r(k_1,\ldots,k_q) \leqslant \binom{k_1+\cdots+k_q-q}{k_1-1,\ldots,k_q-1},$$

where the right-hand side denotes the multinomial coefficient. In particular,

$$r(k;q) < q^{qk}.$$

13.2 Lower bounds on Ramsey numbers

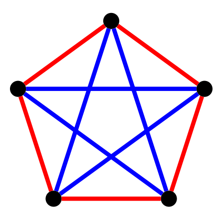
The Erdős–Szekeres bound, Theorem 13.4, gives us the upper bound $r(k) < 4^k$, which improves on Ramsey's earlier bound of $r(k) \leqslant k!$. To understand how good this bound is, we would like to obtain some *lower bounds* on $r(k)$.

Thinking about the definition of Ramsey numbers, we see that proving a lower bound of $r(k) > N$ boils down to exhibiting a 2-coloring of $E(K_N)$ with no monochromatic K_k. Perhaps the simplest such coloring is the *Turán coloring*, which proves the following result.

Proposition 13.6. *For any positive integer k, we have $r(k) > (k-1)^2$.*

Proof. Let $N=(k-1)^2$. We split the vertex set of K_N into $k-1$ parts, each of size $k-1$. We color all edges within a part red, and all edges between parts blue. The red graph is a disjoint union of $k-1$ copies of K_{k-1}, so there is certainly no red K_k. On the other hand, as there are only $k-1$ parts, the pigeonhole principle implies that any set of k vertices must include two vertices in one part; these two vertices span a red edge, and thus there is no blue K_k either. □

Is Proposition 13.6 tight? It's not too hard to see that the answer is no. Indeed, already for $k=3$, Proposition 13.6 implies that $r(3) > 4$, and it is not hard to show that in fact $r(3) > 5$, as witnessed by the following coloring.



Nonetheless, it is not clear how to do much better than Proposition 13.6 in general. Indeed, in the 1940s, Turán believed that the Erdős–Szekeres bound is way off, and that the truth is $r(k) = \Theta(k^2)$ (i.e. that Proposition 13.6 is best possible up to a constant factor). As it turns out, this belief was *way* off.

Theorem 13.7 (Erdős)**.** *For any $k \geqslant 2$, we have $r(k) \geqslant 2^{k/2}$.*

Together with Theorem 13.4, this proves that $r(k)$ really does grow as an exponential function of k, although these theorems do not tell us the precise growth rate. Theorem 13.7 was a major breakthrough not only—or even primarily—because of the result itself. In proving Theorem 13.7, Erdős introduced the so-called *probabilistic method* to combinatorics. This method would quickly become one of the most important tools in combinatorics, and will recur frequently throughout this course.

Proof of Theorem 13.7. Fix k, and let[6] $N = 2^{k/2}$. The claimed bound is trivial for $k = 2$, so let's assume $k \geqslant 3$. Consider a *random* 2-coloring of $E(K_N)$. Namely, for each edge of K_N, we assign it color red or blue with probability $\frac{1}{2}$, making these choices independently over all edges. We begin by estimating the probability that this coloring contains a monochromatic K_k.

For any fixed set of k vertices, the probability that it forms a monochromatic K_k is precisely $2^{1-\binom{k}{2}}$. This is because we have $\binom{k}{2}$ coin tosses, which we need to all agree, and we have two options for the shared outcome (hence the extra $+1$ in the exponent). Moreover, there are exactly $\binom{N}{k}$ possible k-sets we need to consider. Therefore,

$$\Pr(\text{there is a monochromatic } K_k) \leqslant \binom{N}{k} 2^{1-\binom{k}{2}},$$

where we have applied the *union bound* $\binom{N}{k}$ times; this is the bound that says that the probability that A *or* B happens is at most the sum of the probability that A happens and the probability that B happens.

Note that $\binom{N}{k} < N^k/k!$ and that $k! > 2^{1+k/2}$ for all $k \geqslant 3$. Therefore, we have

$$\binom{N}{k} 2^{1-\binom{k}{2}} < \frac{N^k}{k!} \cdot 2^{1-\frac{k^2-k}{2}} < \frac{N^k}{2^{1+\frac{k}{2}}} \cdot 2^{1+\frac{k}{2}-\frac{k^2}{2}} = \left(N \cdot 2^{-\frac{k}{2}}\right)^k = 1, \tag{5}$$

where the final equality is our choice of N.

Putting this all together, we find that in this random coloring, the probability that there is a monochromatic K_k is *strictly less than one*. Therefore, there must exist *some* coloring of $E(K_N)$ with no monochromatic K_k, as if such a coloring did not exist, the probability above would be exactly one. This completes the proof. □

[6] The astute reader will notice that $2^{k/2}$ is not an integer unless k is even. Thus, we should really write here $N = \lceil 2^{k/2} \rceil$. However, once the computations we do become more complicated, keeping track of such floor and ceiling signs becomes not just annoying, but actively confusing. Therefore, for the rest of the course, we'll omit floor and ceiling signs unless they are actually crucial, and it will be understood that any quantity that should be an integer but doesn't look like one should be rounded up or down to an integer.

It's worth stressing the miraculous magic trick that takes place in the proof of Theorem 13.7. Unlike in Proposition 13.6, Erdős does not give any sort of explicit description of a coloring on $2^{k/2}$ vertices with no monochromatic K_k. Instead, he argues that such a coloring must exist for probabilistic reasons, but this argument gives absolutely no indication of what such a coloring looks like. In fact, the following remains a major open problem.

Open problem 13.8 (Erdős)**.** *For some $\varepsilon > 0$ and all sufficiently large k, explicitly construct a 2-coloring on $(1+\varepsilon)^k$ vertices with no monochromatic K_k.*

There was a great deal of partial progress over the years, much of it exploiting a deep and surprising connection to the topic of *randomness extraction* in theoretical computer science. Very recently, there was a major breakthrough on this problem.

Theorem 13.9 (Li)**.** *For some absolute constant $\varepsilon > 0$ and all sufficiently large k, there is an explicit 2-coloring on 2^{k^ε} vertices with no monochromatic K_k.*

The central open problem in Ramsey theory is to narrow the gap between the lower and upper bounds $2^{k/2} \leqslant r(k) \leqslant 4^k$. For over 75 years, there was a great deal of interest in this question, and while there were several important developments, none of them were able to improve either of the constants in the bases of the exponents. But very recently, there was a huge breakthrough on this problem.

Theorem 13.10 (Campos–Griffiths–Morris–Sahasrabudhe)**.** *There is an absolute constant $\delta > 0$ such that $r(k) \leqslant (4-\delta)^k$ for all k.*

Their original proof showed roughly that $r(k) \leqslant 3.993^k$. A later result by Gupta, Ndiaye, Norin, and Wei, which both optimized the original technique and introduced beautiful new ideas, shows that $r(k) \leqslant 3.8^k$, which remains the current record. The proof of Theorem 13.10 is far too complex to cover in this course, but I have written an exposition of it that you can find on my website.

Additionally, just *yesterday*, there was another breakthrough, due to Ma, Shen, and Xie, this time on the lower bound. Erdős's random argument naturally extends to the off-diagonal setting, and proves that

$$r(k, Ck) \geqslant (f(C) + o(1))^k,$$

for some explicit function f, where we think of $C \geqslant 1$ as an absolute constant and let $k \to \infty$. Ma, Shen, and Xie improved this bound for any $C > 1$, proving that for any $C > 1$, there exists some $\varepsilon > 0$ such that

$$r(k, Ck) \geqslant (f(C) + \varepsilon + o(1))^k.$$

They also use the probabilistic method, but consider a different probability distribution coming from high-dimensional geometry. Unfortunately, their technique does not, at the moment, give any improvement on the *diagonal* Ramsey number $r(k)$.

14 Hypergraph Ramsey numbers

14.1 The hypergraph Ramsey theorem

We saw in the proof of the Erdős–Stone–Simonovits theorem that it can be very useful to study hypergraph analogues of graph-theoretic resutls. In the present context, there is a natural analogue of Ramsey's theorem for hypergraphs, which was also proved by Ramsey. Recall that $K_N^{(t)}$ denotes the complete t-uniform hypergraph on N vertices.

Theorem 14.1 (Ramsey). *For all integers $k \geqslant t \geqslant 2, q \geqslant 2$, there exists some N such that the following holds. In any q-coloring $\chi : E(K_N^{(t)}) \to [\![q]\!]$, there is a monochromatic copy of $K_k^{(t)}$. In other words, there exist k vertices such that each of the $\binom{k}{t}$ t-tuples among them receive the same color under χ.*

Continuing our earlier practice, we define the *t-uniform Ramsey number* $r_t(k; q)$ to be the least N for which Theorem 14.1 is true, and we use the shorthand $r_t(k)$ when $q = 2$. We also define the *off-diagonal t-uniform Ramsey number* $r_t(k_1, \ldots, k_q)$ to be the least N so that in any q-coloring of $E(K_N^{(t)})$, there is a monochromatic copy of $K_{k_i}^{(t)}$ in color i, for some $i \in [\![q]\!]$. Similarly, for any t-uniform hypergraphs $\mathcal{H}_1, \ldots, \mathcal{H}_q$, we denote by $r_t(\mathcal{H}_1, \ldots, \mathcal{H}_q)$ the least N such that any q-coloring of $K_N^{(t)}$ contains a monochromatic copy of $\mathcal{H}_i$ in color i, for some $i \in [\![q]\!]$, and write $r_t(\mathcal{H}; q)$ for shorthand if $\mathcal{H}_1 = \cdots = \mathcal{H}_q = \mathcal{H}$.

Probably the most natural way to prove Theorem 14.1 is via the following argument, directly mimicking the proof of Theorem 13.4.

Proof of Theorem 14.1. Let us only deal with the case $q = 2$. We prove by induction on t the statement that $r_t(k, \ell)$ exists for all $k, \ell \geqslant t$, and for any fixed t we prove this statement by induction on $k + \ell$. Note that the base case $t = 2$ is already done by Theorem 13.1, so we fix some $t \geqslant 3$ and assume the result has been proved for $t - 1$. For this fixed t, the base case $k = t$ or $\ell = t$ is trivial, so we may assume the result has been proved for the pairs $(k - 1, \ell)$ and $(k, \ell - 1)$.

The key claim is that the following recursive bound holds, analogously to (4):

$$r_t(k, \ell) \leqslant r_{t-1}(r_t(k - 1, \ell), r_t(k, \ell - 1)) + 1. \tag{6}$$

Note that we are done if we prove (6), since by induction, we know that the numbers $a \coloneqq r_t(k - 1, \ell)$ and $b \coloneqq r_t(k, \ell - 1)$ are finite, as is the number $r_{t-1}(a, b)$. Thus, (6) implies Theorem 14.1, at least in the case $q = 2$.

To prove (6), let $N = r_{t-1}(r_t(k - 1, \ell), r_t(k, \ell - 1)) + 1$, and consider any 2-coloring $\chi : E(K_N^{(t)}) \to \{\text{red}, \text{blue}\}$. Fix a vertex $v \in V(K_N^{(t)})$. There is a bijection between hyperedges containing v and $(t - 1)$-tuples of vertices in $V(K_N^{(t)}) \setminus \{v\}$. That is, we can use χ to define a coloring $\psi : E(K_{N-1}^{(t-1)}) \to \{\text{red}, \text{blue}\}$, by setting

$$\psi(\{w_1, \ldots, w_{t-1}\}) \coloneqq \chi(\{w_1, \ldots, w_{t-1}, v\}).$$

By the definition of N, we know that ψ contains a monochromatic red clique of order $r_t(k-1, \ell)$, or a monochromatic blue clique of order $r_t(k, \ell-1)$. The two cases are symmetric, so let us assume we are in the first. Looking at χ on these $r_t(k-1, \ell)$ vertices, we can either find a monochromatic blue $K_\ell^{(t)}$, or a monochromatic red $K_{k-1}^{(t)}$. In the first case we are done. In the second case, we have $k-1$ vertices, such that each of the t-tuples among them are colored red. Moreover, by the definition of ψ, if we combine any $(t-1)$-tuple from this set with v, we obtain another t-tuple that is colored red by χ. That is, we have found a monochromatic red $K_k^{(t)}$, showing that we are done in this case as well. □

Remark. While this proof is clearly reminiscient of the proof of Theorem 13.4, you might think that some things are different. For example, (6) is a bit different from (4), in that the former has this strange r_{t-1} term, whereas the latter simply has a sum. It is worth pondering what a 1-*uniform hypergraph* should be, and what the 1-uniform version of Theorem 14.1 should say. If you think about this enough, you'll come to realize that the proof above really is nothing more than a generalization of the proof of Theorem 13.4.

14.2 A geometric application

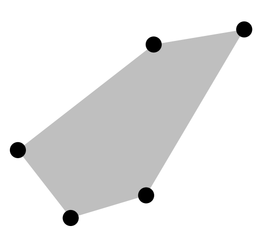
The paper of Erdős and Szekeres in which they proved Theorem 13.4—one of the most influential and foundational papers in the field—was titled "A combinatorial problem in geometry". We will now study this geometric problem, and see how it relates to Ramsey theory.

Definition 14.2. Let $p_1, \ldots, p_k$ be points in $\mathbb{R}^d$. A point $p \in \mathbb{R}^d$ is *in their convex hull* if there exist numbers $\lambda_1, \ldots, \lambda_k \geqslant 0$ with $\sum_{i=1}^k \lambda_i = 1$ such that

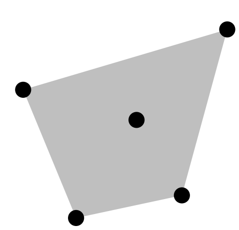
$$p = \sum_{i=1}^{k} \lambda_i p_i.$$

That is, p is in the convex hull of $p_1, \ldots, p_k$ if p is a weighted average of them.

Definition 14.3. A collection $p_1, \ldots, p_k$ of points in $\mathbb{R}^d$ is *in convex position* if no p_i is in the convex hull of $p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_k$.



Five points in convex position
(the gray region is their convex hull)

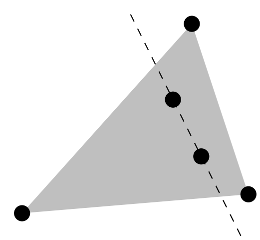


Five points *not* in convex position
(the gray region is their convex hull)

The question studied by Erdős and Szekeres begins with a simple observation of Klein.

Proposition 14.4 (Klein)**.** *Among any five points in $\mathbb{R}^2$, no three of them collinear, there are four points in convex position.*

Proof. Consider the convex hull of the five points. It is a polygon with at most five vertices. If it has four or five vertices, then four of these vertices yield our four desired points in convex position. So we may assume that the convex hull is a triangle, meaning that the final two points lie inside the triangle, as shown in the following picture.



Consider the line through the two interior points. Since no three points are collinear, two of the vertices of the triangle must lie on one side of this line. But then these two vertices, plus the two interior points, yield four points in convex position. □

Although this was before Ramsey theory really existed, Klein realized that there was a Ramsey-theoretic flavor to this result. She asked Erdős and Szekeres whether Proposition 14.4 could be generalized to finding arbitrarily large collections of points in convex position. Erdős and Szekeres proved that the answer is yes.

Theorem 14.5 (Erdős–Szekeres)**.** *For every $k \geqslant 4$, there exists some N such that the following holds. Among any N points in $\mathbb{R}^2$, no three of them collinear, there are k points in convex position.*

We will see two proofs of this theorem (and another proof is in the homework); the first is the original proof of Erdős and Szekeres.

Erdős and Szekeres's proof of Theorem 14.5. We will show that the theorem holds with $N = r_4(5, k)$. Fix N points $p_1, \ldots, p_N$ in $\mathbb{R}^2$, no three of them collinear. We identify $V(K_N^{(4)})$ with $\{p_1, \ldots, p_N\}$, and define a two-coloring of $E(K_N^{(4)})$ as follows. Given a 4-tuple $\{p_a, p_b, p_c, p_d\}$, we color it blue if these four points are in convex position, and red otherwise.

The first observation is that we cannot have a monochromatic red $K_5^{(4)}$. Indeed, this would correspond to five points in the plane, no three collinear, such that *every* 4-tuple among them is not in convex position. Proposition 14.4 says that such a configuration cannot exist.

Therefore, by the choice of N, there must exist k points, say $p_1, \ldots, p_k$, such that each hyperedge among them is colored blue. That is, every 4-tuple among them *is* in convex position. To complete the proof, we require the following simple lemma.

Lemma 14.6 (Carathéodory's theorem)**.** *Let $p_1, \ldots, p_k$ be a collection of points in $\mathbb{R}^2$, such that each 4-tuple among them is in convex position. Then $p_1, \ldots, p_k$ are in convex position.*

In a moment, we will give a formal proof of Lemma 14.6, but the intuitive proof is the following. Suppose for contradiction that $p_1, \ldots, p_k$ are not in convex position, and say without loss of generality that p_k is in the convex hull of $p_1, \ldots, p_{k-1}$, and call this convex hull P. Then P is a convex polygon, whose vertices are (some subset of) $p_1, \ldots, p_{k-1}$. Pick an arbitrary triangulation of P, that is, a partition of P into triangles whose vertices are vertices of P itself. Since $p_k \in P$, we must have that p_k is contained in one of the triangles of the triangulation. But that means that p_k is in the convex hull of three vertices of P; this yields four points out of $p_1, \ldots, p_k$ which are not in convex position.

Given Lemma 14.6, the proof is complete: we have found k points from our original collection that are in convex position. □

While the geometric proof sketch presented above can be made rigorous, there is also a fairly simple linear-algebraic proof of Lemma 14.6, which we now present.

Proof of Lemma 14.6. We may assume that $k \geqslant 5$, for otherwise there is nothing to prove. Suppose for contradiction that one of the points, say p_k, is in the convex hull of of the remaining points. This means that there exist numbers $\lambda_1, \ldots, \lambda_{k-1} \geqslant 0$ with $\sum \lambda_i = 1$ and

$$p_k = \sum_{i=1}^{k-1} \lambda_i p_i.$$

Let us fix such a collection $\lambda_1, \ldots, \lambda_{k-1}$ with the fewest number of non-zero elements. That is, we may assume by renaming the points that $\lambda_1, \ldots, \lambda_t > 0$, that $\lambda_{t+1}, \ldots, \lambda_{k-1} = 0$, and that no such representation is possible with fewer than t non-zero coefficients.

If $t \leqslant 3$, then we have shown that the points p_1, p_2, p_3, p_k are not in convex position (since p_k is in the convex hull of p_1, p_2, p_3), contradicting our assumption that all 4-tuples are in convex position. Therefore we may assume that $t \geqslant 4$. Consider the vectors

$$v_1 \coloneqq p_1 - p_t, \qquad v_2 \coloneqq p_2 - p_t, \qquad \ldots, \qquad v_{t-1} \coloneqq p_{t-1} - p_t.$$

These are $t - 1 \geqslant 3$ vectors in $\mathbb{R}^2$, so they must be linearly dependent. That is, there exist $\alpha_1, \ldots, \alpha_{t-1} \in \mathbb{R}$, at least one of which is non-zero, such that $\sum_{i=1}^{t-1} \alpha_i v_i = 0$. Now note that for any $\varepsilon \geqslant 0$, we have

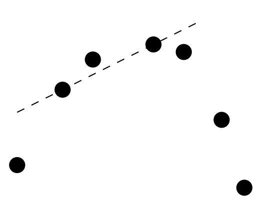
$$\begin{aligned} p_k &= \sum_{i=1}^{t} \lambda_i p_i \\ &= \lambda_t p_t + \sum_{i=1}^{t-1} \lambda_i p_i + \sum_{i=1}^{t-1} \varepsilon \alpha_i v_i \\ &= \lambda_t p_t + \sum_{i=1}^{t-1} [(\lambda_i + \varepsilon \alpha_i) p_i - \varepsilon \alpha_i p_t] \\ &= \sum_{i=1}^{t-1} (\lambda_i + \varepsilon \alpha_i) p_i + \left(\lambda_t - \varepsilon \sum_{i=1}^{t-1} \alpha_i \right) p_t \\ &=: \sum_{i=1}^{t-1} \mu_i(\varepsilon) p_i + \mu_t(\varepsilon) p_t. \end{aligned}$$

Notice that each $\mu_i(\varepsilon)$ is a continuous (in fact, linear) function of ε. Also, by assumption, we have that $\mu_i(0) > 0$ for all $i \in [\![t]\!]$. Also, by construction, we have that $\sum_i \mu_i(\varepsilon) = 1$ for all ε. However, since one of the α_i is non-zero, we see that in the limit $\varepsilon \to \infty$, at least one of the $\mu_i(\varepsilon)$ must become negative. Therefore, there is some smallest value ε^* such that $\mu_i(\varepsilon^*) = 0$ for at least one i, and $\mu_j(\varepsilon^*) \geqslant 0$ for all $j \neq i$. However, this gives us a new representation of p_k as a convex combination of $p_1, \ldots, p_{k-1}$ with fewer non-zero coefficients, contradicting our choice of $\lambda_1, \ldots, \lambda_{k-1}$. □

An alternative proof of Theorem 14.5 was found by Tarsi, who showed how to obtain the same result by using a diagonal 3-uniform Ramsey theorem, rather than the off-diagonal 4-uniform Ramsey theorem used by Erdős and Szekeres.

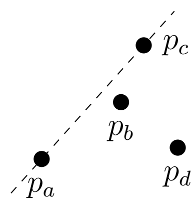
Tarsi's proof of Theorem 14.5. Let $N = r_3(k)$, and fix points $p_1, \ldots, p_N$ in $\mathbb{R}^2$. By rotating the plane if necessary, we may assume that all the points $p_1, \ldots, p_N$ have distinct x-coordinates. Let us also relabel them so that they are sorted by x-coordinate, that is, so that p_1 is to the left of p_2, which is to the left of p_3, and so on. We identify $V(K_N^{(3)})$ with $\{p_1, \ldots, p_N\}$, and color $E(K_N^{(3)})$ as follows. For $1 \leqslant i < j < \ell \leqslant N$, we color the hyperedge $\{p_i, p_j, p_\ell\}$ red if p_j lies above the line $p_i p_\ell$, and blue if p_j lies below the line $p_i p_\ell$.

By the choice of N, there is a monochromatic $K_k^{(3)}$, say $p_{i_1}, \ldots, p_{i_k}$, where $i_1 < \cdots < i_k$. Let us suppose this $K_k^{(3)}$ is red. This means that every point in this set lies above the line between its neighbors on the left and right; intuitively, this means that the points need to look like this:



In particular, the points $p_{i_1}, \dots, p_{i_k}$ are in convex position, as is hopefully intuitive from the picture. This is in fact true, and is a discrete version of the well-known fact that a function with non-positive second derivative is concave.

To prove that $p_{i_1}, \dots, p_{i_k}$ are in convex position, it suffices by Lemma 14.6 to show that any four of them are in convex position. So let p_a, p_b, p_c, p_d be four points, ordered from left to right, with the property that each of the triples they define is red, that is, that each point lies above the line connecting its two neighbors. If they are not in convex position, then one of them must be in the convex hull of the other three, and it is not hard to see that the interior point must be either p_b or p_c (p_a and p_d are necessarily extreme points because they minimize and maximize, respectively, the x-coordinate among these four points). If, say, p_b is in the convex hull of p_a, p_c, p_d, then we see that p_b lies below the line between p_a and p_c, a contradiction.



Similarly, if p_c is an interior point, it lies below the line joining p_b, p_d, another contradiction. This shows that all 4-tuples are indeed in convex position, and thus we have found our desired k-set in convex position by Lemma 14.6. In case $\{p_{i_1}, \dots, p_{i_k}\}$ form a blue clique, the same argument works after vertically reflecting the whole picture. □

14.3 Bounds on hypergraph Ramsey numbers

The proof we saw of Theorem 14.1 shows that $r_t(k, \ell)$ is finite for all t, k, ℓ. However, the bound it gives is absolutely enormous. For example, just trying to upper-bound $r_3(k, k)$, we find from (6) that

$$r_3(k) \leqslant r_2(r_3(k-1, k), r_3(k, k-1)) + 1.$$

Plugging in our bound $r_2(a) < 4^a$, this implies that

$$r_3(k) \leqslant 4^{r_3(k-1,k)}.$$

That is, a single step of the recursion has cost us an exponential! Continuing in this way, this proof yields a bound roughly of the form

$$r_3(k) \leqslant 4^{4^{\cdot^{\cdot^{\cdot^{4}}}}} \Big\} {\scriptstyle 2k \text{ times}}.$$

But then the bound in uniformity 4 is then much worse—a single step of the recursion (6) for $t = 4$ shows that $r_4(k)$ is bounded as a tower-type function of $r_4(k-1, k)$. That is, this proof yields a wowzer-type bound on $r_4(k)$, and in general, the bounds it gives for uniformity t are at the $(t-1)$th level of the Ackermann hierarchy.

Are such abysmal bounds necessary? At first glance, one might suspect that they are—exponential bounds really are the truth for $r_2(k)$, so the argument above is not particularly wasteful for uniformity 2. However, Erdős and Rado discovered an alternative proof of Theorem 14.1, which gives a much stronger bound.

Theorem 14.7 (Erdős–Rado)**.** *For all integers $t \geqslant 3, q \geqslant 2$, and $k_1, \ldots, k_q > t$, we have*

$$r_t(k_1, \ldots, k_q) \leqslant q^{1+\binom{r_{t-1}(k_1-1,\ldots,k_q-1)}{t-1}}.$$

In particular,

$$r_t(k; q) \leqslant q^{1+\binom{r_{t-1}(k-1)}{t-1}}.$$

Theorem 14.7 is sometimes called the *stepping-down* argument; it shows that we can bound a t-uniform Ramsey number by (an exponential function of) a $(t-1)$-uniform Ramsey number, that is, we step down one level in the uniformity. As an immediate consequence, we obtain much stronger bounds on hypergraph Ramsey numbers: for any fixed t, the bound is a fixed tower of 2s.

Corollary 14.8. *We have*

$$r_3(k; q) \leqslant 2^{2^{(Cq \log q)k}}$$

for some absolute constant $C > 0$. Similarly,

$$r_4(k; q) \leqslant 2^{2^{2^{(C'q \log q)k}}},$$

and in general,

$$r_t(k; q) \leqslant \left.2^{2^{\cdot^{\cdot^{2^{(C_t q \log q)k}}}}}\right\} t-1 \text{ twos},$$

for some constant C_t depending only on t.

Additionally, there is a beautiful argument, called the *stepping-up lemma* of Erdős–Hajnal–Rado, which yields nearly matching lower bounds. At a high level, it allows us to convert a lower bound for $r_{t-1}(k/2; q)$ into a lower bound for $r_t(k; q)$ which is *exponentially* larger. In particular, it "should" allow us to close the gap above, by acting in concert with the stepping-down argument Theorem 14.7, as the two yield upper and lower bounds on $r_t(k; q)$ which are exponential in the $(t-1)$-uniform Ramsey number. However, there is an important catch: the stepping-up lemma only works if we start with a construction in uniformity 3 or above.

Theorem 14.9 (Erdős–Hajnal–Rado)**.** *For every $k \geqslant t \geqslant 3, q \geqslant 2$, we have*

$$r_{t+1}(2k + t - 4; q) > 2^{r_t(k;q)-1}.$$

As a corollary, we get a lower bound which "almost" matches Corollary 14.8, but there is a gap of 1 in the height of the tower.

Corollary 14.10. *We have*

$$r_4(k) \geqslant 2^{2^{ck^2}},$$

for some absolute constant $c > 0$. In general, for every $t \geqslant 4$, there is a constant $c_t > 0$ such that

$$r_t(k) \geqslant \left.2^{2^{\cdot^{\cdot^{2^{c_t k^2}}}}}\right\} t-2 \text{ twos}.$$

The most important open problem about hypergraph Ramsey numbers is to close this exponential gap. Note that if one closes this gap for any uniformity $t \geqslant 3$, then one automatically closes it for all higher uniformities, thanks to the stepping-down and stepping-up lemmas, Theorems 14.7 and 14.9. In particular, closing the gap for uniformity 3 would close it for all uniformities. It is generally believed that the upper bound is closer to the truth.

Conjecture 14.11 (Erdős–Hajnal–Rado)**.** *There exists an absolute constant $c > 0$ such that $r_3(k) \geqslant 2^{2^{ck}}$. As a consequence, for every $t \geqslant 3$, there exist constants $c_t, C_t > 0$ such that*

$$t-1 \text{ twos}\left\{2^{2^{\cdot^{\cdot^{2^{c_t k}}}}}\right. \leqslant r_t(k) \leqslant \left.2^{2^{\cdot^{\cdot^{2^{C_t k}}}}}\right\} t-1 \text{ twos}.$$

One important reason to believe this conjecture is that it is known to be true once the number of colors is at least four, via a variant of the stepping-up lemma due to Hajnal.

Theorem 14.12 (Hajnal)**.** *For every $k, q \geqslant 2$, we have*

$$r_3(k; 2q) > 2^{r_2(k-1;q)-1}.$$

In particular,

$$r_3(k; 4) > 2^{2^{ck}}$$

for some absolute constant $c > 0$.

15 Graph Ramsey numbers

15.1 Introduction

We will now move away to a more general topic than we have considered so far, that of *graph Ramsey numbers.*

Definition 15.1. Given graphs $H_1, \dots, H_q$, their *Ramsey number* $r(H_1, \dots, H_q)$ is defined as the minimum N such that any q-coloring of $E(K_N)$ contains a monochromatic copy of H_i in color i, for some $i \in [\![q]\!]$. Here, by a monochromatic copy, we mean a subgraph of K_N isomorphic to H_i, all of whose edges receive color i.

In case $H_1 = \cdots = H_q = H$, we denote this Ramsey number by $r(H; q)$. In case $q = 2$, we use the shorthand $r(H) := r(H; 2)$.

Of course, everything we have studied so far is a special case of these more general graph Ramsey numbers, as $r(k)$ is simply $r(K_k)$, and $r(k, \ell) = r(K_k, K_\ell)$, etc. However, it turns out that there is an extremely rich theory of Ramsey numbers of graphs H which are not necessarily complete graphs; moreover, most of the interesting results actually arise when H is extremely far from being a complete graph.

We begin with a simple observation, which is that if H_i is a subgraph of H_i', then $r(H_1, \ldots, H_q) \leqslant r(H_1', \ldots, H_q')$, since any monochromatic copy of H_i' also yields a monochromatic copy of H_i. Thus, $r(H) \leqslant r(H')$ whenever $H \subseteq H'$. Since every n-vertex graph is a subgraph of K_n, we conlude that

$$r(H) \leqslant r(K_n) < 4^n \qquad \text{for every } n\text{-vertex graph } H.$$

Thus, in the worst case, an n-vertex graph may have Ramsey number that is exponential in n.

On the other hand, the most general lower bound we can get is that $r(H) \geqslant n$ if H is an n-vertex graph. Indeed, we need at least n vertices to be able to "fit" a copy of H. Moreover, this trivial lower bound is best possible in general, for if H has no edges (or even one edge), then $r(H) = n$.

Thus, for a general n-vertex graph H, we know $n \leqslant r(H) \leqslant 4^n$, and both behaviors—linear in n and exponential in n—are possible, for the empty graph and the complete graph, respectively. Based on our experience for cliques, we might expect that the exponential bound should be closer to the truth for most graphs. However, the striking result that we will see is that for many "natural" classes of graphs—and, in fact, for all *sparse* graphs—the lower bound is much closer to the truth.

15.2 Ramsey numbers of trees

Let us begin with the following simple result, which was probably first observed by Erdős and Graham; it says that the lower bound is close to tight for trees.

Theorem 15.2. *If T is an n-vertex tree, then $r(T) \leqslant 4n - 3$.*

To prove this, we will use a simple lemma from graph theory, which you already saw on the homework.

Lemma 15.3. *Let T be an n-vertex tree. If G is a graph with minimum degree at least $n-1$, then $T \subseteq G$.*

Proof. We proceed by induction on n, with the base case $n = 1$ being trivial since the only 1-vertex tree is a subgraph of every non-empty graph. Inductively, suppose this is true for all $(n-1)$-vertex trees. Let T' be obtained from T by deleting a leaf v, and let u be the unique neighbor of v in T. By the inductive hypothesis, $T' \subseteq G$, so let us pick a copy of T' in G, and let w be the vertex of G filling the role of u. As G has minimum degree at least $n-1$, w has at least $n-1$ neighbors, and at most $n-2$ of these neighbors were used in embedding the other $n-2$ vertices of T'. Thus, there is at least one unused neighbor of w, which means that we can extend the T'-copy to a T-copy by adding in this unused neighbor. □

With this lemma, it is straightforward to prove Theorem 15.2.

Proof of Theorem 15.2. Let $N = 4n - 3$, and fix a 2-coloring of $E(K_N)$. Without loss of generality, we may assume that at least half the edges are red. Let $G \subseteq K_N$ be the graph comprising the red edges, so

$$e(G) \geqslant \frac{1}{2}\binom{N}{2} = \frac{1}{2} \cdot \frac{(4n-3)(4n-4)}{2} = N(n-1).$$

By Lemma 10.3, there is a subgraph $G' \subseteq G$ of minimum degree at least $n-1$. By Lemma 15.3, we have $T \subseteq G'$, and this yields a monochromatic red copy of T. □

15.3 Ramsey numbers of complete bipartite graphs

We now turn our attention to complete bipartite graphs $K_{s,t}$.

Theorem 15.4. *For any $s \leqslant t$, we have*

$$r(K_{s,t}) \leqslant 2^{s+1}t.$$

Note that, if we plug in $s = t = n$, then we obtain that $r(K_{n,n}) = O(n2^n)$. Since $K_{n,n}$ has $2n$ vertices, this is a much better, although still exponential, bound than the naïve one of

$$r(K_{n,n}) \leqslant r(2n) < 4^{2n} = 16^n.$$

We remark that $r(K_{n,n})$ really does grow exponentially in n, and that the lower bound

$$r(K_{n,n}) > 2^{\frac{n-1}{2}}$$

will follow from a more general result, Proposition 15.5, which we will prove shortly. On the other hand, if we think of s as a constant, we obtain that $r(K_{s,t}) = O_s(t)$ as $t \to \infty$. Since $K_{s,t}$ has $s + t \leqslant 2t$ vertices, this shows that for fixed s, $K_{s,t}$ has a Ramsey number which is linear in its number of vertices—the same behavior as we saw for trees.

The proof of Theorem 15.4 uses essentially the same strategy we used in proving the Kővári–Sós–Turán theorem.

Proof of Theorem 15.4. The case $s = 1$ follows from a homework problem; it also follows, up to an additive constant of 1, from Theorem 15.2, since $K_{1,t}$ is a tree. We henceforth assume that $t \geqslant s \geqslant 2$.

Let $N = 2^{s+1}t$, and fix a red/blue coloring of $E(K_N)$. For every vertex $v \in V(K_N)$, let $\deg_R(v), \deg_B(v)$ denote the red and blue degrees, respectively, of v. Let S denote the number of monochromatic copies of $K_{1,s}$ in the coloring. We can count S by summing over all N choices for the central vertex, and then picking s distinct neighbors; this shows that

$$S = \sum_{v \in V(K_N)} \left(\binom{\deg_R(v)}{s} + \binom{\deg_B(v)}{s} \right).$$

Note that $\deg_R(v) + \deg_B(v) = N - 1$ for every v, and that the sum $\binom{x}{s} + \binom{N-1-x}{s}$ is minimized[7] when $x = N - 1 - x$, i.e. $x = \frac{N-1}{2}$. Therefore, we find that

$$S \geqslant N \cdot 2\binom{\frac{N-1}{2}}{s}.$$

On the other hand, another way of counting S is by counting over all options for the s leaves of the star. Let us assume for contradiction that there is no monochromatic $K_{s,t}$. Then every s-set of vertices forms the set of leaves of fewer than t red stars $K_{1,s}$, and of fewer than t blue stars $K_{1,s}$. Thus,

$$S < 2t\binom{N}{s}.$$

Comparing our lower and upper bounds on S, we find that

$$2t\binom{N}{s} > 2N\binom{\frac{N-1}{2}}{s}$$

or equivalently

$$t \cdot N(N-1)\cdots(N-s+1) > N \cdot \frac{N-1}{2}\left(\frac{N-1}{2} - 1\right)\cdots\left(\frac{N-1}{2} - s + 1\right).$$

Rearranging, this is equivalent to

$$\frac{2^s t}{N} > \left(\frac{N-1}{N}\right)\left(\frac{N-3}{N-1}\right)\left(\frac{N-5}{N-2}\right)\cdots\left(\frac{N-2s+1}{N-s+1}\right) = \prod_{i=0}^{s-1}\frac{N-2i-1}{N-i}.$$

However, we have that

$$\prod_{i=0}^{s-1}\frac{N-2i-1}{N-i} = \prod_{i=0}^{s-1}\left(1 - \frac{i+1}{N-i}\right) \geqslant 1 - \sum_{i=0}^{s-1}\frac{i+1}{N-i} \geqslant 1 - \frac{2\binom{s+1}{2}}{N} \geqslant \frac{1}{2},$$

where the second inequality uses that $N \geqslant 2s$, hence $N - i \geqslant N/2$ for all $i \leqslant s - 1$, and the third inequality uses that $2\binom{s+1}{2} = (s+1)s \leqslant (s+1)t \leqslant 2^s t = N/2$, since $2^s \geqslant s + 1$ for all $s \geqslant 2$. Putting this all together, we conclude that

$$\frac{2^s t}{N} > \frac{1}{2},$$

which contradicts our choice of N. This contradiction completes the proof. □

[7] This is again a special case of Jensen's inequality. This special case can also be proved directly without appealing to convexity.

15.4 The Burr–Erdős conjecture

So far, we have seen several examples of graph Ramsey numbers, and observed different growth rates. First, we know that $r(K_n)$ grows exponentially in n. Similarly, $r(K_{n,n})$ grows exponentially in n (and thus in $2n$, which is its number of vertices). On the other hand, all trees, as well as complete bipartite graphs in which one side has constant size, have Ramsey numbers *linear* in the number of vertices. Can we figure out a general rule explaining these extremely different growth rates?

Looking at the examples above, it is natural to guess that the major difference has to do with *density*. Both K_n and $K_{n,n}$ are very dense graphs, namely graphs with a quadratic number of edges. On the other hand, trees and complete bipartite graphs with one side of constant size are very sparse, in that their number of edges is only linear in their number of vertices. Equivalently, the average degree of the former graphs is large—linear in the number of vertices—whereas it is *constant* for the latter graphs. Perhaps this explains the difference in the Ramsey numbers?

As it turns out, this is close to the correct explanation. One direction really is true; if a graph has high average degree, then its Ramsey number is large, as shown in the following simple proposition.

Proposition 15.5. *If H has average degree d, then $r(H) > 2^{\frac{d-1}{2}}$.*

Proof. The proof is very similar to that of Theorem 13.7. Let H have $k \geqslant 2$ vertices, and thus $kd/2$ edges. Let $N = 2^{\frac{d-1}{2}}$, and consider a uniformly random 2-coloring of $E(K_N)$. Every tuple of k vertices in K_N forms a monochromatic copy of H with probability $2^{1-kd/2}$, and there are $k!\binom{N}{k}$ such tuples[8]. Therefore, the probability that the coloring has a monochromatic copy of H is at most

$$k!\binom{N}{k} \cdot 2^{1-\frac{kd}{2}} < N^k \cdot 2^{1-\frac{kd}{2}} = 2^{k\frac{d-1}{2}+1-\frac{kd}{2}} = 2^{1-\frac{k}{2}} \leqslant 1,$$

and thus there must exist a coloring with no monochromatic copies of H. □

Thus, we find that if H has average degree which is linear in its number of vertices $v(H)$, then $r(H)$ is exponential in $v(H)$. Is it possible that the same holds at the opposite extreme, namely that if H has constant average degree, then $r(H)$ is linear in $v(H)$, as happened for trees and complete bipartite graphs? It is not hard to see that the answer is no.

Proposition 15.6. *There exists an n-vertex graph H with average degree at most 1 and with $r(H) > 2^{\sqrt{n}/2}$.*

Proof. Let H be obtained from a complete graph $K_{\sqrt{n}}$ by adding $n - \sqrt{n}$ isolated vertices. Then H has $\binom{\sqrt{n}}{2}$ edges, and thus average degree $\frac{2}{n}\binom{\sqrt{n}}{2} \leqslant 1$. However,

$$r(H) \geqslant r(K_{\sqrt{n}}) > 2^{\sqrt{n}/2},$$

[8] Note that we include an extra factor of $k!$, which was not present in the proof of Theorem 13.7. The reason is that K_k is highly symmetric; for a general H, we need to consider not only the k vertices that can define it, but also the potentially $k!$ different ways of identifying $V(H)$ with these k vertices.

by Theorem 13.7. □

Given this example, it's clear why the naïve conjecture "constant average degree implies linear Ramsey number" cannot be true. Namely, the graph H above has constant average degree, but it contains a subgraph (namely $K_{\sqrt{n}}$) with much higher average degree, and it is this subgraph that really determines $r(H)$. This shows that rather than considering the global average degree, we need to consider a more refined parameter that takes into account subgraphs that are denser than H itself. There are several different ways of formalizing such a parameter, and they end up giving essentially identical results; we will use the following.

Definition 15.7. The *degeneracy* of a graph H is defined as the maximum, over all subgraphs $H' \subseteq H$, of the minimum degree of H'. H is said to be *d-degenerate* if its degeneracy is at most d.

From Lemma 10.3, we see that a d-degenerate graph has average degree at most $2d$. On the other hand, the H in Proposition 15.6 is an example of a graph with constant average degree and degeneracy $\sqrt{n} - 1$. Thus, we see that having bounded degeneracy is a strictly stronger condition than having bounded average degree. In particular, Proposition 15.5 implies that graphs with high degeneracy have large Ramsey numbers, as shown in the following result.

Theorem 15.8. *Let H be a graph of degeneracy d. Then $r(H) > 2^{\frac{d-1}{2}}$.*

Proof. By the definition of degeneracy, there is a subgraph $H' \subseteq H$ with minimum degree at least d, and thus also average degree at least d. Then Proposition 15.5 implies that

$$r(H) \geqslant r(H') > 2^{\frac{d-1}{2}}.$$ □

Given this, we can now amend our naïve conjecture to the following fundamental conjecture of Burr and Erdős.

Conjecture 15.9 (Burr–Erdős)**.** *Graphs of bounded degeneracy have linear Ramsey numbers.*

More precisely, for every $d \geqslant 1$ there exists $C \geqslant 1$ such that the following holds. If an n-vertex graph H is d-degenerate, then $r(H) \leqslant Cn$.

The Burr–Erdős conjecture is now a theorem.

Theorem 15.10 (Lee)**.** *Conjecture 15.9 is true.*

Now that we know that the Burr–Erdős conjecture is true, we can start asking more refined questions. What if the degeneracy is not bounded, but instead grows as a function of n? The following conjecture predicts a fairly precise answer.

Conjecture 15.11 (Conlon–Fox–Sudakov)**.** *If H has n vertices and degeneracy d, then*

$$r(H) = 2^{\Theta(d+\log n)}.$$

If d is much larger than $\log n$, then this conjecture predicts that $r(H)$ is exponential in d, matching the lower bound from Theorem 15.8. On the other hand, if d is much smaller than $\log n$ (e.g. if d is bounded, as in the Burr–Erdős conjecture), then it predicts that $r(H)$ is polynomial in n. While this conjecture remains open, there are a number of partial results that come quite close to proving it, differing from the conjecture by only a polylogarithmic factor in the exponent.

Lee's proof of the Burr–Erdős is far too complicated to cover in this course, but we will try to see a few ideas in its direction. The Burr–Erdős conjecture has a long history, with many important partial results. The first major breakthrough in this direction was a theorem of Chvatál, Rödl, Szemerédi, and Trotter, which established the Burr–Erdős conjecture under the stronger assumption that H has bounded maximum degree.

Theorem 15.12 (Chvatál–Rödl–Szemerédi–Trotter). *Graphs of bounded maximum degree have linear Ramsey numbers.*

More precisely, for every $\Delta \geqslant 1$, there exists $C \geqslant 1$ such that the following holds. If an n-vertex graph H has maximum degree at most Δ, then $r(H) \leqslant Cn$.

This result was extremely important, and so was the proof technique they introduced; this theorem is the first result in Ramsey theory to be proved via the so-called *regularity method*, whose basis is the fundamental *regularity lemma* of Szemerédi. This method has become one of the most important techniques in Ramsey theory and in extremal graph theory more broadly. However, let us remark that this proof technique gives truly enormous bounds on how large C has to be as a function of Δ; namely their proof showed that Theorem 15.12 is true for

$$C = 2^{2^{\cdot^{\cdot^{2}}}} \Big\} 2^{100\Delta}.$$

This enormous bound is one of several reasons why many researchers attempted to find alternative proofs of Theorem 15.12.

There are now (at least) two other techniques known for proving Theorem 15.12, both of which are very important in their own right. One is the *dependent random choice* technique, which you've seen a glimpse of on the homework, and which is also the main technique underlying Lee's proof of Conjecture 15.9. The other is the *greedy embedding technique*, which was developed in this context by Graham, Rödl, and Ruciński, although it goes back in some form at least to much earlier work of Erdős and Hajnal. We will unfortunately not have time to discuss this technique in detail in this course, but let us see a high-level overview of how it works.

Proof sketch of Theorem 15.12 using greedy embedding. Let H be an n-vertex graph of maximum degree at most Δ, and let $N = Cn$ for a large constant $C = C(\Delta)$ that we choose appropriately. Fix a red/blue coloring of $E(K_N)$. Our goal is to attempt to find a red copy of H in a greedy manner; we'll then show that, if we fail, we will be able to find a blue copy of H.

Let us label the vertices of H as $v_1, \ldots, v_n$. Define $V_1 = V_2 = \cdots = V_n = V(K_N)$. We think of V_i as the set of candidate vertices for v_i, and will attempt to embed the vertices of H one at a time, at each step updating the set of candidate vertices. We fix some small parameter $\varepsilon > 0$.

Note that if we pick where to embed v_i into V_i, we need to update our candidate sets. Indeed, since our goal is to build a red copy of H, if we choose where to place v_i, we need to shrink each V_j, for all j such that $v_iv_j \in E(H)$, to only include the red neighbors of the chosen embedding of v_i. Let us call a vertex $w \in V_i$ *prolific* if it has the following property: if we choose to embed v_i as w, then each candidate set shrinks by at most a factor of ε. In other words, w is prolific if its red neighborhood in V_j has size at least $\varepsilon|V_j|$, for every j such that $v_iv_j \in E(H)$.

Our embedding rule is now as follows. If there is a prolific vertex in V_1, we embed v_1 there and update all the candidate sets appropriately. If there is now a prolific vertex in V_2, we embed v_2 there and update all the candidate sets. We continue in this way as long as we can.

If this process gets to the end, that is, if we embed v_n into V_n, then we have found a red copy of H. So we may assume that the process gets stuck at some step i. Note that every candidate set shrinks at most Δ times, since H has maximum degree at most Δ, and moreover every time it shrinks it does so by at most a factor of ε. Thus, when we get stuck, we still have that $|V_j| \geqslant \varepsilon^\Delta N$ for all j. In particular, $|V_i| \geqslant \varepsilon^\Delta N$. Moreover, since we got stuck, there is no prolific vertex in V_i. That is, for every vertex $w \in V_i$, there is some j such that the red neighborhood of w in V_j has size less than $\varepsilon|V_j|$. There are at most Δ options for this choice of j, so by the pigeonhole principle, there is some fixed $j \in [\![n]\!]$ and some set $W_i \subseteq V_i$ with $|W_i| \geqslant \frac{1}{\Delta}|V_i|$ such that every $w \in W_i$ has a red neighborhood in V_j of size less than $\varepsilon|V_j|$.

We have thus proved the following lemma. If this greedy embedding procedure ever gets stuck, we find two sets W_i, V_j with $|W_i| \geqslant \frac{1}{\Delta}\varepsilon^\Delta N$ and $|V_j| \geqslant \varepsilon^\Delta N$, and with the property that the density of red edges between W_i and V_j is less than ε. In other words, we have found two sets A_1, A_2 with $|A_1|, |A_2| \geqslant \frac{1}{\Delta}\varepsilon^\Delta N$, and such that the density of *blue* edges between A_1 and A_2 is at least $1 - \varepsilon$.

We now iterate this lemma, as follows. Inside A_1, we run the same procedure to attempt to greedily embed H in red. If we succeed, we are done. If we fail, we find two sets $A_{11}, A_{12} \subseteq A_1$ with blue density between them at least $1 - \varepsilon$, where $|A_{11}|, |A_{12}| \geqslant (\frac{1}{\Delta}\varepsilon^\Delta)^2 N$. We also run the same procedure inside A_2 to find two such sets A_{21}, A_{22}. Moreover, since the blue density between A_1 and A_2 was at least $1 - \varepsilon$, we can ensure[9] that the blue density between A_{1i} and A_{2j} is at least $1 - \varepsilon$, for all $i, j \in [\![2]\!]$.

In other words, we've now found *four* sets, each of size at least $(\frac{1}{\Delta}\varepsilon^\Delta)^2 N$, such that the blue density between every pair is at least $1 - \varepsilon$. Continuing in this manner k times, we can find 2^k such sets, each with size at least $(\frac{1}{\Delta}\varepsilon^\Delta)^k N$, and with all pairwise blue densities at least $1 - \varepsilon$. We now do this until $2^k \geqslant \Delta + 1$ (i.e. pick $k = \lceil \log(\Delta + 1) \rceil$), and we thus obtain at least $\Delta + 1$ sets, which we rename $B_1, \ldots, B_{\Delta+1}$.

[9]There is some subtlety in doing this step correctly; since A_{1i} and A_{2j} are rather small subsets of A_1, A_2, one needs an extra argument to ensure that the blue density remains high when we restrict to them. The trick to do this is to apply, essentially, Lemma 10.3 to always ensure that the minimum blue degree is high before shrinking.

Since H has maximum degree at most Δ, it is $(\Delta+1)$-colorable, i.e. it can be partitioned into $\Delta+1$ independent sets $C_1, \ldots, C_{\Delta+1}$. Note that

$$|B_i| \geqslant \left(\frac{1}{\Delta}\varepsilon^\Delta\right)^k N \geqslant n,$$

where we can ensure the final inequality by picking C sufficiently large as a function of Δ and ε (and thus k, which is itself a function of Δ). Thus, each set B_i is large enough to accommodate embedding C_i. Moreover, one can check that if ε is sufficiently small (e.g. $\varepsilon = \Delta^{-2}$ suffices), then the greedy embedding strategy we tried for red is now *guaranteed* to work in blue. Namely, we greedily embed H in blue, ensuring that all vertices of C_i get embedded into B_i, and updating all candidate sets at every step. The strong density conditions we know about blue imply that we will never get stuck. □

Examining the proof sketch above, we see that it gives a bound of the form $C \leqslant 2^{O(\Delta(\log\Delta)^2)}$. Moreover, in case H is bipartite, the iteration step is unnecessary, and we can simply take $k = 1$ in the proof above, and thus obtain a bound of $C \leqslant 2^{O(\Delta\log\Delta)}$. In other words, the greedy embedding technique allowed Graham, Rödl, and Ruciński to prove the following more refined version of Theorem 15.12.

Theorem 15.13 (Graham–Rödl–Ruciński)**.** *There exists an absolute constant $M > 0$ such that the following holds. If H is an n-vertex graph with maximum degree at most Δ, then*

$$r(H) \leqslant 2^{M\Delta(\log\Delta)^2} n.$$

Moreover, if H is bipartite, we have the stronger bound

$$r(H) \leqslant 2^{M\Delta\log\Delta} n.$$

Remarkably, Graham, Rödl, and Ruciński also proved that their upper bound is nearly tight, even for bipartite graphs.

Theorem 15.14 (Graham–Rödl–Ruciński)**.** *There exists an absolute constant $c > 0$ such that the following holds. For every $n > \Delta > 1$, there is an n-vertex bipartite graph H with maximum degree Δ which satisfies*

$$r(H) \geqslant 2^{c\Delta} n.$$

Looking back at the greedy embedding proof sketch above, one might be struck by the fact that the colors play such asymmetrical roles; we keep trying, insistently, to embed H in red, and only when we have failed many times do we relent and succeed in embedding it in blue. This asymmetry is in fact a weakness of the proof technique, and Conlon, Fox, and Sudakov were able to improve the bound of Theorem 15.13 to $r(H) \leqslant 2^{O(\Delta\log\Delta)} n$ for every n-vertex graph H with maximum degree Δ, by modifying the greedy embedding technique so that both colors play roughly the same role. Unfortunately, it is still not known if this technique can be used to remove the final logarithmic factor, and thus match the lower bound of Theorem 15.14.

Moreover, this discussion hints at another, more fundamental, weakness of the greedy embedding technique, which is that it is tailor-made for the two-color case. Indeed, the entire upshot of the technique is that *failing* to find H in red tells us something about the blue edges. In case there are three or more colors, it is not at all clear how to obtain useful information from the failure of the first embedding. As far as I am aware, no one has been able to use the greedy embedding technique to prove any results on $r(H; q)$ for any H and any $q \geqslant 3$.

16 Canonical Ramsey theorems

This section covers two somewhat disparate topics, which nonetheless share a thematic connection. The *extremely* high-level idea is the following. Most mathematical objects are endowed with a notion of sub-objects (e.g. subsets, subgraphs, subgroups, subspaces, subschemes, subterfuges…). Certain objects are *canonical*, in the sense that all of their sub-objects "look like" the original object. For example, an elementary result in group theory is that all subgroups of a cyclic group are cyclic; a more pronounced version of the same fact is that any subgroup of $\mathbb{Z}$ is isomorphic to $\mathbb{Z}$. A substantially deeper and more difficult statement along the same lines is that any subgroup of a free group is again free.

One question we are interested in is a full classification of such examples: for any given notion of mathematical object, what is a complete list of the canonical ones? Having accomplished this task (which requires formalizing what we mean by "looking like" the original object), one can turn to proving a Ramsey-theoretic statement, along the lines of "any sufficiently large object must contain an arbitrarily large canonical sub-object".

We can view Ramsey's theorem as an instance of this general philosophy. Indeed, consider the class of graphs, endowed with the sub-object relation of induced subgraphs. Then complete graphs and empty graphs are examples of canonical objects, since any induced subgraph of a complete graph is again complete, and any induced subgraph of an empty graph is empty. Moreover, Ramsey's theorem implies that every sufficiently large graph contains an arbitrarily large complete or empty induced subgraph.

16.1 Monotone sequences

Consider a sequence $a_1, \ldots, a_k$ of distinct real numbers. A natural definition for a "canonical" sequence is a monotone sequence (that is, a sequence which either strictly increasing or strictly decreasing), since any subsequence of an increasing sequence is again increasing, and the same holds for decreasing sequences.

As you might expect, there is a Ramsey-theoretic statement, asserting that every sequence of distinct real numbers contains a long monotone subsequence; this was proved in the same seminal paper of Erdős and Szekeres.

Theorem 16.1 (Erdős–Szekeres). *Given $k \geqslant 2$, let $N = (k-1)^2 + 1$. Then any sequence $a_1, \ldots, a_N$ of distinct real numbers contains a monotone subsequence of length k. That is,* *there exist indices* $1 \leqslant i_1 < \cdots < i_k \leqslant N$ *such that*

$$a_{i_1} < a_{i_2} < \cdots < a_{i_k} \qquad \text{or} \qquad a_{i_1} > a_{i_2} > \cdots > a_{i_k}.$$

There are many known proofs of this theorem; we will show a particularly elegant proof discovered by Seidenberg.

Proof of Theorem 16.1 (Seidenberg). For an index $m \in [\![N]\!]$, let $\delta(m)$ denote the length of the longest decreasing subsequence ending at a_m, and let $\iota(m)$ denote the length of the longest increasing sequence ending at a_m. We wish to prove that $\delta(m) \geqslant k$ or $\iota(m) \geqslant k$ for some $m \in [\![N]\!]$. So suppose for contradiction that this is not the case, that is, that $1 \leqslant \delta(m), \iota(m) \leqslant k-1$; note that we have a lower bound of 1 on both functions, since we can always view a_m itself as both an increasing and a decreasing subsequence ending at a_m.

This means that there are at most $(k-1)^2$ possible values for the pair $(\delta(m), \iota(m))$. Since $N = (k-1)^2+1$, the pigeonhole principle implies that there exists two indices $1 \leqslant \ell < m \leqslant N$ such that $(\delta(\ell), \iota(\ell)) = (\delta(m), \iota(m))$. Since the elements of our sequence are distinct, we have $a_\ell < a_m$ or $a_\ell > a_m$. Suppose first that $a_\ell < a_m$. Then any increasing sequence ending in a_ℓ can be extended by one to obtain an increasing sequence ending at a_m, implying that $\iota(m) > \iota(\ell)$, a contradiction. Similarly, if $a_\ell > a_m$, then $\delta(m) > \delta(\ell)$, another contradiction. In either case we are done. □

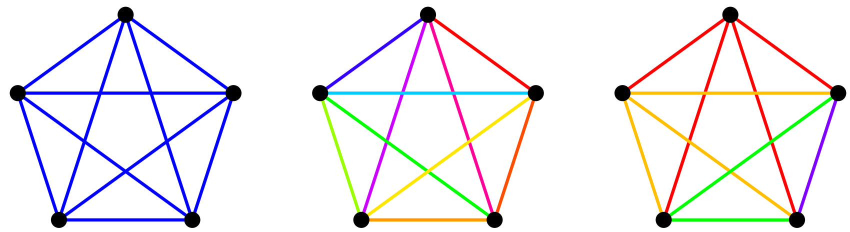
It is not hard to show (as you will do on the homework) that this bound is tight, in that there exist sequences of $(k-1)^2$ distinct real numbers with no monotone subsequence of length k.

16.2 The canonical Ramsey theorem

We now turn to the canonical Ramsey theorem for edge-colorings of the complete graph. Of course, as discussed above, Ramsey's theorem itself is such a statement—any coloring of a complete graph with a fixed number of colors must contain an arbitrarily large monochromatic clique, and monochromatic cliques are clearly canonical, as any subset of a monochromatic clique is another monochromatic clique. However, what if we remove the restriction that the number of colors is fixed?

That is, the question we are asking is the following: we color $E(K_N)$, for a large N, with an arbitrary number of colors. What kinds of subcolorings are canonical, in the sense that all of their induced subgraphs yield colorings of the same type? Certainly, monochromatic cliques are still canonical. On the other hand, once the number of colors is unbounded, we get a new type of canonical coloring: a *rainbow* coloring of K_N, in which each of the edges receives a different color (so $\binom{N}{2}$ colors are used in total).

It is tempting to conjecture that these are the only ones, but this turns out to not be the case. There is a third type of coloring, which we will call *starry*. A coloring of $E(K_N)$ is called starry if there are distinct colors $c_1, \ldots, c_{N-1}$ and if one can sort the vertices as $v_1, \ldots, v_N$, such that the color of the edge $v_i v_j$, where $i < j$, is c_i. In other words, each color class is a star, with the first star centered at v_1, the second centered at v_2 (and not containing v_1), and so on. Note that this is a canonical coloring, as any subset of vertices induces another starry coloring.



monochromatic rainbow starry

As it turns out, these really are the only canonical colorings, in the sense that a canonical Ramsey theorem holds: every sufficiently large edge-colored clique contains an arbitrarily large clique which is monochromatic, rainbow, or starry. This was proved by Erdős and Rado, in a result that is now usually called *the* canonical Ramsey theorem.

Theorem 16.2 (Erdős–Rado)**.** *For every $k \geqslant 2$, there exists some N such that if $E(K_N)$ is colored (with an arbitrary number of colors), there is a K_k which is monochromatic, rainbow, or starry.*

The original proof of Erdős and Rado used a clever reduction to the hypergraph Ramsey theorem in uniformity 4. Namely, for every 4-tuple of vertices, they considered the equivalence relation of colors on the $\binom{4}{2} = 6$ edges. That is, rather than remembering the actual colors on each of these 6 edges, they only record which pairs of edges receive the same color. As it turns out, there are 203 equivalence relations[10] on a set of size 6, so they obtain a 203-coloring of $E(K_N^{(4)})$. By Theorem 14.1, there is a monochromatic $K_k^{(4)}$ in this coloring (assuming N is sufficiently large), and an elementary argument (involving some casework) shows that in each of the 203 cases[11], this monochromatic $K_k^{(4)}$ yields a monochromatic, rainbow, or starry K_k in the original coloring.

However, from a quantitative perspective, the proof of Erdős and Rado is not very good. Letting $\mathrm{ER}(k)$ denote the least N such that Theorem 16.2 holds, the proof of Erdős–Rado only shows that $\mathrm{ER}(k) \leqslant r_4(k; 203) \leqslant 2^{2^{2^{O(k)}}}$, thanks to the bounds on hypergraph Ramsey numbers. A much better bound, with an alternative proof that is also extremely elegant, was found by Lefmann and Rödl.

Theorem 16.3 (Lefmann–Rödl)**.** *We have $\mathrm{ER}(k) \leqslant k^{4k^2}$ for all $k \geqslant 2$.*

In particular, Theorem 16.3 gives a finite bound on $\mathrm{ER}(k)$, thus proving Theorem 16.2. In the course of the proof of Theorem 16.3, we will need the following extremely useful lemma, which allows us to find rainbow cliques in edge-colored graphs where every color class is a graph with bounded maximum degree.

[10]The number of equivalence relations on a set of size n is given by the *Bell number* B_n, and $B_6 = 203$.

[11]In fact, it is not hard to show that most of the 203 cases are actually impossible, so the true number of cases is much smaller.

Lemma 16.4. *Let $k, M \geqslant 2$ be integers, and suppose that $E(K_M)$ is colored so that every vertex is incident to at most M/k^4 edges in every color. Then there is a rainbow K_k in this coloring.*

Proof. Every vertex must be incident to at least one edge of some color, hence no such coloring can exist if $M < k^4$. Thus the statement is vacuously true in these cases, and we may assume henceforth that $M \geqslant k^4$. Also, since every coloring of $E(K_2)$ is rainbow, we may assume henceforth that $k \geqslant 3$. Let χ be the coloring of $E(K_M)$.

Let $v_1, \dots, v_k$ be a uniformly random sequence of k distinct vertices from K_M. That is, we pick a set of k distinct vertices uniformly at random among the $\binom{M}{k}$ options, and then pick a random ordering of that set and label it $v_1, \dots, v_k$. Equivalently, we let v_1 be a uniformly random vertex, v_2 a uniformly random vertex among the remaining vertices, and so on. The key property that we need about this distribution is that if we condition on the outcome of any subset of these vertices, the marginal distribution of any remaining vertex is that of a uniformly random vertex of K_M, apart the ones already picked. Thus, for example, if x, y are two distinct vertices of K_M, and we condition on $v_3 = x, v_7 = y$, the marginal distribution of v_4 is uniformly random on the set $V(K_M) \setminus \{x, y\}$.

For distinct indices $i, j, \ell \in [\![k]\!]$, let $\mathcal{E}_{i,j,\ell}$ be the event that the edges $v_i v_j$ and $v_i v_\ell$ receive the same color. We wish to estimate $\Pr(\mathcal{E}_{i,j,\ell})$. Given two distinct vertices $x, y \in V(K_M)$, we begin by estimating $\Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y)$. Given $v_i = x, v_j = y$, the event $\mathcal{E}_{i,j,\ell}$ is simply the event that $\chi(xv_\ell) = \chi(xy)$, where the only randomness remaining is in the choice of v_ℓ. By assumption, x is incident to at most M/k^4 edges in color $\chi(xy)$, and v_ℓ is a uniformly random vertex in the set $V(K_M) \setminus \{x, y\}$, hence

$$\Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y) \leqslant \frac{1}{M-2} \cdot \frac{M}{k^4} \leqslant \frac{2}{k^4}.$$

Since the same upper bound holds for $\Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y)$ for all x, y, the same bound holds for $\Pr(\mathcal{E}_{i,j,\ell})$. More formally, by the law of total probability, we have

$$\Pr(\mathcal{E}_{i,j,\ell}) = \sum_{x,y} \Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y) \Pr(v_i = x, v_j = y) \leqslant \frac{2}{k^4} \sum_{x,y} \Pr(v_i = x, v_j = y) = \frac{2}{k^4}.$$

Since the events $\mathcal{E}_{i,j,\ell}$ and $\mathcal{E}_{i,\ell,j}$ are the same, there are at most $k^3/2$ such events we need to consider. Hence, by the union bound, the probability that $\mathcal{E}_{i,j,\ell}$ occurs for some triple i, j, ℓ is at most $\frac{k^3}{2} \cdot \frac{2}{k^4} = \frac{1}{k} \leqslant \frac{1}{3}$.

Similarly, for four distinct indices i, j, ℓ, m, let $\mathcal{E}_{i,j,\ell,m}$ be the event that the edges $v_i v_j$ and $v_\ell v_m$ receive the same color. For fixed vertices x, y, z, we now condition on the outcome $v_i = x, v_j = y, v_\ell = z$. By assumption, z has at most M/k^4 neighbors in color $\chi(xy)$. Once we condition, the event $\mathcal{E}_{i,j,\ell,m}$ is just the event that $\chi(zv_m) = \chi(xy)$, where the only randomness is in the choice of v_m, which is uniform on a set of size $M - 3$. So we have

$$\Pr(\mathcal{E}_{i,j,\ell,m} \mid v_i = x, v_j = y, v_\ell = z) \leqslant \frac{1}{M-3} \cdot \frac{M}{k^4} \leqslant \frac{2}{k^4}.$$

Again applying the law of total probability, we conclude that $\Pr(\mathcal{E}_{i,j,\ell,m}) \leqslant \frac{2}{k^4}$. The total number of such events is at most $k^4/4$, since we obtain the same event if we swap i, j or ℓ, m. So by the union bound, the probability that $\mathcal{E}_{i,j,\ell,m}$ happens for some 4-tuple (i, j, ℓ, m) is at most $\frac{k^4}{4} \cdot \frac{2}{k^4} = \frac{1}{2}$.

In total, we find that the probability that $v_1, \ldots, v_k$ span a rainbow K_k is at least $1 - \frac{1}{3} - \frac{1}{2} > 0$, hence there is a rainbow K_k in the coloring. □

Now that we have Lemma 16.4, we can proceed with the proof of Theorem 16.3. Before doing so, it's worth thinking of an alternative way of presenting the proof of Theorem 13.4. To show that $r(k) \leqslant 4^k$, let us fix a 2-coloring of $E(K_N)$, where $N = 4^k = 2^{2k}$. We pick an arbitrary vertex v_1. At least half of its incident edges are of the same color, which we call c_1. We now restrict to the c_1-colored neighborhood S_1 of v_1, and pick from there an arbitrary vertex v_2. At least half of its incident edges in S_1 are of the same color, say c_2. We let S_2 be this neighborhood, and proceed in this fashion. Since

$$|S_{i+1}| \geqslant \left\lceil \frac{|S_i| - 1}{2} \right\rceil$$

for all i, we conclude that $|S_i| \geqslant 2^{2k-i}$ for all i. Hence we can continue this process for at least $2k$ steps, to produce vertices $v_1, \ldots, v_{2k}$ and colors $c_1, \ldots, c_{2k}$. Again by the pigeonhole principle, at least k of these colors must be the same, say $c_{i_1}, \ldots, c_{i_k}$ are all red. But by the way we constructed this sequence, this shows that $v_{i_1}, \ldots, v_{i_k}$ form a red K_k.

The proof of Theorem 16.3 uses a very similar argument, which we will now see.

Proof of Theorem 16.3. Let $N = k^{4k^2}$, and fix an arbitrary coloring of $E(K_N)$. We let $S_0 = V(K_N)$. We now run the following process, for all $i \geqslant 1$.

1. If $|S_{i-1}| < 2$, stop the process.
2. If every vertex in S_{i-1} is incident to at most $|S_{i-1}|/k^4$ edges in each color, we apply Lemma 16.4 to S_{i-1} with $M = |S_{i-1}| \geqslant 2$. We conclude that S_{i-1} contains a rainbow K_k, completing the proof.
3. If not, there is some vertex $v_i \in S_{i-1}$ and some color c_i such that v_i is incident to at least $|S_{i-1}|/k^4$ edges of color c_i in S_{i-1}. We let S_i be the c_i-colored neighborhood of v_i in S_{i-1}.
4. Increment i by 1 and return to step 1.

If we ever find a rainbow K_k in this process, we are done, so we may assume that that never happens. Note that as long as the process continues, we have that $|S_i| \geqslant |S_{i-1}|/k^4$, so by induction we have that $|S_i| \geqslant k^{4(k^2-i)}$. Hence we can continue this process at least until step $i - 1 = k^2 - 1$. In other words, this process produces a sequence $v_1, \ldots, v_{k^2}$ of vertices and $c_1, \ldots, c_{k^2-1}$ of colors, with the property that each v_i is adjacent in color c_i to all v_j with $j > i$.

Suppose first that k of the colors $c_1, \dots, c_{k^2-1}$ are equal, say $c_{i_1}, \dots, c_{i_k}$ are all red. Then $v_{i_1}, \dots, v_{i_k}$ form a monochromatic red K_k, and we are done. But if this does not happen, then at least k different colors must appear in the list $c_1, \dots, c_{k^2-1}$, say $c_{j_1}, \dots, c_{j_k}$ are all distinct. Then $v_{j_1}, \dots, v_{j_k}$ form a starry K_k, and we are again done. □

Theorem 16.3 states that $\mathrm{ER}(k) \leqslant k^{4k^2} = 2^{4k^2 \log k}$. How good is this bound? The best known lower bound is given by the following simple proposition.

Proposition 16.5. *We have*

$$\mathrm{ER}(k) \geqslant r(k; k-2).$$

Proof. Let $N = r(k; k-2) - 1$, and consider a $(k-2)$-coloring χ of $E(K_N)$ with no monochromatic K_k. Note that a starry coloring of K_k must use $k-1$ colors, so there is also no starry K_k in χ, since χ only uses $k-2$ colors. Similarly, a rainbow coloring of K_k must use $\binom{k}{2} > k-2$ colors, hence there is no rainbow K_k in χ either. This shows that $\mathrm{ER}(k) > N$, proving the proposition. □

Note that this construction rules out the existence of starry or rainbow K_k in a pretty silly fashion, by simply using too few colors to allow these structures to appear. However, as far as I know, this is the only technique that anyone has ever found for lower-bounding $\mathrm{ER}(k)$; in particular, no one knows of a "smarter" way of excluding rainbow or starry K_k.

It now remains to find a good lower bound for the multicolor Ramsey number $r(k; k-2)$, or, more generally, for $r(k; q)$. By using a random q-coloring, one can adapt the proof of Theorem 13.7 and prove that for any $k, q \geqslant 3$, we have

$$r(k; q) > q^{k/2}. \tag{7}$$

However, this is not very good, as we recall that our upper bound on multicolor Ramsey numbers, from Theorem 13.5, is $r(k; q) \leqslant q^{qk}$; in particular, the dependence on q is super-exponential in the upper bound, whereas the lower bound in (7) is only polynomial in q. However, there is a simple construction that does substantially better.

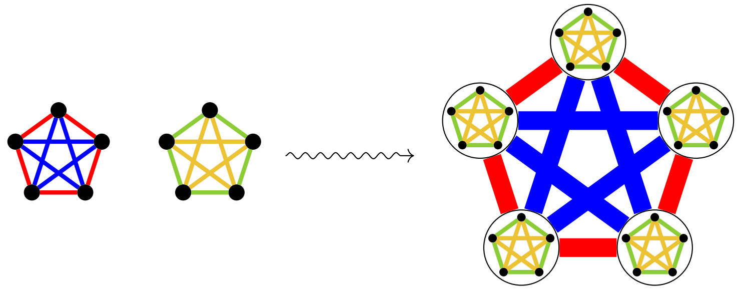
Proposition 16.6 (Abbott)**.** *For all positive integers k, q_1, q_2, we have*

$$r(k; q_1 + q_2) - 1 \geqslant (r(k; q_1) - 1)(r(k; q_2) - 1). \tag{8}$$

As a consequence, we have

$$r(k; q) > 2^{\frac{k}{2}\lfloor \frac{q}{2} \rfloor}.$$

Proof. Let $N_1 = r(k; q_1) - 1$ and $N_2 = r(k; q_2) - 1$. By assumption, we have colorings $\chi_i : V(K_{N_i}) \to [\![q_i]\!]$, for $i = 1, 2$, both of which avoid monochromatic K_k. Let $N = N_1 N_2$, and identify the vertex set of K_N with $V(K_{N_1}) \times V(K_{N_2})$. We can now define a coloring $\chi : E(K_N) \to [\![q_1 + q_2]\!]$ as follows. It is easiest to understand with the following picture, which shows how to convert two 2-colorings of $E(K_5)$ into a 4-coloring of $E(K_{25})$, maintaining the property of having no monochromatic triangle.



Formally, given a pair of vertices $(a_1, b_1), (a_2, b_2) \in V(K_{N_1}) \times V(K_{N_2}) \cong V(K_N)$, we define

$$\chi((a_1, b_1), (a_2, b_2)) = \begin{cases} \chi_1(a_1, a_2) & \text{if } a_1 \neq a_2, \\ q_1 + \chi_2(b_1, b_2) & \text{otherwise.} \end{cases}$$

This is a $(q_1 + q_2)$-coloring of $E(K_N)$, and one can readily verify that there is no monochromatic K_k, as such a monochromatic clique could be used to obtain a monochromatic K_k in either χ_1 or χ_2. Thus proves the claimed inequality (8).

To use it, we recall that we proved in Theorem 13.7 that $r(k; 2) \geqslant 2^{k/2} + 1$. Applying (8) $\lfloor q/2 \rfloor$ times, we conclude that $r(k; q) > (2^{k/2})^{\lfloor q/2 \rfloor}$, as claimed. □

Plugging this result into Proposition 16.5, we find that $\mathrm{ER}(k) \geqslant r(k; k-2) \geqslant 2^{ck^2}$ for a constant $c > 0$. That is, we match the upper bound from Theorem 16.3, apart from a logarithmic gap in the exponent. In fact, we have the same logarithmic gap for multicolor Ramsey numbers, since we now know that $2^{ckq} \leqslant r(k; q) \leqslant 2^{kq \log q}$. It is a very major open problem to close this logarithmic gap in either problem.

Let me remark that in recent years, there have been a number of improvements to the constant factor in Proposition 16.6. Roughly speaking, it says that $r(k; q) \geqslant 2^{\frac{1}{4}kq}$. In 2020, Conlon and Ferber found a new construction that showed, roughly, $r(k; q) \geqslant 2^{\frac{7}{24}kq}$, which is better since $\frac{7}{24} > \frac{1}{4}$. Shortly thereafter, I optimized their technique and improved the lower bound to, roughly, $r(k; q) \geqslant 2^{\frac{3}{8}kq}$, which is again better since $\frac{3}{8} > \frac{7}{24}$. The current record is due to Sawin, who further optimized the technique and proved, roughly, that $r(k; q) \geqslant 2^{0.383796kq}$, which is better since $0.383796 > \frac{3}{8}$. Note that although these improvements are nice and interesting, they do not give any insight into the most important question of whether the logarithmic factor in the exponent is necessary, since they only affect the constant factor in the exponent.

17 Folkman's theorem and beyond

We started this topic with Ramsey's theorem: for every k, there exists an N such that if the edges of K_N are two-colored, then there exists a monochromatic K_k. In Section 15, we generalized the conclusion: rather than finding a monochromatic K_k, we found a monochromatic copy of H, for some not-necessarily-complete graph H. We will now generalize the first part of the statement.

Definition 17.1. Given two graphs G, H, we say that *G is Ramsey for H in q colors* (or *G is q-color Ramsey for H*) if, whenever the edges of G are q-colored, there is a monochromatic copy of H. In case $q = 2$, we simply say that *G is Ramsey for H*.

Thus, Ramsey's theorem simply states that K_N is q-color Ramsey for K_k whenever N is sufficiently large (as a function of q and k).

To gain some intuition for this definition, let's think of the case when $H = K_3$. If G is Ramsey for K_3, then certainly G must contain at least one triangle. But in fact, the definition of G being Ramsey for K_3 tells us that G contains triangles "very robustly". Indeed, another way of saying Definition 17.1 is that, no matter how we try to split G into the union of two subgraphs, we cannot destroy all triangles in G. This idea of robustness is one of the reasons that Definition 17.1 is interesting.

That being said, it's not at all obvious that this definition actually gives us any new information. Indeed, we know that $r(3) = 6$, or equivalently that K_6 is Ramsey for K_3 while K_5 is not. In particular, we find that if G is a graph containing K_6 as a subgraph, then G is Ramsey for K_3. Indeed, given any 2-coloring of $E(G)$, ignore all the edges except for those in the K_6 subgraph; among those $\binom{6}{2}$ edges, we are guaranteed to find a monochromatic triangle, regardless of how the other edges are colored.

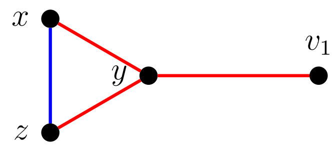
If you spend some time trying to construct graphs that are Ramsey for K_3, you may start to wonder if this is the *only* reason a graph can be Ramsey for K_3. In other words, you might be tempted to conjecture that G is Ramsey for K_3 if and only if $K_6 \subseteq G$. The question of whether this is true was raised by Erdős and Hajnal, and was rapidly answered in the negative independently by Cherlin, Graham, and van Lint. The following slick construction is due independently to Galluccio–Simonovits–Simonyi and to Szabó, and generalizes Graham's original argument. Given two graphs G_1, G_2, their *join*, denoted $G_1 * G_2$, is the graph obtained from their disjoint union by adding all edges with one endpoint in G_1 and one in G_2.

Proposition 17.2 (Galluccio–Simonovits–Simonyi, Szabó)**.** *Let $G = K_3 * C_\ell$, where $\ell \geqslant 3$ is an odd integer. Then G is Ramsey for K_3. Moreover, if $\ell \geqslant 5$, then $K_6 \nsubseteq G$.*

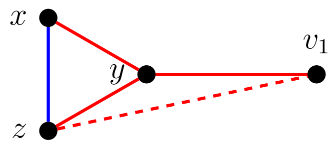
Proof. Let the vertices of G be $x, y, z, v_1, \ldots, v_\ell$, where x, y, z form a triangle, $v_1, \ldots, v_\ell$ form a cycle C_ℓ, and all edges between $\{x, y, z\}$ and $\{v_1, \ldots, v_\ell\}$ are present. Note that if $K_6 \subseteq G$, then at least three of the vertices of this K_6 must come from $v_1, \ldots, v_\ell$ (and they must form a triangle), so the second statement of the proposition is immediate since C_ℓ is triangle-free whenever $\ell \geqslant 5$.

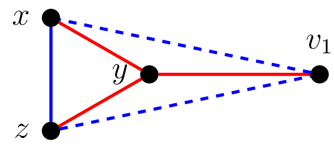
It remains to show that G is Ramsey for K_3, so fix a two-coloring of $E(G)$. If $\{x, y, z\}$ form a monochromatic triangle then we are done, so two of the edges xy, xz, yz receive one color and the third edge receives the other color. Without loss of generality, we may assume that xy, yz are red and xz is blue.

Now consider the edges between $\{x, y, z\}$ and v_1. First suppose yv_1 is red.

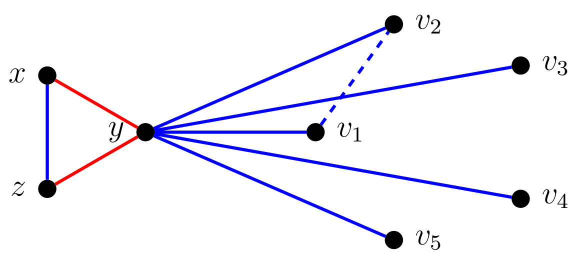


If xv_1 or zv_1 is red, then we close a red triangle xyv_1 or zyv_1, so we may assume that both these edges are blue. But that also creates a blue triangle, xzv_1.

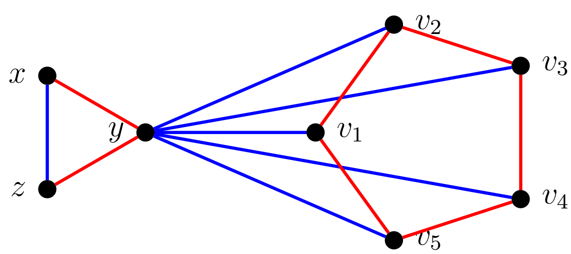




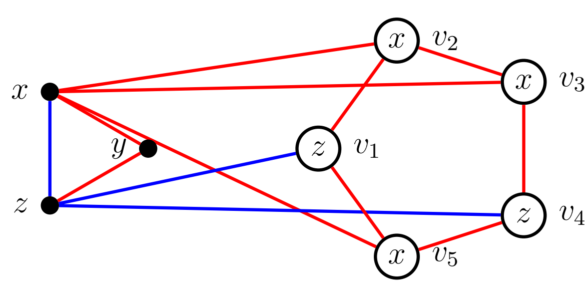
So we may assume that yv_1 is blue. By the same logic, yv_i is blue for all $i \in [\![\ell]\!]$. Note that if any of the edges v_iv_{i+1} in the cycle is blue, then we create a blue triangle yv_iv_{i+1}.



Therefore, we may assume that all the edges in the cycle are red.



Recall that xv_i and zv_i cannot both be blue, as this would create a blue triangle xzv_i. Let us label v_i by either the label x or z, depending on whether xv_i or zv_i is red (picking a label arbitrarily if both are red). By the above, every v_i receives a label.



Since ℓ is odd, the cycle C_ℓ is not bipartite. Hence, two adjacent vertices v_i, v_{i+1} must receive the same label (like v_2 and v_3 in the picture above). But then they create a red triangle together with their label. □

Note that $K_3 * K_3 = K_6$, so this result also gives a new (and more complicated) proof that K_6 is Ramsey for K_3. But it also shows that the set of graphs Ramsey for K_3 is surprisingly rich.

Note that each of the graphs $K_3 * C_\ell$ considered above *does* contain K_5 as a subgraph. So there is a natural weakening of our previous question: does every graph which is Ramsey for K_3 contain K_5 as a subgraph? The answer to this question also turns out to be negative, as proved by Pósa. So we may weaken our question further: does every graph Ramsey for K_3 contain a K_4? The answer to this also turns out to be no, as shown by the following remarkable theorem of Folkman.

Theorem 17.3 (Folkman)**.** *For every $k \geqslant 2$, there exists a graph G such that G is Ramsey for K_k, but $K_{k+1} \nsubseteq G$.*

This is pretty astonishing, even in the case $k = 3$. As discussed above, a graph that is Ramsey for K_3 must contain triangles "very robustly", in the sense that we cannot destroy all the triangles by splitting the graph into two subgraphs. Yet Folkman's theorem shows that such a graph can exist even though, locally, the triangles have almost no overlap.

Folkman's proof only worked for the case of two-colors, but the general case was shortly thereafter established by Nešetřil and Rödl, who proved the following generalization. We denote by $\omega(H)$ the *clique number* of H, that is, the maximum k such that $K_k \subseteq H$.

Theorem 17.4 (Nešetřil–Rödl)**.** *For every graph H and every $q \geqslant 2$, there exists a graph G which is q-color Ramsey for H with $\omega(G) = \omega(H)$.*

In their proof, Nešetřil and Rödl introduced a very powerful technique, called the *partite construction*, which is a very general-purpose way of producing graphs G that are Ramsey for a given graph H, while satisfying certain local sparsity conditions. The partite construction (as well as the earlier construction of Folkman) is completely explicit, so we can get a complete description of what the graph G in Theorem 17.4 looks like. Unfortunately, these constructions are iterative in nature, and each step of the iteration is complicated, so the *size* of the graph G constructed is unbelievably huge.

There is now an alternative approach to constructing such restricted Ramsey graphs, which uses randomness. It has a number of advantages over the partite construction, including giving much better bounds on how large G has to be in results like Theorem 17.4. However, as we will discuss shortly, it also seems to be less flexible than the partite construction, and there are results that the random approach seems incapable of proving.

The main result in this direction is the *random Ramsey theorem* of Rödl and Ruciński. To state it, we recall that the *maximal 2-density* of a graph H is

$$m_2(H) \coloneqq \max_{\substack{J \subseteq H \\ v(J) \geqslant 3}} \frac{e(J) - 1}{v(J) - 2}.$$

Theorem 17.5 (Rödl–Ruciński). *Let H be a graph which is not a forest, and let $q \geqslant 2$. There exist constants $C > c > 0$ such that the following holds. Form an N-vertex graph G by including each edge independently with probability p. Then*

$$\lim_{N\to\infty} \Pr(G \text{ is Ramsey for } H \text{ in } q \text{ colors}) = \begin{cases} 1 & \text{if } p \geqslant CN^{-1/m_2(H)}, \\ 0 & \text{if } p \leqslant cN^{-1/m_2(H)}. \end{cases}$$

In other words, $p \asymp N^{-1/m_2(H)}$ is a *threshold* for the property of G being Ramsey for H. If p is substantially smaller than this value, then G is extremely unlikely to be Ramsey for H, whereas if p is substantially larger than this value, then G is extremely likely to be Ramsey for H. The heuristic reason why this value of p controls the threshold is the following. One can check that at this value, a typical edge of G lies in a constant number of copies of H[12]. Thus, if $p \leqslant cN^{-1/m_2(H)}$ for a small constant c, then the majority of edges of G lie in *zero* copies of H, and thus it is not surprising that G does not "robustly" contain H; we should be able to color $E(G)$ and destroy all copies of H. On the other hand, if $p \geqslant CN^{-1/m_2(H)}$ for a large constant C, then most edges of G lie in very many copies of H, and we expect a great deal of interaction between the copies, such that destroying all of them becomes impossible no matter how we color the edges. While this is a good heuristic explanation, turning it into a proof is substantially harder, and we will not do so in this course.

However, Theorem 17.5 does allow us to easily prove results along the lines of Theorem 17.3. One can actually prove Theorem 17.3 as a consequence of (a more precise version of) Theorem 17.5, but we will content ourselves with proving the following weakening of Theorem 17.3, which generalizes Proposition 17.2 (which corresponds to the case $k = 3, q = 2$).

Proposition 17.6. *For every $k \geqslant 3$ and $q \geqslant 2$, there exists a graph G which is q-color Ramsey for K_k, but $K_{k+3} \not\subseteq G$.*

Proof. We begin by observing that

$$\frac{e(K_k) - 1}{v(K_k) - 2} = \frac{\binom{k}{2} - 1}{k-2} = \frac{k^2 - k - 2}{2(k-2)} = \frac{k+1}{2}.$$

It is not hard to check that $\frac{e(J)-1}{v(J)-2}$ is strictly smaller for any proper subgraph $J \subsetneq K_k$, hence $m_2(K_k) = \frac{k+1}{2}$. By Theorem 17.5, there is a constant $C > 0$ such that the following holds. If we pick an N-vertex graph randomly by including each edge independently with probability $p := CN^{-\frac{2}{k+1}}$, then G is q-color Ramsey for H with probability tending to 1 as $N \to \infty$. In particular, if N is sufficiently large, then this probability is at least $\frac{2}{3}$.

On the other hand, by the union bound, the probability that $K_{k+3} \subseteq G$ is at most

$$\binom{N}{k+3} p^{\binom{k+3}{2}} < C^{\binom{k+3}{2}} \cdot N^{k+3} \cdot N^{-\frac{2}{k+1}\binom{k+3}{2}} = C^{\binom{k+3}{2}} \cdot N^{-\left(\frac{2}{k+1}\binom{k+3}{2} - (k+3)\right)}. \tag{9}$$

[12] I am cheating a bit here; really, I should be counting copies of the subgraph $J \subseteq H$ achieving the maximum in the definition of $m_2(H)$.

We have that

$$\frac{2}{k+1}\binom{k+3}{2} - (k+3) = \frac{(k+3)(k+2)}{k+1} - (k+3) = (k+3)\left(\frac{k+2}{k+1} - 1\right) > 0.$$

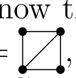
Hence, the exponent on N is negative in (9), so the probability that $K_{k+3} \subseteq G$ tends to 0 as $N \to \infty$. In particular, by picking N sufficiently large, we can ensure that $K_{k+3} \nsubseteq G$ with probability at least $\frac{2}{3}$.

Therefore, with positive probability, G satisfies both the desired properties, proving the claimed result. □

Before ending this section, let us briefly discuss one further recent breakthrough on the structure of restricted Ramsey graphs, due to Reiher and Rödl.

Definition 17.7. Let H be a graph. We say that another graph F is *Ramsey obligatory for H* if the following holds. For every sufficiently large q and every graph G which is q-color Ramsey for H, we have $F \subseteq G$.

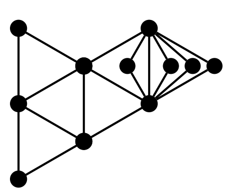
In this language, we can restate Proposition 17.6 as saying that K_{k+3} is not Ramsey obligatory for K_k, and Theorem 17.3 (or more precisely its multicolor extension, which follows from Theorem 17.4) states that K_{k+1} is not Ramsey obligatory for H. On the other hand, we can easily show that certain graphs *are* Ramsey obligatory for H. For example, H itself is Ramsey obligatory for H—if G is Ramsey for H, then certainly G contains H as a subgraph!

To keep things concrete, let's specialize to the case $H = K_3$. Then we know that K_3 is Ramsey obligatory for K_3, but K_4 is not. On the other hand, the graph $F =$ , obtained by gluing two triangles along an edge, is also Ramsey obligatory. Indeed, if G is an F-free graph, then all the triangles in G are edge-disjoint, so certainly we can color $E(G)$ and avoid all monochromatic triangles. More generally, we make the following definition.

Definition 17.8. *Triangle trees* are the class of graphs defined recursively as follows.

- K_3 is a triangle tree.
- Given a triangle tree T, we can obtain a new triangle tree T' by picking an edge of T and gluing a new triangle to it.

A typical triangle tree might look something like the following.



It is not hard to show the following fact; the proof is left for the homework.

Proposition 17.9. *If F is a subgraph of a triangle tree, then F is Ramsey obligatory for K_3.*

The astonishing theorem of Reiher and Rödl is that this sufficient condition is also necessary.

Theorem 17.10 (Reiher–Rödl)**.** *A graph F is Ramsey obligatory for K_3 if and only if F is a subgraph of a triangle tree.*

Said differently, given any graph F which is not a subgraph of a triangle tree, Reiher and Rödl are able to construct a graph G which is q-color Ramsey for K_3, yet does not contain F as a subgraph. In particular, since one can check that K_4 is not a subgraph of a triangle tree, this implies the $k = 3$ case of Theorem 17.3.

In fact, their theorem is vastly more general than this, and implies many strengthenings of Theorem 17.4. Somewhat more surprisingly, it appears that even for proving a result like Theorem 17.10, one actually has to prove these much more general results; their proof is based on a very complicated inductive argument, and in order to make the induction work one has to maintain a very general inductive statement.

18 Book recommendations

If you want to learn more about extremal combinatorics, Ramsey theory, or related topics, here are a few wonderful books.

- László Lovász, *Combinatorial problems and exercises*
- Yufei Zhao, *Graph theory and additive combinatorics*
- Dhruv Mubayi and Jacques Verstraete, *Extremal graph and hypergraph theory* (not yet published, but soon!)
- Ron Graham, Bruce Rothschild, and Joel Spencer, *Ramsey theory*
- Jiří Matoušek, *Thirty-three miniatures*
- Noga Alon and Joel Spencer, *The probabilistic method*
- My lecture notes on Ramsey theory (which were the origin of much of these lecture notes!): `https://n.ethz.ch/~ywigderson/math/static/RamseyTheory2024LectureNotes.pdf`