18.218 — Ramsey Theory

Class by Lisa Sauermann
Notes by Sanjana Das
Spring 2023

Notes for the MIT class **18.218** (Ramsey Theory), taught by Lisa Sauermann. All errors are my responsibility.

Contents

§1 February 7, 2022 — Ramsey's Theorem

§1.1 Ramsey's Theorem

This class is about Ramsey theory. Ramsey theory applies to a large range of combinatorial structures, but its starting point is Ramsey's theorem for graphs.

Theorem 1.1 (Ramsey's Theorem)

Every very large graph must contain a large structured subgraph — a clique or an independent set.

In general in Ramsey theory, for every large combinatorial structure, we find a subpart with very nice structure. In this case, the graphs with the simplest structure are a graph with no edges (i.e., an independent set) or a graph with all edges (i.e., a clique). Ramsey's theorem says that no matter the structure of the original graph, as long as it's large enough, we can find a large clique or independent set.

The example usually used to explain Ramsey theory to non-mathematicians is the following.

Example 1.2

Among 6 people, there must be 3 people who pairwise have met or 3 people who pairwise haven't met.

Remark 1.3. Typically the example is phrased with people knowing or not knowing each other, but having met makes it more clear that the relation is symmetric.

Remark 1.4. In the current room, which has around 25 people, you can actually find 4 people who pairwise know each other or pairwise don't know each other. (In this room, there probably exist much larger cliques of people who've taken classes together before.)

To see the connection, you can interpret people as nodes of a graph, with edges between two people who have met.

Now we'll state Ramsey's theorem more formally. We'll actually phrase it in terms of red edges and blue edges, rather than edges and non-edges (so that we can extend to more colors later, and to make the symmetry clearer).

Theorem 1.5 (Ramsey 1928)

For any integers k, $\ell \geq 2$, there exists a positive integer $R(k, \ell)$ such that in any coloring of the edges of a complete graph on $R(k, \ell)$ vertices, we can find a red clique of size k or a blue clique of size ℓ.

Remark 1.6. In this perspective, we start with all the edges in a complete graph and color the edges with two colors (in any way — of course, the coloring doesn't have to be proper). This is the same as the first perspective if we imagine taking only the red edges for the graph.

Remark 1.7. The *or* in the theorem is not exclusive — it's possible that both occur. But we really do need the two alternatives, because otherwise we could take all edges of the wrong color.

Remark 1.8. If we can find a number $R(k, \ell)$ for which this is true, then it's also true for all larger numbers — because we can just look at a subset of vertices of size $R(k, \ell)$.

Remark 1.9. In the 6-people example, we were looking for 3 people who know each other, or 3 who don't. Why allow k and ℓ to be different? In theory, it doesn't get stronger that way (since you can just take their maximum). But writing it this way is useful for the proof because we're going to use induction, even if we only care about the $k = \ell$ case. Further, there are situations where it's natural to take $k \neq \ell$. Although this theorem doesn't become more general by allowing $k \neq \ell$, we also care not just that $R(k, \ell)$ exists, but also how large it needs to be. Then it does matter — the answer for $(1000, 5)$ may be very different from the answer for $(1000, 1000)$.

Proof of Ramsey's theorem. We use induction on $k + \ell$.

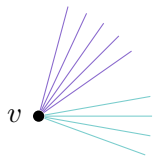
For the base case, the theorem is easy when $k = 2$ or $\ell = 2$. If $k = 2$, we can take $R(2, \ell) = \ell$ — for any red-blue coloring of the edges of a complete graph with ℓ vertices, either we have a red edge (i.e., a red clique of size 2), or the entire graph is blue and forms a blue clique of size ℓ. (The case where $\ell = 2$ is symmetric.)

Now assume that $k, \ell > 2$, and assume that we have already proven the theorem for all $k', \ell' \geq 2$ with $k' + \ell' < k + \ell$. (The most common way to do induction is to induct from n to $n + 1$, and that would probably suffice here; however, Prof. Sauermann wants to encourage us to assume we've proven it for all smaller values instead, since we don't lose anything by doing this and sometimes it's helpful.)

We claim that taking $R(k, \ell) = R(k - 1, \ell) + R(k, \ell - 1)$ works. (Since we assumed that the statement holds for $(k - 1, \ell)$ and $(k, \ell - 1)$, these numbers $R(k - 1, \ell)$ and $R(k, \ell - 1)$ are well-defined; and we claim that we can satisfy the theorem statement for (k, ℓ) by taking $R(k, \ell)$ to be their sum.)

Consider a coloring of the edges of a complete graph with $R(k - 1, \ell) + R(k, \ell - 1)$ vertices with red and blue. We want to show that in this coloring, we can find a red clique of size k or a blue clique of size ℓ.

Fix a vertex v, which we'll use as a starting point of exploration (it won't necessarily be contained in our clique).



There are $R(k - 1, \ell) + R(k, \ell - 1) - 1$ edges out of v, so by the pigeonhole principle, v must have at least $R(k - 1, \ell)$ red neighbors or at least $R(k, \ell - 1)$ blue neighbors.

In the first case, among the red neighbors of v, we can find a red clique of size $k - 1$, or a blue clique of size ℓ (by the inductive hypothesis applied to these neighbors). If we've found a blue clique of size ℓ, then we're done (it's also a blue clique of size ℓ in the original graph). Meanwhile, if we find a red clique of size $k - 1$, then we can add v to it (since all its vertices are red neighbors of v) to get a red clique of size k.

The second case is analogous. So in either case, we can find a red clique of size k or a blue clique of size ℓ in the original graph. □

Remark 1.10. Note that even if we only care about $k = \ell$, the induction requires that we use cases with $k \neq \ell$ as well.

Of course, this is not the end of the entire class — this is just the starting point of Ramsey theory, and there's many directions to go from here.

The first direction is to prove theorems of this flavor in other combinatorial structures. For example, you can consider graphs colored with more than two colors; or hypergraphs (more complex versions of graphs, where edges consist of more than two vertices). But you can also leave the settings of graphs completely,

and prove statements about integers (which we'll see next week). You can also prove Ramsey results of a geometric flavor (starting with points in a plane and trying to find a subcollection with nice geometric properties). Another direction is to try to figure out what the best possible number $R(k,\ell)$ is — we've shown that the number exists, but how large does it really have to be? In this class, we'll explore both avenues.

§1.2 Ramsey Numbers

Definition 1.11. For all $k,\ell \geq 2$, the **Ramsey number** $R(k,\ell)$ is the smallest number $R(k,\ell)$ such that the statement in Ramsey's theorem holds.

Note that we redefined $R(k,\ell)$ from before (where we just required it to be *any* number such that the theorem held); in particular, the statement $R(k,\ell) = R(k-1,\ell) + R(k,\ell-1)$ in the proof above is not true anymore.

Question 1.12. How big is $R(k,\ell)$?

This is an open problem. Determining these numbers exactly is very difficult, even if k and ℓ are very small — $R(2,\ell) = \ell$, $R(3,3) = 6$, and $R(4,4) = 18$ (which was recently determined) are known, but $R(5,5)$ is not known. But in extremal combinatorics, we don't generally care about its *exact* values — instead, we want to find bounds that determine how $R(k,\ell)$ behaves as a function of k and ℓ. (In particular, most people in extremal combinatorics do not care about what $R(5,5)$ is, since they don't care about exact values; instead we want to understand its *growth rate*.) Even understanding the correct behaviour is an open question (we'll see some bounds shortly, but the correct bounds aren't known).

First, the proof of Ramsey's theorem gives recursive bounds:

- $R(2,\ell) = \ell$ and $R(k,2) = k$.
- $R(k,\ell) \leq R(k-1,\ell) + R(k,\ell-1)$ — our proof showed that the theorem holds for the sum on the right-hand side, and since $R(k,\ell)$ is defined as the smallest number such that the theorem holds, it must be at most the sum.

Remark 1.13. Note that $R(k,\ell) = R(\ell,k)$ by symmetry (we can swap the two colors).

Theorem 1.14 (Erdős–Szekeres 1935)
For all $k,\ell \geq 2$, we have $R(k,\ell) \leq \binom{k+\ell-2}{k-1}$.

Remark 1.15. This expression doesn't look symmetric in k and ℓ, but it is — we have $k+\ell-2 = (k-1)+(\ell-1)$, so the two binomial coefficients we'd get are equal.

Proof. We use induction, using the recursive bounds stated earlier.

- We have $R(2,\ell) = \ell = \binom{2+\ell-2}{2-1}$.
- Similarly $R(k,2) = k = \binom{k+2-2}{k-1}$.

Meanwhile, if $k,\ell \geq 3$, then we have $R(k,\ell) \leq R(k-1,\ell) + R(k,\ell-1)$. Applying the inductive hypothesis to bound the two terms on the right, we get

$$R(k,\ell) \leq \binom{k+\ell-3}{k-2} + \binom{k+\ell-3}{k-1} = \binom{k+\ell-2}{k-1}. \qquad \square$$

The most natural case is when $k = \ell$ (i.e., the diagonal Ramsey numbers). For this case, we get the following bound.

Corollary 1.16
We have $R(k,k) \leq \binom{2k-2}{k-1} \leq 4^k$.

More exactly, $\binom{2k-2}{k-1} = O(4^k/\sqrt{k})$; however, $\sqrt{k}$ is much smaller than 4^k, so we don't really care about it.

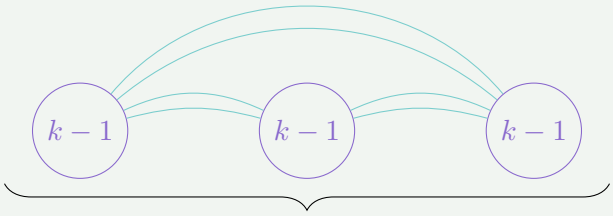
Understanding the growth behavior of $R(k,k)$ is one of the most famous problems in extremal combinatorics, so there's been a series of improvements. The best known bound, due to Ashwin Sah, is of the form 4^k divided by something sub-exponential.

Question 1.17. How close to the truth is 4^k?

To find *lower* bounds, we want to find how large of a graph can we take for which we can find some coloring where we *don't* have a red clique or a blue clique.

Example 1.18
We must have $R(k,k) > k-1$ — *any* graph on $k-1$ vertices can't have a red or blue k-clique.

Example 1.19
We must have $R(k,k) > (k-1)^2$ — as a construction, split the graph into $k-1$ groups of $k-1$ vertices each. Turn each group into a red clique of size $k-1$, and color all edges between groups blue.



It's clear this graph doesn't have a red clique of size k. Meanwhile, if we were to have a blue clique of size k, its vertices would all have to be in distinct groups; but we can't have k vertices in distinct groups, as there are only $k-1$. So this graph has no red or blue clique of size k.

This bound is much smaller than 4^k, but in fact it's extremely difficult to come up with explicit constructions — there's no explicit construction that even gives an exponential lower bound. However, we do have the following bound.

Theorem 1.20 (Erdős 1947)
For all $k \geq 2$, $R(k,k) \geq 2^{k/2}$.

So we do know that the behavior is exponential, though we don't know the correct base ($\sqrt{2}$ and 4 are still the best known bounds).

The reason we have an exponential bound but no exponential construction is that the proof of this theorem

is probabilistic (it takes a random coloring); we will see this proof next lecture.

§2 February 9, 2023 — Ramsey Numbers

§2.1 Bounds on Diagonal Ramsey Numbers

Definition 2.1. For $k \geq 2$, the (diagonal) Ramsey number $R(k,k)$ is the smallest number such that the following holds: In any coloring of the edges of a complete graph on $R(k,k)$ vertices with red and blue, there is a red clique or a blue clique of size k.

We will use *monochromatic clique* of size k to refer to a red clique of size k or a blue clique of size k.

Last class, we saw that $R(k,k) \leq 4^k$ (and we can get $\sqrt{k}$ in the denominator; the best-known bounds are a bit better, but still are of the form 4^k divided by something sub-exponential).

Question 2.2. How close to tight is this?

To find a *lower bound*, we want to come up with a coloring of complete graphs with red and blue such that there are no monochromatic cliques of size k.

Last class, we saw an explicit construction that showed $R(k,k) > (k-1)^2$ (dividing the $(k-1)^2$ vertices into $k-1$ groups of size $k-1$, coloring each group as a red clique, and coloring all other edges blue).

But in fact, a much stronger lower bound holds — $R(k,k)$ is exponential.

Theorem 2.3 (Erdős 1947)
We have $R(k,k) \geq 2^{k/2}$.

No known construction matches, or even gets close to, this behavior (we don't have an explicit construction of any exponential bound). Instead, the proof is probabilistic.

Proof. First we will eliminate the case where k is small (since we'll do bounding later). If $k \leq 4$, then we have $R(k,k) \geq (k-1)^2 \geq 2^{k/2}$ using the construction from last class.

Let $n = \lfloor 2^{k/2} \rfloor$. Consider a complete graph on n vertices, and color edges red or blue *randomly* — each edge is red with probability $\frac{1}{2}$, and blue with probability $\frac{1}{2}$, with all edges independent.

We want to show that with positive probability, this random coloring does not have a monochromatic clique of size k. (In fact, this probability will be very close to 1.)

We'll show that the *expected* number of monochromatic cliques of size k is small. For any given k vertices, the probability they form a monochromatic clique of size k is $2^{-\binom{k}{2}} \cdot 2$ — the probability that all $\binom{k}{2}$ edges between these k vertices are red is $2^{-\binom{k}{2}}$, as is the probability that all these edges are blue.

Then the *expected* number of monochromatic cliques of size k is $\binom{n}{k} \cdot 2^{-\binom{k}{2}+1}$, since there are $\binom{n}{k}$ ways to choose the k vertices.

We will show that this quantity is smaller than 1; this means the *average* number of monochromatic cliques is less than 1, so there must be some coloring where there are no monochromatic cliques. We can use the bound

$$\binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < \frac{n^k}{k!} \cdot 2^{-k^2/2+k/2+1} \leq \frac{2^{k^2/2}}{k!} \cdot 2^{-k^2/2+k/2+1} = \frac{2^{k/2}+1}{k!}.$$

It's clear that this is less than 1 if k is large enough — more explicitly, we have $k! \geq 2^{k-1}$ (since $k! = k(k-1)\cdots 2$ is the product of $k-1$ factors which are all at least 2), which gives us an upper bound of $2^{-k/2+2}$. We dealt with the case $k \leq 4$ earlier, so we're done.

So on average, the number of monochromatic cliques is strictly less than 1; this means there must be at least one outcome of the random coloring where it is 0 (there must be an outcome where it's at most its average, but this number must be an integer), i.e., that there are no monochromatic cliques of size k. □

Remark 2.4. The approximation of $\binom{n}{k} \leq \frac{n^k}{k!}$ is fairly accurate; the step where we were careless was $k! \geq 2^{k-1}$. Optimizing the bounds using Stirling's formula gives the bound

$$R(k,k) > \frac{k}{\sqrt{2}\cdot e} \cdot 2^{k/2}.$$

This bound is better by a linear factor.

Remark 2.5. We could also use the alteration method — color randomly and show that the number of expected cliques is *small*, and then delete vertices to get rid of them. However, Prof. Sauermann doesn't think this improves the bound, except for lower order terms.

Remark 2.6. The best-known lower bound is

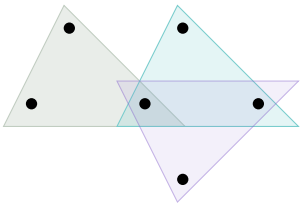
$$R(k,k) > (1+o(1)) \cdot \frac{2k}{\sqrt{2}e} \cdot 2^{k/2},$$

which combines this argument with the Lovász Local Lemma. This was proved by Spencer 1977. Meanwhile, the best-known upper bound is still around 4^k. So we know that $R(k,k)$ is exponential in k, but we don't know the correct base.

§2.2 Ramsey's Theorem for Hypergraphs

Definition 2.7. A **r-uniform hypergraph** is a collection of size-r subsets of a ground set V.

Here the ground set V is the vertex set, and we think of the subsets as edges.



When $r = 2$, this is exactly a graph; so for larger r, it's a generalization of a graph. (It's also possible to look at non-uniform hypergraphs, where the size of the subset is not fixed.)

Definition 2.8. The complete r-uniform hypergraph on V consists of all subsets of V of size r, and a *clique* is a subset $S \subseteq V$ with all possible subsets of S being edges.

As in a graph, the size of a clique is $|S|$. So finding a complete r-uniform hypergraph in a larger graph gives a clique.

We'll now do the exact same thing from last class on hypergraphs. We can again ask for a red clique of size k (i.e., k vertices such that all subsets of size r are colored red) or a blue clique of size ℓ in any coloring of a complete r-uniform hypergraph on sufficiently many vertices with red and blue.

The proof is the same as last class, so we will instead make the question more general by allowing t colors instead.

Theorem 2.9

For any integers $r \geq 2$ and $k_1, \ldots, k_t \geq r$, there exists $R_r(k_1, \ldots, k_t)$ such that the following holds: In any coloring of the edges of a complete r-uniform hypergraph with $R_r(k_1, \ldots, k_t)$ vertices with colors $1, \ldots, t$, there exists a clique of size k_i in color i for size i.

In other words, there exists a color for which we can find a clique of the corresponding size. (It's possible that there are cliques of more than one color.)

Proof. We'll again use induction — we will use double induction on r and $k_1 + \cdots + k_t$. (The outer induction is on r, and the inner induction on $k_1 + \cdots + k_t$. We could also phrase this as taking a minimal counterexample — a counterexample with minimal r, and among all these, one with minimal $k_1 + \cdots + k_t$.)

Suppose first that $k_i = r$ for some i; without loss of generality suppose $k_t = r$. Then by induction, we know that the statement holds for r and $k_1, \ldots, k_{t-1}$ (these have smaller sum). Then for any coloring of the complete r-uniform hypergraph on $R_r(k_1, \ldots, k_{t-1})$ with colors $1, \ldots, t$, either there is an edge of color t — giving a clique of size $k_t = r$ (since a single edge is a clique of size r) — or color t doesn't appear, and we have a coloring with colors $1, \ldots, t-1$, which must contain a clique of size k_i and color i for some $i \in [t-1]$.

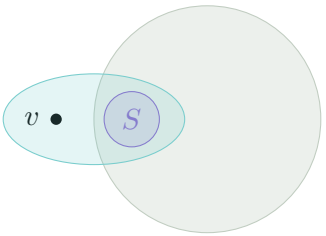
(This is the same as the size-2 case for graphs — the point is that a clique of size r is the same as a single hyperedge, since if $|S| = r$ there is only one subset of size r.)

Now suppose $k_1, \ldots, k_t > r$. Assume that we have already proved the theorem for all smaller values of r (regardless of $k_1 + \cdots + k_t$), and also for the same value of r and all smaller values of $k_1 + \cdots + k_t$.

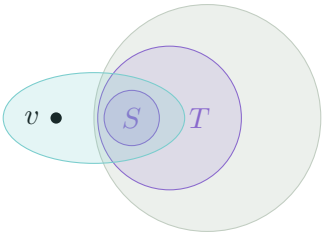
We claim that $R_{r-1}(R_r(k_1 - 1, k_2, \ldots, k_t), R_r(k_1, k_2 - 1, \ldots, k_t), \ldots) + 1$ vertices suffice for the theorem statement. Here the inner quantities $R_r(k_1 - 1, k_2, \ldots, k_t)$ and so on exist by the inner induction, and we can take R_{r-1} of these numbers by the outer induction. (These numbers are horrendously big.)

Consider a r-uniform hypergraph on V with $|V|$ at least this quantity, with the edges colored in colors $1, \ldots, t$.

First, fix a vertex v. Then for every subset $S \subseteq V \setminus \{v\}$ of size $r-1$, we consider the color of $S \cup \{v\}$. This gives a coloring of hte complete $(r-1)$-uniform hypergraph on $V \setminus \{v\}$.



Now we apply the induction hypothesis on this $(r-1)$-uniform hypergraph — for some $i \in [t]$, we can find a clique T of size $R_r(k_1, \ldots, k_i - 1, \ldots, k_t)$ of color i, with respect to the coloring of the complete $(r-1)$-uniform hypergraph on $V \setminus \{v\}$. So T has the property that any $(r-1)$-vertex subset together with v is color i in the original coloring.



Now apply the induction hypothesis to the complete r-uniform hypergraph on T (the original coloring restricted to T — *not* the coloring of the $(r-1)$-uniform hypergraph obtained from removing v). This tells us that there is a clique of one of the specified sizes and colors. There are two possibilities:

- For some $j \neq i$, there is a clique of size k_j in the coloring of the complete hypergraph on T. Then we are done (since this is also a clique in the original graph).
- Otherwise, the clique we find is a clique of size $k_i - 1$ of color i. Then add v to this clique. This gives us k_i vertices, and we claim that they form a clique of color i. Every r-vertex subset *not* involving v has color i by our choice of this subset, while every r-vertex subset which *does* contain v has color i by our construction of T.

So we are done. □

§3 February 14, 2023

§3.1 Ramsey's Theorem for Hypergraphs

On Thursday, we proved Ramsey's theorem for hypergraphs:

Theorem 3.1

For any integers $r \geq 2$ and $k_1, \dots, k_t \geq r$, there exists an integer $R_r(k_1, \dots, k_t)$ such that the following holds: for any coloring of the edges of a complete r-uniform hypergraph on $R_r(k_1, \dots, k_t)$ vertices with colors $1, \dots, t$, there exists a clique of size k_i in color i for some $i \in [t]$.

Taking $r = 2$ gives Ramsey's theorem for graphs. For general r, an *edge* is a r-set.

As in the case of graphs, we define the *Ramsey number* as the smallest number for which this holds.

Definition 3.2. The *Ramsey number* $R_r(k_1, \dots, k_t)$ is the smallest number for which the theorem holds.

As before, we can ask for bounds on the Ramsey numbers. For graphs, we have bounds, but the exact growth is far from understood. For hypergraphs, the bounds are even less understood. If you analyze the proof from last class, the bounds you get are absolutely terrible (the numbers are huge); we will see bounds a few weeks from now.

This week we'll see some other settings in which Ramsey-type questions are commonly studied.

§3.2 A Geometric Setting

The first setting we'll look at is a geometric setting, where we consider points in a plane in convex position. The first result of this type was the Erdős–Szekeres theorem on point sets in convex position.

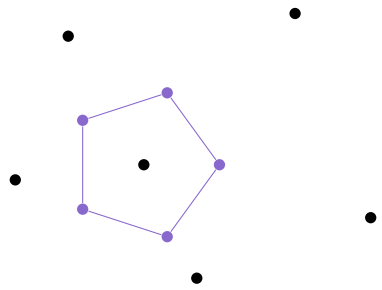
(There is another theorem by Erdős–Szekeres in a different setting, about monotone subsequences.)

In Ramsey results, we're always looking for a certain structure, and the theorem states that we can find it if our original structure is big enough.

Theorem 3.3 (Erdős–Szekeres 1935)

For every $k \geq 3$, there is some K such that the following holds: among any given K points in the plane with the property that no three are collinear, one can find k points forming a convex k-gon.

Here there is no coloring, so we're not trying to find a monochromatic thing; instead we're trying to find k points in convex position.



(The condition that no three points are on a line essentially means that the configuration is *nondegenerate*.)

We'll deduce this theorem from Ramsey's theorem for hypergraphs.

Proof 1. We want to construct a complete r-uniform hypergraph on our points, and color its edges — this means we want to give every r-tuple of vertices a color.

We'll let $r = 4$, and use two colors, red and blue. We color a 4-tuple of points red if they do not form a convex quadrilateral (i.e., if they form a triangle with a point inside), and blue if they do form a convex quadrilateral. This defines a coloring of the complete 4-uniform hypergraph on our vertices.

(Note that we're coloring sets of 4 points, not the points themselves.)

Now we apply Ramsey's theorem for hypergraphs. We want to take the target clique size for the blue points to be $k_{\text{blue}} = k$, and we want to take $k_{\text{red}} = 5$ — because we can't have a red 5-clique (it's not possible to have 5 points such that all 4-tuples are not in convex position, and we'll soon see why).

Then Ramsey's theorem guarantees that we can either find an impossible structure (which definitely won't happen) or one that we want. (We essentially make the red clique impossible to find, and this forces us to find a blue clique.)

So we can now let $K = R_4(5, k)$.

Now we'll prove that this works.

Claim — Among any 5 points in the plane, there must be 4 points forming a convex quadrilateral.

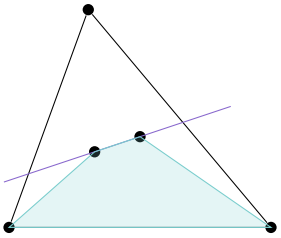
Proof. We can perform a case analysis, depending on what the convex hull of the five points looks like.

If the convex hull is a pentagon, then we're done (as any four points form a convex quadrilateral).

If the convex hull is a quadrilateral, then we are also done (those four points form a convex quadrilateral).

So the only case that requires an argument is the one where the convex hull is a triangle.

In this case, we can take the two points on the inside of the triangle, and draw the line through them. This line cannot pass through any of the vertices of the triangle, so it must intersect two sides of the triangle.



Then we can take the two points on the third side of the triangle along with these two points; the resulting quadrilateral must be convex. □

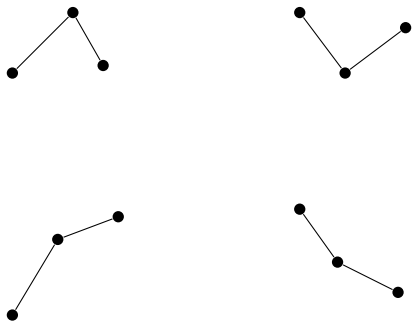
This implies the 4-uniform hypergraph cannot have a red clique of size 5 (as this would cause 5 points such that among the 5 points, any 4 of them are not convex).

So when we apply Ramsey's theorem to our hypergraph on $K = R_4(5, k)$ vertices, we know that we either find a red clique of size 5 or a blue clique of size k. The former is impossible, so we must find a blue clique of size k.

This means we have k points such that any four of them form a convex quadrilateral. This means the k points must themselves form a convex k-gon — if the convex hull didn't have all the points as vertices, then there would be at least one point inside it. Then we could triangulate the convex hull in any way; the point in the interior of the convex hull must lie in the interior of one of the triangles (since no three points are on a line), giving four points which do not form a convex quadrilateral. □

There is another proof, which again applies Ramsey's theorem for hypergraphs; but it'll apply it in the 3-uniform setting instead (where we're coloring *triples* instead). Now it's not clear what our coloring should be — all triangles are convex. Instead, the coloring we'll use will depend on the *direction* of the points — interestingly, it depends on rotations (the theorem statement isn't affected if we rotate the points, but this proof is). This proof gives a much better bound than the one obtained from the first.

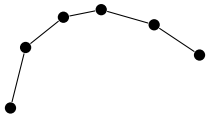
Proof 2. First, we can assume that no line through two points is completely vertical, i.e., parallel to the y-axis. (We can do this because there are only finitely many lines.)



For any three points, draw the lines between the leftmost and middle point, and the middle and rightmost point. If the corresponding graph is concave, we call the three points a *cap*; if the corresponding graph is convex, we call the points a *cup*. (The two configurations on the left form caps, while the two on the right form cups.)

Now let $K = R_3(k, k)$. Color a 3-tuple of points red if they form a cap, and blue if they form a cup. By Ramsey's theorem, we can find a red or blue k-clique — i.e., we can find k points so that any three form a cap, or k points so that any three form a cup.

If we have k points where any three form a cap, then the graph formed by connecting consecutive points from left to right must be concave (since any three points form a cap, the slopes must grow smaller at every step) — i.e., they must form a cap of size k.



Similarly, if any three points form a cup, then the k points must form a cup of size k.

In either case, this means our points must be convex as a configuration of points. □

The bounds we get from these proofs will not be great, because the hypergraph Ramsey bounds are not great. But in fact, once we come up with the approach of finding caps or cups of size k, we can take this idea further and prove a better bound by forgetting about Ramsey's theorem, and proving a statement about caps and cups. (It's very difficult to induct on the property in Erdős–Szekeres — it's difficult to induct on a convex k-gon. But caps and cups glue nicely together, which makes induction work much better.) In particular, this will give a much better bound (this is the primary motivation for studying caps and cups).

Theorem 3.4 (Erdős–Szekeres 1935)

Let $k, \ell \geq 3$, and consider $\binom{k+\ell-4}{k-2} + 1$ points in the plane, no three of which are collinear, and no two of which are on a vertical line. Then we can find a cap of size k or a cup of size ℓ.

(As in Ramsey's theorem for two colors, the induction works much better if we allow different sizes for the cap and cup, even though the case we care about most is the one where they're the same size.)

As with Ramsey's theorem, we need the *or* — if the points we start with all form a cap, then we can't find a cup of any size, and vice versa. The proof is also very similar to Ramsey's theorem (in particular, the bound looks very similar).

Proof. We use induction on $k+\ell$. First we'll deal with the case where one of the numbers is 3. First suppose $k = 3$. Then we have $\binom{\ell-1}{1} + 1 = \ell$ points; so either we have 3 points forming a cap, or all 3-tuples form cups and therefore all ℓ points form a cup of size ℓ. The case where $\ell = 3$ is analogous.

Now assume $k, \ell > 3$, and that we've proven the statement for all smaller $k + \ell$.

Suppose for contradiction that there is no cap of size k or cap of size ℓ. Then by applying the theorem to $(k-1, \ell)$, since we can't find a ℓ-cap, we can find a $(k-1)$-cap. In fact, we can find a lot more than one $(k-1)$-cap (if we find one, then we can take out one of its vertices).

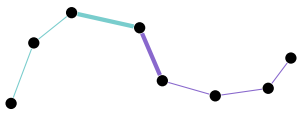
We want to extend our $(k-1)$-caps by gluing them together, and the trick for doing this is to look at the *endpoints* of the $(k-1)$-caps — consider all the points occurring as endpoints (i.e., rightmost points) of a cap of size $k-1$. We claim there are many points with this property: if we delete all these points, then there can't be any $(k-1)$-caps left (all $(k-1)$-caps must have one of their points deleted). But there are also no cups of size ℓ left, because there were no cups of size ℓ to start with. This tells us that the number of points we have left must be at most $\binom{k+\ell-5}{k-3}$ by the induction hypothesis on $(k-1, \ell)$ (as if there were more points, then we could find a $(k-1)$-cap or ℓ-cup).

So there must be at least

$$\binom{k+\ell-4}{k-2} + 1 - \binom{k+\ell-5}{k-3} = \binom{k+\ell-5}{k-2} + 1$$

points occurring as an endpoint of a $(k-1)$-cap. This is exactly $\binom{k+(\ell-1)-4}{k-2} + 1$, so we'll apply the induction hypothesis from $(k, \ell-1)$ to these points; then we can find either a cap of size k (which is impossible by assumption), or a cup of size $\ell - 1$.

Now we focus on the first point (i.e., the leftmost point) of this cup of size $\ell - 1$; we know this point is also the endpoint of a cap of size $k-1$. This gives us a cap of size $k-1$ and a cup of size $\ell - 1$ glued together at this point.



Now compare the slopes of the two edges at this point (shown in bold). If the edge on the left has greater slope than the edge on the right, then we can extend the cap of size $k-1$ to a cap of size k; if it has lesser slope, then we can extend the cup of size $\ell-1$ to a cup of size ℓ. This produces a cap of size k or a cup of size ℓ, so we are done. □

In particular, applying this with $k=\ell$ shows that among any $\binom{2k-4}{k-2}+1$ points, there is a cap of size k or a cup of size k. In particular, there is a convex k-gon. (Not every convex k-gon is a cap or cup, but every cap or cup is a convex k-gon.) So this tells us that in Erdős–Szekeres, we get a bound of

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

This is a much better bound than the ones given by the hypergraph Ramsey numbers from the previous proofs.

In this theorem, the bound $\binom{k-\ell-4}{k-2}+1$ is tight. (There is a recursive construction with $\binom{k-\ell-4}{k-2}$ points with no k-cups or ℓ-cups.) However, this doesn't necessarily imply this is the correct number for Erdős–Szekeres — since we're trying to find a weaker structure.

> **Conjecture 3.5** (Erdős–Szekeres) — The precise best possible number for Erdős–Szekeres is $K = 2^{k-2}+1$.

There exists a construction proving that this value of K is necessary. This conjecture is open, but we do know that $K \leq 2^{k+o(k)}$ (due to Suk 2016) — so the bound is asymptotically almost known, but the precise number isn't. (The problem of determining the best K is called the *Happy Ending Problem* — the people first thinking about the problem were Erdős, Szekeres, and Esther Klein, and Szekeres and Klein eventually got married.)

§4 February 16, 2023

Last class, we saw Ramsey-type results in geometric settings (taking a point set, and looking for convex k-gons or cups or caps). Today we'll see some Ramsey-type results in an arithmetic setting — where we consider colorings of numbers and look for solutions to certain equations (where the numbers have the same color).

§4.1 Schur's Theorem

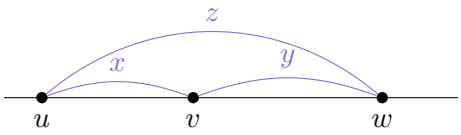
The simplest and oldest result in this vein is a theorem of Schur from 1916. (Note that this even predates Ramsey's theorem.)

> **Theorem 4.1** (Schur 1916)
> For every $t \geq 1$, there exists N such that the following holds: For every coloring of the numbers in $\{1,\dots,N\}$ with t colors, we can find $x,y,z \in \{1,\dots,N\}$ of the same color with $x+y=z$.

Proof. We'll use Ramsey's theorem for graphs. Let $N = R(3, 3, \ldots, 3)$, with t colors.

We now want to construct and color a graph. Our graph will have vertex set $\{1, \ldots, N\}$. We now need to define a coloring — we color each edge vw with the color of $|v - w|$ in the original coloring of $\{1, \ldots, N\}$.

By Ramsey's theorem, in this coloring we can find a monochromatic triangle; let its vertices be $u < v < w$.



Then we can take $x = v - u$, $y = w - v$, and $z = w - u$. Then x, y, and z must be the same color, and $z = x + y$, as desired. □

Remark 4.2. Note that we are allowed to have $x = y$ — if x and y had to be distinct, this proof wouldn't work.

There's a lot of theory about for which equations (in place of $x + y = z$) this is true, and this question was resolved by Rado; we'll see this in a few weeks.

Today we'll look at an extension in a different direction. In additive combinatorics, one of the most natural patterns is *arithmetic progressions* — a three-term arithmetic progression is three numbers x, y, and z such that $y - x = z - y$, or equivalently $x - 2y + z = 0$. More generally, we may look for longer arithmetic progressions — it's natural to look for a k-term arithmetic progression. This was answered by van der Waerden in 1927.

Theorem 4.3 (van der Waerden 1927)

For every $t \geq 1$ and $k \geq 2$, there exists N such that the following holds: In every coloring of $\{1, \ldots, N\}$ with t colors, there exists a monochromatic arithmetic progression fo length k (i.e., there exist distinct $x_1, \ldots, x_k$ in $\{1, \ldots, N\}$ of the same color with $x_2 - x_1 = x_3 - x_2 = \cdots = x_k - x_{k-1}$).

(The cases $t = 1$ or $k = 2$ are trivial.)

Remark 4.4. Note that in both theorems, if the statement is true for some N, then it's true for all larger N as well (as in Ramsey's theorem) — we can simply ignore all the extra numbers.

There is one significant difference between this result and Schur's theorem or Ramsey's theorem — in Schur's theorem, suppose someone gave us a coloring with red, blue, and green, and we were trying to look for a solution. If half the points were red, a quarter were blue, and a quarter were green, then we'd start by trying to find one in red. But although the theorem guarantees that if N is large enough we can find a solution in some color, this color might not be the largest — for example, we could color all the odd numbers red.

The same is true for Ramsey's theorem — it's possible that half the edges are red, but we can't even find a triangle in red (if the red graph is a complete bipartite graph).

So in both cases, just looking at the biggest color class isn't enough. But interestingly, for van der Waerden's theorem it *is* true — if N is large enough, then we can find a progression in the largest color class.

This is called a *density* version of van der Waerden's theorem, because we can now ignore the coloring — the largest color class is simply a subset with at least $\frac{1}{t}$ of all elements.

Theorem 4.5 (Szemerédi 1975)

For every $k \geq 2$ and every $\varepsilon > 0$, for all N which are sufficiently large with respect to k and ε, the following holds: if $A \subseteq \{1, \dots, N\}$ is a subset of size $|A| \geq \varepsilon N$, then A contains an arithmetic progression of length k.

This is stronger than van der Waerden's theorem (it implies van der Waerden's theorem by taking $\varepsilon = \frac{1}{t}$ and considering the largest color class, which must occur at least a fraction $\frac{1}{t}$ of the points), and is much harder to prove.

Note that we stated this for all large N rather than for a specific one; this is because unlike the previous results, there isn't an easy way to go from one N to all larger ones.

Remark 4.6. This is a very fundamental result in additive combinatorics. Szemerédi is a very famous combinatorialist who proved many things; but this is the main theorem that was mentioned in his citation for the Abel prize in 2010 (explaining why his work was important). Szemerédi proved this using regularity theory for hypergraphs (we won't see this in this class, but you can see *graph* regularity in 18.225).

We'll now prove van der Waerden's theorem, but in fact we'll prove it in a more general setting.

§4.2 The Hales–Jewett Theorem

We'll first see a different theorem, the Hales–Jewett theorem; then we'll see that it implies van der Waerden's theorem; and finally, we'll prove the Hales–Jewett theorem, which will imply van der Waerden's theorem as well.

The statement will look a bit technical, so we'll begin by motivating it from van der Waerden's theorem. In van der Waerden's theorem, we're looking for a k-term arithmetic progression. Imagine N is a power of k, and we represent all numbers in $\{1, \dots, N\}$ in base k. For convenience, imagine $k = 10$ (i.e., we want to find a 10-term arithmetic progression).

Imagine that we're not good at subtracting; we can subtract numbers up to 10 (i.e., we can subtract per digit), but we find subtraction hard when the calculation at one digit affects the other digits. So we're going to try to find a 10-term arithmetic progression which is so nice that when we write it down in base 10, even with these constraints it's easy to verify.

Example 4.7

Consider the number 10600108. Then the arithmetic progression

$$10600108, 11601118, 12602128, 13603138, \dots, 19609198$$

is such an arithmetic progression. We'll denote the progression by the *code* $1{*}60{*}1{*}8$; then by replacing the $*$'s with each of the digits 0 through 9, we get a 10-term arithmetic progression.

Informally speaking, the Hales–Jewett theorem says that the statement in van der Waerden's theorem is true even with the additional restriction that we want to find an arithmetic progression given by such a code.

(Typically, we'll do this without 0.)

Notation 4.8. We use $[k]$ to denote $\{1, \dots, k\}$. Then $[k]^n$ is the set of n-tuples with elements in $[k]$.

In the Hales–Jewett theorem, we're coloring such n-tuples (rather than the numbers). Then the arithmetic progressions defined by codes that we're looking for are called *combinatorial lines*.

Definition 4.9. A *combinatorial line* in $[k]^n$ is a subset of $[k]^n$ of size k that can be obtained by some n-tuple λ whose entries are all either numbers in $[k]$ or the symbol $*$, such that there is at least one $*$, in the following way: For each $1 \leq i \leq k$, the ith point of the combinatorial line is obtained by replacing all $*$'s in λ by i.

This is exactly what we did in the example — to form the first point we replace all $*$'s by 1, to form the second we replace all $*$'s by 2, and so on.

Definition 4.10. The *endpoint* of a combinatorial line is the kth point in the line, i.e., the point where $*$'s are replaced by k.

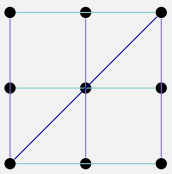
(The definition of a combinatorial line is standard; the definition of its endpoint is not, but we'll use it in the proof.)

Theorem 4.11 (Hales–Jewett 1963)

For every $k \geq 1$ and $t \geq 1$, there exists some n such that for every coloring of the points in $[k]^n$ with t colors, there exists a monochromatic combinatorial line in $[k]^n$.

This is again a case where if it's true for some n, then it automatically holds for all larger n as well.

Remark 4.12. The original reason Hales and Jewett came up with this was to generalize the game of Tic–Tac–Toe — if you think of k as 3 and n as 2, then combinatorial lines describe some of the winning configurations in Tic–Tac–Toe (all but one diagonal).



General combinatorial lines come up when you try to generalize Tic–Tac–Toe to larger grids and more dimensions.

First we'll see why the Hales–Jewett theorem implies van der Waerden's theorem.

Proof of van der Waerden's theorem using Hales–Jewett. Given k and t in van der Waerden's theorem, we'll use the same k and t in Hales–Jewett. It's then possible to take $N = k^n$ and use the argument in our motivation, but we'll see a different proof that gets a better bound.

Given n from the Hales–Jewett theorem, take $N = kn$. Suppose we are given a coloring $\gamma\colon \{1, \dots, N\} \to \{1, \dots, t\}$ (where our colors are named 1, …, t); then we wish to find an arithmetic progression whose γ-values are all equal.

To apply the Hales–Jewett theorem, we first want to construct a coloring of $[k]^n$. From γ, we construct a coloring $\gamma'\colon [k]^n \to \{1, \dots, t\}$ (i.e., a coloring of $[k]^n$ with t colors) such that for all $x_1, \dots, x_n \in [k]$, we set

$$\gamma'(x_1, \dots, x_n) = \gamma(x_1 + \cdots + x_n).$$

(Note that $x_1 + \cdots + x_n \leq kn$, so this gives us a well-defined coloring.)

By the Hales–Jewett theorem, γ' has a monochromatic combinatorial line. Now for each point in the combinatorial line, we look at the sum of its coordinates. This gives a monochromatic k-term arithmetic progression in $\{1, \dots, N\}$ — we've already seen that all these sums are in $\{1, \dots, N\}$, and they're monochromatic because we defined the colors in γ' as the colors of the corresponding sums. Finally, we can check that given any λ and the corresponding combinatorial line, the coordinate sums form an arithmetic progression — the difference between two consecutive terms is exactly the number of $*$'s in λ. □

Finally, we'll prove the Hales–Jewett theorem.

Proof of Hales–Jewett theorem. We'll use induction on k.

First, the case $k = 1$ is easy — $[1]^n$ consists of a single point, and that single point is a combinatorial line. Now fix $k \geq 2$, and assume that we've already proven the theorem for $k-1$ (and *any* number of colors t' — even if t' is much bigger than t, which is fine as we're only inducting on k).

We know that the theorem holds for $k-1$, so we could restrict our coloring to $[k-1]^n$ and find a combinatorial line there. This would give us a combinatorial line of $k-1$ points; in order to get a combinatorial line in $[k]^n$, we'd need to add the endpoint. This point may not be the right color, but the main idea is that we can do this not just for one line, but for many lines. So in the induction step, we'll prove that we can find many combinatorial lines which are monochromatic if we omit their endpoint, and which in fact have the same endpoint. (We will prove this statement by induction as well.)

We'll now fix the number of colors t.

> **Claim —** For all $j = 1, \dots, t$, there exists some n such that for every coloring of $[k]^n$ with t colors, at least one of the following two statements holds: There exists a monochromatic combinatorial line in $[k]^n$, orthere exist j combinatorial lines $\ell_1, \dots, \ell_j$ in $[k]^n$ with the same endpoint x, such that the sets $\ell_1 \setminus \{x\}, \dots, \ell_j \setminus \{x\}$ are each monochromatic and have distinct colors.

So for every coloring, there are two cases — either we're happy (we've found a complete combinatorial line), or we've found many combinatorial lines of distinct colors with the same endpoint, such that all are monochromatic if we omit the endpoint.

Note that this claim gets strictly stronger as we increase j; we state it for all j because we will induct on j. The reason this claim suffices is because by taking $j = t$, either we're done or we have such a line ℓ_i of every color; then x must match one of these colors, giving a monochromatic line.

Proof. We use induction on j. If $j = 1$, then we can take n such that the Hales–Jewett theorem holds for $k-1$ with t colors. Then we restrict our coloring of $[k]^n$ to $[k-1]^n$ and find a monochromatic combinatorial line in $[k-1]^n$. This corresponds to a combinatorial line ℓ in $[k]^n$ with some endpoint x (obtained by plugging in k to the code for the line) such that $\ell \setminus \{x\}$ is monochromatic.

Now we assume that $j \geq 2$, and that we have already proven the claim for $j-1$ (with the same k and t); let n' be as in the claim for $j-1$, and let n'' be as in the Hales–Jewett theorem for $k-1$ and $t^{k^{n'}}$ colors. (This is huge, but it is a well-defined finite number, as n' has already been fixed above.)

Let $n = n' + n''$ (this is gigantic, but it is some number). Consider a coloring on $[k]^n = [k]^{n'} \times [k]^{n''}$ with t colors.

Our first step is to look at the coloring as a coloring simply on $[k]^{n''}$ with $t^{k^{n'}}$ colors — for every point in $[k]^{n''}$, we get an assignment of a list of $k^{n'}$ colors coming from the colors in all the points in $[k]^n$ with this coordinate in $[k]^{n''}$. There are $t^{k^{n'}}$ such possible lists, so this gives a coloring of $[k]^{n''}$ with this many colors.

By our choice of n'', there is a combinatorial line ℓ in $[k]^{n''}$ in this coloring with some endpoint y, such that $\ell \setminus \{y\}$ is monochromatic in this coloring. Let $\lambda \in ([k] \cup \{*\})^{n''}$ be the corresponding code.

The color of this monochromatic set $\ell \setminus \{y\}$ corresponds to a coloring $\varphi\colon [k]^{n'} \to \{1, \dots, t\}$ — more formally, for all $a \in [k]^{n'}$ and $b \in \ell \setminus \{y\}$, the point $(a, b) \in [k]^n$ has color $\varphi(a)$ in the original coloring. (In other words, the line we've found has the property that the color of (a, b) only depends on a, for all $a \in [k]^{n'}$ and all b on the monochromatic set.)

Now apply the claim for $j-1$ to the coloring $\varphi\colon [k]^{n'} \to \{1, \dots, t\}$. This gives combinatorial lines $\ell_1, \dots, \ell_{j-1}$ in $[k]^{n'}$ all with the same endpoint x, such that $\ell_1 \setminus \{x\}, \dots, \ell_{j-1} \setminus \{x\}$ are monochromatic in φ. Let their corresponding words be $\lambda_1, \dots, \lambda_{j-1}$ (which are in $([k] \cup \{*\})^{n-1}$).

Now we look at the j lines corresponding to the codes $(\lambda_1, \lambda), \dots, (\lambda_{j-1}, \lambda)$, and (x, λ) in $([k] \cup \{*\})^n$. (So we append λ to each of our $j-1$ lines, and we get another by appending the common endpoint.) The resulting combinatorial lines all have endpoint (x, y).

By definition, all must be monochromatic if we omit the endpoint. Now each of the lines (λ_i, λ) have distinct colors. If (x, λ) agrees with one of these colors, then this gives a monochromatic line; otherwise we have j lines of distinct colors. □

So then by induction we're done. □

§5 March 2, 2023 — Upper Bounds for Off–Diagonal Ramsey Numbers

Now we'll return to graphs; today we'll see upper bounds on off-diagonal Ramsey numbers, and next week we'll also see lower bounds.

Definition 5.1. Given $k \geq 2$ and $\ell \geq 2$, the Ramsey number $R(k, \ell)$ is the smallest number such that the following holds: in any coloring of a complete graph with $R(k, \ell)$ vertices with red and blue, there is a red clique of size k or a blue clique of size ℓ.

Here the *diagonal* Ramsey numbers are numbers of the form $R(k, k)$ (e.g., the diagonal if we list out all values $R(k, \ell)$ in a table). Today we'll study the case where we think of ℓ as *fixed* — so the numbers we're looking at are far away from the diagonal (a row or column).

Earlier, we proved the following general bound for all k and ℓ:

Theorem 5.2 (Erdős–Szekeres 1935)
We have

$$R(k, \ell) \leq \binom{k+\ell-2}{k-1}.$$

If ℓ is fixed (e.g., $\ell = 3$, which is the smallest interesting case) and k is large, then

$$R(k, \ell) \leq \binom{k+\ell-2}{\ell-1} \lesssim k^{\ell-1}$$

(where the constant is around $\frac{1}{\ell!}$). In some sense, this is a good upper bound — we know $R(k, k)$ grows exponentially with k, while this (for fixed ℓ) is polynomial in k. But we may want a more precise bound.

In fact, one can improve this bound. For $R(k, 3)$ for instance, the true answer is not that $R(k, 3) \asymp k^2$, but rather that

$$R(k, 3) \asymp \frac{k^2}{\log k}.$$

For larger ℓ one can also improve upon the upper bound by logarithmic bounds; in the case $\ell = 3$ this improvement is the right answer, but in the general case the right answer is not known.

Theorem 5.3 (Ajtai–Komlős–Szemerédi 1980)

There is a constant $C > 0$ such that for all $k, \ell \geq 3$, we have

$$R(k, \ell) \leq \frac{(Ck)^{\ell-1}}{(\log k)^{\ell-2}}.$$

For $\ell = 3$, we more precisely have

$$R(k, 3) \leq \frac{k^2}{\log k - 1}.$$

If we think of ℓ as fixed, then $C^{\ell-1}$ is a constant factor, and the main behavior is still $k^{\ell-1}$, but we have a logarithmic improvement in the denominator (and the number of log factors grows with ℓ — this makes sense because when $\ell = 2$ you can't have any log factors, because k is the exact answer).

Remark 5.4. In some sense logarithmic factors may not seem like a big improvement, but it's far from trivial, and it's actually the right answer when $\ell = 3$ and might be the right answer for larger ℓ (which is not known).

Today we'll prove the bound in the case $\ell = 3$; we'll prove the general case next class.

To prove an upper bound, our goal is to find a number of vertices large enough that any coloring will have a red clique of size k or a blue triangle. We may as well assume that the coloring *doesn't* have a blue triangle; so the blue graph is triangle-free. Meanwhile, a red clique of size k corresponds to an independent set of size k in the blue graph. So our goal is actually to bound the independence number of a triangle-free graph (we wish to show that the independence number is at least k).

Definition 5.5. The *independence number* $\alpha(G)$ is the size of the largest independent set of G, i.e., the size of the largest subset U of vertices such that U contains no edges.

In other words, an independent set is a set with no edges; a set of vertices is independent in the blue graph if and only if it's a clique in the red graph.

We'll do this using the following lemma:

Lemma 5.6

Every triangle-free graph G on n vertices with average degree d has independence number

$$\alpha(G) \geq n \cdot \frac{\log d - 1}{d}.$$

Remark 5.7. Without the triangle-free condition, there's a bound of $\frac{n}{d+1}$ — it's true if the *maximum* degree is d by a greedy algorithm,and it's true for the *average* degree by the Caro–Wei theorem (which can be proved probabilistically). Alternatively, if you don't care about constant factors, you can also run a greedy algorithm with the vertices of degree at most $2d$ (there are at least $\frac{n}{2}$ such vertices).

First we'll show why this lemma implies our upper bound.

Proof of Theorem for $\ell = 3$. Suppose that for some $n \geq \frac{k^2}{\log k - 1}$, there is a coloring of the edges of the complete graph on n vertices with red and blue, containing no red clique of size k and no blue triangle.

Now let G be the graph consisting of the blue edges, which we know is triangle-free and $\alpha(G) \leq k - 1$ (i.e., the size of the largest independent set in the blue graph is at most $k - 1$, because an independent set in blue is precisely a clique in red).

Now we'd like to apply the lemma; the lemma involves the average degree, so we'd like to get a handle on the average degree.

For any vertex v, its neighborhood (i.e., set of neighbors) in G must form an independent set (if there were any edges between them, then we would have a triangle consisting of those two vertices and v). This means $\deg(v) \leq \alpha(G) \leq k-1$. So *every* vertex has degree at most $k-1$, which also means G has average degree $d \leq k-1 \leq k$.

For technical reasons, we will also need a lower bound on the average degree. We may also assume that $d \geq 10$ — otherwise we can find an independent set of size at least $\frac{n}{40}$ by our greedy algorithm discussed earlier (or $\frac{n}{11}$ by the Caro–Wei theorem), which is greater than k.

Now by our lemma, we have

$$k-1 \geq \alpha(G) \geq n \cdot \frac{\log d - 1}{d} \geq n \cdot \frac{\log k - 1}{k},$$

using the fact that the average degree is at most k. (Here we're using the fact that $\frac{\log d - 1}{d}$ is monotone decreasing for $d > e^2$; this is why we wanted to assume $d \geq 10$.)

Now plugging in our value of n, we have

$$k-1 \geq n \cdot \frac{\log k - 1}{k} \geq \frac{k^2}{\log k - 1} \cdot \frac{\log k - 1}{k} = k,$$

which is a contradiction. This means it cannot be true that for $n \geq \frac{k^2}{\log k - 1}$ some coloring contains no red clique of size k or blue triangle. So for all $n \geq \frac{k^2}{\log k - 1}$ we can find our red clique or blue triangle, as desired. □

Now we want to prove the lemma, which isn't really about graph colorings anymore, but is instead a statement about graphs in general. We'll see a proof from a few years later, which is basically magic. It proceeds by simple induction — you can't induct on the problem statement, but the proof tweaks the function a little bit (using a slightly bigger function) so that the induction magically works out. The slightly stronger statement we'll prove (in the induction) is the following:

Lemma 5.8

Every triangle-free graph G on n vertices with average degree d has independence number $\alpha(G) \geq n \cdot f(d)$, where $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ is given by

$$f(d) = \frac{d \cdot \log d - d + 1}{(d-1)^2}$$

for $d \neq 0, 1$, and $f(0) = 1$ and $f(1) = \frac{1}{2}$.

(The definitions of $f(0)$ and $f(1)$ come from taking the limits as $d \to 0$ and $d \to 1$.)

First, we need to check that

$$f(d) \geq \frac{\log d - 1}{d},$$

so that this implies our original lemma. We can ignore the cases where $d = 0$ and $d = 1$ (in which case it is clearly true); then

$$f(d) = \frac{d \log d - d + 1}{(d-1)^2} = \frac{(\log d - 1)d + 1}{(d-1)^2} \geq \frac{(\log d - 1)d}{d^2} = \frac{\log d - 1}{d}.$$

It doesn't seem that we've done much — this function is more complicated, and the function is bigger so our statement is harder — but the fact that the function is bigger also gives us a better inductive hypothesis, which actually makes the statement easier to prove.

Proof. First one can check (via relatively straightforward calculation) that f has certain properties:

- f is continuous on $\mathbb{R}_{\geq 0}$;
- $f'(d) < 0$ for all d (i.e., f is monotone decreasing);
- $f''(d) \geq 0$ for all d (i.e., f is convex).

Furthermore, f satisfies the differential equation

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

for all $d \in \mathbb{R}_{\geq 0}$. (Most likely, the proof was found by attempting to induct and seeing that one needs such a property, and then solving the differential equation to find the function f.)

Now we'll induct on n. The base case $n = 0$ is trivial, since both sides are 0.

Our strategy for forming a large independent set is to choose a vertex x that we'll definitely take in the independent set, and remove x and all its neighbors $y \in N(x)$. Then we find an independent set in the remaining graph (using the inductive hypothesis), and add x to it.

Notation 5.9. We use $N(x)$ to denote the *neighborhood* of x, the set of vertices y such that xy is an edge in our graph (or equivalently the set of neighbors of x).

In particular note that $|N(x)| = \deg(x)$. So when we perform this process, the number of vertices we remove is $\deg(x) + 1$.

This tells us the number of vertices we remove in this process. Meanwhile, the number of *edges* we remove is

$$D(x) = \sum_{y \in N(x)} \deg(y).$$

(This is because it's the number of edges incident to either x or any vertex in $N(x)$; but all edges incident to x are also incident to some vertex in $N(x)$. Note that we're using the fact that $N(x)$ is an independent set, since G is triangle-free — if it weren't, then we'd be double-counting edges.)

Now we know the number of edges and vertices we've removed, so we can calculate the average degree of the remaining graph.

For any vertex x, let G_x be the graph obtained from G by removing the vertices in $\{x\} \cup N(x)$. We have

$$|V(G_x)| = n - 1 - \deg(x)$$
$$|E(G_x)| = |E(G)| - D(x) = \frac{dn}{2} - D(x)$$

(where d is the average degree of G). Let d_x be the average degree of the remaining graph; then we have

$$d_x = \frac{|E(G_x)|}{2\,|V(G_x)|} = \frac{dn - 2D(x)}{n - 1 - \deg(x)}.$$

This is a really messy formula (especially if we consider the fact that $D(x)$ is itself a sum), but magically, we *will* be able to keep control over it.

Now by the induction hypothesis (as G_x has fewer vertices than G), the graph G_x contains an independent set of size at least

$$\alpha(G_x) \geq |V(G_x)| \cdot f(d_x) = (n - 1 - \deg(x))f(d_x).$$

We'll now take such an independent set and add x to it; this forms an independent set in G, since all neighbors of x were removed and so cannot be part of the independent set in G_x. So by adding x to this independent set, in G we can find an independent set of size at least

$$\alpha(G_x) + 1 \geq (n - 1 - \deg(x))f(d_x) + 1.$$

This means

$$\alpha(G) \geq (n - 1 - \deg(x))f(d_x) + 1.$$

(This looks hopeless, as our expressions are rather horrendous, but it will actually work out.)

This inequality holds for all $x \in V(G)$. Now we want to make a good choice of x. In fact, we'll choose x *randomly* and take the expectation of the right-hand side — in other words (to phrase the argument non-probabilistically), since $\alpha(G)$ is bounded by this expression for *all* vertices x, it's in particular bounded by the *average* of this expression over all x.

Averaging our bound over all x, we get that

$$\alpha(G) \geq \frac{1}{n} \sum_{x \in V(G)} ((n - 1 - \deg(x))f(d_x) + 1).$$

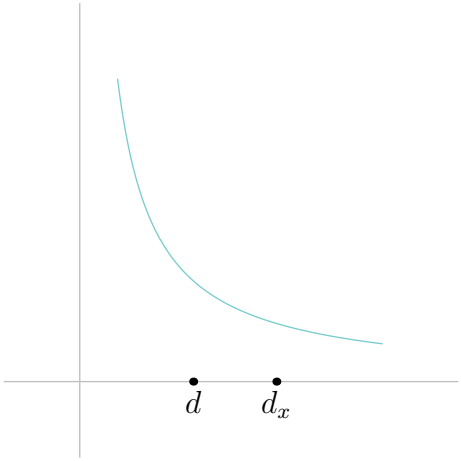
We can pull the 1 out to rewrite this as

$$\alpha(G) \geq 1 + \sum_{x \in V(G)} \frac{n - 1 - \deg(x)}{n} \cdot f(d_x).$$

We now want to check that this expression is at least $nf(d)$.

There's two parts of the sum which depend on x. The first is $f(d_x)$, and the second is $\deg(x)$. It's pretty clear that $f(d_x)$ is the more ugly function, since d_x is an ugly expression, and f is itself ugly. So it's hopeless to try to use $f(d_x)$ directly; this means we'll try to bound $f(d_x)$.

For this, we'll use the convexity of f to relate $f(d_x)$ to $f(d)$.



By convexity (comparing $f(d_x)$ to the tangent line at d), we have

$$f(d_x) \geq f'(d)(d_x - d) + f(d).$$

This term is much nicer because $f(d)$ and $f'(d)$ don't depend on d — so the only x-dependent term is d_x (which we're no longer trying to plug into our function f). So we now have

$$\alpha(G) \geq 1 + \sum_{x \in V(G)} \frac{n - 1 - \deg(x)}{n} \left(f(d) + (d_x - d)f'(d)\right).$$

Now the latter expression has a lot of terms which don't depend on x. Separating these terms, we can write

$$\alpha(G) \geq 1 + \sum_{x \in V(G)} \frac{n-1-\deg(x)}{n}(f(d) - df'(d)) + \frac{1}{n}\sum_{x \in V(G)} (n-1-\deg(x)) \cdot d_x \cdot f'(d).$$

The next-most complicated term is d_x, so we'll now deal with it. Recall that

$$d_x = \frac{dn - 2D(x)}{n-1-\deg(x)},$$

so the denominator cancels out and our summand is simply

$$(n-1-\deg(x))d_x = dn - 2D(x).$$

Meanwhile, we can simplify the first sum as $\sum_{x \in V(G)} \frac{n-1-\deg(x)}{n} = n-1-d$ (since d is the average degree). So then we have

$$\begin{aligned}\alpha(G) &\geq 1 + (n-1-d)(f(d) - df'(d)) + \frac{1}{n}\sum_{x \in V(G)} (d_n - 2D(x))f'(d) \\ &= 1 + (n-1-d)(f(d) - df'(d)) + dnf'(d) - \frac{2f'(d)}{n}\sum_{x \in V(G)} D(x).\end{aligned}$$

Next we'll evaluate this sum — we have

$$\sum_{x \in V(G)} D(x) = \sum_{x \in V(G)} \sum_{y \in N(x)} \deg(y).$$

Our summand only depends on y, so we can swap the sum as

$$\sum_{y \in V(G)} \sum_{x \in N(y)} \deg(y) = \sum_{y \in V(G)} (\deg(y))^2.$$

(For each y, the summand of $\deg(y)$ appears $\deg(y)$ times — once for each neighbor of y — so we get the sum of *squares* of the degrees.) This tells us

$$\sum_{x \in V(G)} D(x) = \sum_{y \in V(G)} (\deg(y))^2 \geq n \cdot d^2$$

by the AM–QM inequality (note that $f'(d) < 0$, so we want a lower bound on this sum). This tells us

$$\begin{aligned}\alpha(G) &\geq 1 + (n-1-d)(f(d) - df'(d)) + dnf'(d) - \frac{2f'(d)}{n} \cdot nd^2 \\ &= (n-1-d)f(d) + 1 + (d-d^2)f'(d).\end{aligned}$$

But by our differential equation on d, we have $1 + (d-d^2)f'(d) = (d+1)f(d)$, which means this comes out to exactly $nf(d)$.

So by induction we're done. □

§6 March 7, 2023 — Upper Bounds for Off–Diagonal Ramsey Numbers

We consider the two-color Ramsey numbers for graphs $R(k, \ell)$. *Off-diagonal* means we're looking at the numbers far from the diagonal — typically this means we fix one number and let the other be large. Since $\ell = 2$ is uninteresting, we fix $\ell \geq 3$ and let k be large; we then want to bound $R(k, \ell)$ in this case.

The best known upper bound for $R(k, \ell)$ for fixed $\ell \geq 3$ (and large k) is as follows.

Theorem 6.1 (Ajta–Komlós—Szemerédi 1980)

There is a constant $C > 0$ such that for all $k, \ell \geq 3$ we have

$$R(k, \ell) \leq \frac{(Ck)^{\ell-1}}{(\ln k)^{\ell-2}}.$$

For $\ell = 3$, this can be strengthened to

$$R(k, 3) \leq \frac{k^2}{\ln k - 1}.$$

You should think of this as a bound when ℓ is fixed and k is large; then this is roughly $k^{\ell-1}$. A bound of $k^{\ell-1}$ itself is not very interesting, since it follows directly from the general bound; but the point is that this upper bound improves on that by a logarithmic factor. It's nice we have exactly $\ell - 2$ logarithms in the denominator, since when $\ell = 2$ the answer is exact and we shoudn't have any logarithms.

(The bound in the $k = 3$ case is a bit better because it removes the absolute constant. This bound is due to Shearer from a few years later, and the proof we saw last time is also due to Shearer. The constant factor doesn't matter too much, but this is still the best constant factor known.)

In the case $\ell = 3$, this bound is also tight up to an absolute constant; but the bound for $\ell > 3$ is not known to be tight (we don't even know if the exponent is).

We proved the $\ell = 3$ case last time, so today we'll prove the $\ell > 3$ case.

Last time, we deduced the $\ell = 3$ case from the following lemma:

Lemma 6.2

Every triangle-free graph G on n vertices with average degree d has independence number

$$\alpha(G) \geq n \cdot \frac{\ln d - 1}{d}.$$

What this lemma tells us is a lower bound for the independence number of a triangle-free graph; this is exactly what we needed for $R(k, 3)$ (since if the blue graph is triangle-free, we want to find an independent set of size at least k).

Proof. We'll use induction on ℓ. Let $C = 5000$ for concreteness (or any sufficiently large value).

For $\ell = 3$, we've already proven that

$$R(k, 3) \leq \frac{k^2}{\ln k - 1} \leq \frac{C^2 k^2}{\ln k}.$$

Now assume that $\ell \geq 4$, and that we've already proven the statement for all $\ell' < \ell$ — i.e., for all $\ell' \in \{3, 4, \ldots, \ell - 1\}$. For a technical reason, we may need to take $\ell' = 2$ as well — in that case, the inequality is trivially true because $R(k, 2) = k$.

Let $k \geq 3$, and assume for contradiction that for

$$n = \left\lfloor \frac{(Ck)^{\ell-1}}{(\log k)^{\ell-2}} \right\rfloor,$$

there is a coloring of the edges of a complete graph on n vertices with red and blue, without a red clique of size k or a blue clique of size ℓ. (This is the correct number of vertices because we want to show that $\frac{(Ck)^{\ell-1}}{(\log k)^{\ell-2}}$ is good enough; this typically isn't an integer, so we need to round down.)

As a first step, similarly to in the $\ell = 3$ case, we will rephrase the problem in terms of edges and non-edges — in other words, let G be the graph consisting of the blue edges. (You can always translate back and forth between colors and the clique and independence number, but in this case it's easier to think about it in this way.)

Then G still has n vertices, and it does *not* have a clique of size ℓ. An independent set in G is precisely the same as a red clique; we know there is no red clique, so $\alpha(G) \leq k - 1 < k$. (If you think of k as large, the difference between $k-1$ and k is not big, so we don't have to be careful of the $+1$; meanwhile ℓ is small, so we do have to be careful with it — replacing ℓ with $\ell + 1$ would significantly change the bound.)

Now we can kind of forget about the original coloring. But the next argument we'll do is very similar to what we did when we first upper-bounded Ramsey numbers (in the very first class, where we proved the bound of $\binom{k+\ell-2}{k-1}$); in the next step of this proof, we will use a similar idea as from then.

In that case, we fixed a vertex and looked at its red neighbors and blue neighbors. We could bound the size of the red neighborhood by the inductive hypothesis, since it has no red clique of size $k-1$ or blue clique of size ℓ; we could do the same with the blue neighborhood.

Here our problem is more asymmetric; we will only look at what corresponds to the blue neighborhood, since we won't get much out of changing k by 1. So we'll focus on the neighborhood of our vertex in G.

For every vertex $v \in V(G)$, the neighborhood $N(v)$ (the set of vertices adjacent to v in G) cannot have a clique of size $\ell - 1$ (since if it did, we could add v and get a clique of size ℓ), and it also has no independent set of size k (because the whole graph has no independent set of size k). Now we'll apply the inductive hypothesis to this neighborhood; by the induction hypothesis,

$$|N(v)| \leq R(k, \ell - 1) \leq \frac{(Ck)^{\ell-2}}{(\ln k)^{\ell-3}}.$$

Now we have an upper bound on the neighborhood of every vertex, which gives us an upper bound on the degree of every vertex. Let $d = \frac{(Ck)^{\ell-2}}{(\ln k)^{\ell-3}}$, so $\deg(v) \leq d$ for all $v \in V(G)$. In particular,

$$|E(G)| \leq \frac{nd}{2}.$$

Now we've found an upper bound on the degree of every vertex, which in particular bounds the average degree. But now we're sort of stuck — we can't apply the lemma, because the graph is not triangle-free. So maybe we need to go further with the induction hypothesis — we'll go one step deeper down, because this will tell us something about the triangles of the graph — which will be useful because then we can try to get rid of the triangles and apply the lemma.

Here we picked one vertex and looked at its neighborhood. To induct one step further, we should look at two adjacent vertices and their common neighborhood.

For any two adjacent vertices $v, w \in V(G)$, we look at their common neighborhood $N(v) \cap N(w)$ (i.e., all vertices which are adjacent to both v and w). This common neighborhood has no clique of size $\ell - 2$ (if we had a clique of size $\ell - 2$ in this common neighborhood, we could add both v and w to get a clique of size ℓ); and it also has no independent set of size k. So by the induction hypothesis, since we have a red-blue coloring without a red clique of size k or blue clique of size $\ell - 2$, we must have

$$|N(v) \cap N(w)| \leq R(k, \ell - 2) \leq \frac{(Ck)^{\ell-3}}{(\ln k)^{\ell-4}}.$$

(Note that we assumed $\ell \geq 4$, which means $\ell - 2 \geq 2$; this is why we needed the technicality that the statement is also true for 2.)

Note that $(\ln k)^4 < d$ (d is smaller when ℓ is smaller, so we can assume $\ell = 4$; then we're comparing some power of $\log k$ with some power of Ck; as soon as C is big enough this is true). Also let $\varepsilon = \frac{1}{\ell-2}$ (since it'll appear a lot in the proof). Note that $(\ell-2)(1-\varepsilon) = \ell - 3$ and $(\ell-3)(1-\varepsilon) = \ell - 4 + \varepsilon$. (These are simple calculations, but it's easier to note them now rather than in the middle of a bigger calculation later, when we will use them.)

Returning to the actual argument, we have a bound on the degree of every vertex, and a bound on $|N(v) \cap N(w)|$. Our strategy was to apply the lemma, but the lemma applies to a triangle-free graph; so we want to make our graph triangle-free.

So our next step is to upper-bound the number of triangles.

> **Claim —** The graph G contains at most $n \cdot d^{2-3\varepsilon/4}$ triangles.

Proof. To choose a triangle in G, we have n choices for the first vertex v, and then $\deg v \le d$ choices for the second vertex w (because it needs to be a neighbor of v). Then the number of choices for the third vertex is

$$|N(v) \cap N(w)| \le \frac{(Ck)^{\ell-3}}{(\ln k)^{\ell-4}}.$$

(In fact, we are counting every triangle 6 times for each of the orders, but this doesn't really matter.)

So the number of triangles is at most

$$n \cdot d \cdot \frac{(Ck)^{\ell-3}}{(\ln k)^{\ell-4}}.$$

Now we'll bound the latter term by $d^{1-3\varepsilon/4}$ — this is

$$nd \cdot \frac{(Ck)^{(\ell-2)(1-\varepsilon)}}{(\ln k)^{(\ell-3)(1-\varepsilon)-\varepsilon}} = nd \cdot d^{1-\varepsilon} \cdot (\ln k)^{\varepsilon}.$$

We stated earlier that $\ln k \le d^{1/4}$, so this is at most $nd \cdot d^{1-\varepsilon} \cdot d^{\varepsilon/4} = nd^{2-3\varepsilon/4}$, as desired. □

So now we have an upper bound on the number of triangles in G. We'd like to apply the lemma on the triangle-free graph. Our graph isn't triangle-free, so we would like to make it triangle-free.

We'd like to remove a vertex from each triangle, but the number of triangles may be greater than the number of vertices. So first we're going to pass to a subgraph of G where the number of triangles is even smaller — we'll do this by choosing the subgraph randomly. Then we'll have a better trade-off between the number of triangles and vertices, and we will actually be able to remove a vertex from each triangle.

First, why does it help us to go to a random subset? If we take every vertex with probability p, then a p fraction of the vertices, p^2 of the edges, and p^3 of the triangles survive — so we thin out the number of triangles more than the other things. (It matters that our bound here is significantly less than n^3 — if it weren't, then we wouldn't be able to choose a good sampling probability.)

With a probabilistic argument, we'll find a triangle-free induced subgraph of G to apply the lemma.

Let our sampling probability be $p = d^{-1+\varepsilon/4}$. (This is trivially between 0 and 1.) (As an exercise, you can check where this number comes from by going through the calculation keeping the variable p, and seeing what value of p makes this work. The 4 is a bit arbitrary (since $\frac{3}{4}$ was arbitrary in the previous claim); the point is we need something a bit bigger than d^{-1}.)

Let $V' \subseteq V$ be a random subset of V, obtained by including every vertex with probability p independently.

First, we have

$$\mathbb{E}[|V'|] = pn = nd^{-1+\varepsilon/4}.$$

Here we need to be a bit more careful — we also need to check that most of the time, the number is not much smaller than this. There's various ways to do this; here we can apply the Chernoff bound, but we can also do this using Chebyshev's inequality. We have

$$\mathrm{Var}[|V'|] = p(1-p)n < pn.$$

So by Chebyshev's inequality, most of the time this number is at least half the expectation — more precisely,

$$\mathbb{P}[|V'| < \frac{1}{2}pn] \leq \frac{pn}{\frac{1}{4}p^2n^2} = \frac{4}{pn} < \frac{1}{4}.$$

(You can check that $pn > 16$, which is true because this is at least $\frac{n}{d}$, which is at least $\frac{(Ck)}{\log k}$ — this bound is very weak, and in reality the probability is actually exponentially small, we just don't need that.)

SO with probability at least $\frac{3}{4}$, we have a reasonable number of vertices in V' (i.e., at least half the expected number).

Let G' be the induced subgraph of G on the vertex set V'. We now want to analyze the relevant parameters of G', namely its number of edges andn triangles. First

$$\mathbb{E}[|E(G')|] = p^2\,|E(G)| \leq p^2 \cdot \frac{nd}{2} = \frac{1}{2}nd^{-1+\varepsilon/2}.$$

We also have

$$\mathbb{E}[\#\text{triangles of } G'] = p^3 \#\text{triangles in } G \leq p^3 \cdot nd^{2-3\varepsilon/4} = \frac{n}{d}$$

(plugging in our definition of p).

In the previous case we had to say that with high likelihood we had enough vertices. But we want the average degree and triangle small, so we want to say that typically they are small; this means we can just apply Markov (and we don't even have to go through the second moment) — by Markov's inequality,

$$\mathbb{P}[|E(G')|] > 4p^2\,|E(G)| \leq \frac{1}{4},$$

and similarly

$$\mathbb{P}[\text{triangles} > 4p^3\text{triangles}] \leq \frac{1}{4}.$$

(This is by Markov's inequality.)

Now we have three unusual events — each happens with probability at most $\frac{1}{4}$ — so by the union bound the probability any of them happens is at most $\frac{3}{4}$, which means with probability at least $\frac{1}{4}$, none of these events happen. So with positive probability — with probability at least $\frac{1}{4}$ — none of these things happen. In particular, there exists some choice of V' for which none of these weird inequalities happens — i.e., there is some outcome for V' and G' satisfying all of the following inequalities:

- $|V'| \geq \frac{1}{2}pn = \frac{1}{2}nd^{-1+\varepsilon/4}$.
- $|E(G')| \leq 4p^2\,|E(G)| \leq 2nd^{-1+\varepsilon/2}$.
- $\#\text{triangles in } G' \leq 4p^3 \#\text{triangles in } G \leq \frac{4n}{d}$.

Now we can forget about the randomness; fix the outcome of V' so that these conditions hold. Our goal is to apply the lemma. Our graph is still not triangle-free, but it has few triangles. But now we can finally do the process of removing one vertex from every triangle in G' — we're in better shape to do this now than we were before, because this number of triangles is way smaller than the number of vertices.

Now let $V'' \subseteq V'$ be obtained by deleting one vertex from every triangle in G' (for each triangle, we pick an arbitrary vertex and delete it). Then

$$|V''| \geq |V'| - \#\text{triangles} \geq \frac{1}{2}nd^{-1+\varepsilon/4} - 4nd^{-1}.$$

(this is an inequality because some vertices may have been double-deleted, but that only helps us). The first quantity is at least twice as large as the second — this requires $d^{\varepsilon/4} > 16$, or in other words $d^{\varepsilon} > 16^4$. But d^{ε} is roughly Ck (up to some logs); if we make C big enough then this is true. SO then

$$|V''| \geq \frac{1}{4} n d^{-1+\varepsilon/4}.$$

(This means we really have not lost much in terms of the number of vertices.)

Let G'' be the induced subgraph on V''. Then G'' is triangle-free (since every triangle in G' was destroyed by removing a vertex, so G'' is definitely triangle-free). We only need an upper bound on the number of edges; we trivially have

$$|E(G'')| \leq |E(G')| \leq 2nd^{-1+\varepsilon/2}.$$

So now we are in pretty good shape to apply the lemma — we have a triangle-free graph on some number of verties (which we have a lower bound for), we know the graph is triangle-free, and we can control its average degree because we have control of the number of edges — the average degree of G'' is

$$\frac{2\,|E(G'')|}{|V''|} \leq 2 \cdot \frac{2nd^{-1+\varepsilon/2}}{\frac{1}{4}nd^{-1+\varepsilon/4}} = 16d^{\varepsilon/4}.$$

Now we can apply the lemma to G'' to lower bound its independence number — by our lemma, (since G'' i triangle-free) G'' has independence number

$$\alpha(G'') \geq |V''| \cdot \frac{\ln(16d^{\varepsilon/4})}{16d^{\varepsilon/4}}$$

(note that the lemma only gives us the bound if we plug in the actual average degree; this is usually fine because the function is monotone starting from 10, so if the average degree is at least 10 then this inequality is justified. However if the average degree is less than 10 (which might be the case), we know that in general the independence number is at least V'' divided by the average degree; and $\frac{V''}{11}$ is also bigger than the term written here, since $d^{\varepsilon/4} > 100$.)

To eat up th e-1, this is at least

$$\geq \frac{1}{5} n d^{-1+\varepsilon/4} \cdot \frac{\varepsilon/4 \cdot \log(d)}{16d^{\varepsilon/4}} \geq \frac{1}{320}\frac{n}{d} \cdot \varepsilon \ln(d).$$

Now this tells us

$$\alpha(G'') \geq \frac{1}{325} \cdot \frac{Ck}{\ln k} \cdot \varepsilon \cdot ((\ell-2)\log(Ck) - (\ell-3)\log\log k).$$

(we change 320 to 325 to account for the rounding on n — then d is the same as n but with both exponents lowered). The ε cancels the $\ell - 2$; we can replace the $\ell - 3$ by $\ell - 2$ as well (since we want to lower-bound) so that also cancels; then This is

$$\geq \frac{1}{325}\frac{Ck}{\ln k}(\ln(Ck) - \ln\ln k).$$

The term on the right is roughly $\ln k$ (since it's $\ln k + \ln C - \ln\ln k$, and $\ln\ln k$ is much smaller than $\ln k$ (unless k is small, in which case it's much smaller than $\ln C$)). Bounding it by $\frac{1}{2}\ln k$, we now get

$$\alpha(G'') \geq \frac{1}{325}\frac{Ck}{\ln k} \cdot \frac{1}{2}\ln k \geq \frac{1}{750}Ck > k.$$

But now we have

$$\alpha(G) \geq \alpha(G'') > k,$$

contradiction (since we only restricted to a vertex subset — G'' is just an induced subgraph, so every independent set in G'' is also one in G). □

This finishes the proof — this coloring can't exist, so we get the desired upper bound on the Ramsey number.

Remark 6.3. For some intuition, why does increasing ℓ by 1 give us an additional log factor? This essentially cojmes from this last step.

Remark 6.4. A lower bound is the following theorem of Spencer (1977): For every fixed $\ell \geq 3$, there is a constant c_ℓ such that

$$R(k,\ell) \geq c_\ell \cdot \left(\frac{k}{\ln k}\right)^{(\ell+1)/2}$$

for all $k \geq 3$. The best-known lower bound improves on this by another log factor — i.e.,

$$c_\ell \left(\frac{k}{\ln k}\right)^{(\ell+1)/2} (\ln k)^{1/(\ell-2)}.$$

Here the main term is $(k/\log k)^{(\ell+1)/2}$, whereas ignoring the constant factors depending on ℓ, our original bound had $k^{\ell-1}/(\ln k)^{\ell-2}$. If we take $\ell = 3$, the upper bound gives $\frac{k^2}{\log k}$ and Spencer's gives $\frac{k^2}{(\log k)^2}$; for a long time it wasn't clear whether it should be log or $\log^2$ in the denominator, a nd this was resolved by Kim in 1995 (who added the factor of $\log k$). We now know that if $\ell = 3$ then the correct answer is indeed $\frac{k^2}{\log k}$. But if we take $\ell > 3$, then even the exponent of k differs (as we have $k^{\ell-1}$ vs. $k^{(\ell+1)/2}$). SO for $\ell > 3$ it' snot just about finding the log power — even the correct exponent isn't known.

§7 Lower bounds for off-diagonal Ramsey numbers

On Tuesday, we proved an upper bound for the off-diagonal Ramsey numbers, of

$$R(k,\ell) \leq \frac{(Ck)^{\ell-1}}{(\ln k)^{\ell-2}}.$$

Off-diagonal Ramsey numbers mean we're interested in the case where ℓ is fixed and k is large. If ℓ is fixed, then C^ℓ is a constant, so we have a bound of the form

$$R(k,\ell) \leq C_\ell \cdot \frac{k^{\ell-1}}{(\ln k)^{\ell-2}}.$$

Today we'll talk about lower bounds; we will see that if $\ell \geq 4$, the lower bound doesn't match the upper bound, so the correct behavior isn't known.

Theorem 7.1 (Spencer 1977)

For every fixed $\ell \geq 3$, there is some constant $c_\ell > 0$ such that

$$R(k,\ell) > c_\ell \left(\frac{k}{\ln k}\right)^{(\ell+1)/2}$$

for all $k \geq 3$.

Comparing this to the upper bound, when $\ell = 3$ the exponent of the k matches (it's k^2 in both cases), but the exponents of the log term don't match. Meanwhile, for $k \geq 4$ even the exponents of k don't match.

For $k = 3$, figuring out the correct exponent of the log was an open problem for a long time. This was resolved by Kim in the mid-90s; it turns out the correct answer is indeed $\frac{k^2}{\log k}$, and not $\frac{k^2}{(\log k)^2}$. This was reproved by Bohman later, and this was then generalized by Bohman and Keevash in 2010 to the bound

$$R(k,\ell) \geq c_\ell \left(\frac{k}{\ln k}\right)^{(\ell+1)/2} \cdot (\ln k)^{1/(\ell-2)}.$$

Now for $\ell = 3$, this matches the upper bound; but for $\ell \geq 4$ even the exponent of k still does not match, so determining the correct behaviour is still wide open.

We won't discuss the improvement (which is very complicated). The approach for Spencer's proof is a probabilistic construction (unsurprisingly, given that our lower bounds for the *diagonal* Ramsey numbers were also probabilistic). If you approach this probabilistically with the same approach as we had for diagonal Ramsey numbers, you only get a weaker power ($\frac{\ell-1}{2}$ instead of $\frac{\ell+1}{2}$). To get the right exponent, we have to do something more involved — applying the Lovász Local Lemma. Still, it's a relatively simple probabilistic proof. The additional improvement is much more complicated. There, the idea is not just to look at a standard random graph, but instead to change the random graph model — rather than defining each pair of vertices to be an edge independently with some probability, they consider the K_ℓ-free process — where we start with a vertex set on the correct number of vertices and no edges; then place the first edge anywhere randomly; then place the second edge randomly among all available edges; but as more and more edges get added, we add the new edge randomly but making sure not to create a K_ℓ. In other words, if we have a pair of vertices where adding an edge would create a K_ℓ, then we forbid adding that edge; at every step among all possible edges to add, we add one uniformly at random. As you can imagine, the edges are now not independent anymore; these dependencies make the process much more complicated to analyze.

§7.1 The Lovász Local Lemma

The proof relies on the Lóvasz Local Lemma, so we'll first state it. As a disclaimer, we'll state it in a form which isn't really the most general form, but is the form it's usually applied in and is probably easiest to parse. (If you have seen the statement before, you may remember there's a notion of a dependency graph, which is kind of subtle — you have to care about *mutual* independence of the non-neighbors, not just pairwise independence — but we'll state the lemma in a specialized form that avoids these subtleties.)

Theorem 7.2 (Lovász Local Lemma)

Let $Z_1, \ldots, Z_m$ be independent random variables, and let $\mathcal{E}_1, \ldots, \mathcal{E}_n$ be events such that each $\mathcal{E}_i$ is determined by the outcomes of the random variables Z_j with $j \in J_i$ for some $J_i \subseteq [m]$ (i.e., each event only depends on a certain subset of the random variables, namely those with indices in the set J_i). Suppose there exist real numbers $x_1, \ldots, x_n$ such that for each $i \in [n]$ we have $0 \leq x_i < 1$ and

$$\mathbb{P}[\mathcal{E}_i] \leq x_i \prod_{k \in [n] \setminus \{i\}, J_i \cap J_k \neq \emptyset} (1 - x_k).$$

(In other words, we take the indices such that the events $\mathcal{E}_i$ and $\mathcal{E}_k$ depend on at least one common variable.) Then the probability that none of the events $\mathcal{E}_1, \ldots, \mathcal{E}_n$ occurs is

$$\mathbb{P}[\overline{\mathcal{E}_1} \cap \cdots \cap \overline{\mathcal{E}_n}] \geq \prod_{i=1}^{n} (1 - x_i) > 0.$$

When we apply the theorem, we'll think of Z_i as the colors of the different edges (or the indicator variables for whether an edge exists). The $\mathcal{E}_i$ will be whether a certain set of vertices forms a monochromatic clique; this only depends on the colors of the edges between those vertices.

If you've never seen this before, it might look a bit strange. But here's some motivation — we want to think of the $\mathcal{E}_i$ as bad events (events we want to avoid). The conclusion of the lemma tells us that with positive probability, we can avoid all the events (which is good because these events are bad; if we can avoid them with positive probability, then there exists some coloring which avoids all these bad events, which is what we want). The precise probability bound typically doesn't matter — we only really care that it's positive, because that tells us that there *is* a good outcome.

What does $x_i \prod(1 - x_k)$ mean? Let's think about the simple toy case where the subsets J_i are all disjoint — so each of the events depends on its own subset of random variables, and they don't interfere at all. This means the events are all independent from each other. In that setting, the second product has no factors, so that tells us $\mathbb{P}[\mathcal{E}_i] \leq x_i$. In that case, the theorem is trivial — it tells us that we have independent events with $\mathbb{P}[\mathcal{E}_i] \leq x_i$, and then the probability none of the events hold is at least $\prod_{i=1}^{n}(1 - x_i)$.

The point of the theorem is that we can still draw a good conclusion if there's some dependencies — the theorem is good to apply when there's only *few* intersections between the sets. If there's too many intersections then the second product becomes really small, and we can't satisfy it. But if there's only a few, then this product doesn't have too many factors, so it doesn't change x_i by too much.

The upshot is that you want to apply Lovász Local Lemma when your events aren't necessarily independent, but there aren't too many dependencies between them.

(We will prove the Lovász Local Lemma on the homework.)

The Lovász Local Lemma also has a simplified symmetric form; this won't be helpful for our problem here, but it's helpful for another homework problem.

(This is essentially what happens when we take all x_i to be equal, and each product has the same number of factors — this works well when the random variables and events have some reasonable amount of symmetry.)

Theorem 7.3 (Symmetric Lovász Local Lemma)

Let $Z_1, \ldots, Z_m$ be independent random variables and let $\mathcal{E}_1, \ldots, \mathcal{E}_n$ be events where for each $i \in [n]$, the event $\mathcal{E}_i$ is determined by the outcomes of the random variables Z_j with $j \in J_i$ for some subset $J_i \in \{1, \ldots, m\}$. Suppose we have $\mathbb{P}[\mathcal{E}_i] \leq p$ for all $i \in [n]$, and also suppose that for each $i \in [n]$, there are at most d indices $k \in [n] \setminus \{i\}$ such that $J_i \cap J_k \neq \emptyset$. Then if $e \cdot p \cdot (d + 1) \leq 1$, then

$$\mathbb{P}[\overline{\mathcal{E}_1} \cap \cdots \cap \overline{\mathcal{E}_n}] > 0.$$

In other words, we're assuming every set J_i intersects at most d other sets (so our product in the original would have at most d factors). Here e is the usual constant ($e = 2.718\cdots$).

Proof. First, if $d = 0$ then the statement is trivial (as then all the events $\mathcal{E}_i$ are independent). Now assume $d \geq 1$, and apply LLL with $x_i = \frac{1}{d+1}$ for all $i \in [n]$. Now we have

$$\mathbb{P}[\mathcal{E}_i] \leq p \leq \frac{1}{d+1} \cdot \frac{1}{e} < \frac{1}{d+1} \cdot \left(1 - \frac{1}{d+1}\right)^d \leq x_i \prod_k (1 - x_k),$$

since the number of factors in the product is at most d. □

We won't use this symmetric version because we have too much asymmetry between our bad events — the bad events we're trying to avoid are red cliques of size k and blue cliques of size ℓ, which behave very differently because $k \gg \ell$. But as you can imagine, if you wanted to lower-bound $R(k, k)$ (which we will do on the homework), then everything does become symmetric and it is easier to apply the symmetric version of LLL.

§7.2 Proof of Spencer's Lower Bound

Now we'll prove Spencer's lower bound for $R(k, \ell)$ (i.e., the above theorem).

Proof. Fix $\ell \geq 3$ (whenever we use asymptotic notation, we think of ℓ as fixed and k as large). It doesn't matter whether we assume $k \geq 3$ or prove the bound for all k sufficiently large with respect to ℓ — if we

prove the bound for all sufficiently large k, then we can change the constant c_ℓ to satisfy the theorem for the finitely many remaining values of k. So we may assume that k is sufficiently large with respect to ℓ.

In particular, we'll use $o(1)$ notation for terms which converge to 0 as $k \to \infty$. (Then at some point they'll be smaller than certain thresholds we care about, so for large k the inequalities we want on those terms will be true.)

Let c_ℓ be smlal enough that we have

$$\ell \geq 4(12\ell)^{\binom{\ell}{2}} c_\ell^{\ell-2}$$

(we will see where this comes from later; but the logic becomes clearer if we fix c_ℓ now. If you were to try to find such a proof, you'd think in your head that c_ℓ is small enough, and determine how small it is later). (Note that the true c_ℓ may be smaller, to account for the finitely many remaining k.)

Now let

$$n = \left\lfloor c_\ell \left(\frac{k}{\ln k}\right)^{(\ell+1)/2} \right\rfloor.$$

We want to show that this number of vertices is *not* good enough for Ramsey's theorem — i.e., we want to show that there exists a coloring of the edges of a complete graph on n vertices with red and blue with no red clique of size k or blue clique of size ℓ. (If we are able to establish that there is such a coloring, that shows n is not good enough in Ramsey's theorem, so we must have $R(k,\ell) > n$, and therefore $R(k,\ell)$ is at least the stated bound.)

Finding such a coloring explicitly is extremely difficult (no one has succeeded at this), so similarly to how we proved lower bounds for *diagonal* Ramsey numbers, we again want to take a random coloring. When we did lower bounds for the diagonal numbers, we set every edge to be red with probability $\frac{1}{2}$, and blue with probability $\frac{1}{2}$. Here, $\frac{1}{2}$ is not the right probability, since k and ℓ are very different. In particular, we want edges to be more likely to be red — since we have much more space with red edges.

Color each edge blue with probability

$$p = 12\ell \frac{\ln k}{k},$$

and red otherwise, independently for all edges. (This will come from if we run the whole argument and then optimize; we will not do this, because it's kind of messy. The 12 is not important, because we don't really care about the constant; the important thing is the $\ell \frac{\ln k}{k}$ behavior.) As a reality check, as k grows this probability is indeed pretty small.

This gives independent random variables $Z_1, \ldots, Z_m$ indicating the colors of the edges (with $m = \binom{n}{2}$). (Technically, you need to label the edges from 1 to $\binom{n}{2}$; but the ordering of the variables doesn't really matter, only which events have variables in common.)

Our bad events $\mathcal{E}_i$ will be the occurrences of cliques we want to avoid — we will have $\binom{n}{k} + \binom{n}{\ell}$ events, in two types. We can compute their probabilities — for any set of ℓ vertices, the probability of forming a blue clique of size ℓ on those vertices is $p^{\binom{\ell}{2}}$ (since we have $\binom{\ell}{2}$ edges which we need to be blue).

Similarly, for any set of k vertices, the probability of forming a a red clique of size k on those vertices is $(1-p)^{\binom{k}{2}}$. (There are $\binom{k}{2}$ edges between them, and each is red with probability $1-p$.)

Since these probabilities will come up quite often, let $q = p^{\binom{\ell}{2}}$ and $q' = (1-p)^{\binom{k}{2}}$ be the probabilities of our bad events.

In order to not blow up our notation, we won't name the events; consider the $\binom{n}{\ell} + \binom{n}{k}$ events where for all $\binom{n}{\ell}$ sets of ℓ vertices we consider the event of forming a blue clique, and for all $\binom{n}{k}$ sets of k vertices we consider the event of forming a red clique. Our goal is to avoid all these events, so we want to show that with positive probability, none of these events occur. We will show this using the Lovász Local Lemma.

Our goal is to find an assignment of x_i which satisfy the conditions of LLL; then it will tell us that with positive probability we avoid all the events simultaneously, and therefore we're done. So our remaining task is to find numbers x_i such that our given inequality is satisfied.

We have some flexibility, because for each of the events we can choose a different x_i. But we only want to choose 2 — the ℓ-clique events are all symmetric, and the k-clique events are symmetric, so we really only want to define two values x and x'.

To figure out the relevant inequalities, let's figure out how many factors of each type we have. To figure this out, we want to figure out for every given i, how many different k there are and which types of events they correspond to.

For a subset S of ℓ vertices (corresponding to one of the blue ℓ-clique events), let $d_{\ell\ell}$ be the number of vertex subsets S' with $|S'| = \ell$ and $|S \cap S'| \geq 2$ (i.e., S and S' contain a common edge). We care about this because this corresponds to the dependencies — for a set of ℓ vertices, the set J of edges is all the $\binom{\ell}{2}$ edges lying between these vertices (their colors are the random variables which our event depends on — to tell whether our ℓ vertices are a blue clique, we only care about the edges between them). Then for two events to have nonempty intersection, we need the two sets of size ℓ to share an edge, which occurs if and only if they have intersection size at least 2. So the number of factors appearing in this product is the number of subsets (of the correct size) which intersect our original set S in at least 2 vertices. So our number $d_{\ell\ell}$ counts the number of factors corresponding to a blue-clique event.

Similarly, when S has size ℓ, we also need to consider the events where $|S'| = k$. Let $d_{\ell k}$ be the number of vertex subsets with $|S'| = k$, and $|S \cap S'| \geq 2$.

We define $d_{k\ell}$ and d_{kk} similarly, for the case where $|S| = k$. (So the first index is the size of the starting set S, and the second index is the size we're considering for S'.)

Then these are precisely the exponents which will appear in our expression. With these definitions, we can now write down the relevant inequalities in the Lovász Local Lemma — in order to be able to apply LLL, we need to find x and x' in $[0, 1)$ such that the following two conditions hold. For the blue-clique events, we need

$$q \leq x \cdot (1-x)^{d_{\ell\ell}} \cdot (1-x')^{d_{\ell k}}.$$

(Our definitions of $d_{\ell\ell}$ and $d_{\ell k}$ precisely track the number of events which interfere with our given ℓ-set, i.e., share an edge.) Similarly, for the red-clique events, we need

$$q' \leq x'(1-x)^{d_{k\ell}}(1-x')^{d_{kk}}.$$

Now that we've defined these numbers, all we need to do is find x and x' which satisfy these conditions; then we've satisfied the conditions of LLL and we are done.

To do this, let's first try to bound the numbers d. Writing out the exact values is pretty ugly, but we can give reasonable upper bounds — note that

$$0 \leq d_{\ell\ell} \leq \binom{\ell}{2} \cdot \binom{n-2}{\ell-2} \leq \frac{1}{2}\ell^2 n^{\ell-2}.$$

(The $\binom{\ell}{2}$ comes from choosing which two vertices S and S' share; then now that we've determined two things to place in S', we need to determine the remaining $\ell - 2$. Note that we need $n-2$ instead of $n-\ell$ because the intersection can be larger than 2; we are overcounting those cases, but this doesn't matter.)

This bound is quite lossy, since we have a factorial term in the denominator, which is usually much bigger than $\frac{1}{2}\ell^2$. But ℓ is fixed, so we don't really care about the factorial in the denominator.

Similarly, we have

$$0 \leq d_{k\ell} \leq \binom{\ell}{2}\binom{n-2}{k-2} \leq \frac{1}{2}n^{k-2}$$

(we have a $(k-2)!$ in the denominator, and we're assuming k is large with respect to ℓ, so the $(k-2)!$ certainly eats the ℓ^2).

Similarly, we have

$$0 \le d_{k\ell} \le \binom{k}{2}\binom{n-2}{\ell-2} \le \frac{1}{2}k^2 n^{\ell-2}$$

(this time we can't omit the k^2 since the factorial in the denominator is not big), and

$$0 \le d_{kk} \le \binom{k}{2}\binom{n}{k-2} \le \frac{1}{2}n^{k-2}$$

(because k is large, so $(k-2)!$ in the denominator eats the k^2.)

(The factors of $\frac{1}{2}$ are not important, but we'll keep them because we'll pick up a factor of 2 later, and keeping them now makes our numbers slightly nicer.)

It is a bit annoying to have $(1-x)$ to some power. The easiest way to deal with such terms is to take logs, but $\log(1-x)$ is also kind of ugly. So we'll use the following fact — for all $x \in \mathbb{R}$ we have $1-x \le e^{-x}$ (imagine a graph — the exponential function always lies above its tangent). We really want the inequality the other way around — if x is close to 0, then this inequality is actually pretty close. So for $0 \le x \le 0.1$ we have $1-x \ge e^{-2x}$. (The 0.1 has a lot of room.) Essentially, this gives us a reverse inequality if we lose a factor of 2.

Let the two conditions we need to satisfy be $(*)$ and $(*')$. First $(*)$ is equivalent to

$$\log \frac{q}{x} \le d_{\ell\ell} \ln(1-x) + d_{\ell k}(1-x').$$

The above inequality gives us the lower bound $\ln(1-x) \ge -2x$. So to prove $(*)$, it is sufficient to show that

$$\ln \frac{q}{x} \le -2d_{\ell\ell}x - 2d_{\ell k}x'.$$

Flipping the signs, this is equivalent to

$$\ln \frac{x}{q} \ge 2d_{\ell\ell}x + 2d_{\ell k}x'.$$

(This suffices as long as $0 \le x, x' \le 0.1$.)

Now we can plug in the bounds we have on $d_{\ell\ell}$ and $d_{k\ell}$ — it suffices to have

$$\ln \frac{x}{q} \ge \ell^2 n^{\ell-2}x + n^{k-2}x'$$

for $(*)$. With the same calculation, to prove $(*')$ it suffices to prove that

$$\ln \frac{x'}{q'} \ge k^2 n^{\ell-2}x + n^{k-2}x'.$$

So now all we need to do is specify values of x and x' which satisfy these two inequalities (and are between 0 and 0.1).

Before we name x and x', we'll calculate our actual expressions for q and q'. Observe that

$$q = p^{\binom{\ell}{2}} = (12\ell)^{\binom{\ell}{2}} \left(\frac{\ln k}{k}\right)^{\binom{\ell}{2}} < 0.05$$

if k is large. Meanwhile, we have

$$q' = (1-p)^{\binom{k}{2}} \le \exp\left(-p\binom{k}{2}\right) = \exp\left(-12\ell\frac{\ln k}{k}\cdot\frac{k(k-1)}{2}\right) \le \exp(-3\ell k\ln k) = k^{-3\ell k}.$$

(The $k-1$ is a bit annoying, so we bound $k-1 \ge \frac{k}{2}$.) (In general, this bound is good to use whenever p is small.)

Now it's time to define x and x'. As a bit of motivation, on the left we have $\log \frac{x}{q}$, so it is natural to define x as something times q and x' as something times q'. So in fact we will take $x = 2q$ (the 2 is not important, but the point is that we want x a bit larger than q so that the left-hand side is a constant). Note that $2q < 0.1$ because q is small enough. We will also take x' to be something times q'. But the right-hand side will be larger than a constant, so we will take $x' = k^{\ell k}q'$. Note that

$$x' \le k^{-2\ell k} < 0.1$$

for k large as well.

Now we need to check our two inequalities. We have

$$n^{\ell-2}x = n^{\ell-2}\cdot 2q \le 2\cdot(12\ell)^{\binom{\ell}{2}}\left(\frac{\ln k}{k}\right)^{\binom{\ell}{2}}\cdot c_\ell^{\ell-2}\cdot\left(\frac{k}{\ln k}\right)^{(\ell+1)(\ell-2)/2}.$$

Here we plugged in the actual definition of n, giving the last two factors; the first few factors come from our bound on q. The $\ln k/k$ terms are very similar — in both cases we have something roughly quadratic — and in fact these two exponents differ by exactly 1. This means the term simplifies to

$$c_\ell^{\ell-2}\cdot 2(12\ell)^{\binom{\ell}{2}}\cdot\frac{\ln k}{k}.$$

Using our assumption on ℓ, the first thing is less than $\frac{\ell}{2}$, which means this thing is

$$< \frac{1}{2}\ell\frac{\ln k}{k}.$$

Now we can bound

$$n^{k-2}x' < (k^\ell)^{k-2}k^{-2k\ell} < k^{-k\ell} = o(1)$$

(here we are using the stupid bound $n < k^\ell$). So we can see that the $n^{k-2}x'$ terms are both $o(1)$, while the first term is more significant.

Now we have

$$\ell^2 n^{\ell-2}x + n^{k-2}x' \le \ell^2\cdot\frac{\ell}{2}\frac{\ln k}{k} + o(1) = o(1),$$

because this term is also $o(1)$ (since ℓ is fixed and k is large). In particular, for large k this is less than $\ln 2 = \ln\frac{x}{q}$.

On the other side, we have

$$k^2 n^{\ell-2}x + n^{k-2}x' \le \frac{\ell}{2}k\ln k + o(1) < k\ell\ln k = \ln k^{\ell k} = \ln\frac{x'}{q'}.$$

That proves both the inequalities we needed, and we're done. □

§8 March 14, 2023 — Bounds for Hypergraph Ramsey Numbers

Today we'll see bounds for *hypergraph* Ramsey numbers.

Definition 8.1. For any integers $r \geq 2$ and $k_1, \dots, k_t \geq r$, the *hypergraph Ramsey number* $R_r(k_1, \dots, k_t)$ is the smallest number such that the following holds: In any coloring of the edges of a complete r-uniform hypergraph on $R_r(k_1, \dots, k_t)$ vertices with colors $1, \dots, t$, there exists a clique of size k_i in color i for some $i \in \{1, \dots, t\}$.

Here r is the uniformity of our hypergraph, the k_i are the target clique sizes (which need to be at least r, because otherwise it doesn't make sense to talk about a clique of that size), and t is the number of colors. (Ordinary Ramsey numbers correspond to the case $r = 2$.)

§8.1 Upper Bounds

We already proved that $R_r(k_1, \dots, k_t)$ indeed exists. In particular, our proof gave the recursive bound

$$R_r(k_1, \dots, k_t) \leq 1 + R_{r-1}(R_r(k_1 - 1, k_2, \dots, k_t), R_r(k_1, k_2 - 1, \dots, k_t), \dots, R_r(k_1, \dots, k_{t-1}, k_t - 1)).$$

As you might imagine, this leads to a *terrible* bound — the numbers $R_r(k_1 - 1, k_2, \dots, k_t)$ and so on are already gigantic, and then you apply R_{r-1} to them, which produces a *really* gigantic number. So this recursion gives an extremely weak upper bound for $R_r(k_1, \dots, k_t)$.

Example 8.2

Consider $R_3(k, k)$ (the 3-uniform case with two colors). For comparison, we know

$$2^{k/2} \leq R_2(k, k) \leq 4^k$$

(so $R_2(k, k)$ grows exponentially with k, although we don't know the base of the exponent — these are the best known bases, though it's possible to improve the lower order terms). Now plugging this into the above recursion, the bound we get for $R_3(k, k)$ is approximately of the form

$$R_3(k, k) \leq 2^{2^{2^{\cdots}}}$$

with a number of 2's linear in k (e.g., 2, $2^2 = 4$, $2^{2^2} = 2^4 = 16$, $2^{2^{2^2}} = 2^{16} = 65536$, $2^{2^{2^{2^2}}} = 2^{65536}$, and so on). This is an extremely fast-growing function, called a *tower-type function* (we'll define the tower function more formally on Thursday, and talk about it more).

(This is because every time we increase k by 1, we plug in the values to R_2, which is exponential; this gives us an iterated exponential.)

This is (unsurprisingly) not the truth; today we'll see a much better bound (which is also recursive, but a much better recursion).

Theorem 8.3 (Erdős–Rado 1952)

For any $k_1, \dots, k_t \geq r$ with $r \geq 3$ and $t \geq 2$, we have

$$R_r(k_1, \dots, k_t) \leq t^{\binom{R_{r-1}(k_1 - 1, \dots, k_t - 1)}{r-1}}.$$

This also looks pretty big — we have a gigantic exponent with a binomial coefficient in the top, which may look even more scary than the previous bound. But it's actually much better. Here we're decreasing the uniformity by 1 (and all the clique sizes); this is actually pretty mild (the binomial coefficient looks scary but is bounded by $R_{r-1}(k_1-1,\dots,k_t-1)^{r-1}$, and though exponentials are large, at least we only have one exponential).

Example 8.4

Applying this theorem to $R_3(k,k)$, we get

$$R_3(k,k) \le 2^{\binom{R_2(k-1,k-1)}{2}} \le 2^{R_2(k-1,k-1)^2} \le 2^{(4^k)^2} = 2^{2^{4k}}.$$

This is much better — rather than having a tower of height linear in k, we only have a double exponential (with the top linear in k). When we build these towers, what's at the top is not that important (whether it's k or k^2); what's most important is the number of times we exponentiate. Here we only exponentiate twice instead of k times, so this is much better (double exponential vs. tower-type).

We can also see what happens for off-diagonal 3-uniform Ramsey numbers.

Example 8.5

For fixed $\ell \ge 4$ (the case $\ell = 3$ is trivial), we get the bound

$$R_3(k,\ell) \le 2^{R(k-1,\ell-1)^2} \le 2^{\left(\frac{c_\ell k^{\ell-2}}{(\log k)^{\ell-3}}\right)^2} = 2^{c_\ell \frac{k^{2\ell-4}}{(\log k)^{2\ell-6}}}.$$

(Here we probably don't care about the log factors, so we could even use the Erdős–Szekeres theorem.) Here we get an expression which is *single*-exponential (with the exponent a polynomial function of k).

Remark 8.6. In the case for fixed ℓ, single-exponential is correct (though we don't know what the exponential should be). In the case $k = \ell$, the correct answer isn't known.

Remark 8.7. The tower height in our bounds is approximately equal to the uniformity. This is much better than a tower height approximately equal to k, since we generally think of the uniformity as fixed and k as large.

Now let's prove the theorem.

Proof. Let $R = R_{r-1}(k_1-1,\dots,k_t-1)$ (for notational convenience), and consider a complete r-uniform hypergraph on $t^{\binom{R}{r-1}}$ vertices where every edge is colored with one of the colors $1,\dots,t$. We need to show that for some $i \in \{1,\dots,t\}$, there exists a clique of size k_i and color i.

In the upper bound proofs we've done so far, our approach of finding a clique was to fix a vertex and look at its neighborhood. This standard approach gives the gigantic bound from earlier; here we'll take a different (though still similar) approach.

We construct a sequence of distinct vertices $v_1,\dots,v_r$ and a nested sequence of vertex sets $S_1 \supseteq S_2 \supseteq \cdots \supseteq S_R$, such that the following three conditions hold:

(1) For any $1 \le j_1 < \cdots < j_{r-1} \le R$, the edges $\{v_{j_1},\dots,v_{j_{r-1}},v\}$ (over v) all have the same color as long as $v \in S_{j_{r-1}}$. In other words, if we choose $r-1$ distinct vertices, then as long as we choose v from a certain candidate set, the color of this edge doesn't depend on v. This essentially associates a color with the $r-1$ vertices $\{v_{j_1},\dots,v_{j_{r-1}}\}$, which is how we'll get a $(r-1)$-uniform hypergraph.

(2) For every $j = 1, \dots, R$, we have $v_{j+1}, \dots, v_R \in S_j$ but $v_1, \dots, v_j \notin S_j$ (i.e., v cannot coincide with any of the previous vertices v_{j_i} in the above point, because that would make the previous point not make sense).

(3) $|S_j| \geq t^{\binom{R}{r-1}-\binom{j}{r-1}} - j$ for all $1 \leq j \leq R-1$. (We need a lower bound on the sizes of the sets, because otherwise the statement would be vacuously true.) Additionally, $|S_R| \geq 1$. (If $j < r-1$ then we nearly have the entire graph, but this makes sense since then S_j can never appear in (1).)

The main condition is the first condition — we want nice sets where the color of the edge $\{v_{j_1}, \dots, v_{j_{r-1}}, v\}$ only depends on the first $r-1$ vertices. Then this gives a coloring of that set of $r-1$ vertices, which is nice because we want to construct a coloring of a $(r-1)$-uniform hypergraph.

Let S_0 be the entire vertex set (so that we can define the sets one by one). This satisfies the necessary condition, as

$$|S_0| = t^{\binom{R}{r-1}} = t^{\binom{R}{r-1}-\binom{0}{r-1}} - 0.$$

For $1 \leq j \leq R$, we construct v_j and S_j as follows (assuming we've already constructed $v_1, \dots, v_{j-1}$ and $S_1 \supseteq \dots \supseteq S_{j-1}$).

The first thing we need to do is construct v_j; we do this simply by picking an arbitrary vertex $v_j \in S_{j-1}$. (To do this, we need S_{j-1} to be nonempty; but the lower bound on $|S_j|$ given by (3) is positive, so this is true.)

Now we need to construct $S_j \subseteq S_{j-1} \setminus \{v_j\}$, satisfying (1).

For every vertex $v \in S_{j-1} \setminus \{v_j\}$, we consider the colors of the edges $\{v_{j_1}, \dots, v_{j_{r-2}}, v_j, v\}$ for each $(r-2)$-tuple of earlier vertices $1 \leq j_1 < \dots < j_{r-2} \leq j-1$ (if we don't take v_j, then there's no new condition — all v are in S_{j-1} so already satisfy the condition). These are the colors we need to analyze for the condition (1).

The number of such $(r-2)$-tuples is $\binom{j-1}{r-2}$, since we need to choose $r-2$ of the first $j-1$ vertices. This means there are $t^{\binom{j-1}{r-2}}$ possibilities for these colors.

In other words, for every v, imagine we record the color associated with each $(r-2)$-tuple. There are $t^{\binom{j-1}{r-2}}$ possibilities of this data (for any given v).

Looking at these different possibilities, we get a partition of $S_{j-1} \setminus \{v_j\}$ based on the different assignments (i.e., we split up our vertices v based on their data, so that a group of the partition consists of vertices with the same data). This partition has $t^{\binom{j-1}{r-2}}$ subsets (one for each set of data). Let S_j be one of these subsets with maximal size.

Now by definition, S_j is a subset of $S_{j-1} \setminus \{v_j\}$ such that this data is the same for all vertices in the subset.

First, (1) holds by construction — for every v in the subset, by the definition of the subset we have the same assignment from $(r-2)$-tuples to colors, i.e., for every $(r-2)$-tuple $(v_{j_1}, \dots, v_{j_{r-2}})$, we get the same color of $\{v_{j_1}, \dots, v_{j_{r-2}}, v_j, v\}$ over all v. So this color doesn't depend on v, as desired. This satisfies (1) if $j_{r-1} = j$ (and the cases where $j_{r-1} < j$ are already satisfied, as they don't involve v_j or S_j).

Now we need to check that (2) and (3) are satisfied. For (2), note that $v_j \notin S_j$ by construction (we took S_j as a subset of $S_{j-1} \setminus \{v_j\}$). But we do have $v_j \in S_{j-1}$ (as we chose v_j from S_{j-1}), and therefore $v_j \in S_i$ for all $i < j$ (as $S_0 \supseteq S_1 \supseteq \dots \supseteq S_R$). So then we have $v_i \in S_j$ if and only if $i > j$, which gives exactly the condition in (2).

Finally, for (3), we need to check that $|S_j|$ is still large enough. The important fact here is that we chose

$|S_j|$ to be the biggest of our subsets, so it's at least their average — i.e., we have

$$\begin{aligned}|S_j| &\geq \frac{|S_{j-1} \setminus \{v_j\}|}{t^{\binom{j-1}{r-2}}} \\ &\geq \frac{|S_{j-1}|}{t^{\binom{j-1}{r-2}}} - 1 \\ &\geq \frac{t^{\binom{R}{r-1}-\binom{j-1}{r-1}} - (j-1)}{t^{\binom{j-1}{r-2}}} - 1 \\ &\geq t^{\binom{R}{r-1}-\binom{j-1}{r-1}-\binom{j-1}{r-2}} - j \\ &= t^{\binom{R}{r-1}-\binom{j}{r-1}} - j,\end{aligned}$$

where the last step is by Pascal's identity (and the intermediate steps are because moving the subtractions out of the fraction can only make the term smaller).

This checks the lower bound for $j \leq R-1$. For the special case $j = R$, we need to check $|S_R| \geq 1$. (Then because of the nested inequalities, we automatically get that all the other sets have size at least 1 as well.) We have

$$\begin{aligned}|S_R| &\geq \frac{|S_{R-1} \setminus \{v_R\}|}{t^{\binom{R-1}{r-2}}} \\ &\geq \frac{t^{\binom{R}{r-1}-\binom{R-1}{r-1}} - (R-1) - 1}{t^{\binom{R-1}{r-2}}} \\ &= \frac{t^{\binom{R-1}{r-2}} - R}{t^{\binom{R-1}{r-2}}}.\end{aligned}$$

We can check that this is positive (we have $t \geq 2$ and $r \geq 3$, so $t^{\binom{R-1}{r-2}} \geq 2^{R-1} > R$); this means $|S_R| > 0$, so we must have $|S_R| \geq 1$ (as it is a set, so its size is an integer). So (3) is satisfied.

Now we've constructed our sequence of vertices and sets, and we now want to use this to construct our cliques.

Consider the complete $(r-1)$-uniform hypergraph with vertices $v_1, \dots, v_R$. Color every edge $\{v_{j_1}, \dots, v_{j_{r-1}}\}$ (ordered with $j_1 < \cdots < j_{r-1}$) with the color appearing in condition (1) — i.e., with the common color of the edges $\{v_{j_1}, \dots, v_{j_{r-1}}, v\}$ in the original coloring for all $v \in S_{j_{r-1}}$. (Any $(r-1)$-edge in this $(r-1)$-uniform hypergraph is of this form for some indices; then we can look at (1) for these indices, which tells us that there's a common color of all edges of this type; we then assign that common color to this $(r-1)$-edge.)

Since $R = R_{r-1}(k_1 - 1, \dots, k_t - 1)$, in this coloring of this $(r-1)$-uniform hypergraph, for some $i \in \{1, \dots, t\}$ we can find a clique of size $k_i - 1$ in color i (applying the definition of the $(r-1)$-uniform hypergraph Ramsey number — this holds for every coloring, so it definitely holds for the coloring we defined here). Let this clique be T.

Now consider $T \cup \{v\}$ for some vertex $v \in S_R$. Then $|T \cup \{v\}| = k_i - 1 + 1 = k_i$. We will show that $T \cup \{v\}$ is a clique of color i in the original coloring of the r-uniform hypergraph. This is good enough, as then we've found a clique of color i and size k_i.

To check this claim, we need to check that for any choice of r vertices from $T \cup \{v\}$, the r-set has color i. There's two possibilities — if our r-set involves v, or if it doesn't.

In the first case, for any vertices $v_{j_1}, \dots, v_{j_{r-1}}$ in T (ordered with $j_1 \leq \cdots \leq j_{r-1}$), the edge $\{v_{j_1}, \dots, v_{j_{r-1}}, v\}$ has color i because of condition (1) and the definition of the hypergraph coloring — we have $v \in S_R \subseteq S_{j_{r-1}}$, so condition (1) requires that our edge must have the color associated to $\{j_1, \dots, j_{r-1}\}$, which must be color i because T was a clique of color i in the $(r-1)$-uniform coloring.

In the second case, suppose we have r vertices $v_{j_1}, \ldots, v_{j_r}$ in T, with $j_1 < \cdots < j_r$. Then the edge $\{v_{j_1}, \ldots, v_{j_r}\}$ has color i — we must have $v_{j_r} \in S_{j_{r-1}}$ (by (2), since $j_r > j_{r-1}$), so again we can apply condition (1) in the same way (taking v_{j_r} in place of v).

So we've checked that every size-r subset of $T \cup \{v\}$ has color i in the original coloring. So it forms a clique of size k_i in color i, and we're done. □

The strategy we saw here is more generally useful. What's going on is that we progressively add vertices and keep a candidate set of future vertices which are well-behaved with respect to the previous vertices (in the sense that they behave uniformly). Then we can make our sequence long enough so that at some point we can look at this set of vertices we've chosen and choose a good set. The idea of having a sequence of vertices and candidate set applies reasonably commonly, and we will see it later in the course.

§8.2 Lower Bounds

Now we've proven the theorem. We already drew some conclusions of the theorem which are much better than the giant bounds obtained from the original recursive proof; now we'll talk about the best-known *lower* bounds.

Theorem 8.8 (Erdős–Hajmal 1972)

There is an absolute constant $c > 0$ such that $R_3(k, 4) \geq 2^{ck}$ for all $k \geq 3$.

In other words, the single-exponential behavior for off-diagonal numbers $R_3(k, \ell)$ (i.e., for fixed ℓ) is correct. (For graph Ramsey numbers, the bound is polynomial, so this is not obvious.) Of course $R_3(k, \ell) \geq R_3(k, 4)$ for any fixed $\ell \geq 4$, so this gives a bound that is exponential in k (the exponent is linear in k rather than polynomial).

(We'll prove this next class.)

The best-known bounds are a bit better, but of a similar behavior — the best bounds for $R_3(k, 4)$ are

$$k^{ck} \leq R_3(k, 4) \leq k^{c'k^2}$$

(this improves the lower bound by a factor of $\log k$ in the exponent, and the upper bound from k^4 to $k^2 \log k$ in the exponent; still, the fundamental behavior is still the same, since we still have a single-exponential with a polynomial gap in the exponent). These bounds are due to Conlon–Fox–Sudakov 2010. More generally, they improved the upper bound to

$$R_3(k, \ell) \leq 2^{c_\ell k^{\ell-2} \ln k},$$

so we've essentially square-rooted the polynomial in the exponent (as we have $k^{\ell-2}$ instead of $k^{2\ell-4}$).

In particular, the conclusion is that $R_3(k, 4)$ is single-exponential in k, but we don't know which polynomial should be in the exponent. For any fixed ℓ the same is true — the answer is single-exponential with some polynomial of k, but we don't know which one.

Meanwhile, for the *diagonal* Ramsey numbers, we have the following theorem (which we will prove on the homework).

Theorem 8.9

There is an absolute constant $c > 0$ such that

$$R_3(k, k) \geq 2^{ck^2}.$$

This lower bound is better than the fixed-ℓ case (where we have a quadratic rather than linear polynomial in the exponent), but it's still far from the double-exponential upper bound — we get

$$2^{ck^2} \leq R_3(k,k) \leq 2^{2^{c'k}}.$$

These are essentially the best-known upper and lower bounds; so it's still unclear whether $R_3(k,k)$ is single-exponential or double-exponential. Erdős conjectured that the true behavior should be double-exponential — i.e., there should be a lower bound of the form $2^{2^{c'k}}$ as well (potentially for a different c'). (He offered \$500 dollars for this conjecture; it is still open.)

All the bounds we've thought about so far are for $r = 3$. Next class we'll see that in some sense, answering these questions for the 3-uniform case is enough — what we've proved here gives us a way to get from an upper bound for $r-1$ to an upper bound for r. Next class we'll see a way to do the same with *lower* bounds (both increase by one exponential). But we can only do this starting with 3-uniform. So 3-uniform is the most interesting case, since it gives all the others automatically.

§9 March 16, 2023 — Bounds for Hypergraph Ramsey Numbers

Definition 9.1. For any integers $r \geq 2$ and $k_1, \ldots, k_t \geq r$, the hypergraph Ramsey number, denoted by $R_r(k_1, \ldots, k_t)$, is the smallest number such that the following holds: in any coloring of the edges of a complete r-uniform hypergraph on $R_r(k_1, \ldots, k_t)$ vertices with colors $1, \ldots, t$, there exists a clique of size k_i in color i for some i.

§9.1 Lower Bounds for the Off-Diagonal Case

Last class, we talked about upper bounds for these hypergraph Ramsey numbers; in particular, we proved the theorem of Erdős and Rado that gives an upper bound for $R_r(k_1, \ldots, k_t)$ in terms of the Ramsey number for one uniformity lower. (Each time we decrease the uniformity by 1, we iterate an exponential.)

Then we started talking about lower bounds.

Theorem 9.2 (Erdős–Hajmanl 1972)

There is an absolute constant $c > 0$ such that $R_3(k, 4) \geq 2^{ck}$ for all $k \geq 3$.

Here we're looking for the off-diagonal hypergraph Ramsey number, where we want a red k-clique or blue 4-clique.

Our upper bound from last class *also* gives us a single-exponential bound for the case $R_3(k, 4)$, though the exponent is a polynomial in k rather than linear — this is interesting because we have an upper and lower bound with the same first-order behavior (single-exponential).

(Note that this also gives a lower bound if we take any value larger than 4 — so this bounds all 3-uniform hypergraph Ramsey numbers that we care about.)

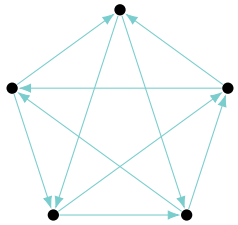
Proof. As usual, we can assume that k is large — if we find some $c > 0$ such that this statement holds for all sufficiently large k (e.g., $k \geq 100$), then we can deal with the finitely many remaining values of k (e.g., $k < 100$) by adapting the constant c. (For every particular k there certainly exists a constant for which this is true, so we can shrink c enough to take care of these finitely many k.)

We will show that the statement is true for $c = \frac{1}{5}$ if k is sufficiently large — consider a complete 3-uniform hypergraph on $n = \lfloor 2^{k/5} \rfloor$ vertices. We need to show that there is a coloring of the edges of this complete 3-uniform hypergraph without a red clique of size k or a blue clique of size 4.

As usual, it is difficult to get good explicit Ramsey colorings; we're going to use a probabilistic approach, where we use randomness to get our coloring. So far, we've looked at every edge and colored it red or blue with a certain probability, and then analyzed this. Here instead of directly coloring the edges randomly, we'll use an elegant trick — we'll use other underlying random objects which we'll then use to define the coloring.

Our random object will be a random *tournament* on the same vertex set.

Definition 9.3. A *tournament* is a directed graph which has exactly one directed edge between any two vertices.



(For example, imagine we have 5 soccer teams; and suppose we organize a tournament where each team plays each other team exactly once. Then our graph records who won each game — $a \to b$ means that team a beat team b. Note that having the edges $a \to b$ and $b \to c$ doesn't tell you anything about the edge $a \to c$.)

First consider a random tournament on the same vertex set — this means between any two vertices, we take a directed edge whose orientation is chosen randomly (and the edges are all independent). (In other words, for every two vertices we have two options for how to draw their edge — $u \to v$ or $v \to u$ — and we flip an independent coin for each edge.)

Now using this tournament, we want to define our coloring — we need to color every 3-tuple of vertices either red or blue.

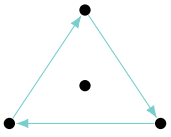
For any three vertices, there are two cases — either we have a cycle, or an arrangement with a superior team, a middle team, and a weak team. We'll color a 3-tuple of vertices red if they form a *transitive tournament* (a configuration with no cycle; such a tournament is called transitive because it has the property that if a beats b and b beats c then a beats c), and blue if they form a cycle.

We now want to check that with positive probability, this has the properties we want. We want to avoid a red clique of size k, and a blue clique of size 4.

Claim — There can *never* be a blue clique of size 4.

In other words, there can never be a tournament on 4 vertices where all triples form cycles (which is what a blue clique of size 4 would be).

Proof. Assume that there is such a tournament, and consider three of the vertices, which must form a cycle.

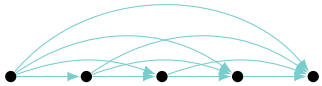


Now consider the fourth vertex. It must form a cycle with each of the other edges; this fixes the orientation of its two edges to every pair of vertices, and these orientations contradict each other. □

(This is somewhat similar to our upper bounds for the Erdős–Szekeres problem, where we used a hypergraph Ramsey argument and took a coloring of 4-tuples, using a lemma that among 5 points we can't have all 4-tuples non-convex. Similarly, here the construction is based on a small configuration we surely cannot have.)

Now we need to find the probability of having a red clique of size k. This is certainly possible — the *whole* tournament could be transitive.

Definition 9.4. A tournament is *transitive* if the vertices can be ordered in a line such that all edges go forwards (i.e., such that the stronger team always beats the weaker team).



(This is actually equivalent to having no cycles of length 3.)

If the entire random tournament is transitive, then all edges would be red, so the entire hypergraph would be a red clique. So this can definitely happen, but it's pretty *unlikely*; we want to show that with high probability there is no red clique of size k.

The probability of having a red clique of size k can be bounded as follows: for any given k vertices, we have a red clique on these k vertices if and only if these k vertices form a transitive tournament (since being a red clique is equivalent to having no blue triples, i.e., no cycles of length 3, and this is equivalent to the transitive property). The probability that k vertices form a transitive tournament is $k!\,2^{-\binom{k}{2}}$ (there are $k!$ possible rankings, and in each there are $\binom{k}{2}$ edges that need to be oriented according to that ranking). We can bound this as

$$k!\,2^{-\binom{k}{2}} \le k^k \cdot 2^{-k^2/2+k/2} \le 2^{-k^2/4}$$

for large k (replacing the denominator of $-\frac{k^2}{2}$ with something a bit larger is enough to eat all the lower-order terms).

This bounds the probability that any *given* k vertices form a red clique. Now we need to consider all the ways of choosing k vertices; so union bounding over all possible choices of the k vertices, the total probability of having a red clique of size k is at most

$$\binom{\lfloor 2^{k/5} \rfloor}{k} \cdot 2^{-k^2/4} \le 2^{k^2/5} \cdot 2^{-k^2/4} < 1.$$

(Note that depending on what we want to bound, sometimes we can just $\binom{n}{k} \le n^k$ and ignore the $k!$, and sometimes we don't want to ignore it. Here we can ignore it.)

So the probability of having a red clique of size k is strictly less than 1, and with positive probability there is no red clique of size k. Meanwhile, there is also necessarily no blue clique of size 4. So there exists a coloring without a red clique of size k and blue clique of size 4. □

This tells us $R_3(k,4) \ge 2^{ck}$, so the off-diagonal Ramsey numbers are at least exponential in k. We also have an upper bound which is exponential in some polynomial of k.

This also gives a lower bound for $R_3(k,k) \ge R_3(k,4)$. But in fact, we can do better — there is an absolute constant $c > 0$ such that $R_3(k,k) \ge 2^{ck^2}$. (This can be done by a relatively straightforward probabilistic argument, and we'll see it in the homework.) This is still single-exponential in k, but the best-known upper

bound for the diagonal case is *double*-exponential in k. So the 3-uniform diagonal Ramsey number is at least single-exponential and at most double-exponential, and it's still open to figure out which is right (it's conjectured that double-exponential is right).

§9.2 Lower Bounds with Multiple Colors

We can also ask the same question if we take more than 2 colors. For any fixed t, the theorem from last time (which works for any number of colors) gives that

$$R_3(k,k,\dots,k) \le 2^{2^{c_t k}}.$$

(E.g. Prof. Sauermann's daughter has 5 favorite colors, so she might be interested in $R_3(k,k,k,k,k)$.) So the 3-uniform Ramsey number is still double-exponential in k, where the constant depends on the number of colors.

Remark 9.5. We generally think of t as fixed. It's also reasonable to fix k and let $t \to \infty$, but this is a different flavor of problem; we'll mention what's known and conjectured in that case later (for graphs). Straightforwardly plugging in the proof we have should give c_t around t^t — we should get

$$t^{\binom{R(k-1,\dots,k-1)}{2}} \le t^{R(k-1,\dots,k-1)^2} \le t^{\binom{t(k-1)}{k-1,\dots,k-1}^2},$$

which should give something on the order of t^t. But this probably isn't the right answer.

The single-exponential lower bound for 2 colors still applies to the case with more than two colors, so it looks like we still have this gap. But interestingly, if the number of colors is at least 4, then double-exponential is correct; this is what we'll see next.

Theorem 9.6 (Erdős–Hajmal)
We have $R_3(k,k,k,k) \ge 2^{R(k-1,k-1)-1}$ for all $k \ge 3$.

Here $R(k-1,k-1)$ is the 2-color Ramsey number for graphs. In particular, by the lower bound for diagonal graph Ramsey numbers, we get

$$R_3(k,k,k,k) \ge 2^{2^{k/2-2}}$$

for all $k \ge 3$. This is double-exponential in k, so our upper bound is correct up to the constant c_t. (In particular, if this bound holds for 4 colors, it also holds for any number of colors at least 4.)

Note that this is really cool — for 2 colors we have a gap, but for 4 or more colors, we know the answer is double-exponential. (This is part of why people believe the answer should be double-exponential even for 2 colors.)

Proof. For notational convenience, define $R = R(k-1,k-1)-1$ — i.e., R is not good enough for the graph Ramsey problem with 2 colors — and consider a coloring of the edges of the complete graph G with vertex set $\{0,\dots,R-1\}$ such that there is no monochromatic red clique or blue clique of size $k-1$ (which exists because R is smaller than the corresponding Ramsey number).

We want to find a coloring of the 3-uniform hypergraph on 2^R vertices with 4 colors, and with no monochromatic k-clique (to show that 2^R vertices is not good enough). So consider the complete 3-uniform hypergraph on the vertex set $\{0,1\}^R$ (which has 2^R vertices, labelled by sequences of 0s and 1s). We need to find a coloring with 4 colors of this 3-uniform hypergraph, such that there is no monochromatic clique of size k.

To describe the coloring, we first want to think of our labels $\{0,1\}^R$ as binary numbers — every vertex $v \in \{0,1\}^R$ can be interpreted as a binary number $v = v_{R-1}v_{R-2}\cdots v_0$ (where v_0 is the coefficient of 2^0, v_1 is the coefficient of 2^1, and so on).

This gives us a natural ordering of these vertices, where for $v, w \in \{0,1\}^R$ we say $v < w$ if $v = v_{R-1}v_{R-2}\cdots v_0$ is a smaller binary number than $w = w_{R-1}w_{R-2}\cdots w_0$ (i.e., if $\sum_{j=0}^{R-1} v_j 2^j < \sum_{j=0}^{R-1} w_j 2^j$).

If we want to compare the size of two binary numbers, we can simply start from the beginning (the left) and check if the two digits are equal; if they're equal, then we move one digit right. The first time we hit an entry where the two digits differ, if v has a 0 and w has a 1, then $v < w$. So we name this index — for $v, w \in \{0,1\}^R$ with $v \neq w$, we define $\delta(v,w) = \max\{j \mid v_j \neq w_j\}$, i.e., $\delta(v,w)$ is the index of the first coordinate where v and w differ. Note that if $v < w$, then for $j = \delta(v,w)$ we have $v_j = 0$ and $w_j = 1$.

Claim — The function δ has the following properties:

- For every $v, w, z \in \{0,1\}^R$ with $v < w < z$, we have $\delta(v,w) \neq \delta(w,z)$.
- For $v^{(1)}, \ldots, v^{(\ell)}$ in $\{0,1\}^R$ with $v^{(1)} < \cdots < v^{(\ell)}$, we have

$$\delta(v^{(1)}, v^{(\ell)}) = \max(\delta(v^{(1)}, v^{(2)}), \delta(v^{(2)}, v^{(3)}), \ldots, \delta(v^{(\ell-1)}, v^{(\ell)})).$$

(We use upper indices to refer to ℓ distinct vectors, and lower indices to refer to coordinates.)

Proof. For the first statement, if both values were equal to j, then we'd need $v_j = 0$ and $w_j = 1$ from $v < w$, but $w_j = 0$ and $z_j = 1$ from $w < z$; this is impossible.

For the second statement, consider the first index j where the $v^{(i)}$ are not all equal. Then at that digit j, some initial segment of the $v^{(i)}$ are 0, and then the rest are 1. In particular, we will have $\delta(v^{(1)}, v^{(\ell)}) = j$. Meanwhile, the terms $\delta(v^{(i)}, v^{(i+1)})$ will all be j or later; and there's exactly one place where we flip from 0 to 1, so there's exactly one term of j. □

Now we can use this coloring and our function δ to define a coloring of our 3-uniform hypergraph $\mathcal{H}$ with vertex set $\{0,1\}^R$ with 4 colors. We need to color every triple of vertices; for any triple $v, w, z \in \{0,1\}^R$ ordered such that $v < w < z$, we define the color of the edge $\{v, w, z\}$ in the following way:

- We color $\{v, w, z\}$ *dark red* if $(\delta(v,w), \delta(w,z))$ is a red edge in G and $\delta(v,w) < \delta(w,z)$. (Note that by the first property these two numbers $\delta(v,w)$ and $\delta(w,z)$ are two different numbers; so they define an edge in the original graph on R vertices, which must be either red or blue. If the edge is red and this relationship holds, then we color the 3-uniform edge dark red.)
- We color the triple *light red* if $(\delta(v,w), \delta(w,z))$ is a red edge and $\delta(v,w) > \delta(w,z)$.
- We color $\{v, w, z\}$ *dark blue* or *light blue* analogously.

In other words, the color we choose depends on the inequality between the two values $\delta(v,w)$ and $\delta(w,z)$ (which are not equal), and the edge that they define in the graph G.

We need to check that this coloring does not have a monochromatic clique of size k. The cases for the four colors are all analogous, so we'll check this for dark red.

Suppose that there is a dark red clique of size k in $\mathcal{H}$. Name the vertices in this clique $v^{(1)} < \cdots < v^{(k)}$.

Since our clique is dark red, every triple of three consecutive elements is dark red, so in particular we have

$$\delta(v^{(1)}, v^{(2)}) < \delta(v^{(2)}, v^{(3)}) < \cdots < \delta(v^{(k-1)}, v^{(k)}).$$

(Each individual inequality is generated by looking at three consecutive vertices; we then string these inequalities together.) This means in particular that for any $1 \leq i < j \leq k-1$ we have

$$\delta(v^{(i+1)}, v^{(j+1)}) = \max\{\delta(v^{(i+1)}, v^{(i+2)}), \delta(v^{(i+2)}, v^{(i+3)}), \ldots, \delta(v^{(j)}, v^{(j+1)})\} = \delta(v^{(j)}, v^{(j+1)})$$

(where the first equality follows from the earlier claim).

Now, we know that the edge $\{v^{(i)}, v^{(i+1)}, v^{(j+1)}\}$ is dark red; this means $(\delta(v^{(i)}, v^{(i+1)}), \delta(v^{(i+1)}, v^{(j+1)}))$ forms a red edge in G. But as we just calculated, the second entry is just $\delta(v^{(j)}, v^{(j+1)})$, so this tells us $(\delta(v^{(i)}, v^{(i+1)}), \delta(v^{(j)}, v^{(j+1)}))$ is a red edge in G.

But this means the vertices $\delta(v^{(1)}, v^{(2)})$, $\dots$, $\delta(v^{(k-1)}, v^{(k)})$ form a red clique of size $k-1$ in G (since the above argument shows that every pair forms a red edge in G; and they are all distinct by the inequality that $\delta(v^{(1)}, v^{(2)}) < \dots < \delta(v^{(k-1)}, v^{(k)})$). This is a contradiction, since we chose our coloring of G to have no red clique of size $k-1$.

We can do the same for all the other colors, so we're done. □

§10 March 21, 2023 — Bounds for Hypergraph Ramsey Numbers

As usual, we'll use r to denote the uniformity of our hypergraph (right now we generally care about $r \geq 3$; $r = 2$ corresponds to a graph). We'll use t to denote our number of colors, and k_1, $\dots$, k_t to denote the target clique sizes.

Last class, we saw the following theorem:

Theorem 10.1 (Erdős–Rado 1952)

For any $k_1, \dots, k_t, r \geq 3$ with $t \geq 2$, we have

$$R_r(k_1, \dots, k_t) \leq t^{\binom{R_{r-1}(k_1-1, \dots, k_t-1)}{r-1}}.$$

The binomial coefficient can be bounded by $R_{r-1}(k_1 - 1, \dots, k_t - 1)^{r-1}$. You should think of t as 2 (since it's fixed).

So this gives a double-exponential bound for the 3-uniform hypergraph Ramsey numbers, and by iterating this we get an upper bound for the r-uniform ones for any fixed r.

Last class, we also started talking about lower bounds; but we only saw various lower bounds for the 3-uniform case. We haven't talked yet about higher uniformity because from 3-uniform lower bounds we can get lower bounds for any higher uniformity, which also grow exponentially (mentioned earlier). The previous theorem lets us transfer an upper bound on Ramsey numbers to one uniformity higher; now we'll see a corresponding way to transfer *lower* bounds one uniformity higher.

Theorem 10.2 (Erdős–Hajnal, 'Stepping Up Lemma')

For any $k_1, \dots, k_t \geq r \geq 4$ with $t \geq 2$, we have

$$R_r(k_1, \dots, k_t) \geq 2^{R_{r-1}\left(\left\lceil \frac{k_1-r+4}{2} \right\rceil, \left\lceil \frac{k_2-r+4}{2} \right\rceil, \dots, \left\lfloor \frac{k_t-r+4}{2} \right\rfloor\right)-1}.$$

The target clique sizes are still roughly k_1, $\dots$, k_t, but we lose a little bit (we subtract a bit and divide by 2). The most natural case is $R_r(k, k)$ for k large with respect to r; then the subtraction doesn't matter. We also lose a factor of 2 for a constant number of steps (r steps), so we're essentially losing a constant, which doesn't really matter.

(You can probably even omit the factors of 2 from the third one on, but it doesn't really matter — the 2 doesn't play a major role.)

Remark 10.3. There's a major caveat. The restriction $t \geq 2$ is trivial — you wouldn't do Ramsey theory with one color — but the requirement $r \geq 4$ is not trivial. It means we can go from 3-uniform to 4-uniform and 4-uniform to 5-uniform and so on, but we can't go from 2-uniform to 3-uniform (unlike with the upper bound, where we *can* go from ordinary to 3-uniform Ramsey numbers). This really means that if we want to understand the hypergraph Ramsey numbers (at least, in the regime $R_r(k,k)$ where r is fixed and k is large), the big question is to understand it for $r = 3$ — because for higher uniformity, we can transfer both bounds. We understand fairly well how $R_2(k,k)$ works, so if we understand it for 3, then we could understand it for higher r. But $R_3(k,k)$ is not well-understood (we just know it's between single and double exponential).

Now we'll prove this.

Proof. The proof strategy is in some sense similar to what we did last class when we proved a lower bound for uniformity 3 with at least 4 colors. (It'll be more complicated, though. In the proof on Thursday, we got a lower bound on uniformity 3 with at least 4 colors in terms of the lower bound for graph Ramsey numbers with 2 colors — so we allowed ourselves to use more colors when going up. Here we can't do that — we don't change the number of colors in the bound, so we can't split red into dark red and light red.)

Let $R = R_{r-1}(\left\lceil \frac{k_1-r+4}{2} \right\rceil, \dots, \left\lceil \frac{k_t-r+4}{2} \right\rceil) - 1$, and consider a coloring of the edges of the complete $(r-1)$-uniform hypergraph $\mathcal{H}$ with vertex set $\{1, \dots, R\}$ with the colors 1, …, t such that there is no clique of size $\left\lfloor \frac{k_i-r+4}{2} \right\rfloor$ in color i for any $i \in \{1, \dots, t\}$ — such a coloring exists, because R is one lower than the corresponding Ramsey number.

We now want to show that $R_r(k_1, \dots, k_t) \geq 2^R$ (our proof will in fact give $2^R + 1$, but the $+1$ is so irrelevant that we didn't write it down). Consider the complete r-uniform hypergraph on the vertex set $\{0,1\}^R$. We are going to find a coloring of the edges of this hypergraph with colors 1, …, t without a clique of size k_i in color i for any $i \in \{1, \dots, t\}$.

As on Thursday, we interpret every vertex $v \in \{0,1\}^R$ as a binary number $v = v_{R-1}v_{R-2}\cdots v_0$. For v and $w \in \{0,1\}^R$ with $v \neq w$, we again define

$$\delta(v,w) = \max\{j \mid v_j \neq w_j\},$$

i.e., the most significant digit where v and w differ (so to tell which one is bigger, we could just look at this digit).

Recall that last class, we saw the following properties of δ:

- If $v, w \in \{0,1\}^R$ with $v < w$, we have $v_j = 0$ and $w_j = 1$.
- For $v, w, z \in \{0,1\}^R$ with $v < w < z$, we must have $\delta(v,w) \neq \delta(w,z)$. (This is because if they were the same index j, then the above point would tell us $v_j = 0$ and $w_j = 1$, while the second would tell us $w_j = 0$ and $z_j = 1$; and w_j cannot be both 0 and 1.)
- If we have ℓ vectors $v^{(1)}, \dots, v^{(\ell)} \in \{0,1\}^R$ with $v^{(1)} < \cdots < v^{(\ell)}$, then

  $$\delta(v^{(1)}, v^{(\ell)}) = \max(\delta(v^{(1)}, v^{(2)}), \delta(v^{(2)}, v^{(3)}), \dots, \delta(v^{(\ell-1)}, v^{(\ell)})).$$

  This is because if we look at the maximal index j where the vectors aren't all equal, then because of the ordering, some initial part of the vectors will have jth coordinate 0, and the later ones will have jth coordinate 1. This tells us that this index j is the first where $v^{(1)}$ and $v^{(\ell)}$ differ — so $\delta(v^{(1)}, v^{(\ell)}) = j$. Meanwhile, there must be some point where we switch from 0 to 1, so one of the δ-values on the right is equal to j; and all are at most j because all the vectors are equal at all indices left of j are all equal.

Now we're ready to define our coloring. We want to define a coloring on the edges of our hypergraph, where every edge consists of r vertices; this means for every r-tuple of vertices we want to define a color. For $v^{(1)}$, $\dots$, $v^{(r)}$ in $\{0,1\}^R$, ordered such that $v^{(1)} < \dots < v^{(r)}$, we define the color of the edge $\{v^{(1)}, \dots, v^{(r)}\}$ as follows: Let $\delta_i = \delta(v^{(i)}, v^{(i+1)})$ for each $i = 1, \dots, r-1$ — i.e., the δ-values between consecutive pairs. Note that $\delta_i \neq \delta_{i+1}$ for all $i = 1, \dots, r-2$ (by the second property stated above — δ_i and δ_{i+1} are defined as $\delta(v^{(i)}, v^{(i+1)})$ and $\delta(v^{(i+1)}, v^{(i+2)})$). Now we have four cases:

- If $\delta_1 < \delta_2 < \dots < \delta_{r-1}$ or if $\delta_1 > \delta_2 > \dots > \delta_{r-2}$, then we color $\{v^{(1)}, \dots, v^{(r)}\}$ in the color of the edge $\{\delta_{(1)}, \dots, \delta_{(r-1)}\}$ in $\mathcal{H}$ (the complete $(r-1)$-uniform hypergraph on the vertex set $\{1, \dots, R\}$ — our δ's are all elements of $\{1, \dots, R\}$, and because of the monotonicity they're all distinct, so this set is a set of $r-1$ distinct elements of $\{1, \dots, R\}$ and so has a well-defined color in $\mathcal{H}$).
- If $\delta_1 < \delta_2$ and $\delta_2 > \delta_3$ (i.e., if there's a local maximum at δ_2), then color $\{v^{(1)}, \dots, v^{(r)}\}$ in color 1 (in this case we don't look at $\mathcal{H}$ at all, and simply color with color 1 — note that this is the same color 1 that appeared in $\mathcal{H}$, since we can't introduce new colors).
- If $\delta_1 > \delta_2$ and $\delta_2 < \delta_3$ (i.e., δ_2 is a local minimum), then we color $\{v^{(1)}, \dots, v^{(r)}\}$ in color 2. (Again color 2 is some color which already appeared in $\mathcal{H}$, different from 1 — for example, imagine 1 is red and 2 is blue.) (*A priori* it seems strange to zone in on these ones — the proof is very asymmetric — but we'll see in the proof that it works.)
- If none of the above cases occurs, then we color $\{v^{(1)}, \dots, v^{(r)}\}$ with an arbitrary color (of course, one of the colors from $1, \dots, t$).

Now we've colored every edge (because every edge not colored after the first three steps is automatically colored in the last step). (This definition is strange, and it's surprising that it works.)

We need to show that this coloring doesn't contain a clique of size k_i in color i for any i. This definition is asymmetric between colors — colors 1 and 2 appear in special places — but to avoid repeating ourselves, checking the condition for 1 and 2 is strictly harder than for the others, and the conditions on 1 and 2 are basically symmetric, so we'll just check it for color 1 (and the rest is analogous).

Suppose there is a clique in color 1 of size k_1, formed by vertices $w^{(1)}, \dots, w^{(k_1)} \in \{0,1\}^R$ ordered such that $w^{(1)} < \dots < w^{(k_1)}$. Unsurprisingly, to analyze this clique we'll need to talk about the δ-values between them; for each $i = 1, \dots, k_1 - 1$ let $\delta'_i = \delta(w^{(i)}, w^{(i+1)})$ (to avoid notation clash with the δ_i defined earlier when defining the color of an *edge*). Note that $\delta'_i \neq \delta'_{i+1}$ for all $i = 1, \dots, k_1 - 2$ (for the same reason as before — applying the second observation on δ to $w^{(i)}$, $w^{(i+1)}$, and $w^{(i+2)}$).

Now we have a sequence $\delta'_1, \dots, \delta'_{k-1}$. So far, we only know that any two consecutive terms are distinct. But the next claim tells us that there will be a block in this sequence which is nicely monotone.

> **Claim —** There are j and j' with $1 \leq j < j' \leq k_1 - r + 3$ and $j' - j \geq \left\lfloor \frac{k_1 - r + 4}{2} \right\rfloor - 1$ such that either $\delta'_j < \delta'_{j+1} < \dots < \delta'_{j'}$ or $\delta'_j > \delta'_{j+1} > \dots > \delta'_{j'}$.

So the whole sequence might do crazy things, but we can find a segment of this sequence which is increasing or decreasing, and this segment is reasonably long.

Proof. To prove this, we want to investigate the monotonicity behavior of this entire sequence. *A priori* it could do all kinds of crazy things. But recall that these vectors form a clique in color 1; so when we pick any r of these vectors, we *never* arrive at the third bullet point leading us to color 2 (we may arrive at any of the others, but not the third). This will help us severely restrict the monotonicity behavior of our sequence — we'll be able to show that it cannot have a local minimum (unless it occurs very late), because that'll force us to be in the color-2 case. Then because there are no local minima, the monotonicity behavior of the sequence will become much easier.

More precisely, there does not exist any $i \in \{1, \dots, k_1 - r + 1\}$ such that $\delta'_i > \delta'_{i+1}$ and $\delta'_{i+1} < \delta'_{i+2}$ (i.e., such that δ'_{i+1} is a local minimum) — if there were such an i, then the edge $\{w^{(i)}, w^{(i+1)}, w^{(i+2)}, \dots, w^{(i+r-1)}\}$ would have color 2 (which it does not, because it's part of the clique of color 1). To check this precisely, first note that this is a valid edge because $i + r - 1 \le k_1$ (so this statement makes sense). When we color this edge, we look at the consecutive δ-values in this sequence, which are exactly δ'_i, δ'_{i+1}, ... Then we're exactly in the case where the second is smaller than the first and third; which means we color the edge with color 2.

So the sequence $\delta'_1, \dots, \delta'_{k_1-r+3}$ cannot have a local minimum. Once this sequence decreases, it must keep decreasing forever — because if we went down and then up, we'd have a local minimum. This means the sequence must have an increasing part, followed by a decreasing part (either part may be empty, but we can't go back and forth between increasing and decreasing) — i.e., it can be split (possibly trivially) into an increasing part first and then a decreasing part.

The sequence has $k_1 - r + 3$ terms, so it contains $k_1 - r + 2$ steps. This means one of the two parts must consist of at least $\left\lfloor \frac{k_1-r+2}{2} \right\rfloor$ steps. So we can find an increasing or decreasing part with at least

$$\left\lfloor \frac{k_1 - r + 2}{2} \right\rfloor = \left\lfloor \frac{k_1 - r + 4}{2} \right\rfloor - 1$$

steps (note that the number of steps in $\delta_j, \dots, \delta_{j'}$ is $j' - j$), which gives us a monotone block of the desired length. □

Remark 10.4. For the colors other than 1 and 2, the entire sequence must be monotone (up to index $k_i - r + 3$) for the same reason, so we can essentially omit the factors of 2; but this doesn't really matter.

First suppose that $\delta'_j > \dots > \delta_{j'}$ in the claim, so the $j' - j + 1 \ge \left\lceil \frac{k_1-r+4}{2} \right\rceil$ vertices $\delta'_j, \dots, \delta'_{j'}$ in $\mathcal{H}$ cannot form a clique of color 1 in $\mathcal{H}$ (the $(r-1)$-uniform hypergraph which we defined to have no clique of this size in color 1). This means among them, we can pick $r - 1$ whose corresponding edge doesn't have color 1 in $\mathcal{H}$ — i.e., there exists indices $j \le i_1 < i_2 < \dots < i_{r-1} \le j'$ such that the edge $\{\delta'_{i_1}, \dots, \delta'_{i_{r-1}}\}$ does not have color 1 in $\mathcal{H}$. We now want to use this information to get a contradiction to the fact that $w^{(1)}, \dots, w^{(k_1)}$ is a clique of color 1 in our original hypergraph — we want to find an edge among these vertices which has the same color as our above edge.

Let's now consider the edge $\{w^{(i_1)}, \dots, w^{(i_{r-1})}, w^{(i_{r-1}+1)}\}$. (This is because we need to add one more vertex to have r of them; we have information about the consecutive δ's, so it makes sense to choose this one.)

Plugging this edge into our color definition, these vectors are already ordered correctly. Their consecutive δ's are

$$\delta(w^{(i_1)}, w^{(i_2)}) = \max\{\delta(w^{(i_1)}, w^{(i_1+1)}, \dots, w^{(i_2-1)}, w^{(i_2)})\} = \max\{\delta'_{i_1}, \delta'_{i_1+1}, \dots, \delta_{i_2-1}\} = \delta_{i_1}$$

(since the δ's are decreasing). Similarly, we have

$$\delta(w^{(i_2)}, w^{(i_3)}) = \delta'_{i_2},$$

and so on, up to

$$\delta(w^{(i_{r-2})}, w^{(i_{r-1})}) = \delta'_{i_{r-2}},$$

and finally we have $\delta(w^{(i_{r-1})}, w^{(i_{r-1}+1)}) = \delta'_{i_{r-1}}$ by definition. These are monotonic decreasing, so we are in the first case; this means the color of this edge $\{w^{(i_1)}, \dots\}$ is the same as the color of $\{\delta'_{i_1}, \dots, \delta'_{i_{r-1}}\}$ in $\mathcal{H}$. But this color is *not* color 1; this is a contradiction, because we've found an edge not of color 1 in what was supposed to be a clique of color 1.

The first case, where we instead have $\delta'_j < \cdots < \delta'_{j'}$, is very similar; but we look at the edge

$$\{w^{(i_1)}, w^{(i_1+1)}, w^{(i_2+1)}, \dots, w^{(i_{r-1}+1)}\}$$

instead (i.e., we pick indices whose δ''s form an edge in $\mathcal{H}$ not of color 1, and the same analysis shows that this edge in our r-uniform hypergraph has the same color as this edge in $\mathcal{H}$).

So in both cases we have a contradiction; this means there cannot be a clique of color 1 of size k_1. The other colors can be done analogously (for color 2, we run the same argument except that in the proof of the claim, we exclude a local *maximum* rather than a local minimum, so we split the sequence into a decreasing part and then an increasing part). For the other colors, we can simply copy-paste the proof (since we can exclude both a local maximum and local minimum, so in particular we can exclude a local minimum). □

Now this means if we have both upper and lower bounds for Ramsey numbers of uniformity r, we can boost them up to higher uniformity.

Definition 10.5. Define the tower function as $t_1(x) = x$, and $t_h(x) = 2^{t_h(x)}$.

So in other words, $t_h(x) = 2^{2^{2^x}}$, where the tower has height h.

Then by iterating our theorems, we obtain that for any $t \geq 2$, we have

$$t_{r-1}(c_{r,t}k^2) \leq R_r(k, \dots,) \leq t_r(c_{r,t}k).$$

So we have an upper and lower bound with tower heights differing by 1; this is exactly because we have such bounds for $r = 3$ (the lower bound is single-exponential with k^2 in the exponent, and the upper is double-exponential). In particular, to close the gap, it's enough to close it for uniformity 3; this is why uniformity 3 is the crucial open problem.

Also, for $t \geq 4$, the gap disappears because we've proven the double-exponential lower bound for R_3 — so for $t \geq 4$ we obtain

$$t_r(c_{r,t}k) \leq R_r(k, \dots, k) \leq t_r(c_{r,t}k).$$

§11 March 23, 2023 — Ramsey Numbers of Graphs

Today we'll talk about Ramsey numbers for graphs. This sounds like the same topic we've discussed before, but it isn't — unlike earlier lectures, we'll look at Ramsey numbers of graphs where the target we're looking for is not a clique, but is instead some other graph.

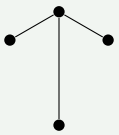
In the ordinary Ramsey theorem, we take t colors and specify target clique sizes for each. But now rather than cliques, we're looking for some specific *other* graphs. (The ordinary Ramsey theorem corresponds to the case where these target graphs are cliques of the corresponding sizes.)

Definition 11.1. For graphs G_1, …, G_t, the Ramsey number $R(G_1, \dots, G_t)$ is the smallest number such that the following holds: In any coloring of the edges of a complete graph on $R(G_1, \dots, G_t)$ vertices with t colors $1, \dots, t$, there exists a subgraph isomorphic to G_i in color i for some $i \in \{1, \dots, t\}$.

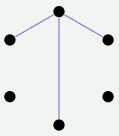
In other words, we're trying to find a copy of G_i which is entirely colored i. (The subgraph does not have to be induced.)

Example 11.2

Let G_1 be the star on four vertices.



Then the following is a copy of G_1 in the complete graph on 6 vertices in purple (it doesn't matter what color the remaining edges are).



We wrote this as a definition rather than a theorem that such a number exists. But in fact we have

$$R(G_1, \dots, G_t) \leq R(|V(G_1)|, \dots, |V(G_t)|)$$

(where the right-hand side is the ordinary Ramsey number where we're looking for cliques), and we know that the t-color Ramsey number exists — if we can find a clique on $|V(G_i)|$ vertices in color i, then in particular this clique contains a copy of G_i.

Cliques are the most intensively studied setting of Ramsey numbers, but there's also been some study of this setting.

§11.1 A Lower Bound

Theorem 11.3 (Chvátal, Havary 1972)

For any two connected graphs G and H, we have

$$R(G, H) \geq (\chi(G) - 1) \cdot (|V(H)| - 1) + 1.$$

Roughly speaking (if we ignore the ± 1's), this is $\chi(G) \cdot |V(H)|$.

Definition 11.4. The *chromatic number* of G, denoted $\chi(G)$, is the minimum number of colors we need to color the vertices of G in such a way that any two adjacent vertices have different colors.

(Note that the chromatic number refers to a *vertex* coloring, rather than an *edge* coloring.)

Example 11.5

If G is a clique, then all vertices must have different colors, so the chromatic number is the number of vertices.

We'll prove this on the homework. (There's a relatively simple explicit construction.)

Question 11.6. Is this a good bound?

In some cases, it's a terrible bound. When G and H are both cliques on k vertices, it gives the bound $R(k, k) \geq k^2$, which is terrible (the Ramsey number is exponential in k). But there are cases where this is a

pretty good bound, and in fact there are even cases where this is tight. The proof is similar to in the first class when we constructed a lower bound for $R(k,k)$ via an explicit construction, so it's natural to assume that the bound is generally terrible, but that's not always true — sometimes it's terrible, but sometimes it's actually pretty good.

§11.2 Ramsey Numbers of Trees vs. Cliques

Now we'll state an example where we can determine the Ramsey number *exactly*, and it matches this lower bound.

Theorem 11.7 (Chvátal 1977)
Let T be a tree with $t \geq 1$ vertices, and let $s \geq 1$. Then

$$R(T, K_s) = (s-1)(t-1)+1.$$

(Of course $R(G,H) = R(H,G)$, since we can swap the names of the colors.)

Notation 11.8. K_s denotes a clique on s vertices.

Amazingly, this works for any s and t (we don't need to assume anything is large). This might seem like cheating because trees are very simple, but the second graph is still a clique, which is the hardest thing to find — and the lower bound is terrible if we take a clique vs. a clique. But if we take a *tree* vs. a clique, we get this beautiful exact formula. There are very few results in Ramsey theory where the answer is known exactly, and this is one of them.

Proof. To prove the lower bound, by the previous theorem, taking $G = K_s$ and $H = T$, we have

$$R(T, K_s) = R(K_s, T) \geq (\chi(K_s)-1)(t-1)+1 = (s-1)(t-1)+1.$$

The more interesting part is to show the upper bound — that this number of vertices is enough.

Consider any coloring of the edges of the complete graph on $(s-1)(t-1)+1$ vertices with red and blue; we want to show that we can find either a red subgraph isomorphic to T, or a blue clique of size s. Suppose that this is not the case.

We'll look at the subgraph formed by the red edges (because then we can think about embedding a tree in the subgraph, and the condition we have no blue clique of size s still translates well) — let H be the graph consisting of the red edges. Then $|V(H)| = (s-1)(t-1)+1$. We assumed that H does not have a subgraph isomorphic to T, and since there's no blue clique of size s, the independence number of H is $\alpha(H) \leq s-1$. (We saw this same translation in the proof about off-diagonal Ramsey numbers.)

We're interested in the chromatic number of H. We have the inequality

$$\chi(H) \geq \frac{|V(H)|}{\alpha(H)}$$

(there exists a coloring of H with $\chi(H)$ colors, and each color class is an independent set, since two neighbors must have different colors. So $\chi(H)$ color classes, all of size at most $\alpha(H)$, cover all the vertices). This means

$$\chi(H) \geq \frac{|V(H)|}{\alpha(H)} \geq \frac{(s-1)(t-1)+1}{s-1} > t-1,$$

which means we must have $\chi(H) \geq t$.

Now we know there exists a subgraph of H (namely H itself) with chromatic number at least t. The next trick is to consider a *minimal subgraph* with chromatic number at least t — let H' be a minimal subgraph of H with the property that $\chi(H') \geq t$ (we needed the above bound to be sure that such an H' exists). (Here we don't require the subgraph to be induced — minimality means we can't delete any vertices *or* edges — but here it doesn't matter. There are other contexts where edge-minimality does matter, though.)

The reason we want to do this is that the minimality property (with respect to deleting vertices) gives us strong information about all the degrees in H'.

> **Claim —** For every vertex $v \in H'$, we have $\deg_{H'}(v) \geq t-1$.

This is in general a very good trick — it tells us that H' is a very nice subgraph of H, in the sense that all its vertices have high degree. This is pretty useful in being able to find a tree isomorphic to T (after proving this claim we'll be done). The point of doing this is not that we care about the chromatic number of H or H' — it's really to get the degree property, and the chromatic number is a vehicle for getting us a subgraph with this strong degree condition.

Proof. Assume for contradiction that $\deg_{H'}(v) \leq t-2$ for some vertex $v \in H'$. Now we delete it and use the minimality property — let H'' be the graph obtained from H' by deleting v. Then by the minimality of H', we have $\chi(H'') \leq t-1$ (we chose H' to be the smallest graph with $\chi(H') \geq t$, so since H'' is smaller than H', it can't satisfy the same condition).

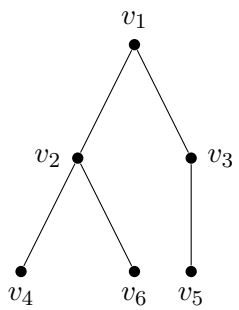
Now take a coloring of H'' with at most $t-1$ colors, such that no two adjacent vertices have the same color. But now we can extend this coloring from H'' to H' by assigning v a color which is different from the colors of all its neighbors — we can do this because it has $\deg_{H'}(v) \leq t-2$ neighbors, so there are at most $t-2$ colors which are forbidden, and we have $t-1$ colors at our disposal so there's at least one valid choice we can make.

So then we've found a valid coloring of H' with at most $t-1$ colors. This means $\chi(H') \leq t-1$, which is a contradiction. □

(We could slightly strengthen this — we could find a subgraph of H where all degrees are at least $\chi(H)-1$, since we haven't used t in any other way yet.)

Now we want to use H' to find a subgraph isomorphic to T in H (we'll actually find one in H'). We can do this greedily — at every step, we just take a new edge that works at that step, and the degree condition means that we'll never get stuck.

First, since $\chi(H') \geq t > 0$, H' is nonempty. We'll embed T as a subgraph into H'. To formally describe how we'll do this, we want to label the vertices of T in some reasonable way so that we don't jump around the tree (we can do either depth-first or breadth-first search) — let the vertices of T be $v_1, \ldots, v_T$ labelled in such a way that for every $i \in \{2, \ldots, t\}$, the vertex v_i has an edge to exactly one of $v_1, \ldots, v_{i-1}$. (Essentially, we're having the tree growing naturally, rather than having two parts which join later.) We can think of v_1 as the root of our tree.



We can now embed T into H' greedily, starting by embedding v_1. Whenever we want to embed a new vertex v_i for $i \geq 2$, it has exactly one edge to the previous vertices, if v_6 has an edge to v_4, we need to select a new neighbor of v_4 at which we embed v_6. The number of neighbors of v_4 is at least $t-1$. We can't take a vertex we've already chosen, so some of these may be taken up by the previous images; but the number of previous images is at most $t-2$ (since there's the vertex which is still missing we're trying to embed, and v_4 itself). So at most $t-2$ vertices are blocked, which means among the $t-1$ vertices, we can find one to use.

We can first embed v_1 because H' is nonempty. For every $i = 2, \ldots, t$, assume that $v_1, \ldots, v_{i-1}$ have already been embedded, and let v_j be the unique neighbor of v_i in $\{v_1, \ldots, v_{i-1}\}$ in T. Then the image w_j of v_j in H' has degree at least $t-1$ in H', and among the at least $t-1$ neighbors of w_j in H', at most $t-2$ have already been used up by being the images of vertices in $\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_{i-1}\}$ (since $i \leq t$ and v_j is missing, there's at most $t-2$ vertices in this list). So there exists a vertex w_i in H' adjacent to w_j which is not already in the image of one of the previous vertices $v_1, \ldots, v_{i-1}$.

Now w_i is a suitable image for v_i in this embedding — it's not equal to any of the previous vertices, and it's adjacent to w_j — so we can map v_i to w_i, and continue the embedding.

This argument tells us that we can embed T into H' — so T occurs as a subgraph of H', and therefore of H. This is a contradiction, because we assumed H does not have a subgraph isomorphic to T. □

Remark 11.9. As a philosophical remark, why did we need H'? H might have some vertices with really low degree; in order to run this argument, we need this condition on *every* vertex. In H many vertices may have high degree, but if we just embed greedily without thinking ahead, we have to be careful that we don't accidentally use a low-degree vertex where we get stuck later. So for a greedy argument to work we need a minimum-degree condition on *every* vertex; this is why we pass from H to H'.

§11.3 Another Upper Bound

In some sense, this answers the question on how good our bound for $R(G, H)$ is — in this case, it's exactly tight. However, there are other cases to consider as well. As a thought experiment, assume that G has bounded chromatic number. Then this gives a lower bound which is linear in the number of vertices in H; this is not particularly interesting, because clearly $R(G, H) \geq |V(H)|$. We've seen a bunch of standard clique Ramsey numbers which were exponential in the diagonal case, and polynomial in the off-diagonal case but still with exponents at least quadratic. From far away, this looks like a quadratic function. So all Ramsey numbers we've encountered are at least quadratic; having a bound that is linear in the number of vertices seems totally off. We'll see one example on the homework released today. But here's a more difficult statement (resolving a conjecture of Erdős and Burr). This theorem is really cool because it tells us that some Ramsey numbers are linear, and that's *really* surprising.

(This theorem is in the 'diagonal' regime, where we take $R(H, H)$; it gives a linear bound for $R(H, H)$ in terms of its number of vertices under some condition.)

Theorem 11.10 (Chvátal, Rödl, Szemeredi, Trotter)

For any graph H with k vertices and maximum degree d, we have

$$R(H, H) \leq C_d \cdot k,$$

where C_d is a constant only depending on d.

This tells us that the diagonal Ramsey numbers for H are linear in the number of vertices, where the linearity factor depends on the maximum degree — so in other words, for graphs with bounded maximum

degree, we have a *linear* bound on $R(H, H)$. (Note that the number of vertices in H can be much larger than k.)

We'll prove this after spring break. For a few historical remarks, this was originally proven with the regularity lemma. We'll prove it with a different proof (you'll learn about the regularity lemma in 18.225).

§11.4 Announcements

Some logistical announcements:

- Homework 2 is due today (at 8 p.m.); the new homework will come out today.
- Next week is spring break, so there will be no class or office hours. In particular, the next homework (which is released today) will be due three weeks from now instead of two, on April 13.
- The grades for Homework 1 have been published on Gradescope and Canvas; if there are grading issues we should email Yannick Yao at yyao1@mit.edu.

And here is a mathematical announcement (which is more interesting than the logistical ones). Recall that as we discussed, the diagonal graph Ramsey numbers $R(k, k)$ are the most fundamental numbers in Ramsey theory. In the first two weeks of class, we saw bounds roughly of

$$\sqrt{2}^k \leq R(k, k) \leq 4^k$$

(where the upper bound came from a binomial coefficient roughly $\binom{2k}{k}$, and the lower bound from a probabilistic argument). Three weeks ago, Prof. Sauermann told us that the improvement on the loewr bound is what we've seen in the homework; for the upper bound there'd been improvements on lower order terms, but the 4 was still the best-known exponent.

This was true 3 weeks ago, but there is now a new upper bound, of

$$R(k, k) \leq (4 - \varepsilon)^k,$$

where $\varepsilon \approx \frac{1}{128}$ (the important thing is that $\varepsilon > 0$). There's been a paper by Campos, Griffiths, Morris, and Sahasrabudhe making this gigantic breakthrough improving the upper bound from 4 to $4 - \varepsilon$; this is a gigantic breakthrough, because the barrier of 4 came out of the simple inductive argument we saw in class, and there's been a lot of work where no one managed to push beyond 4 until exactly a week ago.

Prof. Sauermann will try to reshuffle some of the content in the class so that we can hopefully talk a bit about what this proof is doing, towards later in the class.

§12 April 6, 2023

Before spring break, we started talking about Ramsey numbers of graphs. The version we stated earlier involved t colors, but here we'll only be concerned with two colors.

Definition 12.1. For graphs H and H', the Ramsey number $R(H, H')$ is the smallest number such that in any coloring of the edges of a complete graph on $R(H, H')$ vertices with red and blue, there is a red copy of H or a blue copy of H'.

If H and H' are cliques, then this is the same as the usual Ramsey number; in general we're not looking for monochromatic cliques, but rather monochromatic copies of a given graph.

Last class, we stated the following theorem, which gives a bound for the Ramsey number in the most interesting case — where $H = H'$ — for graphs of bounded degree:

Theorem 12.2 (Chvátal, Rödl, Szemerédi, Trotter 1983, resolving a conjecture of Erdős, Burr 1975)
For any graph H with k vertices and maximum degree d, we have $R(H, H) \leq C_d \cdot k$, where C_d is a constant only depending on d.

We should think of d as fixed and k as large. This is surprising because if we think of d as fixed, this bound is *linear* in the number k of vertices. Of course we also have a linear lower bound, since we certainly need at least k vertices to find a copy of H. So this theorem tells us that $R(H, H)$ is actually linear in the number of vertices if d is fixed. This is really surprising because most of the results we've seen for Ramsey numbers have been huge — we have terribly fast growing numbers for hypergraphs, and the graph Ramsey numbers are exponential. Then we saw that for trees vs. cliques we had roughly st-behavior, but that's still quadratic-looking. Meanwhile here we have something linear — that's really surprising, and why this is such an important result.

Remark 12.3. The original proof used Szemerédi's regularity lemma, which at that point was a rather new tool — it appeared in the 8 years between when the conjecture was phrased and when it was proven, and this was one of the first applications of Szemerédi's regularity lemma beyond Szemerédi's original work introducing it. But the proof we'll see doesn't use the regularity lemma (because not all of us know the regularity lemma — some of us may know it from 18.225 — and it takes a while to discuss). We'll see an alternative proof due to Graham, Rödl, and Rucinsky. In particular, because the proof doesn't rely on the regularity lemma, it gives better quantitative dependencies on C_d, although we don't really care about that.

The proof is long, but uses several interesting ideas that also appear in other places. (So it won't fit into today's class.)

As usual, we want to take a complete graph colored with red and blue, and we want to find a red copy of H or a blue copy of H. The proof finds this copy of H in different ways depending on the density of the red edges and the density of the blue edges in several subsets. So we'll split up into several lemmas that cover different cases of how we can find this red copy or blue copy of H.

The first lemma tells us that we can find a copy of H in some color if that color is *very* dense in some subset.

Lemma 12.4
Let H be a grpah with k vertices and maximum degree d, and let G be a graph with at least $4k$ vertices and density at least $1 - \frac{1}{8d}$. Then G contains H as a subgraph.

Later, we'll apply this lemma with the same H, taking G to be the edges of some color in some vertex *subset* of our graph.

How should we interpret these two conditions? The first assumption is just a technicality — if G has too few vertices (e.g., less than k), then we definitely can't find H as a subgraph. The important assumption is the density assumption — $1 - \frac{1}{8d}$ is really close to 1, meaning that G has almost all the possible edges. In the proof of the theorem, we'll see various density assumptions; but here we should emphasize that this density is really big (much bigger than $\frac{1}{2}$).

We'll now write down all the density definitions we need.

Definition 12.5. For a graph G and disjoint vertex sets X and Y in $V(G)$, the *density* between X and Y is defined as

$$d(X,Y) = \frac{e(X,Y)}{|X| \cdot |Y|},$$

where $e(X,Y)$ denotes the number of edges between X and Y. Meanwhile, the density of X is

$$d(X) = \frac{e(X)}{\binom{|X|}{2}},$$

where $e(X)$ is the number of edges inside X. The density of G is

$$d(G) = d(V(G)) = \frac{|E(G)|}{\binom{|V(G)|}{2}}.$$

In other words, in both cases we're counting the number of edges which G actually has, and dividing by the number of edges which could possibly exist — this is what you'd intuitively expect density to mean.

Proof. Our goal is to embed H as a subgraph in G — this means we want to find distinct images of all the vertices of H in G such that all vertices are mapped to edges. Most things in G are edges, which is good; so we just need to be careful to not hit a non-edge. Taking the complement of G (where we interchange edges and non-edges, so this complement has few edges and they create conflicts if we try to embed H), we want to make sure to avoid the vertices which have high degree in the complement of G.

We'll start by arguing that there's only a small number of such bad vertices. Let $\overline{G}$ be the complement of G, i.e., the graph on $V(G)$ with edges between precisely the pairs of vertices which are *not* edges in G. Then

$$d(\overline{G}) = 1 - d(G) \leq \frac{1}{8d}.$$

Let $n = |V(G)| = \left|V(\overline{G})\right|$, so that $\overline{G}$ has average degree at most $\frac{n}{8d}$. (In fact, it's at most $\frac{n-1}{8d}$, but this doesn't really matter.)

We want to consider vertices of high degree to be bad, and we'll use the threshold for 'high degree' to be twice the average, or $\frac{n}{4d}$; then $\overline{G}$ has at most $\frac{n}{2}$ vertices of degree larger than $\frac{n}{4d}$.

Now we kick these bad vertices out; let G' be the graph obtained from $\overline{G}$ by deleting all vertices $v \in V(\overline{G})$ with $\deg_{\overline{G}}(v) > \frac{n}{4d}$. Then $|V(G')| \geq \frac{n}{2}$, and $\deg_{G'}(v) \leq \frac{n}{4d}$ for all $v \in V(G')$.

We now want to embed H into G only using the vertices in G' — and now we can do this greedily, since we've already kicked out all the bad vertices. (This is similar to the greedy argument we saw before spring break.) Embedding greedily means that we embed one vertex of H at a time (in the tree case we had to be careful about the labelling of H; here we don't, and we can do it arbitrarily), being careful to map edges to edges.

More precisely, let $v_1, \dots, v_k$ be the vertices of H, and let us map these vertices to vertices $w_1, \dots, w_k$ of G inside the vertex set $V(G')$. (It's important that at every step, we take our image to be in the good set $V(G')$ — because if we go outside it, we might create too many conflicts for later steps.)

We need to check that we can do this — suppose that for some j, we've already defined the images $w_1, \dots, w_{j-1}$ in $V(G')$, and now we want to find $w_j \in V(G')$ with the desired edges (i.e., we need to respect the edge relationships from the new vertex to the vertices we've already defined, so w_j should have edges to all w_i for which v_j has an edge to v_i).

There are at most d different vertices w_i with $1 \le i \le j-1$ for which w_j needs to have an edge to w_i (since H has maximum degree d — these are precisely the indices such that v_j is a neighbor of v_i). For each of these w_i, the number of *forbidden* vertices in $V(G')$ — i.e., vertices that don't have an edge to w_i, meaning that we can't choose them for w_j — is at most $\deg_{G'}(w_i) \le \frac{n}{4d}$, since the edges in G' are precisely the non-edges in G (inside our vertex set $V(G')$). So each w_i excludes at most $\frac{n}{4d}$ choices for w_j. In total, this excludes at most $d \cdot \frac{n}{4d} = \frac{n}{4}$ vertices in $V(G')$ as choices for w_j (since there's at most d choices of i for which we need to take care of this condition) — and as long as we avoid these at most $\frac{n}{4}$ vertices with missing edges, our w_j is allowed. But there are at most $\frac{n}{4}$ forbidden vertices, and the set of vertices we can choose from has size

$$|V(G') \setminus \{w_1, \dots, w_{j-1}\}| \ge \frac{n}{2} - (k-1) \ge \frac{n}{4} + 1$$

(since we're not allowed to choose any of the previously chosen vertices). So it is indeed possible to choose w_j as desired.

This tells us we can always choose our new vertex w_j respecting the desired edge relationships. We can do this for each $j = 1, \dots, k$, which means we can indeed embed all our vertices (i.e., find distinct images for all our vertices such that edges are mapped to edges), and so we have found an image of H in G. (We've even found H in G', but we don't care about that in the end — we just needed that to maintain the condition in our greedy embedding that the number of forbidden vertices is sufficiently small.) □

This tells us that if we have a graph which is very dense, then we can find H as a subgraph. In particular, this tells us that if either the red graph or blue graph is very, very dense, then we are happy. Most likely, that won't happen though (since in most colorings, both will have density around $\frac{1}{2}$).

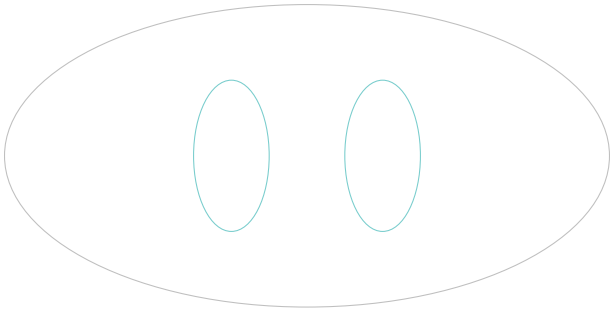
But in fact, this tells us more — we can also pass to a *subset* of the red or blue graph. So if within some *subset* the red edges are very dense, then we're still happy.

Now suppose that's not the case; so in every subset, there must be at least some small but reasonable amount of blue edges. This will tell us that in some sense, the blue edges are well-spread over the whole graph (they don't leave any patches empty). In that case, we'll use a different lemma to allow us to find a blue copy of H.

To more formally state this, we need a good notion of the blue edges being well-spread out. We'll need a definition which is a bit ad hoc for this proof (unlike the above density definitions, it's not a standard term).

Definition 12.6. We say that a graph G is bi-(ρ, δ)-dense for $0 \le \rho \le 1$ and $0 \le \delta \le 1$ if for all pairs of disjiont subsets $X, Y \subseteq V(G)$ with $|X| \ge \rho\,|V(G)|$ and $|Y| \ge \rho\,|V(G)|$ we have $d(X,Y) \ge \delta$.

In other words, for any two vertex subsets which are not too small, the density between them is at least δ; then ρ quantifies what 'not too small' means.



The next lemma tells us that under this assumption, we can also find a copy of H.

Lemma 12.7

Let H be a graph with k vertices and maximum degree d, and let $0 < \delta \leq 1$. If g is a bi-$(\frac{\delta^d}{8d^2}, \delta)$-dense graph on at least $8\delta^{-d} \cdot k$ vertices, then G contains H as a subgraph.

Here H is a given graph with k vertices (k may be very large) and fixed maximum degree d, and this lemma gives another sufficient criterion for G to contain H as a subgraph (different from the criterion above). Here our criterion is that G satisfies a bi-(ρ, δ)-dense assumption; but what δ and ρ do we use? This lemma says that we can take δ to be anything, and then if we choose ρ accordingly then we get such a criterion. (Of course, we also need G to have enough vertices, or else it would be hopeless to find a copy of H; this necessary number of vertices is again linear in k for fixed d and δ.) In the end, δ will be some function of d (e.g., $\frac{1}{10d}$) to combine with the first lemma.

Proof. The idea is again to embed H greedily into G, but we have to be more careful in how we do this than in the last lemma.

Again, we label the vertices of H when we embed them one by one. But this time we need to keep track for the allowed candidates of the images of all the other vertices (i.e., the vertices with the necessary edges to the currently embedded vertices). We'll need to maintain the condition that these candidate sets are big enough.

As a preprocessing step, we also want these candidate sets to be disjoint for any two vertices for which we need an edge, since the hypothesis only gives us a lower bound on edges between disjoint subsets. Obviously we can't make the sets disjoint for *all* the vertices, since the number of vertices is only linear in k — that would make the candidate sets tiny, and we want them to have size linear in k — but we want the candidate sets to be disjoint between all vertices for which we need edges.

To do this, we first take a proper vertex-coloring of H and then make disjoint candidate subsets for each of the colors; this will have the desired property.

Note that $\chi(H) \leq d + 1$, so the vertices of H can be colored with colors $1, \ldots, d + 1$ such that any two adjacent vertices have different colors. (This is because we can greedily color the vertices of H with $d + 1$ colors — whenever we want to assign a new vertex a color, it has at most d neighbors, so there are at most d forbidden colors, and therefore there's a color we can choose.) We now use these color classes to get a partition of G — partition $V(G) = U_1 \cup U_2 \cup \cdots \cup U_{d+1}$ into parts of almost equal sizes, i.e., so that the sizes differ by at most 1. Then

$$|U_i| \geq \left\lfloor \frac{|V(G)|}{d+1} \right\rfloor \geq \frac{V(G)}{4d} \geq \delta^{-d} k.$$

Now in the beginning, our candidate set for every vertex of H is obtained by looking at its color, and taking the corresponding candidate set — this gives us an initialization for the candidate sets. Now we greedily go on to embed the vertices of H; we have to be careful that the newly chosen vertex has edges to all the

vertices it's supposed to have edges to. In order to do that, we track this by adjusting our candidate sets — so whenever we choose a new vertex, we adjust the candidate sets for the future vertices so they only consist of vertices which have the correct edge relationships to be the embeddings of that vertex. (In particular, the candidate sets for vertices of the same color start out the same, but they may eventually change because the vertices have different edge-relationships.)

Let $v_1, \dots, v_k$ be the vertices of H. Our goal is to find a copy of H in G where $v_1, \dots, v_k$ are mapped to $w_1, \dots, w_k$ in G such that $w_i \in U_h$ for every vertex v_i of color h.

The next claim says that we can run this greedy embedding procedure with these candidate sets (which we update at every step and maintain to have large enough size).

Claim 12.8 — For all $\ell = 0, \dots, k$, we can find distinct images $w_1, \dots, w_\ell$ of $v_1, \dots, v_\ell$ in G with $w_i \in U_h$ for v_i of color h, such that:

- $w_i w_j$ is an edge whenever $v_i v_j$ is an edge;
- For each $\ell < i \le k$, the following holds: let h denotes the color of v_i, and let $N_\ell(i) \subseteq \{1, \dots, \ell\}$ be the set of indices $j \in \{1, \dots, \ell\}$ such that v_j is adjacent to v_i. Then there are at least $\delta^{|N_\ell(i)|} |U_h|$ vertices in U_h which are adjacent to w_j for all $j \in N_\ell(i)$.

In other words, the second condition tells us that the candidate sets for the future vertices are not too small — $N_\ell(i)$ records the neighbors of v_i among the vertices $v_1, \dots, v_\ell$ that we've already defined, i.e., the vertices for which we need to have an edge in H. So this is saying we look at the edges in U_h which are candidates of v_i — which requires them to have edges to all vertices in $N_\ell(i)$. The condition tells us that the number of such vertices is at least this number. This condition is reasonable because for every relationship we have, the number of such vertices would be expected to go down by a factor of δ.

The proof is roughly as long as the statement — once you've stated the conditions correctly, the claim basically inducts on its own once we've set it up right.

Proof. We use induction on ℓ. The case $\ell = 0$ is trivial — the first two conditions are trivially satisfied since there are no w_i embedded yet, and for the third condition there are no conditions since there are no embedded vertices, so the candidate sets are just the sets U_h.

Now suppose $\ell \ge 1$, and suppose that the images $w_1, \dots, w_{\ell-1}$ of $v_1, \dots, v_{\ell-1}$ are already defined by the induction hypothesis (i.e., such that the conditions hold). All we need to do is find a good image for w_ℓ in the correct set U_h, which has edges to the correct w_i we've already defined — these two conditions are automatically satisfied by being in the candidate set we currently have for it at that point — such that when we update the other candidate sets accordingly, they don't get too small.

Let $W_\ell, \dots, W_k$ be the current candidate sets for the vertices $w_\ell, \dots, w_k$ in the last condition of the claim. Then for all $i = \ell, \dots, k$, we have that

$$|W_i| \ge \delta^{N_{\ell-1}(i)} \cdot \frac{|V(G)|}{4d} \ge \frac{\delta^d}{4d} |V(G)|.$$

(using the bound $|U_h| \ge \frac{V(G)}{4d}$). We now want to find a vertex $w_\ell \in W_\ell$ — which automatically satisfies the first two conditions — such that $|N(w_\ell) \cap W_i| \ge \delta |W_i|$ for all $i = \ell + 1, \dots, k$ such that v_i is adjacent to v_ℓ. The choice of w_ℓ doesn't influence any of the candidate sets for v_i not adjacent to v_ℓ, so we don't need to update their candidate sets; but for the ones which are adjacent, we need to update the sets to intersect them with the neighborhood of w_ℓ, since we need those embeddings to be neighbors of w_ℓ. In those cases, we've also increased the size of $|N_\ell(i)|$ by 1. So we are allowed one extra factor of δ, which means if we can do this then we're happy.

First, there are at most d such indices i, since v_ℓ has at most d neighbors. For each of them, we claim that there are at most $\frac{\delta^d}{8d^2} |V(G)|$ choices for $w_\ell \in W_\ell$ which *fail* the condition — this is because the set of choices

which fails the condition has density less than δ to the set W_i. (We take the set of all choices failing the condition and think of it as a vertex set, and look at its density to W_i. That density is less than δ by the definition of failing the condition — each failing vertex has density less than δ, so this remains true when we combined them. Our lemma has a bi-density assumption, and W_i itself has size greater than a $\frac{\delta^d}{8d^2}$-fraction of vertices, since we saw it has at least a $\frac{1}{4d}$-fraction. So if there were at least this many failing vertices, then these two sets would contradict the bi-(ρ, δ)-density assumption; this means we can have at most this many failing vertices.)

So for each i, we get such a condition on w_ℓ. Then the total number of forbidden choices is at most

$$\frac{\delta^d}{8d} |V(G)| \leq \frac{1}{2} |W_\ell|$$

(using the lower bound for W_ℓ). So at most half of the choices for W_ℓ are forbidden, which means we can make a valid choice for w_ℓ. (You again have to check that you can also choose one which is distinct from the previous images; this follows from the fact that $\frac{1}{2} |W_\ell|$ is greater than the number of previously chosen vertices.) □

Then the lemma trivially follows from the claim taking $\ell = k$, since when $\ell = k$ we've found distinct images for all the vertices in G with the correct edge relationships, which means we have embedded H into G. □

§13 April 11, 2023 — Ramsey Numbers of Bounded-Degree Graphs

Today we'll continue discussing Ramsey numbers of bounded-degree graphs. Recall that for the last few classes, we've been looking at Ramsey numbers where our target graphs are not cliques, but arbitrary graphs.

Definition 13.1. For graphs H and H', the *Ramsey number* $R(H, H')$ is the smallest number such that in any coloring of the edges of a complete graph on $R(H, H')$ vertices with red and blue, there is a red subgraph isomorphic to H or a blue subgraph isomorphic to H'.

Last time, we started the proof of the following theorem, which considers the 'diagonal' case where $H = H'$ (which is probably the most interesting case).

Theorem 13.2 (Chvátal, Rödl, Szemerédi, Trotter 1983, resolving a conjecture of Erdős, Burr 1975)
For any graph H with k certices and maximum degree d, we have $R(H, H) \leq C_d k$, where C_d is a constant only depending on d.

Last time we saw most of the proof; today we'll finish the proof. This theorem is amazing because the bound $C_d k$ is *linear* in the number of vertices of H (we think of d as fixed and k of large) — this is very surprising because the Ramsey numbers are typically *much* bigger than the size of the original object.

The proof we're seeing isn't the original proof (which used the regularity lemma), but a later proof that gives a better dependence of C_d on d (our proof will give something like d^{d^2}). Our proof might seem somewhat complicated, but it's not much more complicated than the regularity lemma proof.

Recall that on Thursday, we proved two lemmas.

Lemma 13.3
Let H be a graph with k vertices and maximum degree d. Let G be a graph with at least $4k$ vertices and density at least $1 - \frac{1}{8d}$. Then G contains H as a subgraph.

This lemma tells us that if a graph is very, very dense, then it must have H as a subgraph. This gives one sufficient criteria for us to find H in some graph. We then proved another lemma that gives another sufficient criterion, relying on the definition of bi-(ρ, δ)-density.

Definition 13.4. We say that a graph G is bi-(ρ, δ)-dense for $0 \leq \rho \leq \frac{1}{2}$ and $0 \leq \delta \leq 1$ if for all pairs of disjoint subsets $X, Y \subseteq V(G)$ with $|X| \geq \rho |V(G)|$ and $|Y| \geq \rho |V(G)|$, we have $d(X, Y) \geq \delta$.

(We can allow any $0 \leq \rho \leq 1$, but the definition becomes pointless if $\rho > \frac{1}{2}$ — every graph is then bi-(ρ, δ)-dense, since no such X and Y exist.)

In other words, bi-(ρ, δ)-density tells us that whenever we take two vertex subsets which are not too small, we have a reasonable density of edges between them.

The second lemma gives us a sufficient condition for a graph to contain H as a subgraph using bi-(ρ, δ)-density.

Lemma 13.5

Let H be a graph with k vertices and maximum degree d, and let $\delta > 0$. If G is a bi-$(\frac{\delta^d}{8d^2}, \delta)$-dense graph on at least $8\delta^{-d}dk$ vertices, then G contains H as a subgraph.

This lemma has the same conclusion, but a different assumption on G. Rather than requiring the graph G to be very dense, we only require a moderate density of δ (which we should think of as close to 0). But the point is that we don't require this density just on the graph as a whole, but between *any* two vertex subsets. (The term $\frac{\delta^d}{8d^2}$ is just a particular ρ we needed in order to make the proof work; it is not too important.)

Both proofs of these lemmas found a greedy embedding of H. The first lemma did this easily — we just kick out low-degree vertices, and have no problems with the remaining. The second was more tricky — we kept candidate sets, and chose vertices under the condition that the candidate sets don't shrink too much in size.

Now we'll show that these lemmas together imply the theorem. The point is that we have two different sufficient criteria to find H as a subgraph in G. The idea is to take a graph G on the given number of vertices with edges colored red and blue, and try to find a red or blue copy of H using lemma 1 or 2. We'll apply the first lemma to red and the second to blue; but so far it's not obvious that one of the two lemmas always applies. To apply lemma 1 we need a subset where red is very, very dense. If this doesn't hold, then *within* every subset, we have a reasonable density of blue edges. But lemma 2 doesn't require reasonable density *within* a subset, but rather *between* two subsets. The next lemma tells us that we *will* always be able to apply one of the two lemmas.

Lemma 13.6

Let $0 \leq \rho \leq \frac{1}{2}$ and $0 \leq \delta \leq \frac{1}{3}$, and let $s \geq \delta^{-1}$ be an integer. Then for every graph G, at least one of the following two statements holds:

(a) G has a bi-(ρ, δ)-dense subgraph on at least $(\frac{\rho}{2})^s |V(G)|$ vertices.

(b) There is a vertex subset $U \subseteq V(G)$ of size at least $(\frac{\rho}{2})^s |V(G)|$ with density $d(U) \leq 3\delta$.

So in the first case, we can find a nice bi-(ρ, δ)-dense subgraph whose number of vertices is linear in the number of vertices of G. Meanwhile, in the second case we can find a pretty sparse subset of a certain size. (We lose a constant factor in the density, but it doesn't really matter.) It's tautological that you can find a sparse subset or a dense subset; the content of this lemma is that you can find a sparse subset or a *bi*-dense subset.

The upshot of this lemma is that it tells us that in our coloring of the complete graph, the blue subgraph will fall into one of these two cases. If the blue graph falls into (a), then we have a blue bi-(ρ, δ)-dense subgraph, and we can apply Lemma 2. If it falls into (b), then we have a subgraph where the blue edges are very sparse, so the red edges are very dense and we can apply Lemma 1 on the red graph.

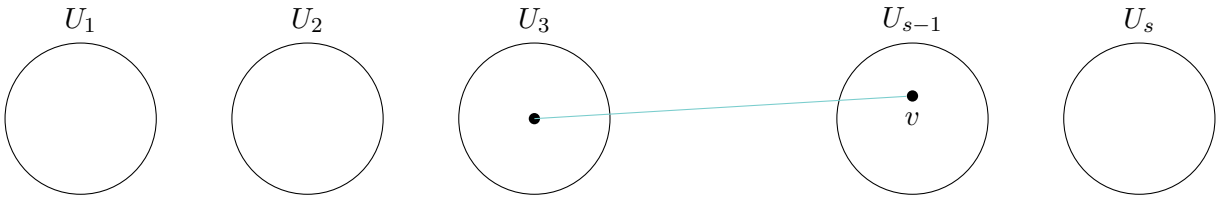
Before we do this, we'll prove the lemma.

Proof. Let $t = \lceil (\frac{\rho}{2})^s |V(G)| \rceil$ (so t is the minimum size of the subgraph we're trying to find).

Assume that (a) does not hold; this means every subgraph of G on at least t vertices is *not* bi-(ρ, δ)-dense. Our goal is to now show that (b) holds, i.e., to find a subset of at least t vertices which is very sparse.

The idea for how to construct this subset is roughly as follows: if we don't have a bi-(ρ, δ)-dense subgraph, we first look at G as a whole. This graph is not bi-(ρ, δ)-dense, so we can find disjoint X and Y of reasonable size with density less than δ. Our goal is to then find a single subset in which there is low density. We might first try to take $X \cup Y$, but this isn't good enough (that graph might have density up to $\frac{1}{2}$, because of edges inside X and Y). But the idea is to take a piece of X and then continue in Y — repeatedly applying this operation to construct pieces such that between any two pieces we have low density. If we construct a lot of such pieces and patch them together, that'll be good enough.

In other words, we'll construct disjoint vertex subsets $U_1, \ldots, U_s$ of sizes $|U_1| = \cdots = |U_s| = t$ such that $U_1 \cup U_2 \cup \cdots \cup U_s$ satisfies (b). The idea is to construct these in such a way that we have low density between any two of them. In fact, we'll construct these subsets in such a way that for all $1 \leq i \leq t$, every vertex $v \in U_{i+1} \cup \cdots \cup U_s$ has at most $2\delta t$ edges to U_i.



This condition tells us that if we take e.g., U_3 and then look at any vertex v in one of the future sets, then v has at most $2\delta t$ edges to U_3. In particular, this will tell us that between any two subsets, the density is at most δ (since we can count density by summing over vertices in the later subset).

We will construct these recursively. (Any recursive argument can be written down with a lot of indices, or by writing down a formal claim to induct over; Prof. Sauermann prefers the latter because it involves fewer indices.) The claim we'll induct over is the following.

> **Claim —** For every $\ell = 0, \ldots, s$, we can find disjoint subsets $U_1, \ldots, U_\ell$, $W \subseteq V(G)$ of sizes $|U_1| = \cdots = |U_\ell| = t$ and $|W| \geq (\frac{\rho}{2})^\ell |V(G)|$ such that for all $i = 1, \ldots, \ell$, every vertex in $U_{i+1} \cup \cdots \cup U_\ell \cup W$ has at most $2\delta t$ edges to U_i.

Here W will be our set of candidates for the later vertex subsets; we want W to be big enough, since it's our pool of future vertices. (Our condition shouldn't just hold for the sets themselves, but also for the pool of later vertices.) This claim just formalizes our goal in a recursive fashion, where W is the set of 'candidate vertices' that satisfy the condition, and therefore from which we can choose our future vertices.

Proof. We use induction on ℓ. The case $\ell = 0$ is vacuously true — we don't need to specify any sets U_i, and we can take W to be the set of all vertices (since there are no conditions on W).

Now suppose $\ell \geq 1$ and that we've already constructed $U_1, \ldots, U_{\ell-1}$, W'. We now want to find disjoint sets $U_\ell, W \subseteq W'$ such that $|U_\ell| = t$, $|W| \geq (\frac{\rho}{2}) |W'|$, and every vertex in W has at most $2\delta t$ edges to U_ℓ. The reason we have so few conditions here are that all others are automatically guaranteed if we take $U_\ell, W \subseteq W'$

— for any of the sets U_i with $i \leq \ell-1$ and any vertex in either a later set or W', that vertex will have at most $2\delta t$ edges to this set. So if we pick U_ℓ and W to be subsets of W', then we automatically get this condition for free. This means the only pair we need to check is U_ℓ and W — all other conditions are automatically satisfied. For the size conditions, if we take $|W| \geq \frac{\rho}{2}|W'|$, then since we had $|W'| \geq (\frac{\rho}{2})^{\ell-1}|V(G)|$, the condition on $|W|$ is automatically satisfied as well.

The way we'll find U_ℓ and W is by using the fact that W' is not bi-(ρ,δ)-dense — since $|W'| \geq (\frac{\rho}{2})^{\ell-1}|V(G)| \geq (\frac{\rho}{2})^s|V(G)|$ we must have $|W'| \geq t$ (since t is simply the ceiling of the right-hand side), then the induced subgraph $G[W']$ is not bi-(ρ,δ)-dense (since we assumed that there is no bi-(ρ,δ)-dense subgraph on at least t vertices). This tells us that there are two disjoint subsets $X, Y \subseteq W'$ of size $|X| \geq \rho|W'|$ and $|Y| \geq \rho|W'|$ such that $d(X,Y) \leq \delta$.

We want to take U_ℓ and W to be subsets of these sets. But these sets are much bigger than the ones we want — U_ℓ should have size *exactly* t, while X may be much bigger than this. So we would like to pass to a vertex subset of X which still satisfies this degree condition. We can do this by an averaging or random argument — we can restrict X to a smaller subset of itself which still has the same density condition to Y.

To check that X really is big enough, note that

$$|X| \geq \rho|W'| \geq \rho\left(\frac{\rho}{2}\right)^{\ell-1}|V(G)| \geq \left(\frac{\rho}{2}\right)^s|V(G)|,$$

which means $|X| \geq t$. So X really is big enough to find a subset of size t, and by averaging over all subsets of X of size t, we can find a particular subset $U_\ell \subseteq X \subseteq W'$ of size $|U_\ell| = t$ which satisfies the same density condition to Y — i.e., $d(U_\ell, Y) \leq \delta$ (since the average density over all these subsets is simply the density between X and Y).

We now need to choose $W \subseteq Y$. We can't just take W to be Y because we need an upper bound for the number of edges from *every* vertex of W. On *average* the vertices of Y have less than δt edges to U_ℓ, but there may be some vertices with too high degree.

But we can deal with this by simply deleting the vertices of too high degree — the vertices $v \in Y$ have *on average* less than δt edges to U_ℓ. So there are at most $\frac{1}{2}|Y|$ vertices $v \in W$ with at least $2\delta t$ edges to U_ℓ. We then kick out those vertices; let $W \subseteq Y \subseteq W'$ be the set of vertices $v \in Y$ with at most $2\delta t$ vertices to U_ℓ. Since we kicked out at most half of Y, we have $|W| \geq \frac{1}{2}|Y| \geq (\frac{\rho}{2})|W'|$ (since $|Y| \geq \rho|W'|$).

This means we've found W (which is disjoint from U_ℓ, since X and Y were disjoint), proving the claim. □

Now taking $\ell = s$ in the claim, we've found our sets $U_1, \ldots, U_s$ with the condition we wanted (as well as a leftover set of candidates, but we don't care about those anymore — the candidates were just for the purpose of the induction) — i.e., we have disjoint subsets $U_1, \ldots, U_s$ of $V(G)$, each of size $|U_1| = \cdots = |U_s| = t$, such that our edge-count condition holds — so in particular, such that for all $1 \leq i < j \leq s$, there are at most $2\delta t^2$ edges between U_i and U_j. (We know *every* vertex in U_j has at most $2\delta t$ edges to U_i, and $|U_j| = t$, so summing over all vertices in U_j gives at most $2\delta t^2$ edges.)

We now take the set $U = U_1 \cup \cdots \cup U_s$; we want to check that it satisfies (b). First we have $|U| = st \geq t \geq (\frac{\rho}{2})^s|V(G)|$, so U is indeed big enough. Now we need to check it has sufficiently low density — the density in U is

$$d(U) = \frac{e(U)}{\binom{|U|}{2}}.$$

To bound $e(U)$, there are two types of edges in U — the edges *inside* the sets $U_1, \ldots, U_s$, and the edges between them. There are s sets and each has at most $\binom{t}{2}$ edges, so there are at most $s\binom{t}{2}$ edges living inside the sets. Meanwhile there are at most $2\delta t^2\binom{s}{2}$ edges between two sets (since there are $\binom{s}{2}$ pairs (i,j), each of which has at most $2\delta t^2$). So then

$$d(U) \leq \frac{s\binom{t}{2} + \binom{s}{2}\cdot 2\delta t^2}{\binom{st}{2}} = \frac{st(t-1) + s(s-1)2\delta t^2}{st(st-1)} = \frac{t-1}{st-1} + 2\delta\frac{(s-1)t}{st-1} \leq \frac{1}{s} + 2\delta \leq 3\delta.$$

(Geometrically, we have blobs, and we can control the density between the blobs to be at most 2δ; this in total contributes a density of at most 2δ. Then the question is how much the density is increased if we add in the edges inside the blobs. This gives at most a $\frac{1}{s}$ contribution (since at most $\frac{1}{s}$ edges live inside the blobs).)

So then U satisfies (b), and we are done. □

We are now equipped with these three lemmas, and our goal is to now deduce the theorem. All the work is basically done; we just have to put the puzzle pieces together now.

Proof of Theorem. The idea is that we want to apply Lemma 3 to the blue graph, so that we fall into cases (a) or (b). In (a), we find a blue bi-(ρ,δ)-dense subgraph and apply Lemma 2 to the blue graph; in (b) we find a very dense red subgraph and apply Lemma 1. Now the question is which ρ and δ we want to use.

We want $3\delta \le \frac{1}{8d}$ (for Lemma 1), so we take $\delta = \frac{1}{24d}$. Meanwhile, we take $\rho = \frac{\delta^d}{8d^2} = \frac{1}{8\cdot 24^d d^{d+2}}$ (for Lemma 2); note that this only depends on d. Finally, we take $s = \frac{1}{\delta} = 24d$.

We will take

$$C_d = \left(\frac{2}{\rho}\right)^{25d} = \left(16\cdot 24^d \cdot d^{d+2}\right)^{25d}$$

(so the main behavior is δ^{δ^2}). Consider a complete graph on $C_d \cdot k$ vertices with edges colored red and blue, so we want to find a red or blue copy of H.

Let G be the graph formed by the blue edges. Then the size bound in Lemma 3 is

$$\left(\frac{\rho}{2}\right)^s |V(G)| \ge \frac{2}{\rho}\cdot k \ge 8\delta^{-d}dk$$

(we even have a d in the exponent left, but it doesn't matter). Now applying Lemma 3 to G, we conclude that at least one of the following two cases holds:

(a) The blue graph G has a bi-(ρ,δ)-dense subgraph (for these particular values of ρ and δ) on at least $8\delta^{-d}dk$ vertices. Then we are in the situation where we can apply Lemma 2 (the parameters were chosen so that the conditions hold), so we can find a blue copy of H.

(b) There is a vertex subset $U \subseteq V(G)$ of size $|U| \ge 8\delta^{-d}dk \ge 4k$ with blue density $d(U) \le 3\delta = \frac{1}{8d}$. Then the red density inside U is at least $1 - \frac{1}{8d}$ (since the red and blue densities together add up to 1). So we are in the situation to apply Lemma 1 to the red graph; then by Lemma 1 applied to the red graph, we can find a red copy of H.

So in either case, we have found a monochromatic copy of H, and we are done. □

Remark 13.7. The value of C_d we get from this proof is roughly d^{25d^2}. This is not the best-known bound. The original proof gave a much *worse* C_d, because it applied the regularity lemma (giving a tower-type function, which is terrible as a quantitative dependence). This is much better, but in fact better dependencies are known. A slightly more careful version of the proof we just saw already improves this dependence a bit; it was later improved by Conlon, Fox, and Sudakov to one of the form $\exp(cd\log d)$ (i.e., it removes one d from the exponent — it's of the form d^{ad} rather than d^{ad^2}). This is the best-known bound.

The theorem can also be strengthened qualitiatively — this theorem is amazing because it gives a *linear* dependence in the number of vertices. The qualitative strengthening, proven relatively recently, weakens the maximum degree assumption on H. Obviously we need *some* assumption (since the behavior for cliques is very different). But rather than saying something about the *maximum* degree, there's a theorem by Lee (also conjectured by Erdős and Burr) which widened the class to *d-degenerate graphs*.

Theorem 13.8 (Lee 2017, resolving a more general conjecture of Erdős and Burr)
We have $R(H,H) \leq C_d k$ for every d-degenerate graph H on k vertices.

Definition 13.9. A d-degenerate graph is a graph which can be constructed by the following operations: we start with the empty graph. We then add vertices one at a time, and for each vertex we add, we can draw at most d edges to the previous vertices.

More formally, it's a graph where the vertices can be labelled so that every vertex has degree at most 3 to the preceding vertices.

(Lemma 1 in our proof still works — we can choose which order to embed our vertices, so we can take the given ordering. For Lemma 2 however, the proof we saw breaks down because we had to use the maximum degree condition going the other way — for every new vertex we chose we had to make sure the candidate sets don't shrink too much, and there we used the condition on at most d neighbors for the *not* yet embedded vertices.)

In particular, every graph of maximum degree d is degenerate (this is trivially true, since we can take any order to add the vertices in; if the maximum degree is at most d). So this is a strengthening of the original theorem.

§14 April 13, 2023 — Multi-Color Ramsey Numbers

Last two classes, we've looked at Ramsey numbers of graphs, where we're looking for graphs that are not necessarily cliques. We proved some surprising results in the bounded-degree setting. Today we'll go back to the classical Ramsey setting with cliques. We've already talked quite intensively about the known bounds for Ramsey numbers, especially diagonal Ramsey numbers, in the most classical 2-color setting. Today we'll talk about bounds for the diagonal Ramsey numbers with many colors (where many means more than 2).

Definition 14.1. Given $k_1, \dots, k_t \geq 2$, the *Ramsey number* $R(k_1, \dots, k_t)$ is the smallest number such that in any coloring of the edges of a complete graph on $R(k_1, \dots, k_t)$ vertices with colors $1, \dots, t$, there is a monochromatic clique of size k in color i for some $i \in \{1, \dots, t\}$.

As usual, we have t colors, and $k_1, \dots, k_t$ are our target clique sizes. The most natural case to study is when $k_1 = \cdots = k_t = k$, so that we're just looking for a monochromatic clique of size k.

We already proved that $R(k_1, \dots, k_t)$ exists, and our proof gave an upper bound on this number; we'll start by recalling the bounds that we already know. Our simple proof of existence from the first week of the course gives a recursive upper bound

$$R(k_1, \dots, k_t) \leq \sum_{j=1}^{t} R(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_t)$$

(where we sum over all the different colors, and in each term, we reduce one target clique size by 1). (In fact, the bound is a tiny bit better because we can subtract 1 from each of these terms and add 1; but this doesn't really matter.) By repeatedly plugging this in and solving the recurrence, we get an explicit bound of the form

$$R(k_1, \dots, k_t) \leq \binom{k_1 + \cdots + k_t - t}{k_1 - 1, \dots, k_t - 1}.$$

(This notation denotes a multinomial coefficient — it's the number of ways to take the numbers $1, \dots, k_t - t$ and color $k_1 - 1$ of them with color 1, $k_2 - 1$ of them with color 2, and so on.)

(We may not have written this recursive formula down directly, since we did the proof in the r-uniform hypergraph setting; but we could get this expression by either specializing our proof to $r = 2$, or by generalizing our 2-color proof in the obvious way.)

We care most about the setting when all our target clique sizes are the same; in that case, we get the relatively easy upper bound

$$R(k, \dots, k) \leq \binom{tk - t}{k - 1, \dots, k - 1} \leq t^{kt}$$

(where t is the number of colors).

This is an easy upper bound, but it's unclear whether its behavior is close to the truth. Understanding $R(k, \dots, k)$ is one of the biggest open problems in graph Ramsey theory. We have two parameters k and t; so we can consider the two extremes where we fix K and let t be large, or fix t and let k be large (this is the most natural first step in most extremal combinatorics problems with two parameters).

§14.1 The Fixed-k Setting

We'll first consider what happens when we fix k and let t be large (so we have many colors). When $k = 2$, the problem is not interesting. But even when we fix $k = 3$, the answer isn't understood.

Conjecture 14.2 (Erdős, $100) — There is an absolute constant $C > 0$ such that $R(3, 3, \dots, 3) \leq C^t$.

(This is smaller than the bound we got above, which is exponential in $t \log t$.)

The best-known bounds are of the form

$$C_1^t \leq R(3, \dots, 3) \leq C_2 t!$$

for absolute constants C_1 and C_2 (we'll see these bounds on the homework, though not with the best constants known). It's conjectured that the true answer is exponential in t, but this isn't known. (Note that $t!$ is roughly like t^t, so in some sense it's not so much better than our above bound.)

§14.2 The Fixed-t Setting

Now we'll look at the setting where the number of colors is fixed and k is large. This might be more natural in the context of what we've discussed before, since we usually operate with a fixed number of colors (e.g., we've talked extensively for the known bounds for $R(k, k)$).

For a fixed number $t \geq 2$ of colors and large k, the function

$$R(k, k, \dots, k) \leq t^{kt}$$

is exponential in k. We also have a lower bound of $\sqrt{2}^k$ (since this is true even in the case where we have 2 colors). So this quantity really is exponential in k, but the next question is, what is the correct base of the exponent?

This is in some sense unsurprising, since it's a similar state of what we know for $k = 2$ — the lower bound is essentially $\sqrt{2}^k$, and when we discussed the upper bound it was 4^k (and now it's $(4 - \varepsilon)^k$), but we still have the question of determining the correct exponential base. This problem was also stuck for a long time without any movement, but in the last few years, more progress has been made.

Question 14.3. What's the best easy lower bound we can get for fixed t?

We can easily get a lower bound of $\sqrt{2}^k$ from the $R(k, k)$ case. But we can do better. In the case where $t = 2$ we took a random coloring (where we colored each edge red or blue uniformly at random); similarly, here it's natural to try to color every edge one of $1, \dots, t$ uniformly at random. We won't do the calculation; but this gives a lower bound of roughly $\sqrt{t}^k$ (this is perhaps unsurprising because it's the same as for $t = 2$).

This is nice, but if you think about it, if you start with 2 colors then you get $\sqrt{2}^k$. If we go to 4 colors, or 5 colors, or 6 colors, the base grows rather slowly with t — as we double our colors, the base only is multiplied by a factor of $\sqrt{2}$. But having more colors should give us in some sense *way* more power — for example, in our proof of the stepping-up lemma for 3-uniform hypergraphs, having 4 colors was much more powerful than having just 2.

And this is in fact *not* the best we can do. For $t \geq 4$, a better bound can be obtained by a *product construction* (which takes the construction from $t = 2$, but then makes use of the additional colors in a better way).

Lemma 14.4

We have $R(k, \dots, k) - 1 \geq (R(k, \dots, k) - 1)(R(k, \dots, k) - 1)$, where the term on the left-hand side has $t_1 + t_2$ colors, and the terms on the right-hand side have t_1 and t_2 colors.

In other words, given a construction of a graph with t_1 colors with no monochromatic clique, and a construction for t_2 colors, we can obtain a construction for $t_1 + t_2$ colors on their product number of vertices.

Proof. The idea is to take the second construction and repeat it inside each vertex of the first construction. This is called a *blow-up* construction — we take the first construction and blow up each vertex (turning it into a blob), and inside each blob we place a copy of the second construction (using t_1 different colors). Suppose we start with the following first construction.

Then we replace each vertex with a blob, and we color all the edges between any two blobs with the original color of the edge between those two vertices.

We make each blob have $R(k,\dots,k)-1$ vertices (with t_2 k's), and we place a copy of the first construction in each blob.

If there were a monochromatic clique of size k in some color, then it'd have to be in either one of the first t_1 colors, or one of the latter t_2. If it were in one of the first t_1 colors, then it could only have one vertex from each blob (since the edge between any two vertices in the same blob is of one of the last t_2 colors), so it'd be a monochromatic clique in the original graph, which we assumed doesn't exist.

Similarly, if there were a clique in one of the last t_2 colors, then since the blobs aren't connected by that color, it'd have to live entirely in a blob; but our blobs don't have any cliques by construction, so this is again a contradiction. □

Together with $R(k,k)\geq\sqrt{2}^k$, iterating the lemma now gives

$$R(k,\dots,k)\geq 2^{kt/4}$$

for even t. (If t is even, then we can split the t colors into pairs and iterate this lemma; we get a product of $\frac{t}{2}$ factors each of this size.) This is significantly better than $\sqrt{t}^k$. If t is odd, we have to be more careful; if we pair then we get one factor left, which we could just waste, but that's a bit wasteful. We could instead group into 3's.

But in fact pairs aren't the most efficient — we get something better if we use groups of 3 and use the bound of $\sqrt{3}^k$. So combining the lemma with $R(k,k,k)\geq\sqrt{3}^k$, we get the bound

$$R(k,k,\dots,k)\geq 3^{kt/6}$$

if t is divisible by 3. (We can check that $3^{1/6}$ is a bit larger than $2^{1/4}$.)

Then if t is not divisible by 3, we can combine the leftover colors in pairs (if we have two leftover colors we pair them; if we'd have one, then we instead form two pairs).

These bounds were known for several decades, and this is where they'd been stuck for a long time (similarly to the two-color case, where we had an exponential upper bound and an exponential lower bound, but the bases were different).

But then there was movement a few years ago, with a pretty surprising paper that improved this lower bound. The improvement doesn't look drastic, but it's significant because the bound was stuck for a long time, and it's surprising that there's such a nice and simple idea that no one had found before.

Theorem 14.5 (Conlon–Ferber, Wigderson 2020)

For any fixed $t\geq 2$, we have

$$R(\underbrace{k,k,\dots,k}_{t\text{ times}})\geq 2^{(\frac{3}{8}t-\frac{1}{4}-o(1))k}.$$

Again this bound is exponential in k, but the exponent (which depends on t) is now of this shape. This is better than the previous bound — for $t = 2$ this is the same as the $\sqrt{2}^k$ bound, but for all greater t it's an improvement.

Remark 14.6. This theorem is a bit hard to attribute — a paper by Conlon and Ferber proved the bound for $t = 3$, but had a slightly different bound for $t = 4$ which was a bit weaker, and then said that we can get an improvement for the remaining t by the product construction; this gives an improvement from the previously known bound, but not one as strong as the one we stated above. The above bound was then established by Wigderson, who took the Conlon–Ferber construction but found a better way to generalize to more than t colors, rather than the product construction. A year later, Will Sawin improved the bound further to

$$R(\underbrace{k, \dots, k}_{t \text{ times}}) \geq 2^{(0.383796(t-2)+\frac{1}{2}-o(1))k}.$$

Note that if in place of this decimal we had $\frac{3}{8} = 0.375$, then we'd get the expression as above; since 0.383796 is a bit larger than $\frac{3}{8}$, this is a bit better. To prove this, Sawin took Wigderson's generalization, but kicked out the original part of Conlon–Ferber's construction and replaced it with another random argument. We will prove the bound with $\frac{3}{8}$, but using Will Sawin's approach, and then it won't be so hard to see what we need to do differently to get the slightly better constant (but we won't do this, to avoid carrying around ugly terms).

The argument has two parts. Half of the argument is basically the same as the part that Wigderson added when he generalized Conlon–Ferber the right way to more colors.

A central notion in the proof, which connects the two parts, is the following definition:

Definition 14.7. For a graph G and some $k \geq 2$, we define $q_G(k)$ as the probability

$$q_G(k) = \mathbb{P}[\{v_1, \dots, v_k\} \text{ is an independent set in } G],$$

where we take $v_1, \dots, v_k$ to be independent uniformly random vertices (note that these may not be distinct).

In other words, we choose k independent random vertices in G, and define $q_G(k)$ as the probability that these vertices (which may not be distinct) form an independent set.

The idea of the following lemma is due to Wigderson, though it was phrased in this way by Sawin.

Lemma 14.8 (Wigderson 2020)

For every $k \geq 2$ and $t \geq 2$, and every graph G not containing a clique of size k, we have

$$R(\underbrace{k, \dots, k}_{t \text{ times}}) \geq \left(\frac{1}{q_G(k)}\right)^{(t-2)/k} \cdot 2^{(k-1)/2}.$$

So this gives a lower bound for the Ramsey number we care about from any graph G with no clique of size k. (Note that $q_G(k) < 1$, so taking an exponential of it is a *good* thing.) *A priori* this expression may seem a bit strange, but we'll see that this term comes up naturally from the proof.

Proof. As usual, we'll let

$$n = \left\lfloor \left(\frac{1}{q_G(k)}\right)^{(t-2)/k} \cdot 2^{(k-1)/2} \right\rfloor.$$

Our goal is to find a coloring of the edges of a complete graph on n vertices with t colors and no monochromatic clique of size k; then the Ramsey number must be strictly greater than n, so in particular it must be at least the above quantity. So it suffices to show that the edges of a complete graph with vertex set $\{1, \dots, n\}$ can be colored with t colors in such a way that there is no monochromatic clique of size k.

As you can imagine (since we have a bound in terms of a random experiment), we should take a random coloring; but this random coloring should have something to do with the graph G. The way we want to do this is that we want to in some sense pull back the graph G onto our vertex set $\{1, \dots, n\}$ in a random way, and use this to obtain one color class. We'll do this roughly as many times as there are colors. This means we need some function that maps our vertices to G. This is where randomness comes in — we'll consider random functions that do this.

We consider $t-2$ independent and uniformly random functions $f_1, \dots, f_{t-2} \colon \{1, \dots, n\} \to V(G)$. (We'll see why we have $t-2$ functions soon.) (Each of these functions maps each vertex in $\{1, \dots, n\}$ to a random vertex of G, and everything is uniform and independent.)

Now to define color 1, if we have two vertices i and j such that f_1 maps them to an edge in G, we give the edge ij color 1. This guarantees that color 1 has no cliques of size k, since G has no clique of size k. We define color 2 the same way for f_2 (if this causes an edge to have color 1 and color 2, we can choose which one to use arbitrarily — it doesn't really matter, since if there's an edge where we could use either color 1 or color 2, then we are even more happy). We then define color 3 the same way, and so on.

The reason we don't want to do this with all t colors is that at the end of this random process, we're not sure that every edge has been colored. We could do this up to $t-1$ and give all the remaining edges the last color t, but it turns out it's better to only do it up to $t-2$; and then for all of the remaining edges, we color them randomly with color 1 and 2. (You could also go up to $t-3$ and do the last three randomly, or so on; there isn't a clear reason to see why 2 is the best, but if you try it you'll see that it is.)

More explicitly, for $x, y \in \{1, \dots, n\}$ with $x \neq y$, we color the edge xy as follows:

- If for some $i \in \{1, \dots, t-2\}$, the map f_i maps xy to an edge of G (i.e., if it maps x and y to distinct vertices with an edge between them), then we color xy in color i. (If there are several such i, then we pick one arbitrarily, e.g., we can choose the smallest i or choose randomly.)
- Otherwise, we color xy randomly with either the color $t-1$ or t (with probability $\frac{1}{2}$ each, and independently of all the other random choices).

So our coloring depends on randomness in two different ways — the random functions, and the random choices we're making here.

We now want to show that with positive probability, there is no monochromatic clique of size k — in other words, we need to prove that the probability this coloring has a monochromatic clique of size k is strictly less than 1.

We've already seen we can't have a monochromatic clique of size k in one of the first $t-2$ colors (such a clique would correspond to a clique of size k in G, which by assumption does not exist). This means we only need to bound the probability that there is a monochromatic clique of size k in colors $t-1$ or t.

As usual, we'll look at k particular vertices, and bound the probability that they form a clique. For any given subset of $\{1, \dots, n\}$ of size k, the probability that this given subset forms a monochromatic clique of color $t-1$ or color t can be found as follows — none of the edges between our k vertices can be assigned one of the first $t-2$ colors, so each of our functions $f_1, \dots, f_{t-2}$ must map our k vertices to an independent set in G (since an edge in G would produce an edge in our subset). This has probability exactly $q_G(k)$ of happening for each color, so since they're all independent, this has probability $q_G(k)^{t-2}$. Then we also need to make the same color choice out of $t-1$ and t for all the edges; this gives us a bound of

$$(q_G(k))^{t-2} \cdot 2 \cdot 2^{-k(k-1)/2}.$$

Now we can union-bound over all possible subsets — so the total probability of having a monochromatic clique in color $t-1$ or t of size k is at most

$$\binom{n}{k} \cdot (q_G(k))^{t-2} \cdot 2 \cdot 2^{-k(k-1)/2}.$$

Now we can do some computations — first we have $\binom{n}{k} < \frac{n^k}{k!}$. Now since $k \geq 2$, we have $k! \geq 2$, so the $k!$ cancels the 2, and we get a (strict) bound of

$$n^k \cdot (q_G(k))^{t-2} \cdot 2^{-k(k-1)/2}.$$

Now plugging in our definition of n, we perfectly cancel out both terms and get 1. So our total probability of having a monochromatic clique in colors $t-1$ or t is strictly less than 1; and since we can't have a monochromatic clique in any of the other colors, we're done. □

Remark 14.9. Note that being an independent set is easier if we allow repetition — so the probability $q_G(k)$ is larger than it would be if we demanded $v_1, \ldots, v_k$ to be distinct. This wouldn't make sense though if $k \geq |V(G)|$, which is important because in fact our number n will be much bigger than $|V(G)|$ (though k may not be), so we can't try to force our maps $f_1, \ldots, f_{t-2}$ to be injective.

Now we'll see the second lemma.

Lemma 14.10 (Sawin 2021, Conlon–Ferber 2020)

For every $k \geq 2$, there is a graph G not containing a clique of size k such that

$$q_G(k) \leq 2^{-(\frac{3}{8}-o(1))k^2}.$$

Our goal in the end is to get a lower bound for $R(k, \ldots, k)$, and we have a lower bound in terms of $q_G(k)$. So we'd like a graph G where $q_G(k)^{-(t-2)/k}$ is pretty big, meaning that $q_G(k)$ is pretty small.

Remark 14.11. In this form, the lemma is due to Sawin. However, Conlon–Ferber in their construction basically also gave a graph with this property; but their graph really only has $\frac{3}{8}$, while in fact it is possible to restate the lemma with 0.383796 using Sawin's approach. So who proved the lemma first is up to debate, but the proof we will see is due to Sawin.

We do not have enough time to prove the lemma right now, so we will first prove the theorem from the two lemmas.

Proof of Theorem. Choosing G as in Lemma 2, by Lemma 1 we get that

$$R(\underbrace{k, \ldots, k}_{t}) \geq \left(\frac{1}{q_G(k)}\right)^{(t-2)/k} \cdot 2^{(k-1)/2} \geq 2^{(\frac{3}{8}-o(1))(t-2)k} 2^{(k-1)/2}.$$

Remembering that t is fixed and k is large, the $o(1)$ quantity may depend on t (it's just important that for fixed t it goes to 0 as $k \to \infty$), so we may pull t inside it. We may also absorb $-\frac{1}{2}$ into the $o(1) \cdot k$, giving a bound of

$$2^{(\frac{3}{8}(t-2)+\frac{1}{2}-o(1))k} = 2^{(\frac{3}{8}t-\frac{1}{4}-o(1))k}.$$

(Here we can see that if we improve Lemma 2 to replace $\frac{3}{8}$ with the $0.38\cdots$ constant, then the same calculation will give us Sawin's stronger bound stated above.) □

We will prove the lemma next Tuesday.

(There isn't a strong reason to believe what the true constant in place of $\frac{3}{8}$ should be. This lower bound is nice, but it's still quite different from the upper bound of t^{tk}.)

> **Remark 14.12.** Will Sawin also finds lower bounds for $q_G(k)$ in his paper, though they are pretty far away from this upper bound.

§15 April 18, 2023 — Induced Ramsey Theorem for Graphs

Today we'll first talk about a new topic (since its proof is long and doesn't compartmentalize well); if we have time we'll see the remaining piece from last Thursday, and otherwise we'll see it next Thursday.

We've known Ramsey's theorem for graphs since the first class. Here it's written in a slightly different, but equivalent, form:

> **Theorem 15.1**
> For every graph H there exists a graph G such that the following holds: For every cloring of the edges of G with red and blue, there exists a subgraph of G that is isomorphic to H and such that all of its edges have the same color.

Here H is the target graph. This isn't exactly how we phrased it — we phrased it as 'there exists a number such that if G is a complete graph with that number of vertices, then the theorem holds.' Here we don't gain anything by allowing G to be a non-complete graph, so we might as well assume G is a complete graph (adding additional edges can't hurt us). However, as the title suggests, today we will talk about what happens if we're looking for *induced* subgraphs.

Imagine that we replace 'subgraph' in this theorem with *induced subgraph*. Then it's important to allow ourselves the added flexibility of taking G to be a non-complete graph (because if H is not a clique, then this won't hold if G is a clique). This makes the statement much harder — the statement that $R(k,k)$ exists takes 15 minutes to prove, but the proof we'll see today is more complicated. (Without 'induced' we may as well assume H is a clique and G is a clique, but in this setting if H is a clique then G cannot be a clique. If H is a clique, then this theorem is equivalent to the previous one; but if H is not a clique, then this becomes harder, and it was proven several decades later.)

> **Theorem 15.2** (Deuber, Erdős–Hajnal–Pesa, Rödl 1973)
> For every graph H there exists a graph G such that the following holds: For every cloring of the edges of G with red and blue, there exists an *induced* subgraph of G that is isomorphic to H and such that all of its edges have the same color.

(We're allowed to choose G depending on H. So we want to make sure that G has many induced subgraphs isomorphic to H, arranged in such a way that for every coloring with red and blue, we can still find an induced subgraph in one color.)

The proof in the non-induced version — equivalently, the proof that $R(k,k)$ exists — was by some sort of inductive argument. Our recursion to prove that $R(k,k)$ exists went via also considering off-diagonal Ramsey numbers $R(k,\ell)$ (we needed the added flexibility to look for different clique sizes). We'll do a similar thing here — we'll prove a version of this where we're looking for different induced subgraphs in red and blue.

Theorem 15.3

Given two graphs H_r and H_b, there exists a graph G such that for every coloring of the edges of G with red and blue, at least one of the following holds:

- There exists an induced subgraph of G isomorphic to H_r all of whose edges are red.
- There exists an induced subgraph of G isomorphic to H_b all of whose edges are blue.

Here H_r is the graph we're looking for in red and H_b the graph we're looking for in blue (which is the reason for the indices).

It's pretty clear that this theorem implies the original (since we can take H_b and H_r to both be H). The converse is also true, i.e., the diagonal case also implies the case with distinct graphs (it's closely connected to one of the problems on the homework, so we won't explain it now), but we don't need that here.

§15.1 Proof of the Theorem

We're going to use induction (which is why it's important that we have the non-diagonal version); we induct on $|V(H_r)|+|V(H_b)|$ (similarly to in the non-induced version of Ramsey's theorem, where we induct on the sum of the clique sizes).

First we'll deal with some trivial cases. The theorem is trivial if $|V(H_r)| = 1$ or $|V(H_b)| = 1$, since we can just take G to be a single vertex, so we assume that $|V(H_r)| \geq 2$ and $|V(H_b)| \geq 2$.

If H_r has an isolated vertex v, then this is also easy — we can apply the theorem for $H_r \setminus \{v\}$ and H_b (using the inductive hypothesis, since these graphs have a smaller total number of vertices) to obtain a graph G, and add on an isolated vertex to G. Then we can find an induced copy in G of $H_r \setminus \{v\}$ in red or H_b in blue; in the second case we're done, while in the first case we can add our new isolated vertex to obtain an induced copy of H_r.

The same is true if H_b has an isolated vertex.

(These cases may not be strictly necessary to separate, but they will cause some sets to be empty, so it's clearest to just get rid of them.)

Now for the more interesting cases, we want to apply induction by getting to a graph with 1 fewer vertex. The most natural way to do this is to kick out a vertex; so let's fix a vertex v_r in H_r and v_b in H_b, and let $H_r' = H_r \setminus \{v_r\}$ and $H_b' = H_b \setminus \{v_b\}$. We should also keep track of the neighborhoods of v_r and v_b (in the non-induced version, we added an additional vertex connected to everything — here we need to get one more vertex, but again with the right neighborhood), so let $N(v_r) \subseteq V(H_r')$ be the neighborhood of v_r in H_r — i.e., the set of vertices in H_r which are adjacent to v_r — and define $N(v_b) \subseteq V(H_b')$ similarly, as the neighborhood of v_b in H_b.

We now win if we either find an induced copy of H_r with all edges red, an induced copy of H_b with all edges blue, *or* an induced copy of H_r' in red along with an additional vertex with the right connecting edges to H_r', all of red as well (or the same with blue).

We want to use the inductive hypothesis on (H_r', H_b) and (H_r, H_b') (we don't want to use it on (H_r', H_b'), since this weakens us twice) — by our induction hypothesis, there exists a graph $G(H_r, H_b')$ satisfying the theorem for H_r and H_b', which we call G_1, and similarly there exists a graph $G(H_r', H_b)$ satisfying the theorem for H_r' and H_b.

Our strategy for the proof is to first look at $G_1 = G(H_r, H_b')$, which has the property that when we color it red and blue, we find an induced copy of H_r with all edges red (in which case we're done), or an induced copy of H_b with all edges blue (which is not so good, since we're looking for H_b and not H_b'). It's likely that this graph actually has many different copies of H_b', and our blue copy of H_b' that we found might be

any one of these copies. The strategy of the proof is to modify this graph G_1 step by step (making it more complicated each step) so that with every iteration, we exclude one more possible induced subgraph to be the blue copy (unless we find one of the two objects we're looking for).

Consider our graph G_1; we now want to look at all the induced subgraphs of G_1 isomorphic to H_b' (since these are the possible copies). Let $W_1, \dots, W_n$ be the list of all vertex subsets of $V(G_1)$ such that the induced subgraph $G_1[_W j]$ is isomorphic to H_b'. For each $j \in \{1, \dots, n\}$, choose a particular isomorphism from H_b' to $G_1[W_j]$, and let $W_j' \subseteq W_j$ be the image of $N(v_b) \subseteq V(H_b')$. (So W_j' records which vertices correspond to the neighbors of our particular vertex v_b — the vertices at which we need to attach an additional vertex in order to make H_b.) We want to record this because we want to extend W_j to a copy of H_b, and to do that we need to attach a vertex at W_j'.

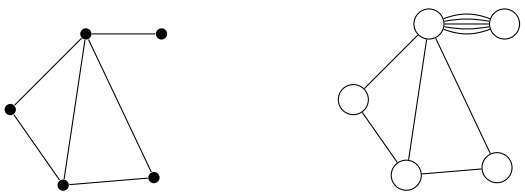
The way we'll prove the theorem is a n-step recursion through all these copies W_j. In order to avoid too many indices, we'll summarize what we want to maintain throughout the recursion in a claim which we'll prove by induction on j. As usual, applying it to the last possible j (here $j = n+1$) will then imply what we want (i.e., that the graph G satisfying the theorem for H_r and H_b exists).

Claim 15.4 — For $j \in \{1, \dots, n+1\}$, there exists a graph G_j and disjoint nonempty vertex subsets $U(v) \subseteq V(G_j)$ for all $v \in V(G_1)$, satisfying the following conditions:

- For distinct vertices $v, v' \in V(G_1)$, we have that uu' is an edge of G_j if and only if vv' is an edge of G_1 for all $u \in U(v)$ and $u' \in U(v')$.
- For any coloring of the edges of G_j with red and blue, at least one of the following holds:
  (a) There is an induced subgraph of G_j isomorphic to H_r with all edges red.
  (b) There is an induced subgraph of G_j isomorphic to H_b with all edges blue.
  (c) For some $k \in \{j, \dots, n\}$, there is an induced subgraph of G_j consisting of exactly one vertex in each of the sets $U(v)$ for $v \in W_k$ such that all edges of this induced subgraph are blue.

So we're looking at the vertices associated with our *original* graph G_1, and for each of them we should have a distinguished subset of G_j corresponding to that vertex. We'll construct this by starting with G_1 and then doing some operations (blowing up certain vertices in certain ways), and the set $U(v)$ will be everything that resulted from blowing up v.

- The first condition says that if we look at two different vertices v and v' in $V(G_1)$ and all their associated vertices $U(v)$ and $U(v')$, the edge relationships between these vertices are exactly the same as the edge relationships between v and v' — so we have a sort of blow-up structure on the associated vertices.



  So between the sets $U(v)$ we either have all edges or no edges (corresponding to the edges in $V(G)$); we don't yet specify what happens inside the sets $U(v)$.
- In the second condition, the first two options are the two we're looking for, and the third is an option depending on j which will be trivially satisfied for $j = 1$ (which gets more restrictive as j increases). This third option says that we can find an induced subgraph consisting of choosing exactly one of the vertices in the subsets corresponding to W_k, such that all edges of this induced subgraph are blue. Such a subgraph has to be isomorphic to H_b' — between the sets $U(v)$ we have the same edge relationships as between the corresponding vertices in G_1, and the vertices in W_k form an induced subgraph in G_1

isomorphic to H_b'. But in fact, the condition is much stronger, because we also have a condition on where it lies and what shape it has.

(We first choose k, which tells us what W_k is. Then W_k is some vertex subset in G_1, and we're saying that for the corresponding blobs in G_j, we choose one vertex from each of these blobs. This forms an induced subgraph which is automatically isomorphic to the induced subgraph on W_k, which is by definition isomorphic to H_b'.)

Of course we have to prove this claim, but once we do this and apply it for $j = n+1$, we're done — we only care about the second condition in the end, and there the first two options mean we've won (we can find an induced subgraph isomorphic to H_r in red or H_b in blue), and the third option cannot hold when $j = n+1$ because there are no indices $k \in \{j, \dots, n\}$. So we're forced to be in (a) or (b), and we are happy.

(You can think of this proof as — the condition (c) talks about all possible options W_k, and we exclude them one by one at each step of the induction. In the end the options are all gone, so option (c) is not viable anymore.)

So our task is now to prove the claim, which we'll do by induction on j.

Proof. We use induction on j. In the case $j = 1$, we take the graph G_1 which we've already defined, and the vertex subsets $U(v) = \{v\}$ for all $v \in V(G_1)$. (We start out with each vertex as its own individual blob.) The first condition is tautologically true. For the second condition, we know that G_1 works in the theorem for H_r and H_b', so in any coloring we will either find an induced copy of H_r with all edges red (in which case we're in case (a) and therefore happy), or an induced copy of H_b' with all edges blue — which must then be one of $W_1, \dots, W_n$, since this is a list of all possible such subgraphs — and therefore (c) is satisfied.

Now suppose that $2 \le j \le n+1$, and that we already have such a graph G_{j-1} for $j-1$. Our goal is to now construct G_j from G_{j-1}.

We'll construct G_j from G_{j-1} in two steps, as follows.

First, let $U = \bigcup_{v \in W_{j-1}'} U(v) \subseteq V(G_{j-1})$ (G_{j-1} has everything we want except hte possibility that our index set in (c) is W_{j-1}; so our goal in the induction is to exclude W_{j-1} as a possibility). We're looking at W_{j-1}' because we'll do something, but we only want the new stuff we do to be connected to W_{j-1}' (since we want a vertex connected to W_{j-1}' in order to form a copy of H_b).

For the first step, we replace every vertex of G_{j-1} in U by a copy of the graph $G(H_r', H_b)$ (so we're taking our graph, and blowing up each vertex to something — and this something is the other graph obtained from the inductive hypothesis that we haven't used yet), and every vertex in $V(G_{j-1}) \setminus U$ stays as is. This defines our new graph. We define the sets $U(v)$ in the way we'd expect — when we blow up a vertex, the new vertices it became are part of the same set $U(v)$. In other words, for every vertex in some set $U(v)$ with $v \in W_{j-1}'$, the entire associated copy of $G(H_r', H_b)$ still belongs to $U(v)$. (So we're not changing the boundaries of our blobs, we're just growing things within the blobs — you can think of these blobs as being gardening plots, where each of us students has a plot and grows vegetables in our plots; these blobs are those plots, and the seeds inside the plot grow into bigger graphs, but never leave the plot.) We call the resulting graph G_{j-1}^*.

Now we'll do the second step. Let m be the number of labelled induced subgraphs of $G(H_r', H_b)$ (i.e., the graph we put into each of the blown-up vertices) isomorphic to H_r'. (This isn't strictly necessary, but it'll give us some intuition — 'labelled' means we're also fixing an isomorphism.) For every $u \in U$, let Φ_u be the collection of all embeddings $\varphi_u \colon H_r' \to G_{j-1}^*$ whose image is an induced subgraph of G_{j-1}^* isomorphic to H_r' that lies in the copy of $G(H_r', H_b)$ corresponding to u in G_{j-1}. This sounds complicated, but it's saying that for each of the vertices which we blew up in the previous step, we look at the number of different ways to map H_r' into the thing we blew it up into (i.e., the small plant that grew here) with the property that this embedding maps H_r' to an induced subgraph isomorphic to H_r' inside this thing. The number of such mappings for every u is precisely m (we won't need this, but to keep things organized in your head, you

can think that there's m different mappings from H_r' into our graph $G(H_r', H_b)$, and since each of our new things is a copy of this graph, we also have m mappings for each u).

We now need to attach these additional vertices — for every choice of a tuple of maps $(\varphi_u)_{u\in U}$ with $\varphi_u \in \Phi_u$ (i.e., over all possibilities to choose a map φ_u for each $u \in U$ — note that this gives us $m^{|U|}$ such tuples), we add an additional vertex to G_{j-1}^* with edges to all vertices in $\bigcup_{u\in U} \varphi_u(N(v_r))$ — i.e., for each of these maps φ_u, we look at the image of our set $N(v_r) \subseteq V(H_r')$ (the set of vertices we want to attach something at to make H_r' into H_r), and we look at these images over all φ_u, and connect an additional vertex to this particular subset. (In particular, this means we add $m^{|U|}$ vertices in total in this step, because there are this many ways of choosing such tuples.)

These newly added vertices are not part of any of our sets $U(v)$.

We now let G_j be the resulting graph (with $U(v)$ defined as above). We need to check the conditions for G_j and the sets $U(v)$.

The first condition (on the blowup-like structure between the sets $U(v)$) is by construction, since in the first step we blew up certain vertices without changing the structure of our sets, and the newly added vertices don't affect this condition.

The more interesting thing is the second condition. For the second condition, we need to consider a coloring of the edges of G_j with red and blue; we need to show that this coloring satisfies at least one of the three conditions (a), (b), or (c).

Let's assume that our coloring doesn't satisfy (a) or (b); we need to show that it satisfies (c).

Recall that in the first step, for every vertex $u \in U$, we replaced u in G_{j-1} by a copy of $G(H_r', H_b)$. We know the coloring of this copy has a red H_r' — by definition it must have an induced copy of H_b with all edges blue, or an induced copy of H_r' with all edges red. The first can't hold by assumption, so the graph must contain an induced subgraph isomorphic to H_r' with all edges red.

This induced subgraph must be of the form $\varphi_u(H_r')$ for some $\varphi_u \in \Phi_u$ (by definition — Φ_u just lists all the possible ways to find such an induced subgraph in this graph isomorphic to H_r').

So for every $u \in U$, we've found such a φ_u; together, these form a tuple. So we now have one special tuple $(\varphi_u)_{u\in U}$ associated to our coloring. For this tuple, we added a particular vertex in the second step; let $x \in V(G_j)$ be the vertex added for this tuple.

Now for every $u \in U$, x together with $\varphi_u(H_r')$ forms an induced subgraph of G_j isomorphic to H_r (since x has exactly the correct adjacencies — the ones corresponding to $N(v_r)$.) Within $\varphi_u(H_r')$, all the edges are already red; if all edges from x to $\varphi_u(N(v_r))$ are red as well, then we get a red induced H_r. So at least one edge from x to $\varphi_u(N(v_r))$ must be blue.

This means for every $u \in U$, there is a vertex $f(u) \in \varphi_u(N(v_r))$ such that the edge from x to $f(u)$ is blue. Meanwhile for $u \in V(G_{j-1}) \setminus U$, let $f(u) = u$.

Because of the way we've chosen these vertices, together the vertices $f(u)$ for $u \in V(G_{j-1})$ form an induced subgraph isomorphic to G_{j-1}. But this subgraph is also colored, so we can apply the second condition on G_{j-1} — if the coloring on G_{j-1} satisfies (a) or (b) then we have a red H_r or blue H_b (which we assumed doesn't happen), so it must satisfy (c) for some $k \in \{j-1, \dots, n\}$. If $k \geq j$ then we are done (since G_j then satisfies (c) for that k). So our only problem is if $k = j-1$.

But if $k = j-1$, then we can look at the induced copy of H_b' as in (c), and add x to it — we can add x because it has the specified blue edges and no other. Then adding x to this blue copy precisely forms a blue copy of H_b, and we are done. □

§16 April 20, 2023

§16.1 Continuation about Multi-Color Ramsey Numbers

Today we'll start by filling in the missing piece of our proof from last week. Recall that we were discussing the following lower bound on multi-color Ramsey numbers — the setup here is that we have a fixed number t of colors and some large k, and we're looking at the diagonal t-color Ramsey number. Conlon–Ferber proved this lower bound; if we plug in $t = 2$ we get the lower bound we already knew of $2^{k/2}$ (which we obtained from the probabilistic method). But for small fixed t (e.g., 3 or 4 or 5), this improves the previously known lower bounds.

Theorem 16.1 (Conlon–Ferber, Wigderson 2020)
For any fixed $t \geq 2$, we have

$$R(\underbrace{k, k, \ldots, k}_{t \text{ times}}) \geq 2^{(\frac{3}{8}(t-2)+\frac{1}{2}-o(1))k}.$$

Will Sawin improved $\frac{3}{8} = 0.375$ to 0.383796.

We saw most of the proof of this result, but we had left the proof of one lemma; we will fill in that proof now.

The lemma doesn't concern the Ramsey numbers anymore. Instead, it concerns the following definition.

Definition 16.2. For a graph G and some $k \geq 2$, we define

$$q_G(k) = \mathbb{P}[\{v_1, \ldots, v_k\} \text{ is an independent set in } G],$$

where $v_1, \ldots, v_k$ are independent uniformly random vertices of G.

Note that since the vertices are independent and uniformly random, so some vertices can be repeated (which means the set could have smaller size than k). The reason for this is that when we apply this lemma to prove our theorem, we can't assume our maps are injective (since we will take a larger graph and map it to a smaller graph); so it's important that we allow repetition.

Last Thursday, we gave a lower bound for the Ramsey number in terms of something depending on $q_G(k)$. We then saw the following lemma; combining these proved the theorem.

Lemma 16.3 (Sawin 2021, Conlon–Ferber 2020)
For every $k \geq 2$, there is a graph G not containing a clique of size k such that

$$q_G(k) \leq 2^{-(\frac{3}{8}-o(1))k^2}.$$

The graph with this property was already established by Conlon–Ferber with an explicit construction; the lemma was only stated as such by Sawin, who replaced $\frac{3}{8}$ with the slightly better constant. Today we'll do the proof with $\frac{3}{8}$, and briefly comment what to adapt to get the larger constant.

We can now forget about the context, and just prove the lemma — our task is to find a graph G without a clique of size k with this upper bound on the probability of having an independent set in our random sampling.

Note that this lemma is not entirely trivial — the fact that there can't be a clique of size k heuristically tells us that there can't be too many edges in G, while upper-bounding the probability of an independent set goes the other way (having more edges would help make the probability smaller). So we have two constraints in opposite directions, and the statement says we can find a graph G that satifies both at the same time.

Proof. We'll construct the graph randomly; for each edge, we'll independently decide or not whether to include it. In order to get $\frac{3}{8}$, we can include each edge with probability $\frac{1}{2}$ (to get the better constant, you need to adjust the probability).

Let $m = \lceil 2^{(k-1)/2} \rceil$ be our number of vertices, and consider a random graph G on m vertices, where every edge is present with probability $\frac{1}{2}$ (independently for all edges). Now we're interested in getting an outcome of this random graph which doesn't have a clique of size k and has this bound on $q_G(k)$ — so we need to deal with two events at the same time (the restriction that G can't have a clique of size k, and the restriction on $q_G(k)$). We only have a certain number of allowed outcomes (ones without a k-clique), and from these, we want to find one with small $q_G(k)$. So it's natural to consider the *expectation* of $q_G(k)$, but only over the allowed G (i.e., the ones without a clique of size k) — this means we want to consider

$$\mathbb{E}[q_G(k) \mid G \text{ has no clique of size } k].$$

Our goal is to show this conditional expectation is at most $2^{-(3/8-o(1))k^2}$; then we can pick one of the good graphs with small $q_G(k)$.

Conditional expectations are hard to deal with, so we'll first remove the conditioning — we have

$$\mathbb{E}[q_G(k) \mid G \text{ has no clique of size } k] \le \frac{\mathbb{E}[q_G(k)]}{\mathbb{P}[G \text{ has no clique of size } k]}.$$

This is a general true fact, as long as the quantity we're considering is always nonnegative — the easiest way to see this is by clearing demoniators, so that the statement then says $\mathbb{E}[q_G(k)] \ge \mathbb{E}[q_G(k) \mid G \text{ has no clique of size } k]\mathbb{P}[G \text{ has no clique of size } k]$. This is the contribution to the expectation from graphs G with no clique of size k, and the contribution from graphs with a clique of size k is nonnegative.

So now it suffices to show that

$$\frac{\mathbb{E}[q_G(k)]}{\mathbb{P}[G \text{ has no clique of size } k]} \le 2^{-(3/8-o(1))k^2}.$$

We first want to lower-bound the probability that G has no clique of size k, which means we want to upper-bound the probability of G having a clique of size k. We can use the union bound — there's $\binom{m}{k}$ possible k-cliques, and for each the probability that it is a clique is $2^{-\binom{k}{2}}$, so we obtain

$$\mathbb{P}[G \text{ has a } k\text{-clique}] \le \binom{m}{k} \cdot 2^{-\binom{k}{2}} \le \frac{m^k}{k!} \cdot 2^{-k(k-1)/2} \le \frac{(2^{(k-1)/2}+1)^k}{k!} 2^{-k(k-1)/2}.$$

The main term cancels out (the $+1$ doesn't contribute much — the $k!$ more than makes up for it), so this is certainly at most $\frac{1}{2}$ if k is large enough. (It might even be $\frac{1}{2}$ for all k, but since we have an $o(1)$, we can just assume k is large.)

So we've shown that the probability that G *has* a clique of size k is at most $\frac{1}{2}$, which means the probability it has *no* clique of size k is *at least* $\frac{1}{2}$. This is good, because it means the denominator is at least $\frac{1}{2}$ and therefore contributes a factor of at most 2. This is eaten by the $o(1)$, so all we have to show is that

$$\mathbb{E}[q_G(k)] \le 2^{-(3/8-o(1))k^2}.$$

Recall that $q_G(k)$ is defined by taking a random experiment where we choose k independent random vertices of G; our outer expectation is over the randomly chosen G, while $q_G(k)$ is also a probability (or expectation) over choosing k random vertices. So we can put our randomness together to make this into one single probability — letting $v_1, \dots, v_k$ be independent uniform random vertices of G, we have

$$\mathbb{E}_G[q_G(k)] = \mathbb{P}_{G,v_1,\dots,v_k}[\{v_1, \dots, v_k\} \text{ is an independent set}].$$

We can first look at the outcomes of $v_1, \dots, v_k$, and suppose that we've chosen ℓ distinct vertices (where $\ell = k$ if they're all distinct, and $\ell = 1$ if they're all the same). Then we can sum over all possibilities of ℓ; this gives us

$$\sum_{\ell=1}^{k} \mathbb{P}[|\{v_1, \dots, v_k\}| = \ell] \cdot 2^{-\binom{\ell}{2}}$$

(where $2^{-\binom{\ell}{2}}$ is the probability that there are no edges between the ℓ chosen vertices).

Now we'll upper-bound this. (Note that heuristically we can't assume that all the vertices are distinct — we're choosing k vertices from roughly $2^{k/2}$. So the probability for anygiven two of the vertices to collide is roughly $2^{-k/2}$. There are roughly k^2 pairs, so in theory the probability for two vertices to collide is $k^2 \cdot 2^{k/2}$. It's true that this happens rarely, but it turns out to not be good enough for the bound.)

In order to bound this probability, we can look at all possibilities of choosing ℓ vertices from our ground set of m vertices, and then find the probability that each of our vertices is mapped into these ℓ; this gives us a bound of

$$\sum_{\ell=1}^{k} \binom{m}{\ell} \cdot \left(\frac{\ell}{m}\right)^k \cdot 2^{-\binom{\ell}{2}} \leq \sum_{\ell=1}^{k} m^\ell \cdot \frac{\ell^k}{m^k} \cdot 2^{-\ell(\ell-1)/2}.$$

(There's an explicit expression which counts this probability, a Stirling number; you could use the asymptotics on that number to get a tighter expression, but we don't need that here.) Here we nicely have a m^ℓ and m^k, which is nice. The term of ℓ^k is a bit annoying, but we can bound $\ell^k \leq k^k$ and pull it out (note that a k^k factor may sound gigantic, but it gets absorbed into the $o(1)$ since $k^k \approx 2^{k \log k}$, and $k \log k = o(k^2)$, so we don't need to worry about it); this gives a bound of

$$k^k \cdot \sum_{\ell=1}^{k} m^{\ell-k} \cdot 2^{-\ell(\ell-1)/2}.$$

Plugging in our definition of m, we get an upper bound of

$$k^k \cdot \sum_{\ell=1}^{k} 2^{(k-1)(\ell-k)/2} \cdot 2^{-\ell(\ell-1)/2}.$$

In our exponents, we can break up the first as $\frac{\ell(k-1)}{2} - \frac{k(k-1)}{2}$. The latter doesn't depend on ℓ, so we can pull it out of the sum; then combining the remaining terms in the exponent gives

$$k^k \cdot 2^{-k(k-1)/2} \cdot \sum_{\ell=1}^{k} 2^{\ell(k-\ell)/2}.$$

We can bound the summation by taking k times the maximum term; we have $\frac{\ell(k-\ell)}{2} \leq \frac{k^2}{8}$, so this is at most

$$k^k \cdot 2^{-\frac{1}{2}k^2 + \frac{1}{2}k} \cdot k \cdot 2^{k/8} = 2^{-(3/8 - o(1))k^2}$$

(the only terms which matter are the ones with k^2 in the exponent, and we have $-\frac{1}{2} + \frac{1}{8} = -\frac{3}{8}$). □

Remark 16.4. Sawin proved a version of the lemma with $\frac{3}{8}$ replaced by $0.383796\dots$ — it turns out that if you choose a probability p a tiny bit less than $\frac{1}{2}$ (namely $0.454997\cdots$, obtained by leaving it as a variable and then optimizing for p in the end), and take $m = p^{-(k-1)/2}$. We'll have p and $1-p$ instead of 2 and 2, so the ideas are the same, but the calculations become more complicated.

Remark 16.5. Conlon–Ferber found an explicit graph with this property. We'll describe the graph, but not prove that it has the property — they took the graph G with vertex set $\{v \in \mathbb{F}_2^k \mid v \cdot v = 0\}$ (i.e., all isotropic vectors in $\mathbb{F}_2^k$) and edges given by vw if $v \cdot w = 1$. We can check that this graph doesn't have a clique of size k (this is a linear algebra argument similar to the odd-town argument taught in Prof. Sauermann's course last year), and that the graph happens to have the property on $q_G(k)$ (they don't write it down this way, but the idea is the same). This gives an explicit construction for $\frac{3}{8}$, but it only gives $\frac{3}{8}$; for the improved bound, you probably have to use a probabilistic construction.

Remark 16.6. Sawin also showed in his paper that for every graph G not containing a clique of size k, we have

$$q_G(k) \geq \left(\frac{3\sqrt{3}}{2}\right)^{-k^2} \approx 2^{-1.37744k^2}.$$

This number is quite far from $0.38\cdots$ — so we don't know what the correct constant is — but it shows that the rough behavior is 2^{-ck^2}.

How is this possible? The densest we can make a graph without having a clique of size k is essentially $1 - \frac{1}{k}$ (by Turan's theorem). However, the graphs which are this dense but don't have a clique of size k have pretty strong structure — they have huge independent sets, which cause large $q_G(k)$. So it's not a contradiction to have a graph with pretty high density but still large $q_G(k)$.

§16.2 Ramsey Graphs

Our next topic is Ramsey graphs. (We have seen a bunch of topics which are permutations of the words Ramsey, graph, and number, so it's maybe surprising there's a permutation that we haven't seen yet, but there is.)

In the first week of class, we saw the following theorem (in slightly different phrasing):

Theorem 16.7 (Erdős–Szekeres)

Every graph on n vertices contains a clique or an independent set of size at least $\frac{1}{2}\log_2 n$.

(We didn't state it in this form, but it's true because we can color the edges red and the non-edges blue; then this says that for every complete graph on n vertices with edges colored red or blue, we have a monochromatic clique of this size. We also usually stated the theorem in terms of an upper bound on Ramsey numbers instead; but these are equivalent — if $k = \lceil \frac{1}{2}\log_2 n \rceil$, then we saw that $R(k,k) \leq \binom{2k-2}{k-1} \leq 2^{2k-2} \leq n$. So since $R(k,k) \leq n$, if we take the described coloring then we must have a monochromatic clique of size k.)

Remark 16.8. The constant $\frac{1}{2}$ stood for about 80 years, until it was improved about a month ago, by a recent result by CGMS; we'll talk about this more in two weeks.

In the other direction, we also discussed the following theorem:

Theorem 16.9 (Erdős 1947)

For $n \geq 3$, there exist n-vertex graphs with no clique or independent set of size at least $2\log_2 n$.

(This is because taking $k = \lfloor 2\log_2 n \rfloor$, we saw that $R(k,k) > 2^{k/2} \geq n$, so n isn't good enough for Ramsey's theorem for k; this means there must be a red-blue coloring with no monochromatic clique, giving a graph with no clique or independent set of size k.)

We've known both statements since the first or second week of the class (given a graph on n vertices, we're guaranteed to find a clique of size $\frac{1}{2}\log_2 n$, but we can't guarantee a bound of $2\log_2 n$).

The proof of the latter relied on a probabilistic method (it was one of the first uses of the probabilistic method, and it very much popularized probabilistic methods). Even after many decades and a lot of attention, we still don't know any non-probabilistic proofs of this — in other words, we don't know any explicit construction of such a graph. (We know they exist, and even that most graphs have this property, but we can't name an explicit one with this property.) This is intensively studied in computer science as well (there's a connection to randomness extraction), so a lot of people have tried very hard, with no success. In fact, there isn't even success if we replace 2 with any constant (replacing 2 with a bigger number, e.g., 10 or 1000, makes the task easier) — in other words, no explicit constructions of such graphs are known, even replacing 2 by some other constant $C > 2$.

Definition 16.10. For fixed $C > 0$, an n-vertex graph is called a *C-Ramsey graph* if it does not contain a clique or independent set of size at least $C\log_2 n$.

In other words, a C-Ramsey graph is a graph with this property where we replace 2 by C. The property gets weaker and weaker as we increase C. We can restate the second theorem as saying that there exist 2-Ramsey graphs (for $n \geq 3$). This means we know there exist C-Ramsey graphs for all $C \geq 2$, but we don't know of any. (We also know they don't exist for $C < \frac{1}{2}$, but we should think of C as large.)

Remark 16.11. What do we mean by 'explicit' construction? Obviously if we write down the graph on a paper then it's explicit, but that doesn't make sense for a construction for large n. Most definitions say that there's an algorithm which can describe our graph in a reasonable amount of time. There are also number-theoretic things — for example, take the vertices of the graph to be $\mathbb{F}_p$, and draw an edge if and only if a number-theoretic property holds; these number-theoretic properties may be hard to check. For some people this would count as explicit; for others the notion of 'explicit' depends on the runtime of calculating whether something is an edge. But a reasonable-time algorithm constructing such a graph would count as explicit in nearly all definitions — in particular, we don't have any.

It's known for $C \geq 2$ almost all graphs will be C-Ramsey, but we can't name any; for this reason, we'd like to understand what C-Ramsey graphs look like in order to get a grasp of the unstructuredness that C-Ramsey graphs must have — the heuristic is that they have to be sort of random-looking, which is why it's heuristically hard to find them — since something we write down will have lots of structure.

People have proven various results, and there have been various conjectures about these graphs (many of which have been resolved). In particular, here's a very nice theorem about these C-Ramsey graphs.

Theorem 16.12 (Erdős–Szemeredi 1972)

For every $C > 0$, there exists $\varepsilon > 0$ such that for every sufficiently large n, every C-Ramsey graph on n vertices has density at least ε and at most $1 - \varepsilon$.

Roughly speaking, this theorem says that C-Ramsey graphs can't be extremely sparse or extremely dense — so their density is bounded away from 0 and 1.

A priori this sounds like an extremely weak statement, since it sounds like it tells us nothing about the graph except the overall density. But in fact this statement is much stronger than it looks — consider a C-Ramsey graph on n vertices, and pick any $\sqrt{n}$ vertices. Then on these $\sqrt{n}$ vertices we *still* don't have a clique or independent set of this size, and so we get a $2C$-Ramsey graph on those vertices (as $C\log_2 n = 2c\log_2\sqrt{n}$). This means we can apply this theorem on those $\sqrt{n}$ vertices. So we can apply this theorem not just to the graph itself, but also to all induced subgraphs with a reasonable number of vertices. This gives us a very strong statement about the distribution of edges — we can't have any reasonably large subset which is very sparse or very dense.

Note that 'at least ε' and 'at most $1-\varepsilon$' are equivalent, since the complement of a C-Ramsey graph is also C-Ramsey; so in order to prove this, we only have to prove 'at least ε' (and the other direction follows automatically).

We will not start the proof yet (since it won't fit in 10 minutes). But we will see some preparation for the proof of this.

§16.3 Binary Entropy

Definition 16.13. For $0 \le x \le 1$, we define the *binary entropy* of x by

$$H(x) = x \log \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$$

if $0 < x < 1$, and $H(x) = 0$ if $x \in \{0, 1\}$.

This defines a function; we don't have to worry about why it is called entropy.

Then H is a continuous function on $[0,1]$ (we can check that the limits of the first expression as $x \to 0$ or $x \to 1$ are indeed 0), and it is monotone increasing on $[0, \frac{1}{2}]$ and monotone decreasing on $[\frac{1}{2}, 1]$.

When we plug in $\frac{1}{2}$, we obtain $H(\frac{1}{2}) = 1$; so then $H(x) \in [0,1]$. We can also see that H is symmetric about $\frac{1}{2}$, and that it is concave (we won't need this for the argument we'll make, but it's sometimes useful).

The relevance of this is the following — for any $0 \le k \le n$, we can use entropy to bound binomial coefficients as

$$\binom{n}{k} \le 2^{H(k/n) \cdot n}.$$

The most natural way to think about this is that if $k = \frac{n}{2}$, then we know $\binom{n}{n/2}$ is asymptotically roughly the same as 2^n, which is what we get here — but at some point, you might ask yourself what is $\binom{n}{n/3}$ or $\binom{n}{n/4}$ or so on? That's exactly what the entropy function tells us. (This bound is asymptotically tight — the same lower bound holds up to $o(n)$ in the exponent.) So if we ever have a combinatorial argument where we run into e.g., $\binom{n}{n/10}$, bounding it by 2^n is quite wasteful; this gives the correct bound.

We'll now prove this fact.

Proof. Stirling's formula tells us that

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

(we generally don't care about the constant factors). Then we can write

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \le \frac{e\sqrt{n}\left(\frac{n}{2}\right)^n}{\sqrt{2\pi} \cdot \sqrt{k} \cdot \left(\frac{k}{e}\right)^k \cdot \sqrt{2\pi} \cdot \sqrt{n-k} \cdot \left(\frac{n-k}{e}\right)^{n-k}}.$$

Conveniently $e < 2\pi$, $\sqrt{n} \leq \sqrt{k(n-k)}$, and the exponentials of e cancel out; so we get an upper bound of

$$\binom{n}{k} \leq \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} = 2^{H(n/k)\cdot n}. \qquad \square$$

This is how the terms $H(x)$ naturally appear — they're intimately connected to binomial coefficients.

§17 April 25, 2023 — Ramsey Graphs

Today we'll continue talking about Ramsey graphs.

Definition 17.1. For fixed $C > 0$, a n-vertex graph is called a C-Ramsey graph if it does not contain a clique or independent set of size at least $C \log_2 n$.

The larger we make C, the looser the condition gets — so the easier it is to find a C-Ramsey graph. Still, there's no fixed C for which we have explicit constructions of C-Ramsey graphs (for large n) — as discussed last time, we know C-Ramsey graphs exist for all $C \geq 2$ (and $n \geq 3$), by the probabilistic argument we saw in the first or second week of class; but no explicit constructions are known.

We want to think of C-Ramsey graphs as the graphs which behave sort of optimally for Ramsey's theorem, up to a constant factor C.

Note that a graph G is C-Ramsey if and only if its complement is C-Ramsey (this is a very simple but useful observation).

§17.1 Density Bounds on Ramsey Graphs

Last time, we stated (but didn't prove) the following theorem:

Theorem 17.2 (Erdős–Szemerédi 1972)

For every $C > 0$, there exists $\varepsilon > 0$ such that for all sufficiently large n, every C-Ramsey graph on n vertices has density at least ε and at most $1 - \varepsilon$.

The point is that for every fixed C, the C-Ramsey graphs have density bounded away from 0 and 1 (where the bounds only depend on C). *A priori* this statement doesn't seem so strong — it basically just gives us a statement about the number of edges — but as mentioned last time, the true strength of this theorem comes from applying this not just to the C-Ramsey graph itself, but to all induced subgraphs on sufficiently many vertices. For example, we can apply it to each induced subgraph on at least $n^{0.1}$ vertices; such graphs are $10C$-Ramsey graphs, and applying the theorem to them gives us a statement about the densities in these subgraphs — so we really get a result about the distribution of edges.

Last class, we defined teh entropy function

$$H(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x},$$

and saw that for any $0 \leq k \leq n$, we have

$$\binom{n}{k} \leq 2^{H(k/n) \cdot n}.$$

(This is what we'll use it for.)

Proof of Theorem. Let $0 < \varepsilon < \frac{1}{16}$ be such that

$$H(8\varepsilon) < \frac{1}{4C}.$$

(Given some C, we can always find such an ε — $\frac{1}{4C}$ is a small positive number, so we can make ε small enough that $H(8\varepsilon)$ is even smaller.)

We need to show that in any C-Ramsey graph on n vertices, the density is at least ε and at most $1-\varepsilon$. First it suffices to show that the density is at least ε; then we automatically get that it's at most $1-\varepsilon$ by applying this to the complement (the complement is also C-Ramsey, so the density of the complement will be at least ε, so the original graph will have density at most $1-\varepsilon$).

Suppose that G is a C-Ramsey graph on n vertices where n is sufficiently large — more precisely, such that $C\log_2 n \leq \frac{n}{8}$ and $n \geq 2^{12}$ — and assume for contradiction that the density of G is less than ε. (We'll see later in the proof where we need these size conditions.)

Right now, we know that G is a C-Ramsey graph with density less than ε. If the density of G is less than ε, then its average degree is less than εn (since the density can be calculated by averaging the degrees and dividing by $n-1$). This means at most half the vertices — i.e., at most $\frac{n}{2}$ vertices — can have degree at least $2\varepsilon n$. These vertices are bad because they have abnormally large degree; as usual, we'll just kick them out. Let $U \subseteq V(G)$ be the set of vertices of G with degree at most $2\varepsilon n$, so that $|U| \geq \frac{n}{2}$.

We now want to use the fact that G is C-Ramsey — so there is no large clique or independent set. Since we have a graph of low density, the most naive thing to do is to try to find a large independent set — so let's try to find a large independent set in U. *A priori* it's not clear how we'll do this, but the idea is to pick a maximum-sized independent set in U and analyze it, and somehow get a contradiction to its maximality.

Let $I \subseteq U$ be an independent set inside U of maximum size. Then $|I| < C\log_2 n$ (we know G is C-Ramsey, so all independent sets in G have size smaller than $C\log_2 n$, and I is an independent set in G). In particular, this implies $|I| \leq \frac{n}{8}$ (this bound is very loose, but we'll need it in a moment).

We'll now analyze the edges between I and $U \setminus I$ — we have

$$e(I, U\setminus I) \leq 2\varepsilon n \cdot |I|\,.$$

(because we can upper-bound this quantity by $\sum_{v\in I}\deg(v)$ — every vertex $v \in I$ contributes at most $\deg(v)$ edges — and $I \subseteq U$, so by definition every vertex in I has degree at most $2\varepsilon n$). This means a typical vertex in $U \setminus I$ can't have too many edges to I (since $U \setminus I$ has on the order of n vertices) — more precisely, there are at most $\frac{n}{4}$ vertices in $U \setminus I$ that have at least $8\varepsilon\,|I|$ edges to I (since if there were more, then they'd contribute a total of more than $\frac{n}{4}\cdot 8\varepsilon\,|I| > 2\varepsilon\,|I|$ edges between them). As before, we will kick them out — let $W \subseteq U \setminus I$ be the set of vertices in $U \setminus I$ with at most $s = \lfloor 8\varepsilon\,|I|\rfloor$ edges to I. Then

$$|W| \geq |U| - |I| - \frac{n}{4} \geq \frac{n}{2} - \frac{n}{8} - \frac{n}{4} = \frac{n}{8}$$

(this is what we needed the loose bound of $|I| \leq \frac{n}{8}$ for).

So far, we've pruned out the high-degree vertices of G, then looked at a maximal-size independent subset I and pruned away the vertices outside it that had abnormally many edges into I. Now we have a set W on $\frac{n}{8}$ vertices, where each only has a few edges to I — so it has a neighborhood in I which is at most an 8ε-fraction of I. Then we can analyze these neighborhoods using the binomial coefficient bound, which bounds the number of possible neighborhoods.

For every vertex $w \in W$, we choose a subset $S(w) \subseteq I$ of size s, such that $N(w)\cap I \subseteq S(w)$ — this is possible because W has at most s neighbors in I. (If w has exactly s neighbors, then we define $S(w)$ to be the set of these s neighbors; if it has fewer, then we define $S(w)$ to be a set of size s consisting of those neighbors and add extra vertices for padding. Note that $s < |I|$, so this is possible.)The point is that we want to say there are many vertices in W which have the same set $S(w)$, since there aren't too many possibilities for $S(w)$.

The total number of subsets of I of size s is

$$\binom{|I|}{s} \leq 2^{H(|I|/s)\cdot|I|} \leq 2^{H(8\varepsilon)\cdot|I|}$$

(we defined s so that $\frac{|I|}{s} \leq 8\varepsilon$, and by the monotonicity of the entropy function — since we defined $\varepsilon < \frac{1}{16}$, so $8\varepsilon < \frac{1}{2}$ — we can use this to upper-bound the entropy). We defined ε so that $H(8\varepsilon) < \frac{1}{4C}$, so then

$$\binom{|I|}{s} \leq 2^{\frac{1}{4C} \cdot C \log_2 n} = n^{1/4}.$$

Now for every vertex in W, we've associated to it a set of size s in I, and there's at most $n^{1/4}$ possibilities for this set, so many vertices in W must have been associated to the same set — there is a subset $W' \subseteq W$ and a subset $S \subseteq I$ of size s such that

$$|W'| \geq \frac{|W|}{n^{1/4}} \geq \frac{n}{8n^{1/4}} = \frac{n^{3/4}}{8}$$

and $S(w) = S$ for all $w \in W'$. (The point is that there's at least $\frac{n}{8}$ vertices in W and each gets associated a set $S(w)$, but there's at most $n^{1/4}$ choices; so at least $\frac{n}{8n^{1/4}}$ vertices must have the same set $S(w)$, and we call that same set S, and let W' be the set of these vertices w with that set.) Then all edges from W' to I have their endpoints in S — so there are no edges from W' to $I \setminus S$.

Claim — W' does not contain an independent set of size $s+1$.

Proof. Suppose that $J \subseteq W'$ is an independent set of size $s+1$. Then $(I \setminus S) \cup J$ — is an independent set — $I \setminus S$ is an independent set (because I is), J is an independent set (by assumption), and there are no edges between $I \setminus S$ and J because $J \subseteq W'$ and there are no edges between $I \setminus S$ and W'. But $(I \setminus S) \cup J$ has size

$$|I| - s + (s+1) = |I| + 1.$$

But it is also a subset of U (since all relevant sets occurring here are subsets of U — the vertices we kicked out at the beginning are already gone); this contradicts the choice of I as an independent set in U of maximum size. □

Now we have a bound on the independence number of W'. Meanwhile, we also know that W' does not contain a clique of size at least $C \log_2 n$ (by the Ramsey graph property — since G is a C-Ramsey graph, it doesn't have a clique of size $C \log_2 n$). Let $t = \lceil C \log_2 n \rceil - 1$; then we know that W' does not contain a clique of size $t+1$. Also note that $|I| \leq t$ (since $|I| < C \log_2 n$).

Now we have an upper bound for $|I|$, the cliques in W', and the independent sets in W'. We'd first like to combine the upper bounds for the sizes of cliques and independent sets in W' — so we can use Ramsey's theorem to obtain an upper bound on W', which gives

$$|W'| \leq R(s+1, t+1) \leq \binom{s+t}{s} \leq 2^{H(s/(s+t))(s+t)}$$

(since W' is a graph with no independent set of size $s+1$ or clique of size $t+1$). We have

$$\frac{s}{s+t} \leq \frac{s}{t} \leq \frac{8\varepsilon\,|I|}{t} \leq 8\varepsilon$$

(using the fact that $|I| \leq t$), so then this gives

$$|W'| \leq 2^{H(8\varepsilon) \cdot 2t}$$

(since $s \leq t$ — note that t is approximately $C \log_2 n$, while s is around $8\varepsilon\,|I| \leq 8\varepsilon t$). Then plugging in our bounds, we obtain

$$|W'| \leq 2^{\frac{1}{4C} \cdot 2C \log_2 n} \leq \sqrt{n} < \frac{n^{3/4}}{8} \leq |W'|,$$

which gives a contradiction. □

Remark 17.3. Note that we had two assumptions — G doesn't have a large clique or independent set — and we wanted to prove that the density of G is at least ε (so we wanted to exclude the case that the graph is very sparse). You might first think that in order to do this, it only matters that there is no large independent set (since you'd naively expect a very sparse graph to have a large independent set) — i.e., that just having no large independent set would imply density at least ε. But this is not true — our proof did use the fact that there's no large clique. In fact, this shouldn't be that surprising — by Turan's theorem, in order to show that the density is at least ε just using independent sets, you'd need to exclude independent sets of a much smaller size.

§17.2 Ramsey Graphs and Induced Subgraphs

The next thing we'll discuss is the following theorem, proving a conjecture of Erdős and Hajnal (which was open for around a decade).

Theorem 17.4 (Prömel–Rödl 1999)

For every $C > 0$, there exists $\lambda > 0$ such that every C-Ramsey graph on n vertices contains every graph on at most $\lambda \log_2 n$ as an induced subgraph.

The intuition is that it's hard to find explicit constructions of C-Ramsey graphs because we heuristically believe they should be random-like or very unstructured. This is sort of a concrete incidence of this unstructuredness — such graphs must have induced subgraphs of all shapes (as long as they're not too big).

In some sense this should be very surprising — the definition of a C-Ramsey graph is that it does not contain a clique or independent set of size at least $C \log_2 n$. But the conclusion is on subsets of the same order of magnitude (just changing C to λ) — and looking at hose subsets, we can find *everything*. (We can definitely find cliques or independent sets, which is maybe not that surprising but is definitely *somewhat* surprising because Ramsey's theorem tells us we can find *either* a clique or independent set, but this tells us we can find both; and anything else we might wish for.)

We'll now prove this. The first ingredient in the proof is a modification of a lemma that we saw before. First, here's the form of the lemma we saw before (in the proof of the bound for Ramsey numbers of bounded-degree graphs — that $R(H, H)$ is linear in the number of vertices of H if H has bounded degree).

Lemma 17.5

Suppose that $0 \leq \rho \leq \frac{1}{2}$ and $0 \leq \delta \leq \frac{1}{4}$, and let $s \geq \frac{1}{\delta}$ be an integer. Then for every graph G, at least one of the following two statements holds:

(a) There is a vertex subset $U \subseteq V(G)$ of size $|U| \geq (\frac{\rho}{2})^s |V(G)|$, such that for all pairs of disjoint subsets $X, Y \subseteq U$ of size $|X| \geq \rho |U|$ and $|Y| \geq \rho |U|$, we have $d(X, Y) \geq \delta$.

(b) There is a vertex subset $U \subseteq V(G)$ of size $|U| \geq (\frac{\rho}{2})^s |V(G)|$ with density $d(U) \leq 3\delta$.

Last time, for (a) we just said that there's a vertex subset U such that the induced subgraph on U is bi-(ρ, δ)-dense; the condition written here is just the definition of what it means to be bi-(ρ, δ)-dense, which we won't repeat since we'll modify it shortly.

This isn't exactly the statement we need — (a) only tells us that the density between pairs is *at least* something, but if we want to find induced subgraphs, we also want that there are not too many edges (since we need both non-edges and edges to behave correctly). So we'll modify it as follows:

Lemma 17.6

Suppose that $0 \le \rho \le \frac{1}{2}$ and $0 \le \delta \le \frac{1}{4}$, and let $s \ge \frac{2}{\delta}$ be an integer. Then for every graph G, at least one of the following two statements holds:

(a) There is a vertex subset $U \subseteq V(G)$ of size $|U| \ge (\frac{\rho}{2})^s |V(G)|$, such that for all pairs of disjoint subsets $X, Y \subseteq U$ of size $|X| \ge \rho |U|$ and $|Y| \ge \rho |U|$, we have $\delta \le d(X,Y) \le 1-\delta$.

(b) There is a vertex subset $U \subseteq V(G)$ of size $|U| \ge (\frac{\rho}{2})^s |V(G)|$ with density $d(U) \le 3\delta$ or $d(U) \ge 1-3\delta$.

In (a) we added an upper bound of $1-\delta$. We have to pay for this — first, we change the bound on s by a constant factor (which doesn't really matter). We also get another exit case where the density of U is extremely high; this gives the second condition in (b). So this is a more symmetric version in terms of edges and non-edges than the version we had before — we basically took all edge conditions and added the analogous non-edge conditions. (Note that (a) gives both bounds at the same time, while (b) has an 'or.')

As you'd expect, we can prove this similarly to how we proved the lemma without the modification; this will be on the homework (we have to tweak the proof a little bit to get this — Prof. Sauermann doesn't see a way to black-box the previous theorem and apply it twice on edges and then non-edges because the parameters don't work out if we apply it twice, but if we find one then we can).

We'll now use this to prove the theorem (or at least, to begin the proof).

Proof of Theorem. The proof will apply the theorem of Erdős–Szemerédi. First note that we can assume n is sufficiently large with respect to C — if we are able to prove the theorem for all n which are at least some threshold only depending on C, then we get some λ which works for all n above that threshold. There are only finitely many n below that threshold, and for all of them the statement trivially holds if we take λ to be around $\frac{1}{\log_2 n}$ — so we can just artificially lower λ so that $\lambda \log_2 n \le 1$ for all such n, and the theorem will hold for them as well.

Now let G be a C-Ramsey graph on n vertices; we will assume that n is large enough (in order to satisfy the threshold for the Erdős–Szemerédi theorem as well as a few other inequalities we'd want).

The theorem tells us something about the density of G, but that's rather weak (it just tells us something about the total number of edges). But as mentioned earlier, the true power of the theorem lies in applying it to smaller induced subgraphs — so we will apply it to induced subgraphs of size at least $\sqrt{n}$. Note that every induced subgraph of G on at least $\sqrt{n}$ vertices is a $2C$-Ramsey graph (it still can't have a clique or independent set of size at least $C \log_2 n$, and $C \log_2 n = 2C \log_2 \sqrt{n}$, so if it has $m \le \sqrt{n}$ vertices then it can't have a clique or independent set of size $2C \log_2 m$).

Now let $\varepsilon > 0$ be such that the previous theorem holds for $2C$-Ramsey graphs (with sufficiently many vertices). Then every induced subgraph of G on at least $\sqrt{n}$ vertices must have density at least ε and at most $1-\varepsilon$.

Now let $\delta = \frac{\varepsilon}{4}$ and $s = \left\lceil \frac{2}{\delta} \right\rceil$ (so that $3\delta < \varepsilon$ and $s \ge \frac{2}{\varepsilon}$); note that ε, δ, and s only depend on C (and not on n). Now we can apply the above lemma with these values of δ and s, and some ρ — let $\rho = 2n^{-1/(2s)}$, so

$$\left(\frac{\rho}{2}\right)^s = n^{-1/2}.$$

Then applying the lemma to our C-Ramsey graph G, we know that either (a) or (b) must hold. But (b) is not possible, since we've already shown that all vertex subsets $U \subseteq V(G)$ of size $|U| \ge \sqrt{n} = (\frac{\rho}{2})^s n$, we have

$$3\delta < \varepsilon \le d(U) \le 1-\varepsilon < 1-3\delta.$$

So then option (b) is impossible, which means option (a) must happen, and we obtain a vertex subset $U \subseteq V(G)$ of size $|U| \ge \sqrt{n}$ such that for all pairs of disjoint subsets $X, Y \subseteq U$ of size $|X| \ge \rho |U|$ and

$|Y| \geq \rho |U|$, we have $\delta \leq d(X,Y) \leq 1-\delta$ — so we've found a very nice vertex subset of size at least $\sqrt{n}$, such that between any two reasonably sized subsets we have density between δ and $1-\delta$. Next time we'll use this to embed all graphs of size at most $\lambda \log_2 n$ in this subset. □

§18 April 27, 2023 — Ramsey Graphs

Today we'll finish the proof of the theorem about Ramsey graphs we saw last time. Recall the definition of C-Ramsey graphs:

Definition 18.1. For fixed $C > 0$, a n-vertex graph is called a *C-Ramsey graph* if it does not contain a clique or an independent set of size at least $C \log_2 n$.

In other words, a C-Ramsey graph behaves optimally with respect to Ramsey's theorem, up to a constant factor.

As we saw last time, Ramsey graphs have a number of properties; the following one is quite surprising.

Theorem 18.2 (Prömel, Rödl 1999, proving a conjecture of Erdős, Hajnal)
For every $C > 0$, there exists $\lambda > 0$ (depending only on C) such that every C-Ramsey graph on n vertices contains every graph on at most $\lambda \log_2 n$ vertices as an induced subgraph.

So locally our Ramsey graph has all possible shapes. It's surprising that the assumption and conclusion are both on vertex subsets of size $\log_2 n$, though with different constants (it would be less surprising if we had something of much lower order, but $\lambda \log_2 n$ is really quite large given our assumption).

Today we'll continue the proof of this result.

Proof. We may assume that n is sufficiently large with respect to C. Let G be a C-Ramsey graph on n vertices, so we want to show we can find every graph on $\lambda \log_2 n$ as an induced subgraph.

Last time, we defined $\varepsilon > 0$ (which had to do with the earlier result from Erdős–Szemeredi for $2C$-Ramsey graphs), $\delta > 0$ (as $\varepsilon/4$), and $s \in \mathbb{N}$ (as $\lceil 2/\delta \rceil$) only depending on C, and we defined $\rho = 2n^{-1/(2s)}$ (which depends on both n and C).

We proved that in our C-Ramsey graph G, there is a vertex subset $U \subseteq V(G)$ of size $|U| \geq \sqrt{n}$ such that U has the following nice property: for all pairs of disjoint subsets $X, Y \subseteq U$ with $|X| \geq \rho |U|$ and $|Y| \geq \rho |U|$, we have $\delta \leq d(X,Y) \leq 1-\delta$. In other words, whenever we find disjoint X and Y which are not too small (each is at least a ρ-fraction of U), we can't have too extreme of a density between them. You can imagine such a property is very helpful in order to embed things as induced subgraphs — for induced subgraphs we need to guarantee both edges and non-edges, and if between any two reasonably large subsets we have a reasonable number of both edges and non-edges, it seems reasonable that we can succeed at our embedding task.

Now we'll choose λ — let $\lambda > 0$ be chosen (only depending on C) such that $\lambda \log_2 \frac{1}{\delta} < \frac{1}{3s}$. Since δ and s only depend on C, then λ only needs to depend on C as well (both sides are positive numbers, so we can make λ small enough so that this inequality is true). Observe that then

$$\delta^{\lambda \log_2 n} = n^{-\lambda \log_2(1/\delta)} \geq n^{-1/(3s)}.$$

(In particular, this is quite a bit bigger than $\rho = n^{-1/(2s)}$.)

Our task is to show that every graph on at most $\lambda \log_2 n$ vertices can be found as an induced subgraph in G, so let H be a graph with $m \leq \lambda \log_2 n$ vertices, and let the vertices of H be named $v_1, \ldots, v_m$. Our

goal is to embed the vertices $v_1, \dots, v_m$ into G such that both edges and non-edges are preserved. The idea is to do this greedily in the nice subset U we found earlier — we'll greedily map $v_1, \dots, v_m$ to distinct vertices $w_1, \dots, w_m$ in U, such that their images form an induced subgraph isomorphic to H. The following claim will let us do this — we'll inductively embed the vertices one by one, and the claim will captre the properties we'll keep track of during the process; then taking $j = m$ (the last step of the process) will give the desired embedding.

As we embed vertices one by one, we'll look at the candidates we have left for the remaining vertices; as earlier, we want to make sure at every step that the size of these candidates don't go down too much.

Claim — For every $0 \le j \le m$, there exists distinct vertices $w_1, \dots, w_j$ in U and disjoint subsets $W_{j+1}, \dots, W_m$ of $U \setminus \{w_1, \dots, w_j\}$, each of size at least $\frac{\delta^j}{2m} \cdot |U|$, such that:

- the adjacencies between $w_1, \dots, w_j$ are the same as between $v_1, \dots, v_j$ in H (i.e., $w_k w_\ell$ is an edge if and only if $v_k v_\ell$ is an edge).
- For every $i = j+1, \dots, m$, every vertex in W_i has the same adjacencies to $w_1, \dots, w_j$ as v_i has to $v_1, \dots, v_j$ in H.

Here $W_{j+1}, \dots, W_m$ are our candidate sets; we need to exclude the previously chosen vertices from them, and we want to make sure that each of these candidate sets are big enough. The second condition captures what it means to be a suitable candidate set — for each single candidate set, all its vertices should have the right adjacencies to the already chosen vertices. (We don't have assumptions on the edges *between* the candidate sets.) So for example, if v_i has an edge to v_2, then every vertex in W_i should have an edge to w_2.

If we prove this claim, then we're done — taking $j = m$, this tells us that there are vertices $w_1, \dots, w_m$ with the same adjacencies as $v_1, \dots, v_m$, and we're done (since this forms an induced copy of H).

Before we prove the claim, we'll do the following side calculation (which will come up repeatedly in the proof) — note that

$$\frac{\delta^m}{(2m)^2} \ge \frac{\delta^{\lambda \log_2 n}}{(2\lambda \log_2 n)^2} \ge \frac{n^{-1/(3s)}}{(2\lambda \log_2 n)^2}.$$

(this comes from calculating how small the sets become, with an exxtra factor of $2m$ thrown in). Recall that λ only depends on C and not n, so the n-dependence is $n^{-1/(3s)}/(\log n)^2$; in particular, if n is sufficiently large, then this is greater than $n^{-1/(2s)} = \rho$ (since $\frac{1}{3s} < \frac{1}{2s}$). So the point is that these set sizes are all at least ρ, even if we introduce an extra factor of $2m$; this is important because our property on U only applies to sets of size at least $\rho\,|U|$, and this calculation will ensure we can apply it to our relevant sets.

Proof of Claim. We'll use induction on j. The base case is when $j = 0$. Then we don't need to specify any vertices, and we just need to specify m disjoint subsets of U — so we can just take $W_1, \dots, W_m$ to be a partition of U into sets of size at least $\left\lfloor \frac{|U|}{m} \right\rfloor \ge \frac{|U|}{2m}$ (note that $|U| \ge \sqrt{n}$ and m is only logarithmic in n, so $\frac{|U|}{m} > 1$ and therefore rounding it down loses a factor of at most 2 — the 2 isn't a real thing, and just comes from the rounding).

Note that *a priori* our candidates would be all of U, so it might be surprising our candidate sets only have this size. But the point is that we want to ensure our candidate sets are disjoint (since the condition on U is only for disjoint sets, so it's easier this way); and we can easily afford this because losing a factor of $2m$ doesn't really hurt us — so we just split up U in the beginning into disjoint sets which we take as our candidates.

Now assume that $1 \le j \le m$, and that we have already proven the claim for $j-1$. This means we have distinct $w_1, \dots, w_{j-1}$ in U, and disjoint subsets $W_j', W_{j+1}', \dots, W_m'$ of $U \setminus \{w_1, \dots, w_{j-1}\}$ (the primes are there to distinguish from the sets we'll obtain for j, since we need to make these sets smaller) with the conditions in the claim (for $j-1$).

Now we need to prove the claim for j — this means we need to produce the next vertex w_j which should come from W_j' — this automatically makes it distinct from the previous vertices, and automatically ensures it has the right adjacencies to w_1, …, w_{j-1}. But picking a vertex $w_j \in W_j$ will restrict our remaining candidate sets (since we need to ensure the correct edges to w_j); we need to make sure that this doesn't make the candidate sets too small.

In other words, we need to find a subset $w_j \in W_j'$ such that for all $i \in \{j+1, \dots, m\}$ with $v_j v_i \in E(H)$, the vertex w_j has at least $\delta |W_i'|$ neighbors in W_i', and for all i with $v_j v_i \notin E(H)$, the vrtex w_j has at *most* $(1-\delta)|W_i'|$ neighbors in W_i'. (This just says that we want to find a vertex $w_j \in W_j'$ such that when we update each of the remaining candidate sets — which we update by for each i, taking the subset of W_i' with the correct adjacencies to the newly picked vertex w_j — we shrink the set by a factor of at least δ. If v_i and v_j are adjacent, then we need to update W_i' to the subset consisting of the neighbors of w_j in W_i', so we need the set of neighbors to be at least $\delta |W_i'|$; if $v_i v_j$ is not an edge then we need to update W_i' to the *non-neighbors*, so in order to have at least $\delta |W_i'|$ non-neighbors we need at most $(1-\delta)|W_i'|$ neighbors.)

If we can do this, then we're done — we automatically have the right adjacencies (taking W_i to be these subsets of neighbors or non-neighbors in W_i'), and they're still disjoint and have the right sizes.

In order to show that there is such a vertex in W_j', for every i we can count the number of choices of w_j which *fail* the condition, using the nice property of U. For each $i \in \{j+1, \dots, m\}$, there are fewer than $\rho \cdot |U|$ vertices $w_j \in W_j'$ which violate the condition for i — fix i and look at all the vertices in W_j' which violate the condition. Then we can take this set to be X and W_i' to be Y in the condition on U, as we have a density condition between X and W_i'. If we have an edge, then a violating vertex has density less than δ to W_i'; so the density of each violating vertex to W_i' is less than δ, which means the set of *all* violating vertices to W_i' is also less than δ. Similarly, if we have a non-edge, then the density is more than $1-\delta$. So then we've found two disjoint subsets with density less than δ or more than $1-\delta$, and the only way this is possible is if one of these sets has size less than $\rho |U|$. But we already know W_i' has size at least $\rho |U|$ by the earlier calculation, as

$$|W_i'| \geq \frac{\delta^{j-1}}{2m} |U| \geq \rho |U|.$$

So the only way this is possible is if the set of violating vertices has size strictly less than $\rho |U|$ (otherwise we would have two reasonably large sets of vertices with density at most δ or at least $1-\delta$, which is not possible).

Now we can just count — for every i there are less than $\rho |U|$ vertices which fail, so in total, there are at most

$$(m-j)\rho |U| \leq m\rho |U| < \frac{\delta^m}{4m} |U| \leq \frac{1}{2} \frac{\delta^{j-1}}{2m} |U| \leq \frac{1}{2} \left|W_j'\right|$$

vertices which violate one of the conditions (using our calculation from earlier). So at most half of the vertices in W_j' are bad, which means there must be at least one good vertex, and we can find $w_j \in W_j'$ as desired. □

The point of this proof is that our strong condition on U was good enough to run this greedy embedding. □

§18.1 Rado's Theorem

We've talked about graphs for the last eight weeks; now we'll come back to a topic which is *not* about graphs, and is instead about colored numbers and solving equations. We've seen a bunch of results along these lines earlier, which we'll now recall:

Theorem 18.3 (Schur 1916)

For every $t \geq 1$, there exists some N such that the following holds: for every coloring of the numbers in the set $\{1, \dots, N\}$ with t colors, we can find x, y, z in $\{1, \dots, N\}$ of the same color with $z = x + y$.

This says that if we fix a number of colors t, then if we take N large enough, whenever we take the ground set $\{1, \dots, N\}$ and color the numbers with t colors, we'll always be able to find a solution to $z = x + y$ where x, y, and z are (not necessarily distinct) elements of the set with the same color.

Here are two other statements of this type, which we've also seen (on the homework). Let $\mathbb{N} = \{1, 2, \dots\}$ denote the positive integers.

Proposition 18.4 (Homework 1, Problem 2(a))

For every $d \geq 1$ and $t \geq 1$, there exists some N such that the following holds: for every coloring of $\{1, \dots, N\}$ with t colors, there exists x, y_1, …, $y_d \in \mathbb{N}$ such that the *affine cube*

$$\{x + s_1 y_1 + \cdots + s_d y_d \mid s_1, \dots, s_d \in \{0, 1\}\}$$

is a monochromatic subset of $\{1, \dots, N\}$.

(The homework problem asked for a bit more — it also required for the 2^d numbers in the above set to be distinct — but it's more natural to omit the condition for today's topic.)

Proposition 18.5 (Homework 2, Problem 1, generalization of van der Waerden's theorem)

For every $t \geq 1$ and $k \geq 1$, there exists some N such that the following holds: for every coloring of $\{1, \dots, N\}$ with t colors, there exist x_1, …, x_k, $y \in \{1, \dots, N\}$ of the same color such that x_1, …, x_k form a k-term arithmetic progression with common difference y — i.e.,

$$x_2 - x_1 = x_3 - x_2 = \cdots = x_k - x_{k-1} = y.$$

In particular this tells us x_1, …, x_k is a monochromatic arithmetic progression (which is van der Waerden's theorem); in the homework we proved the strengthening that we can also make the common difference y of the same color.

These three statements have the same common pattern — we fix some number of colors and some system of (linear, homogeneous) equations, and we say that there's some large enough N such that for every coloring of $\{1, \dots, N\}$ with our t colors, we can solve our system of equations monochromatically.

(This is obvious in the first and third results; for the second, we can introduce 2^d additional variables for each of the elements, and enforce the equations $x_{00\cdots} = x$, $x_{010\cdots} = x + y_1$, and so on; this is a bit stronger, as it tells us that x and the y_i are also of the same color.)

In each of these statements, we are given a homogeneous system of equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1k}x_k &= 0 \\ \vdots \qquad & \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mk}x_k &= 0. \end{aligned}$$

(*Homogeneous* means that we have all 0's on the right-hand side; here we have m equations, and the a_{ij} are generic coefficients.) In each of these problems, we are looking for a solution $(x_1, \dots, x_n) \in \mathbb{N}^k$ with x_1, …, x_k of the same color.

We can write our system of equations more concisely using matrices — let $A \in \mathbb{Z}^{m \times k}$ be the coefficient matrix (consisting of all our coefficients a_{ij}) and let $x = (x_1, \dots, x_k)^\intercal \in \mathbb{N}^k$ be the column vector recording our solution. Then we're looking for a solution to $Ax = 0$ (where A is a $m \times k$ matrix, x a $k \times 1$ vector, and 0 is the 0 vector), where $x = (x_1, \dots, x_k) \in \mathbb{N}^k$ is such that x_1, …, x_k are all the same color.

(So far we haven't really said anything, but we've compressed all three statements into this more abstract framework.)

Remark 18.6. In all equations we're considering, we have 0 and not a constant (on the right-hand side of the equations). The results start getting weird if we put constants instead. For example, if we look at 3-term progressions, then we're looking at $x - 2y + z = 0$. But if we consider $x - 2y + z = 1$, it's easy to define a coloring with no solutions (taking a parity coloring); if we take $x - 2y + z = 2$ then we can take a mod 3 coloring, and so on. So when we put constants, we start running into weird issues with modular constraints; we don't run into these issues with 0's, so in that sense it's more natural to have 0's.

§18.1.1 Finite vs. Infinite

It isn't *always* possible to find such an x. Before we start discussing this, we'll make an observation that reduces the complexity of our theorem statements.

Our theorem statements have been written in terms of taking t and stating there exists N such that the theorem holds on $\{1, \dots, N\}$. But the N part is kind of obsolete in the following sense (as suggested by our rephrasing, which didn't involve N) — we can also ask the same question for colorings of *all* positive integers (instead of stopping at N, you could consider colorings of all of $\mathbb{N}$). If we know there is N such that this holds if we stop at N, then in particular it also holds for a coloring of $\mathbb{N}$. The converse direction is less obvious, but it also holds; this means the 'there exists N' part of our statements is not really necessary, unless we care about the quantitative aspects of what N is. (It can make sense to ask for the best N, but if we just qualitatively care about finding *some* solution, then N isn't necessary.)

Lemma 18.7

For any matrix $A \in \mathbb{Z}^{m \times k}$ and any integer $t \geq 1$, the following two statements are equivalent:

(i) There exists some $N \in \mathbb{N}$ such that for every coloring of the elements of $\{1, \dots, N\}$ with t colors, there is a solution to $Ax = 0$, where $x = (x_1, \dots, x_k)^\intercal \in \{1, \dots, N\}^k$ such that $x_1, \dots, x_k$ have the same color.

(ii) For every coloring of $\mathbb{N} = \{1, 2, \dots\}$ with t colors, there is a solution to $Ax = 0$ where $x = (x_1, \dots, x_k)^\intercal \in \mathbb{N}^k$ such that $x_1, \dots, x_k$ have the same color.

It's obvious that (i) implies (ii) — if we have (i), then we can just look at the numbers $\{1, \dots, N\}$ in $\mathbb{N}$. The fact that (ii) implies (i) is not tautological — it states that if we know the result for all of $\mathbb{N}$, then we only need to look up to a certain point. This is certainly true for any *given* coloring, but there are infinitely many colorings, so we can't just take their maximum; so this is not entirely obvious. We'll prove this on the homework (it depends on the axiom of choice, but so does a large fraction of mathematics).

This means that when we study for which equations we can do this, we might as well look at the simpler statement of coloring all positive integers.

§18.1.2 Partition Regularity

This still doesn't tell us for which equations we actually succeed — it just states that the task coloring only the integers up to some point is equivalent to coloring all of them, but doesn't tell us for which ones it is possible.

Definition 18.8. The system $Ax = 0$ is *partition regular* if (i) and (ii) above hold for every t (i.e., for any coloring of $\mathbb{N}$ with a finite number of colors, we can find a solution).

In other words, a system is partition regular if for every number of colors, these statements are true. Schur's theorem tells us that the equation $z = x + y$ is partition regular; the second statement tells us that the homogeneous system corresponding to an affine cube is partition regular; the third statement tells us that the system $x_2 - x_2 = \cdots = x_k - x_{k-1} = y$ is partition regular.

Remark 18.9. These are called *partition regular* because you can think of the coloring as a partition of $\mathbb{N}$ into color classes, and then we're trying to solve our equation in one of the sets in the partition.

Remark 18.10. It's important $\mathbb{N}$ starts with 1 and not 0, as otherwise every system would be trivially solvable (taking $x = 0$).

This still doesn't answer our question — it just names the systems which work — but a beautiful theorem of Rado exactly characterizes which homogeneous systems of equations are partition regular.

Theorem 18.11 (Rado 1933)

Let $A \in \mathbb{Z}^{m\times k}$ be an integer matrix, and let $a^{(1)}, \ldots, a^{(k)} \in \mathbb{Z}^m$ be its columns. Then the system $Ax = 0$ is partition regular if and only if the following condition holds: there exists a partition $\{1, \ldots, k\} = I_0 \cup \cdots \cup I_\ell$ of the column indices (for some ℓ) such that

$$\sum_{i\in I_0} a^{(i)} = 0$$

(note that both sides are vectors in $\mathbb{Z}^m$) and for every $j = 1, \ldots, \ell$,

$$\sum_{i\in I_j} a^{(i)} \in \mathrm{Span}_{\mathbb{Q}}\left(\{a^{(h)} \mid h \in I_0 \cup \cdots \cup I_{j-1}\}\right).$$

(This is from Rado's PhD thesis, which had the name *Studies of Combinatorics*. His supervisor was Schur, so he generalized Schur's theorem from seventeen years earlier.)

(The partition in Rado's theorem has nothing to do with the name 'partition regular,' which makes the name a bit unfortunate.)

The first condition tells us that the sum of columns in our first set I_0 of the partition is 0, and for each of the future sets I_j in the partition, the sum of its column vectors is in the span of the columns in the previous sets. (The first condition is essentially just the second condition for $j = 0$, since the span of the empty set is 0.)

To get a bit more familiar with what this condition means, we'll try to understand it for our examples.

Example 18.12

Consider the equation $x + y = z$; this corresponds to the matrix $A = \begin{bmatrix}1 & 1 & -1\end{bmatrix}$. Then we can take the partition $I_0 = \{2, 3\}$ and $I_1 = \{1\}$; then our condition on I_0 is satisfied as $1 + (-1) = 0$, and our condition on I_1 is satisfied as $1 \in \mathrm{Span}_{\mathbb{Q}}(1, -1)$.

Example 18.13

In the third example $x_2 - x_1 = \cdots = x_k - x_{k-1} = y$, we can write this as the matrix

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

(where the last column is all 1's, and the rows consist of a -1 and a 1 with 0's everywhere else, moving diagonally). We can then take I_0 to be all the columns except the last (each row has a 1 and a -1, so these columns sum to 0), and I_1 to consist of the last column (which is in our span).

Example 18.14

In the second example with $d = 2$, we have the system

$$\begin{aligned} z_{00} &= x \\ z_{10} &= x + y_1 \\ z_{01} &= x + y_2 \\ z_{11} &= x + y_1 + y_2. \end{aligned}$$

Taking our columns to correspond to x, y_1, y_2, z_{00}, z_{01}, z_{10}, z_{11}, we get

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

We can take $I_0 = \{1, 4, 5, 6, 7\}$. Then the last four vectors 4, 5, 6, 7 span everything, so we can take $I_1 = \{1, 2\}$.

§19 May 2, 2023 — Rado's Theorem

Recall that we use the convention $\mathbb{N} = \{1, 2, \ldots\}$ (a lot of what we do will be pointless if we include 0).

Last class, we started considering the following problem: consider a given homogeneous system of linear equations with integer coefficients (with m equations and k variables). We can write such a system as $Ax = 0$, where $A \in \mathbb{Z}^{m \times k}$ is the coefficient matrix and $x = (x_1, \ldots, x_k)^\intercal$.

Definition 19.1. For a matrix $A \in \mathbb{Z}^{n \times k}$, the system $Ax = 0$ is called *partition regular* if for every integer $t \geq 1$ and every coloring of $\mathbb{N}$ with t colors, there is a solution to $Ax = 0$ where $x = (x_1, \ldots, x_k)^\intercal$ for $x_1, \ldots, x_k$ of the same color.

As discussed last class, we've seen several instances of problems of this shape (e.g., van der Waerden's theorem).

Note that the property requires us to find a monochromatic solution for *every* coloring with *any* finite number of colors. (Obviously the definition would be pointless if we allowed infinitely many colors, since we could color every integer with a different color.)

Remark 19.2. Are there examples of systems which have this property for $t = 2$ colors, but not for any t? The answer is yes — for example, the system described by the equation

$$x_1 + x_2 - 3x_3 = 0.$$

You can always solve this with two colors, but by the theorem we'll discuss next, it's not possible for all numbers of colors.

Rado's theorem characterizes the systems which have this property.

Theorem 19.3 (Rado 1933)

Let $A \in \mathbb{Z}^{m \times k}$ be an integer matrix, and let $a^{(1)}, \dots, a^{(k)} \in \mathbb{Z}^m$ be its columns. Then the system $Ax = 0$ is partition regular if and only if the following condition holds: there exists a partition of the column indices into non-empty subsets $\{1, \dots, k\} = I_0 \cup \dots \cup I_\ell$ such that

$$\sum_{i \in I_0} a^{(i)} = 0$$

and for every $j = 1, \dots, \ell$,

$$\sum_{i \in I_j} a^{(i)} \in \mathrm{Span}_{\mathbb{Q}}(a^{(h)} \mid h \in I_0 \cup \dots \cup I_{j-1}).$$

Remark 19.4. Rado proved this in his PhD thesis; his advisor was Schur, and this is a generalization of Schur's theorem.

Remark 19.5. The first condition is really the same as the second condition with $j = 0$ — the second condition would tell us that $\sum_{i \in I_0} a^{(i)}$ is in the span of nothing, which is 0. We only wrote it separately because it might be confusing to consider the span of the empty set.

Remark 19.6. On Thursday, we wrote down the partitions for a few examples; in those examples, we only needed the sets I_0 and I_1. But it is sometimes necessary to take more sets — it's not always possible to move everything to I_1. This is because taking more sets gives you more power as you go along — in I_1 we have the condition that $\sum a^{(i)}$ needs to be in the span of $a^{(h)}$ for $h \in I_0$. But then when we go to I_2, in our span we have both I_0 and I_1 at our disposal. So we gain more power as we go on, and that's necessary sometimes. A concrete example where a partition exists but we can't find one just with I_1 is on the current homework. (There's a problem where we essentially need to show something is partition-regular, which you can solve by writing down the conditions as a system of linear equations and applying Rado's theorem to show that the system is partition-regular. When we exhibit that this condition holds, we'll find that we need to go further than just I_1 — the problem doesn't ask us to prove that, but it is true.)

Remark 19.7. Note that the x_i may not be distinct; they only need to have the same color.

Now we will prove Rado's theorem. We probably won't finish today, so we'll do this today and also on Thursday.

Rado's theorem is an 'if and only if' statement, so there are two directions. We'll first prove that if the system is partition-regular, then the funny condition holds; then we'll start talking about the reverse direction, which we'll finish on Thursday.

§19.1 Proof of Necessity

Suppose that our system is partition-regular; we want to show that it then has the given property. For this, we'll need the following lemma (we'll see soon how it relates to our problem).

Lemma 19.8

Let $v_1, \dots, v_\ell \in \mathbb{Z}^m$ and $v \in \mathbb{Z}^m$ be vectors such that $v \notin \mathrm{Span}_{\mathbb{Q}}(v_1, \dots, v_\ell)$. Then there are finitely many primes p such that the following statement holds: we have

$$p^z v \equiv \lambda_1 v_1 + \cdots + \lambda_\ell v_\ell \pmod{p^{z+1}}$$

for some integer $z \geq 0$ and some integer coefficients $\lambda_1, \dots, \lambda_\ell$.

(The congruence means coordinate-wise congruence — both sides are two integer vectors, and we say they're congruent mod p^{z+1} if the first coordinates of both vectors have the same remainder mod p^{z+1}, the second coordinates have the same remainder, and so on.)

The statement says that whenever such a relation holds for any z and integer coefficients, this constrains p to a finite list of primes (which does not depend on z or the λ_i).

Proof. We're given $v \notin \mathrm{Span}_{\mathbb{Q}}(v_1, \dots, v_\ell)$, so we'll find a vector which witnesses this (meaning that it's orthogonal to the span but not to v) — there exists a vector $w \in \mathbb{Q}^m$ such that $w \cdot v \neq 0$ but $w \cdot v_i = 0$ for all $i = 1, \dots, \ell$. (This is a simple linear algebra fact.) Of course, this property isn't affected by rescaling w, so we may assume that $w \in \mathbb{Z}^m$.

Now we want to show that whenever a relationship of the form above holds, the prime p must be special. Assume that for some prime p we have the relationship

$$p^z v \equiv \lambda_1 v_1 + \cdots + \lambda_\ell v_\ell \pmod{p^{z+1}}$$

for some $z \geq 0$ and z, $\lambda_i \in \mathbb{Z}$. Then we can take the dot product of both sides with w (this is why it's important that $w \in \mathbb{Z}^m$, because otherwise the calculation with mods would be weird). This gives

$$p^z (w \cdot v) \equiv \lambda_1 (w \cdot v_1) + \cdots + \lambda_\ell (w \cdot v_\ell) \pmod{p^{z+1}}.$$

The right-hand side is 0 (since we assumed $w \cdot v_i = 0$ for all i), so then

$$p^z (w \cdot v) \equiv 0 \pmod{p^{z+1}},$$

which means that

$$p \mid w \cdot v.$$

(If $p^z (w \cdot v)$ is divisible by p^{z+1}, then $w \cdot v$ has to be divisible by p.) But w is fixed (we chose it only depending on v and the v_i, i.e., it doesn't depend on z or the λ_i), so $w \cdot v$ is simply a fixed integer (which is nonzero by assumption), and therefore it has finitely many prime divisors. So there are only finitely many primes for which such a relation can hold. □

Now we'll return to Rado's theorem, and prove that any partition-regular system $Ax = 0$ must have the partition property given by Rado's theorem.

Proof of necessity in Rado's theorem. Assume that $A \in \mathbb{Z}^{m \times k}$ is an integer matrix with columns $a^{(1)}, \dots, a^{(k)} \in \mathbb{Z}^m$ such that the system $Ax = 0$ is partition regular, i.e., that for every coloring of $\mathbb{N}$ with finitely many colors, we can find a monochromatic solution to the system. We want to show that we can then partition the column set in this way.

We'll first use the above lemma. We want to use the fact that $Ax = 0$ is partition-regular by considering a particular coloring, and then using the fact that we must be able to find a monochromatic solution in this coloring. This particular coloring will depend on a well-chosen prime p; the above lemma will ensure that we can choose a prime with the properties we want.

> **Claim —** We can find a prime p satisfying the following condition: for any two subsets $I, J \subseteq \{1, \dots, k\}$ (not necessarily nonempty or disjoint) such that $\sum_{j \in J} a^{(j)} \notin \text{Span}(a^{(i)} \mid i \in I)$, we do *not* have
>
> $$p^z \sum_{j \in J} a^{(j)} \equiv \sum_{i \in I} \lambda_i a^{(i)} \pmod{p^{z+1}}$$
>
> for any integers $z \geq 0$ and λ_i for $i \in I$.

In other words, whenever we have two subsets I and J such that the sum of the $a^{(j)}$ vectors (for $j \in J$) is not in the rational span of the $a^{(i)}$ vectors (for $i \in I$), we are *not* able to write a congruence where p^z times the sum of the $a^{(j)}$ is congruent to some integer linear combination of the $a^{(i)}$ mod p^{z+1}.

Proof. There are only finitely many choices for I and J, so it suffices to check that any given I and J only exclude finitely many primes (note that this isn't tautologically obvious because even after fixing I and J, there are infinitely many choices of z and λ_i). But this is implied by the lemma — once we've fixed I and J, the lemma precisely tells us that there are only finitely many primes p which are excluded by I and J (taking $v = \sum a^{(j)}$).

So by the lemma, for any fixed choice of $I, J \subseteq \{1, \dots, k\}$, only finitely many primes are excluded by the lemma; then since there are finitely many I and J, there are only finitely many forbidden primes in total, which means there is some prime which isn't forbidden (since there are infinitely many primes). □

Of course it could be the case that for some choices of I and J the sum *is* in the span, but we don't have a condition from those I and J.

We now have a prime p with the above (maybe weird-looking) property. We want to use this prime p to define a coloring of $\mathbb{N}$; then we'll apply the partition regularity of $Ax = 0$ to that coloring, and use it to find our desired partition.

Consider the following coloring of $\mathbb{N} = \{1, 2, \dots\}$ with colors $1, \dots, p-1$ — given some $x \in \mathbb{N}$, we remove all factors of p from it to get a number not divisible by p, and then color it with its remainder mod p. In other words, every $x \in \mathbb{N}$ can be written (uniquely) as

$$x = p^{z(x)}(p \cdot m(x) + r(x))$$

for integers $z(x) \geq 0$, $m(x) \geq 0$, and $r(x) \in \{1, \dots, p-1\}$ (where $z(x)$ is the exponent of p in x, and after we remove all these factors of p, then $r(x) \in \{1, \dots, p-1\}$ is the remainder of the leftover part mod p). We let the color of each $x \in \mathbb{N}$ be $r(x) \in \{1, \dots, p-1\}$. (We wrote the coloring in this way because we will need these quantities in a future calculation.)

(Note that it's important we excluded 0, because 0 can't be written in this way.)

Since $Ax = 0$ is partition regular, for every coloring of $\mathbb{N}$ with finitely many colors we can find a monochromatic solution to $Ax = 0$. Here we have a coloring with finitely many (namely $p-1$) colors, so there exists a solution $x = (x_1, \dots, x_k)^\intercal \in \mathbb{N}^k$ to $Ax = 0$ such that $x_1, \dots, x_k$ have the same color. This tells us that $r(x_1) = \cdots = r(x_k)$; let $r \in \{1, \dots, p-1\}$ be this common value.

We now have a solution to $Ax = 0$ where $r(x_1) = \cdots = r(x_k) = r$. We'll now introduce some variables for the values appearing as $z(x_i)$ — let $0 \leq z_0 < z_1 < \cdots < z_\ell$ be the different values appearing among $z(x_1)$, $\dots$, $z(x_k)$. (Some of these values may be equal and some may not be, but there's certainly a finite list of values that appear among them; we call these values $z_0, \dots, z_\ell$, ordered to be strictly increasing.)

We'll then define our partition based on these $z_0 < \cdots < z_\ell$ (it's not a coincidence that we use the same index set) — for each $j = 0, \ldots, \ell$, we let $I_j \subseteq \{1, \ldots, k\}$ be the set of indices $h \in \{1, \ldots, k\}$ such that $z(x_h) = z_j$. (Each $h \in \{1, \ldots, k\}$ has $z(x_h)$ equal to exactly one of the values z_j, and we put h into the corresponding set.) Then $\{1, \ldots, k\} = I_0 \cup \cdots \cup I_\ell$ is a partition. (These subsets are also nonempty, since we chose the values z_j such that all appear.)

We now need to check that our conditions hold. The first condition is secretly the same as the second for $j = 0$, so to avoid repeating the same argument twice, we'll just check the second condition for all $j = 0$, $\ldots$, ℓ.

In order to check the condition for j, first recall that x is a solution to $Ax = 0$ (this is an equation in $\mathbb{Z}^m$). We want to relate this to the column vectors; we can expand out our system as

$$x_1 a^{(1)} + \cdots + x_k a^{(k)} = 0. \tag{$*$}$$

Now we can plug in what we know about the x_h — recall that $x_h = p^{z(x_h)}(p \cdot m(x_h) + r(x_h))$. We know $r(x_h) = r$, and whenever $h \in I_j$, we have $z(x_h) = z_j$ (by the definition of I_j), and therefore

$$x_h = p^{z_j}(p \cdot m(x_h) + r)$$

for all $h \in I_j$.

Now fix $j = 0, \ldots, \ell$, and consider ($*$) mod p^{z_j+1}. Then the equation simplifies — note that p^{z_j+1} divides x_h for all $h \in I_{j+1} \cup \cdots \cup I_\ell$ (i.e., for all of the index sets coming after I_j), because their corresponding values of $z_{j'}$ are at least $z_j + 1$. So then all the future terms disappear.

Furthermore, for $h \in I_j$, we know

$$x_h = p^{z_j}(p \cdot m(x) + r) \equiv p^{z_j} \cdot r \pmod{p^{z_j+1}}$$

(the first term disappears). So now taking ($*$) mod p^{z_j+1}, we get

$$\sum_{h \in I_0 \cup \cdots \cup I_{j-1}} x_h a^{(h)} + \sum_{h \in I_j} r p^{z_j} a^{(h)} \equiv 0 \pmod{p^{z_j+1}}$$

(since all the remaining summands with $h \in I_{j+1} \cup \cdots \cup I_\ell$ all disappear).

Now we're almost done; we almost have the condition we wanted to prove, and the only slightly annoying thing is r. But we can remove r by multiplying by its inverse — since $r \in \{1, \ldots, p-1\}$, there is some $s \in \{1, \ldots, p-1\}$ such that $rs = mp + 1$ for some $m \in \mathbb{Z}$ (i.e., r has a multiplicative inverse mod p). Then multiplying the above equation by s, we obtain

$$\sum_{h \in I_0 \cup \cdots \cup I_{j-1}} s x_h a^{(h)} + \sum_{h \in I_j} (mp + 1) p^{z_j} a^{(h)} \equiv 0 \pmod{p^{z_j+1}},$$

and since $mp \cdot p^{z_j} \equiv 0 \pmod{p^{z_j+1}}$ we can simply omit this term, so we can rewrite this equation as

$$p^{z_j} \sum_{h \in I_j} a^{(h)} \equiv \sum_{h \in I_0 \cup \cdots \cup I_{j-1}} (-s x_h) a^{(h)} \pmod{p^{z_j+1}}.$$

So we now have the property that p^{z_j} times the sum of the $a^{(h)}$ can be written as an integer linear combination of certain $a^{(h)}$'s mod p^{z_j+1}.

We chose p with the property that whenever for some sets I and J, and integers z and λ_i, we have the relationship

$$p^z \sum_{j \in J} a^{(j)} \equiv \sum_{i \in I} \lambda_i a^{(i)} \pmod{p^{z+1}},$$

we must have $\sum_{j\in J} a^{(j)}$ in the rational span of the $a^{(i)}$ (since we chose p such that for any I and J where this vector is *not* in the span there do not exist any z and λ_i for which such a relationship holds). Here we have such a relationship, so by our choice of p, this means that we must have

$$\sum_{i\in I_j} a^{(i)} \in \mathrm{Span}_{\mathbb{Q}}(a^{(h)} \mid h \in I_0 \cup \cdots \cup I_{j-1}).$$

(If we didn't have such a containment, then by our choice of p we couldn't have a relationship of the form that we do have.) This is exactly the condition that we wanted to prove. (In particular, applying this for $j = 0$, we get that $\sum_{i\in I_0} a^{(i)} \in \mathrm{Span}_{\mathbb{Q}}(\emptyset) = \{0\}$, which means $\sum_{i\in I_0} a^{(i)} = 0$ — this gives the first condition as well.)

So we have produced a partition $I_0 \cup \cdots \cup I_\ell$ with the desired property. □

§19.2 Proof of Sufficiency

We've now proven one direction of Rado's theorem — if a system is partition-regular, then this condition holds. We now need to prove the converse — if this condition holds, then the system is partition-regular. (This is maybe more interesting, since we saw examples where we worked rather hard to show this property for a particular system; this direction of Rado's theorem gives us all those results at once for free.)

We'll use an argument not due to Rado (Rado's original proof used van der Waerden's theorem). This proof will in fact also give stronger side results we'll comment on tomorrow.

The proof will rely on the notion of *Deuber sets*.

Definition 19.9. For integers ℓ, d, $c > 0$, an (ℓ, d, c)-*Deuber set* is a subset $M \subseteq \mathbb{N}$ such that there exist $y_0, \ldots, y_\ell \in \mathbb{N}$ such that M consists of all integers of the form

$$cy_j + \lambda_{j+1}y_{j+1} + \lambda_{j+2}y_{j+2} + \cdots + \lambda_\ell y_\ell$$

for some $j \in \{0, \ldots, \ell\}$ and $\lambda_{j+1}, \ldots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]$.

So the notion of Deuber sets depends on three parameters ℓ, d, and c (which are positive integers). Here $\ell+1$ is the number of y_i we have, c is the coefficient of the first y_j, and the other coefficients must be integers between $-d$ and d (so they're integers of absolute value at most d).

Given $y_0, \ldots, y_\ell$, there are quite a few such combinations of them we can get in this way. If all these combinations are positive integers (this won't always be the case) and we put all of them into a set M, then we call M a Deuber set.

In other words, $M \subseteq \mathbb{N}$ is an (ℓ, d, c)-Deuber set if

$$M = M_{d,c}(y_0, \ldots, y_\ell) = \{cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_\ell y_\ell \mid j \in \{0, \ldots, \ell\}, \lambda_i \in \mathbb{Z} \cap [-d, d]\}$$

for some $y_0, \ldots, y_\ell$ in $\mathbb{N}$. (In other words, for every choice of $y_0, \ldots, y_\ell$ — for which these elements are all positive — we get a (ℓ, d, c)-Deuber set.)

Remark 19.10. An (ℓ, d, c)-Deuber set exists for all ℓ, d, and c — if we first choose $y_1, \ldots, y_\ell$, all we need is to ensure that all the coefficients are positive. We can ensure this is the case by taking $y_{\ell-1}$ way bigger than y_ℓ, and so on (for example $y_\ell = 1$, $y_{\ell-1} = d+1$, $y_{\ell-2} = (d+1)^2$, ... — note that $c \geq 1$, so this will work). So it's not a problem to find parameters when these sets *exist*.

These sets are nice because we can find all linear combinations of this shape inside them, and this will help us solve equations of the shape $Ax = 0$.

There isn't enough time to finish the proof in the remaining seven minutes, but we'll now see a roadmap of how the proof works and how Deuber sets come in. The plan is to prove two lemmas.

Lemma 19.11

Suppose that $A \subseteq \mathbb{Z}^{m \times k}$ is a matrix whose columns $a^{(1)}, \ldots, a^{(k)}$ satisfy the the condition in Rado's theorem. Then there exist integers ℓ, d, $c > 0$ such that every (ℓ, d, c)-Deuber set $M \subseteq \mathbb{N}$ contains a solution to $Ax = 0$ with $x = (x_1, \ldots, x_k)^\intercal \in M^k$.

So we consider a fixed system $Ax = 0$ satisfying the assumption $Ax = 0$, and our goal in the end is to show it's partition-regular. The first lemma tells us that whenever we have a system with this assumption, we can find a solution to the system in every (ℓ, d, c)-Deuber set for ℓ, d, c chosen appropriately (depending on A).

This will be nice, because it already gives us some indication of why (ℓ, d, c)-Deuber sets are useful — we can solve these equations in them. This proof isn't very deep — the definition of such sets means they contain lots of linear combinations, and we choose particular linear combinations for $x_1, \ldots, x_k$ such that our equation gets solved.

Then we'll know that we can solve the system in every Deuber set; but our goal was to solve the system in some color class for every coloring. So it's clear what the missing piece is — that we can find a monochromatic (ℓ, d, c)-Deuber set in any coloring.

Theorem 19.12 (Deuber 1973)

Let ℓ, d, c, $t > 0$ be integers. Then in every coloring of $\mathbb{N}$ with t colors, we can find a monochromatic (ℓ, d, c)-Deuber set.

Then putting together these two pieces gives our desired implication — if the assumption is satisfied, then the first lemma lets us find fixed ℓ, d, c such that every (ℓ, d, c)-Deuber set contains a solution. Then applying the theorem to these ℓ, d, c and our particular number of colors for which we're checking partition regularity, we know that for every coloring we can find a monochromatic (ℓ, d, c)-Deuber set, and in that Deuber set we can find a solution.

(We'll in fact prove a stronger version of this theorem, which will have more amazing consequences we'll talk about next class.)

§20 May 4, 2023

Today we will continue discussing the proof of Rado's theorem.

Recall that $\mathbb{N} = \{1, 2, \ldots\}$.

Definition 20.1. For a matrix $A \in \mathbb{Z}^{n \times k}$, the system of equations $Ax = 0$ is *partition regular* if for every integer $t \geq 1$ and every coloring of $\mathbb{N} = \{1, 2, \ldots\}$ with t colors, there is a solution to $Ax = 0$ with $x = (x_1, \ldots, x_k)^\intercal \in \mathbb{N}$ such that $x_1, \ldots, x_k$ have the same color.

Here our system of equations has m equations and t variables; it is partition regular if for every coloring of $\mathbb{N}$ with finitely many colors (i.e., partition of $\mathbb{N}$ into finitely many subsets) we can find a monochromatic solution. We've already seen a few special instances, but you can try studying partition regularity in general. It'll be true for some systems and not others, and Rado's theorem gives an exact characterization of when this is true.

Theorem 20.2 (Rado, 1933)

Let $A \in \mathbb{Z}^{m \times k}$ be an integer matrix and let $a^{(1)}, \ldots, a^{(k)} \in \mathbb{Z}^m$ be its columns. Then the system $Ax = 0$ is partition regular if and only if the following condition holds: There exists a partition $\{1, \ldots, k\} = I_0 \cup \cdots \cup I_\ell$ into nonempty subsets such that

$$\sum_{i \in I_0} a^{(i)} = 0$$

and for every $j = 1, \ldots, \ell$,

$$\sum_{i \in I_j} a^{(i)} \in \mathrm{Span}_{\mathbb{Q}}(a^{(h)} \mid h \in I_0 \cup \cdots \cup I_{j-1}).$$

We will denote the above condition by $(*)$. On Tuesday, we proved one direction — that if $Ax = 0$ being partition regular then $(*)$ holds, i.e., we can find a partition of the columns into subsets satisfying these conditions. So our remaining task is to prove the backwards direction — if the columns of A have this property, then for every coloring with finitely many colors we can find a monochromatic solution.

§20.1 Deuber Sets

Our proof will follow an idea of Deuber, and in particular the notion of Deuber sets.

Definition 20.3. For positive integers ℓ, d, c, an (ℓ, d, c)-Deuber set is a subset $M \subseteq \mathbb{N}$ such that there exist $y_0, \ldots, y_\ell \in \mathbb{N}$ such that M consists precisely of all integers of the form

$$cy_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_\ell y_\ell$$

with $j \in \{0, \ldots, \ell\}$ and $\lambda_{j+1}, \ldots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]$.

In other words, M should consist of all integer linear combinations of the y_i where the first nonzero coefficient is c, and all future coefficients have absolute value at most d. There are a lot of such linear combinations; if M is precisely the set of all such integers, then M is a (ℓ, d, c)-Deuber set.

In other words, $M \subseteq \mathbb{N}$ is an (ℓ, d, c)-Deuber set if $M = M_{d,c}(y_0, \ldots, y_\ell)$ for some choice of $y_0, \ldots, y_\ell$, where

$$M_{d,c}(y_0, \ldots, y_\ell) := \{cy_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_\ell y_\ell \mid j \in \{0, \ldots, \ell\}, \lambda_i \in \mathbb{Z} \cap [-d, d]\}.$$

Remark 20.4. For all ℓ, d, $c > 0$, there exists an (ℓ, d, c)-Deuber set. This is not entirely obvious, because if you choose $y_0, \dots, y_\ell$ badly, these combinations may not all be positive integers. But you can choose $y_0, \dots, y_\ell$ well enough that these combinations are all positive integers — e.g., $y_0 = (d+1)^\ell$, $y_1 = (d+1)^{\ell-1}$, $\dots$, $y_\ell = 1$ — and if you can do this, then you obtain an (ℓ, d, c)-Deuber set.

Remark 20.5. If you want, you can make better choices and ensure all the values are distinct — if $y_i > (2d+1)y_{i+1}$ for all $i = 0, \dots, \ell - 1$ and $y_0, \dots, y_\ell \in \mathbb{N}$, then $M_{d,c}(y_0, \dots, y_\ell)$ is indeed a Deuber set (i.e., it is a subset of $\mathbb{N}$), and all the terms $cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_\ell y_\ell$ are distinct. But it's not a requirement for Deuber sets that this holds (this is not really relevant unless you care about finding solutions with distinct variables).

Remark 20.6. If $M_{(2d+1)c,c}(y_0, \dots, y_\ell)$ is a Deuber set (i.e., if it is a subset of $\mathbb{N}$), then we automatically have the condition $y_i > (2d+1)y_{i+1}$ for all i, as above — so by forcing the value of d to be larger, we can obtain additional properties on the y_i.

It's unclear what this has to do with Rado's theorem; the answer to this was revealed last class. We want to show that $(*)$ implies we can always solve our system monochromatically. This will follow from the combinatino of two statements.

Lemma 20.7

If $A \in \mathbb{Z}^{m \times k}$ satisfies $(*)$, then there exists ℓ, d, c such that every (ℓ, d, c)-Deuber set contains a solution to $Ax = 0$ with $x = (x_1, \dots, x_k)^\intercal \in M^k$.

Theorem 20.8 (Deuber)

For every fixed ℓ, d, c, t, for every coloring of $\mathbb{N}$ with t colors, we can find a monochromatic (ℓ, d, c)-Deuber set.

Combining the lemma and the theorem shows the backwards direction of Rado's theorem — i.e., that $(*)$ implies that $Ax = 0$ is partition regular. This is because beginning with A satisfying $(*)$, we can apply the lemma to obtain (ℓ, d, c) such that every (ℓ, d, c)-Deuber set contains a solution. Then we can apply the theorem to these values; for every coloring we can find a monochromatic (ℓ, d, c)-Deuber set, and this Deuber set will contain a solution to our system.

So our remaining tasks are now to prove the lemma and the theorem. We'll start by proving the lemma (which is the easier part).

§20.2 Proof of Lemma

By the assumption on A, there is a partition $\{1, \dots, k\} = I_0 \cup \cdots \cup I_\ell$ into nonempty subsets with the properties described in Rado's theorem, i.e., such that $\sum_{i \in I_j} a^{(i)} \in \mathrm{Span}_{\mathbb{Q}}(a^{(h)} \mid h \in I_0 \cup \cdots \cup I_{j-1})$.

This condition tells us we can write each $\sum_{i \in I_j} a^{(i)}$ as a linear combination of the previous vectors with rational coefficients. Rational coefficients are a bit annoying (since we have integers), so we'll clear denominators — by taking a common multiple of *all* occurring denominators (over all values of j and all terms in our linear combinations), we can find an integer $c > 0$ such that for all $j = 1, \dots, \ell$, we can write

$$c \sum_{i \in I_j} a^{(i)} = \sum_{h \in I_0 \cup \cdots \cup I_{j-1}} \lambda_{j,h} a^{(h)}$$

where our coefficients $\lambda_{j,h}$ are now *integers*.

Our goal was to find ℓ, d, and c so that every (ℓ, d, c)-Deuber set contains a solution to this system. We've already suggestively named our common multiple c, and ℓ will also be the same value as from the partition, so it just remains to find d.

Let $d > 0$ be the maximum absolute value among our integer coefficients $\lambda_{j,h}$ for $j = 1, \dots, \ell$ and $h \in I_0 \cup \dots \cup I_{j-1}$ (i.e., we take all the coefficients λ which appear above, and take d to be larger than all of them in absolute value).

We've now defined ℓ, d, and c, so our goal is to show that every (ℓ, d, c)-Deuber set contains a solution, so let M be an (ℓ, d, c)-Deuber set, i.e.,

$$M = \{cy_j + \lambda_{j+1}y_{j+1} + \dots + \lambda_\ell y_\ell \mid j \in \{0, \dots, \ell\}, \lambda_i \in \mathbb{Z} \cap [-d, d]\}.$$

We need to show that M contains a solution to $Ax = 0$.

We will write down a solution explicitly — we want to define each of $x_1, \dots, x_k$ to be a term of the form $cy_j + \lambda_{j+1}y_{j+1} + \dots + \lambda_\ell y_\ell$. This definition will depend on which set the term we're defining belongs to in the partition — for every $j = 0, \dots, \ell$ and every $i \in I_j$, we define

$$x_i = cy_j - \lambda_{j+1,i}y_{j+1} - \dots - \lambda_{\ell,i}y_\ell.$$

To check that this is a valid term of M, the coefficients $\lambda_{j,h}$ are defined whenever h is in a set with smaller index than j. We have $i \in I_j$, which is strictly smaller than each of the first indices, so these λ's are well-defined. We can then see that this term is of the necessary form — each of the λ's is of absolute value at most d (since we chose d such that this is true).

So this is a valid definition, and we need to check that $Ax = 0$, i.e., that

$$\sum_{i=1}^{k} x_i a^{(i)} = 0.$$

This is a direct calculation — in order to see this, we have

$$\begin{aligned}\sum_{i=1}^{k} x_i a^{(i)} &= \sum_{j=0}^{\ell} \sum_{i \in I_j} x_i a^{(i)} \\ &= \sum_{j=0}^{\ell} \sum_{i \in I_j} (cy_j - \lambda_{j+1,i}y_{j+1} - \dots - \lambda_{\ell,i}y_\ell) a^{(i)}.\end{aligned}$$

We'll reorder the sum so that we can look at the coefficient on each y_j separately — y_j appears with a coefficient of c in the term we see above, but it'll appear with other coefficients from the other terms. The reason we want to do this is that we know nothing about the y_j's (there's no special relationship between them), so the only way we can hope for everything to cancel is if the coefficients cancel separately for every y_j.

So we'll rewrite this as

$$\sum_{j=0}^{\ell} \left(\sum_{i \in I_j} cy_j a^{(i)} - \sum_{i \in I_0 \cup \dots \cup I_{j-1}} \lambda_{j,i} y_j a^{(i)} \right)$$

(y_j appears for every i in $I_{j'}$ with $j' < j$). We can rewrite this as

$$\sum_{j=0}^{\ell} y_j \left(c \sum_{i \in I_j} a^{(i)} - \sum_{i \in I_0 \cup \dots \cup I_{j-1}} \lambda_{j,i} a^{(i)} \right).$$

This equation is 0, exactly by our definition of the c and $\lambda_{j,h}$ (of course, we chose the coefficients this way to make this step work out), so this sum is indeed 0.

So these choices of $x_1, \dots, x_k$ form a solution to $Ax = 0$, and they are indeed in M; this proves the lemma.

Remark 20.9. The proof looks complicated, but isn't particularly deep — it can be quickly summarized by saying that you choose c and d well enough so that then just using linear combinations of this shape, you can form a solution to the system. The much harder part is to prove the theorem (which is why the first is called a lemma and the second a theorem).

§20.3 Proof of Theorem

We've now proved the lemma, so all that remains (to prove Rado's theorem) is to prove the theorem. Before we prove the theorem, we'll first generalize it — in fact, Deuber actually proved something stronger than Rado's theorem. The following theorem (also due to Deuber, from 1973) is a strengthening of the first.

Remark 20.10. Deuber gave this new approach in his PhD thesis; this is funny because Rado also proved the theorem in his PhD thesis.

Theorem 20.11

Let ℓ, d, c, t be positive integers. Then there exist positive integers ℓ^*, d^*, c^* such that for every (ℓ^*, d^*, c^*)-Deuber set $M^* \subseteq \mathbb{N}$ and every coloring of M^* with t colors, there exists a monochromatic (ℓ, d, c)-Deuber set $M \subseteq M^*$.

The initial theorem told us we can find a monochromatic (ℓ, d, c)-Deuber set if we color *all* the positive integers. This new theorem says it's not necessary to color all the positive integers; it suffices to color an (ℓ^*, d^*, c^*)-Deuber set (for some suitable choice of ℓ^*, d^*, c^* depending on ℓ, d, c, t) — i.e., for every such Deuber set M^*, colorign M^* with t colors is enough to find a monochromatic (ℓ, d, c)-Deuber set. This is stronger than the initial theorem, since for every (ℓ^*, d^*, c^*) we can find a (ℓ^*, d^*, c^*)-Deuber set $M^* \subseteq \mathbb{N}$, and then we can look at our coloring of $\mathbb{N}$ only on M^*.

Remark 20.12. We'll now comment on finding *distinct* variable solutions. Note that in the definition of partition regularity, we want to find a solution with all variables of the same color; we can ask what happens if we want the variables to also be distinct (i.e., we want a nondegenerate solution). For some systems, this will obviously be hopeless (because the system could contain the equation $x_1 = x_2$, or a list of equations that together imply $x_1 = x_2$). So for systems which inherently require two variables to be equal, you can't do this.

But you might hope that if you demand $(*)$ and the additional condition that the system doesn't force two variables to be equal, then you can find a monochromatic solution with distinct variables. This is indeed true. We won't show this explicitly, but all you need to do in order to show this is to prove a slightly different version of the lemma, where we assume $(*)$ and the fact that $Ax = 0$ doesn't imply any two variables are equal, and then prove that in a (ℓ, d, c)-Deuber set we can find a solution with distinct $x_1, \ldots, x_k$. It's not very hard to modify our proof to get this additional version — we write the linear combinations the way we already did, take ℓ to be one bigger (i.e., take one more variable $y_{\ell+1}$ in our Deuber set). By assumption $Ax = 0$ has some fixed solution where all variables are distinct (not necessarily in our set M), and we'll take these integers in the solution as additional coefficients for $y_{\ell+1}$. Then when we perform the calculation, again everything will cancel, because all the coefficients we get on $y_{\ell+1}$ will be from that solution.

Then we have to do one additional thing — we still need to guarantee the x_i are distinct. To do this, we can use the remark from earlier (on $M_{(2d+1)c,c}$) — we can make d artificially larger to guarantee the property that in our Deuber set, all the combinations we actually take will end up being distinct.

§20.3.1 Consequences of Stronger Theorem

Before we prove our new, we'll talk about some additional consequences of this new theorem — in particular, it answers a question that Rado left open.

To motivate this, we'll state another theorem (which follows from Rado's theorem).

Theorem 20.13

For every coloring of $\mathbb{N}$ with a finite number of colors, there exists a color class containing solutions to $Ax = 0$ for *all* partition-regular systems $Ax = 0$.

By the definition of partition regularity, whenever we have a partition regular system, we have a solution to that system in some color class. But this theorem says that we can choose a color class which achieves a solution *simultaneously* for all these systems. (It's tautological from the definition that for every system we can find a color class solving it, but the point here is that we can find a color class solving all of them simultaneously.) This may be surprising at first, but it can be deduced from Rado's theorem. (We won't discus show, since we'll see something stronger soon.)

So in other words, whenever we color $\mathbb{N}$ with finitely many colors, one of these colors will be an amazing set, in the sense that it solves all partition regular systems. So then we can ask the following question:

Question 20.14. Suppose we start with an amazing set (which solves every system), and then color just that set with finitely many colors. Is it still true that there's a color class which still solves all systems?

This does *not* tautologically follow from the above theorem; and it's what Rado conjectured but couldn't prove, and Deuber proved in his PhD thesis using the stronger result.

Theorem 20.15 (Deuber 1973, solving a conjecture of Rado)

Let $M \subseteq \mathbb{N}$ be a set containing solutions to all partition regular systems $Ax = 0$. Then for every coloring of M with finitely many colors, there exists a color class that also contains solutions to all partition regular systems.

So if we start with a set that's amazing (in the sense that we can solve all partition regular systems) and split it into finitely many pieces (color classes), then one of those pieces still solves all partition regular systems; this is quite cool.

To see how this follows from Deuber's stronger theorem, the idea (we won't write down all the details for the sake of time) is the following:

Claim — There must be a color class containing a $(z, z \cdot z!\,, z!\,)$-Deuber sets for infinitely many z.

Proof. For any particular z, we can look at Deuber sets with these parameters, and take ℓ^*, d^*, c^* as in the stronger theorem. Now (ℓ^*, d^*, c^*)-Deuber sets are described by some complicated system of equations. This system will be partition regular (e.g., because of the second theorem, which states that we can find them in any coloring of $\mathbb{N}$), so there will exist a solution to it in M.

That (ℓ^*, d^*, c^*)-Deuber set has been colored with t colors, so by our theorem, one color class must contain a Deuber set with our parameters $(z, z \cdot z!\,, z!\,)$.

But there are only finitely many color classes, so there must be some color class containing such a Deuber set for infinitely many z. □

Then this color class will be the one we want — this color class will contain an (ℓ, d, c)-Deuber set for all (ℓ, d, c), since if $z > \ell$, d, c then every $(z, z \cdot z!, z!)$-Deuber set will contain an (ℓ, d, c)-Deuber set. (Deuber sets aren't really monotone in the sense that just making ℓ, d, c bigger doesn't give you containment — at least, for c — but this statement is true.)

Now we have a color class containing (ℓ, d, c)-Deuber sets for all ℓ, d, c, which means that it contains a solution to all systems.

Remark 20.16. Theorem 4 is strictly weaker than Theorem 5, since you can solve every partition regular system in $\mathbb{N}$. But we can get Theorem 4 directly from Rado's theorem — suppose for contradiction that for every color class, there exists a particular system that is not solvable in that color class. Now we have t different systems, and we can take a system consisting of all of these systems (on disjoint variables, so our different pieces don't interact). This will still be a linear system, and it will still be partition regular (since it was satisfied for each of the individual pieces), so we must be able to solve our gigantic system in some color class. But that's a contradiction, since we can't solve the piece of it corresponding to that color class.

Remark 20.17. The goal of the class is to get a formal proof of Rado's theorem, which relies on the lemma and Theorem 2. Theorem 3 is a stronger version of Theorem 2 (which we will soon prove). Theorems 4 and 5 are there in order to show us why Theorem 3 is cooler than Theorem 2 (by showing a conjecture of Rado's that can be proven using Theorem 3).

Remark 20.18. As mentioned earlier, (ℓ, d, c)-Deuber sets don't behave as monotonously as you might hope for. If you take an (ℓ, d, c)-Deuber set, it'll still *contain* an (ℓ', d', c)-Deuber set for all $\ell' \leq \ell$ and $d' \leq d$. But for c, the monotonicity behaviour is weird (reducing ℓ just means omitting some variables, which is fine, and reducing d just means allowing fewer λ's, which is also fine, but changing c really changes the structure of our terms, which is bad). But in c there is some containment based on divisibility, and this is why we had $z!$. If $c = c'f$, then we need to multiply all the y's by f; this then affects the threshold for d — this is the reason why we have another $z!$ term in d as well. The formal statement should be that every (ℓ, df, c)-Deuber set should contain a $(\ell, d, c/f)$-Deuber set for $f \mid c$.

§20.3.2 Proving the Stronger Theorem

We'll now say a little bit about the proof of Theorem 3 (we can now forget basically everything about systems, partition regularity, and so on; we just need to think about Deuber sets and colorings).

As we can see, the terms in a Deuber set naturally fall into groups, where for each fixed j we get a particular group. So as a first step towards Theorem 3, we'll prove something weaker — rather than finding a (ℓ, d, c)-Deuber set where *everything* has the same color, we'll find a (ℓ, d, c)-Deuber set where each *group* is monochromatic (i.e., for each j).

We'll do this inductively — we'll first take $j = 0$ and make that set monochromatic, and then we'll take $j = 1$ and make that set monochromatic, and so on.

So to prove Theorem 3, we will show the following by induction on h.

Proposition 20.19

Let ℓ, d, c, t be positive integers, and let $h \in \{0, \dots, \ell\}$. Then there exists integers ℓ^*, d^*, $c^* > 0$ (depending on ℓ, d, c, t, and h) such that for every (ℓ^*, d^*, c^*)-Deuber set $M^* \subseteq \mathbb{N}$ and every coloring of M^* with t colors, there exists an (ℓ, d, c)-Deuber set $M = M_{d,c}(y_0, \dots, y_\ell) \subseteq M^*$ such that for each $j = 0, \dots, h$, the set

$$\{cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_\ell y_\ell \mid \lambda_i \in \mathbb{Z} \cap [-d, d]\}$$

is monochromatic.

So this lets us find a Deuber set in M^* which is not necessarily altogether monochromatic, but every layer of it (each corresponding to a value of j) is monochromatic (there's no requirement that different layers have the same color). The role of h is how far we can go; obviously this statement becomes stronger as we increase h, and we'll eventually apply this statement for $h = \ell$. But we'll prove the statement by induction on h (which is why we have h here).

Once we have the statement for $h = \ell$, this still isn't the statement we wanted — we wanted an (ℓ, d, c)-Deuber set which is altogether monochromatic, not just one where every group is monochromatic. But once we've shown this, we can get Theorem 3 in the following way: right now, each of our groups has a different color, and we only want to keep the groups with the same color. By Pigeonhole, at least a $1/t$-fraction of the groups should have the same color. So applying the proposition with ℓt in place of ℓ, we'll be able to find $\ell + 1$ groups which have the same color; then we'll only take the variables y_j for which this is the case. (We'll do this more formally next class — we'll prove the proposition, which is the hard part, and then go from the proposition to the theorem.)

In order to prove the proposition, we'll use Hales–Jewett — or rather, the multidimensional version of Hales–Jewett on Homework 1.

§21 May 9, 2023

Today we'll finish up the proof of Rado's theorem that we started in the previous two lectures. Recall that $\mathbb{N} = \{1, 2, \ldots\}$ is the set of positive integers (i.e., it doesn't include 0).

Last time, we reduced the proof of Rado's theorem to a particular theorem about Deuber sets (so we don't need to remember the statement about Rado's theorem now). As we've seen earlier, the definition of Deuber sets is as follows.

Definition 21.1. For integers ℓ, d, $c > 0$, an (ℓ, d, c)-*Deuber set* is a subset $M \subseteq \mathbb{N}$ such that there exist $y_0, \ldots, y_\ell \in \mathbb{N}$ such that M consists of precisely all integers of the form

$$cy_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_\ell y_\ell$$

with $j \in \{0, \ldots, \ell\}$ and $\lambda_{j+1}, \ldots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]$.

In other words, $M \subseteq \mathbb{N}$ is an (ℓ, d, c)-Deuber set if

$$M = M_{d,c}(y_0, \ldots, y_\ell) = \{cy_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_\ell y_\ell \mid j \in \{0, \ldots, \ell\}, \lambda_i \in \mathbb{Z} \cap [-d, d]\}$$

for some $y_0, \ldots, y_\ell \in \mathbb{N}$. Here we start with positive integers $y_0, \ldots, y_\ell$, and take all linear combinations where the first coefficient is c (here j is the coefficient where we start), and the other coefficients are all of magnitude at most d. All elements of this form must be in M, and all elements of M have to be of this form; note that M must be a subset of $\mathbb{N}$, so only certain choices of $y_0, \ldots, y_\ell$ are allowed (we've seen that $M \subseteq \mathbb{N}$ if each y_i is sufficiently larger than the later ones).

The missing piece in the proof of Rado's theorem is the following theorem ('Theorem 3' from last class).

Theorem 21.2 (Deuber 1973)

Let ℓ, d, c, $t > 0$ be integers. Then there exist integers ℓ^*, d^*, $c^* > 0$ such that for every (ℓ^*, d^*, c^*)-Deuber set $M^* \subseteq \mathbb{N}$ and every coloring of M^* with t colors, there is a monochromatic (ℓ, d, c)-Deuber set $M \subseteq M^*$.

(Of course, ℓ^*, d^*, c^* depend on ℓ, d, c, and t.)

We saw last time that in order to get Rado's theorem, we only need something weaker than this — we only need this statement for colorings of all positive integers, rather than colorings of M^*. But the proof will actually prove this stronger statement, and this stronger statement also has some cool consequences we saw last class (in particular, Deuber used this to resolve a conjecture of Rado which at that point was 40 years old).

As mentioned last time, to prove this we will use the multidimensional version of the Hales–Jewett theorem. We proved this on Homework 1, Problem 3 (we deduced it from the ordinary version of the Hales–Jewett theorem).

Theorem 21.3

For any finite set S and any $t, m \in \mathbb{N}$, there exists $n \in \mathbb{N}$ such that for every coloring of S^n with t colors, there is a monochromatic m-parameter subset of S^n.

Here t is our number of colors, and we're coloring all n-tuples of elements of S. (The ordinary Hales–Jewett statement is about monochromatic combinatorial lines, which are precisely 1-parameter subsets; this is a multidimensional version because we find a monochromatic subset of greater dimension.) This is a

typical Ramsey–type statement — for n large enough, in every coloring we can find a monochromatic nice substructure.

The definition of a m-parameter subset is not very complicated, but a bit annoying to write down.

Definition 21.4. A m-parameter subset of S^n is a subset obtained in the following way: we take some $w \in (S \cup \{*_1, \dots, *_m\})^n$ containing each of $*_1, \dots, *_m$ at least once. Then for each of the $|S|^m$ ways to replace each of the symbols $*_1, \dots, *_m$ with an element of S, we consider the resulting n-tuple obtained from w (by making these replacements). Then (for fixed w and all the different possible replacements), we take the set of all $|S|^m$ n-tuples obtained in this way; we call this an n-parameter subset.

Remark 21.5. On the homework problem we had $S = \{1, \dots, q\}$; but the changed notation fits better into the framework we have here.

In class, we saw this in the case where there's just one special symbol $*$ — then we get a combinatorial line (w was a code with some fixed entries and some $*$'s, and we looked at all the ways to replace $*$ by an element of S; here it's the same, but we have multiple different variables $*$).

To prove the theorem, we will first show the following proposition (showing this proposition is the main work of proving the theorem — the deduction from the proposition to the theorem is pretty easy by pigeonhole, as we stated last time).

Proposition 21.6

Let $\ell, d, c, t > 0$ be integers, and let $h \in \{0, \dots, \ell\}$. Then there exist integers ℓ^*, d^*, c^* (depending on ℓ, d, c, t, h) such that for every (ℓ^*, d^*, c^*)-Deuber set M^* and every coloring of M^* with t colors, there exists an (ℓ, d, c)-Deuber set

$$M = M_{d,c}(y_0, \dots, y_\ell) \subseteq M^*,$$

such that for each $j = 0, \dots, h$, the set

$$\{cy_j + \lambda_{j+1}y_{j+1} + \dots + \lambda_\ell y_\ell \mid \lambda_{j+1}, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]\}$$

is monochromatic.

This proposition starts the same way as our theorem, but we also fix some h. Then conclusion is that for every (ℓ^*, d^*, c^*)-Deuber set M^* and every coloring, there exists an (ℓ, d, c)-Deuber set inside M^* which we're not requiring to fully be monochromatic, but which instead satisfies the condition that for every $j = 0, \dots, h$, the jth 'level' of M — the linear combinations in M coming from our particular choice of j — form a monochromatic set.

This proposition is clearly the strongest if we take $h = \ell$; the point of writing it down for general h is that we want to use induction on h. (In the end, we will only use it for $h = \ell$.)

We'll first see how to deduce the theorem from this proposition.

Proof of Theorem assuming Proposition. In order to go from the proposition to the theorem, first applying the proposition for $h = \ell$, we know that each of our levels is monochromatic, but they don't all have the same color. We'll then make ℓ bigger (taking ℓt instead), so that we have $\ell t + 1$ different levels; this will guarantee that $\ell + 1$ have the same color, which will give what we want.

In the theorem, we're given certain ℓ, d, c, t. We then apply the proposition to ℓt, d, c, t, and $h = \ell t$. Then we obtain certain ℓ^*, d^*, c^* such that the proposition holds — i.e., for every (ℓ^*, d^*, c^*)-Deuber set $M^* \subseteq \mathbb{N}$ and every coloring of M^* with t colors, there exists an $(\ell t, d, c)$-Deuber set

$$M_{d,c}(y_0, \dots, y_{\ell t}) \subseteq M^*$$

such that for each $j = 0, \dots, \ell t$, the set

$$\{cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_{\ell t}y_{\ell t} \mid \lambda_{j+1}, \dots, \lambda_{\ell t} \in \mathbb{Z} \cap [-d, d]\}$$

is monochromatic (so the set is monochromatic for every j, but the color may be different for different j). Now we'll use the pigeonhole principle — there are $\ell t + 1$ possible j and only t colors, so by the pigeonhole principle there are $\ell + 1$ values of j with the same color — in other words, there are indices $0 \le j_0 < j_1 < \cdots < j_\ell \le \ell t$ such that the sets corresponding to $j_0, \dots, j_\ell$ all have the same color.

Then we can take

$$M = M_{d,c}(y_{j_0}, \dots, y_{j_\ell}) \subseteq M_{d,c}(y_0, \dots, y_{\ell t}) \subseteq M^*,$$

which is a monochromatic (ℓ, d, c)-Deuber set. In other words, we try to write down $M_{d,c}(y_{j_0}, \dots, Y_{j_\ell})$. Our variables here are a subset of the original variables. We can check that in the definition of a Deuber set $M_{d,c}(y_0, \dots, y_\ell)$, if we only take a subset of our original variables, then the resulting $M_{d,c}$ is a subset of the original — this is because we can simply set the extra λ's to 0. So then $M_{d,c}(y_{j_0}, \dots, y_{j_\ell})$ is a subset of $M_{d,c}(y_0, \dots, y_{\ell t})$, and therefore a subset of M^*; in particular it's a subset of $\mathbb{N}$, so it is indeed a Deuber set. It's monochromatic because we chose $j_0, \dots, j_\ell$ such that each of the layers corresponding to these indices in $M_{d,c}(y_0, \dots, y_{\ell t})$ was monochromatic, and of the same color. All the linear combinations in the layers of our new Deuber set are *subsets* of these layers (since again, we can set the coefficients of the variables we don't use to 0), so they're monochromatic and of the same color as well; this means M is indeed monochromatic. □

Now the only missing piece of the proof of Rado's theorem is this proposition, which we'll now prove.

Proof of Proposition. First, we can make an easy but useful observation: last time, we mentioned that (ℓ, d, c)-Deuber sets are not entirely monotonous because of issues with c. However, with d they *are* monotonous — so the proposition is monotone in d, in the sense that the bigger we make d, the harder the proposition gets. So if we can show the theorem for larger d then we automatically get it for smaller d (this is because if we take $M_{d,c}$ and shrink d, the resulting set becomes a subset, since we only have fewer choices for the λ's — and having fewer linear combinations we need to check makes our job easier). So we may assume that $d \ge c$.

We'll now prove the proposition by induction on h. The base case is when $h = 0$. In many inductive proofs, the base case is trivial; in this case, it's actually not (it's roughly as hard as the inductive step; both are not super difficult to deduce from the multidimensional Hales–Jewett theorem, but they're a bit technical and may be confusing).

We will apply the multidimensional Hales–Jewett theorem. In the $h = 0$ case, the only condition we have is for $j = 0$ — we want to find an (ℓ, d, c)–Deuber set such that the set $\{cy_0 + \lambda_1 y_1 + \cdots + \lambda_\ell y_\ell\}$ is monochromatic. The first term c is fixed, while the λ_i vary. So the set S (i.e., the 'alphabet') we want to take is the set of choices for these λ_i — so we will take $S = \mathbb{Z} \cap [-d, d]$. We'll take t (the number of colors) as given. Finally, we'll take $m = \ell$ (the reason is that looking at our statement for $j = 0$, there are ℓ different λ's which vary, so we want m different parameters which we can vary here). This gives some $n \in \mathbb{N}$ such that the statement in the (multidimensional) Hales–Jewett theorem holds.

Now let $\ell^* = n$ (this is because ℓ^* is basically the number of variables we have at our disposal in the *original* Deuber set M^* — we'll take $j = 0$ and look at our original set M^*, where we have ℓ^* different variables to vary; this corresponds to the dimension in the Hales–Jewett theorem). We also take $d^* = cd$ and $c^* = c^2$ (these choices aren't obvious; we'll later see the reason for them).

Now fix $M^* = M_{d^*,c^*}(y_0, \dots, y_n)$ to be an (ℓ^*, d^*, c^*)-Deuber set, and consider a coloring of M^* with t colors.

In order to apply the Hales–Jewett theorem, we want to consider a coloring of S^n, which should come from the coloring of M^* with t colors. We do this as follows — the coloring of M^* gives a coloring of $S^n = (\mathbb{Z} \cap [-d, d])^n$ with t colors, where we color any $(\lambda_1, \dots, \lambda_n) \in S^n$ with the color of

$$c^2 y_0 + \lambda_1 c y_1 + \lambda_2 c y_2 + \cdots + \lambda_n c y_n \in M^*.$$

In other words, we're looking at the linear combination with coefficients c^2, $\lambda_1 c$, …, $\lambda_\ell c$ — this is in M^* because M^* was a (ℓ^*, d^*, c^*)-Deuber set with $c^* = c^2$ and $d^* = cd$ — so our first coefficient is correct, and $|\lambda_i c| \le dc = d^*$, so these are allowed coefficients. This means these terms have well-defined colors in M^*, so we can perform this coloring.

(It is right now unclear why we have these additional c factors; this will become clear later. But as a hint, the point is that when we apply multidimensional Hales–Jewett theorem, we'll get a monochromatic m-parameter subset described by some word w; and w will have a fixed part, and parts which vary. We need that fixed part to all go into the first term cy_0 when defining our Deuber set, which means we need to force it to be divisible by c.)

This defines a coloring on S^n; we now apply the statement of Hales–Jewett to our coloring. By our choice of n (so that Hales–Jewett holds), there exists a monochromatic ℓ-parameter subset (recall that $m = \ell$) of S^n described by some word $w = (w(1), \dots, w(n)) \in ((\mathbb{Z} \cap [-d, d]) \cup \{*_1, \dots, *_\ell\})^n$, such that each $*_i$ appears at least once. We don't know which order in which they appear; but the definition is symmetric in the roles of $*_1$, …, $*_m$, which means we can interchange their roles; so (by interchanging the symbols) we can assume that the first occurrence of $*_1$ is before the first occurrence of $*_2$, which is before the first occurrence of $*_3$, and so on. (In other words, the first $*_i$ we see is $*_1$, the next new one we see is $*_2$, and so on.)

Now we have a word w which (using our replacement operation) gives an ℓ-parameter subset of S^n which is monochromatic. Our aim was to find an (ℓ, d, c)-Deuber set $M = M_{d,c}(y_0, \dots, y_\ell)$ such that for $j = 0$, the set $\{cy_j + \lambda_{j+1} y_{j+1} + \dots + \lambda_\ell y_\ell\}$ is monochromatic. We'll use the letters x_i instead of y_i (since we're already using y_i for the variables generating our original Deuber set).

We'll define our variables x_0, …, x_n as follows — we want cx_0 to start with $c^2 y_0$, and then contain all the fixed terms in w, so we define

$$x_0 = cy_0 + \sum_{\substack{i \in \{1,\dots,n\} \\ w(i) \in \mathbb{Z} \cap [-d,d]}} w(i) y_i.$$

(In other words, we're considering all i where $w(i)$ is not one of the $*$'s.)

Meanwhile for x_j, we want to take all the contributions of $*_j$ (in other words, x_1 is associated with $*_1$, x_2 with $*_2$, and so on). So we take

$$x_j = \sum_{\substack{i \in \{1,\dots,n\} \\ w(i) = *_j}} cy_i.$$

(We take the y_i associated with the indices i where we have a $*_j$ in w at that position.)

We'll now make a few claims.

> **Claim —** We have $M_{d,c}(x_0, \dots, x_\ell) \subseteq M^* = M_{d^*,c^*}(y_0, \dots, y_n)$.

Proof. This is basically true by definition — we need to look through the terms appearing in our definition of an (ℓ, d, c)-Deuber set, i.e., $cx_j + \lambda_{j+1} x_{j+1} + \dots + \lambda_\ell x_\ell$. If we write out all the terms, we'll get something of the correct form. (Note that when we rewrite each of the x_i's in terms of the y_i's, the first term will have a c^2 — we have one factor of c from the definition of the x's, and another from the multiplication. Meanwhile each of the other terms with the λ's can be checked as well (x_0 can't be multiplied by a λ, and for each of the future terms, we end up with coefficients of λc, which has absolute value at most dc). But this is why we needed the funny c factors — the point is that x_0 is basically $1/c$ of the fixed term, which is why we needed the fixed term to be divisible by c.) □

Note that this in particular implies our set is a subset of $\mathbb{N}$, and therefore $M_{d,c}(x_0, \dots, x_\ell)$ is an (ℓ, d, c)-Deuber set. We now need to check that the set

$$\{cx_0 + \lambda_1 x_1 + \dots + \lambda_\ell x_\ell \mid \lambda_1, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]\}$$

is monochromatic. This will come from the same calculation we alluded to earlier — plugging in the definitions of our terms, this is equal to

$$\left\{ c^2 y_0 + \sum_{\substack{i \in \{1,\dots,n\} \\ w(i) \in \mathbb{Z} \cap [-d,d]}} w(i) c y_i + \sum_{\substack{i \in \{1,\dots,n\} \\ j \in \{1,\dots,\ell\} \\ w(i) = *_j}} \lambda_j c y_i \right\}$$

(here we combined the sum over j with the sum over i, and plugged in the definition of each x_j — we have cy_i from the definition of x_j, and we're multiplying by λ_j). We can rewrite this in the following way — we can think of this as a sum of y_i with different coefficients. Each comes with a c; for the first term, we look at all i for which $w(i)$ is in S and take $w(i)$ itself as a coefficient; for all where we have $*_j$, we put a λ_j. So the coefficient is precisely what happens to w if we replace all $*_j$'s by λ_j's. So we can rewrite this as

$$\{c^2 y_0 + \sum_{i \in \{1,\dots,n\}} w_{\lambda_1,\dots,\lambda_\ell}(i) c y_i \mid \lambda_1, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d,d]\}$$

where $w_{\lambda_1,\dots,\lambda_\ell}$ is the word obtained from w by replacing $*_1, \dots, *_\ell$ with $\lambda_1, \dots, \lambda_\ell$.

By our choice of w, this is monochromatic — we chose w so that the corresponding ℓ-parameter subset of S^n is monochromatic in the coloring, and the colors in our coloring of S^n correspond to these terms.

We've now done the base case of the induction. The inductive step is essentially the same; we'll sketch it but not check all the details for the sake of time (they're roughly the same as what we did here).

For the inductive step, let $h \geq 1$, and assume that we have proven the statement for $h-1$ — this means in M^*, we can always find some M such that the sets we have are monochromatic for all $j = 0, \dots, h-1$, and our goal is to make the last level monochromatic as well. To do this, we'll again apply the multidimensional version of Hales–Jewett.

Apply the Hales–Jewett theorem again with $S = \mathbb{Z} \cap [-d,d]$ and $m = \ell - h$ (the reason we have $\ell - h$ now is because we only need to work hard to get the condition for $j = h$ — smaller h come from the inductive assumption — and for $j = h$, the varying λ's are $\lambda_{h+1}, \dots, \lambda_\ell$, so we only have $\ell - h$ parameters). This gives some $n \in \mathbb{N}$ such that the Hales–Jewett theorem holds.

Meanwhile, we apply the induction assumption for $h-1$ replacing ℓ with $n+h$ (because at the last step, we want to have n variables left after we put away the first h), d with cd^2, c with c^2, and the same value of t, to obtain ℓ^*, d^*, $c^* > 0$. (These are the ℓ^*, d^*, c^* for which we'll show the theorem holds.)

Now let $M^* \subseteq \mathbb{N}$ be an (ℓ^*, d^*, c^*)-Deuber set, and consider a coloring of M^* with t colors.

The first thing we'll do is apply the induction assumption; then there exists an $(n+h, cd^2, c^2)$-Deuber set

$$M' = M_{cd^2,c^2}(y_0, \dots, y_{n+h}) \subseteq M^*$$

such that for each $j = 0, \dots, h-1$, the set

$$\{c^2 y_j + \lambda_{j+1} y_{j+1} + \dots + \lambda_{n+h} y_{n+h} \mid \lambda_{j+1}, \dots, \lambda_{n+h} \in \mathbb{Z} \cap [-cd^2, cd^2]\}$$

is monochromatic.

As before, we consider the coloring of $S^n = (\mathbb{Z} \cap [-d,d])^n$ with t colors, where we color any

$$(\lambda_{h+1}, \dots, \lambda_{h+n}) in S^n$$

(note that we're using the indices $h+1, \dots, h+n$ for our n-tuples in S^n, for reasons we'll soon see) with the color of

$$c^2 y_h + \lambda_{h+1} c y_{h+1} + \dots + \lambda_{h+n} c y_{h+n} \in M' \subseteq M^*$$

(this is the same idea as before — with c^2 for the first term, and λc for the remaining ones — except that we're starting at h instead of 0 because we care about getting the statement for $j = h$). As before, we can check that this is a well-defined linear combination appearing in M', and therefore M^* (so it has a well-defined coloring).

The strategy is the same as before — from our original coloring we get a coloring of S^n (by coloring with the element of M^* corresponding to these coefficients), and apply the Hales–Hewett theorem to it. By the choice of n, there exists a monochromatic $(\ell - h)$-parameter subset (since we chose $m = \ell - h$) described by some word $w = (w(h+1), \dots, w(h+n)) \in ((\mathbb{Z} \cap [-d, d]) \cup \{*_1, \dots, *_{\ell-h}\})^n$. By interchanging the order of the symbols, we can again assume that the first occurrence of $*_1$ is before the first occurrence of $*_2$, and so on.

We again want to define the $x_0, \dots, x_\ell$ we'll use to define our final set (whose layers are each monochromatic). We'll define them as follows — we define

$$x_h = cy_h + \sum_{\substack{i \in h+1, \dots, h+n \\ w(i) \in \mathbb{Z} \cap [-d,d]}} w(i) y_i$$

In other words, we're taking cy_h, and all indices i for which $w(i)$ is in the fixed part of our word. We also define $x_0 = cy_0, \dots, x_h = cy_{h-1}$.

Finally, for the remaining terms, to define x_{h+j} for each $j = 1, \dots, \ell - h$, we set

$$x_{h+j} = \sum_{\substack{i \in \{h+1, \dots, n+h\} \\ w(i) = *_j}} cy_i.$$

(This is the same as before, except that we only start at x_h, and we take the previous terms x_i to just be c times the y_i.)

By the same reasoning as earlier, it is still true that

$$M_{d,c}(x_0, \dots, x_\ell) \subseteq M_{cd^2, c^2}(y_0, \dots, y_{n+h}) = M' \subseteq M^*.$$

(This again comes from writing out all terms and checking they're of the correct shape.) This in particular means it's a set of positive integers, and is therefore an (ℓ, d, c)-Deuber set.

Now we need to check that a bunch of things are monochromatic. First, for each $j = 0, \dots, h-1$, we have

$$\{cx_j + \lambda_{j+1} x_{j+1} + \cdots + \lambda_\ell x_\ell \mid \lambda_{j+1}, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]\}$$

is a subset of

$$\{c^2 y_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_{h+n} y_{h+n} \mid \lambda_{j+1}, \dots, \lambda_{h+n} \in \mathbb{Z} \cap [-cd^2, cd^2]\}$$

(you can check that the coefficients work out — everything of the first form can be written in the second). The sets of the latter type are all monochromatic by our choice of M', so the former sets are all monochromatic as well.

This means the only set we need to check is for $j = h$, and this is the same calculation from earlier — we consider the set

$$\{cx_h + \lambda_{h+1} x_{h+1} + \cdots + \lambda_\ell x_\ell \mid \lambda_{h+1}, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]\}$$

as the set

$$\left\{c^2 y_h + \sum w(i) c y_i + \sum y_j c y_i\right\},$$

which we can rewrite as

$$\{c^2 y_h + \sum_{i \in \{1, \dots, n\}} w_{\lambda_{h+1} \cdots \lambda_\ell}(i) c y_i \mid \lambda_i \in \mathbb{Z} \cap [-d, d]\}.$$

This is exactly the set we chose to be monochromatic, so that means we've checked the condition for $j = h$; and we're done. □

§22 May 11, 2023 — New Upper Bound for Diagonal Ramsey Numbers

Today and next class, we'll discuss the breakthrough new upper bound for diagonal Ramsey numbers. This is a very recent work which is very exciting to everyone in the field. We can't see the whole proof in detail, because it's 40 pages long, but the goal is to understand roughly what the strategy is, and how it differs from what was known before — and to roughly give us an idea of the outline of the proof. (The paper is very well-written and readable, so if we want more details we should read the paper.)

For context, in the first two weeks of class we discussed the diagonal Ramsey numbers $R(k,k)$ — the number of vertices we need in a complete graph to ensure that for every edge-coloring with red and blue we can find a monochromatic k-clique. In the first two weeks of class, we proved the following upper and lower bounds for $R(k,k)$.

Theorem 22.1 (Erdős–Szekeres 1935 and Erdős 1947)
For all $k \geq 2$, we have

$$2^{k/2} \leq R(k,k) \leq 4^k.$$

Closing this gap has been maybe the most important problem in Ramsey theory since then. In over 75 years, there's been no progress to improve either $\sqrt{2}$ or 4 — there's been a lot of work on this which improved lower order terms (the lower bound was improved by a factor of 2, and the upper bound was improved by subexponential (though better than constant) terms).

More generally, the Erdős–Szekeres argument gives $R(k,\ell) \leq \binom{k+\ell-2}{k-1}$.

The work we'll discuss (which came out about two months ago) is such a big breakthrough because it improves on the constant 4.

Theorem 22.2 (Campos, Griffiths, Morris, Sahasrabudhe 2023+)
There exists a constant $c > 0$ such that

$$R(k,k) \leq (4-c)^k$$

for all $k \geq 2$.

Today we'll discuss the proof outline. The proof also gives an exponential improvement for the off-diagonal Ramsey numbers (if the ratio between k and ℓ is bounded), but for time reasons we won't state exactly what the improvement is.

We need to show that for any edge-coloring of complete graph on $(4-c)^k$ vertices, we can find a red or blue k-clique. We'll frequently use the following notation (since we'll refer to which vertices are connected in which colors):

Notation 22.3. In a graph with edges colored red and blue, for a vertex x, let $N_R(x)$ denote the red neighborhood of x and $N_B(x)$ the blue neighborhood of x.

§22.1 An 'Algorithmic' Proof of the Upper Bound

As a warm-up, and to put the approach of the new work into context, let's redo a proof of roughly $R(k,\ell) \leq \binom{k+\ell-2}{k-1}$ (up to some lower-order factors), phrased in a somewhat unusual way that will motivate the new approach. The approach in this paper is to in some sense write down an algorithm that will lead us to find

a monochromatic clique of size k. (It's not an algorithm in the sense that it is efficient to perform.) Along these lines, we'll do a basic 'algorithmic' proof of the original upper bound.

We'll prove the bound

$$R(k,\ell) \leq e(k+\ell)\frac{(k+\ell)^{k+\ell}}{k^k \ell^\ell}$$

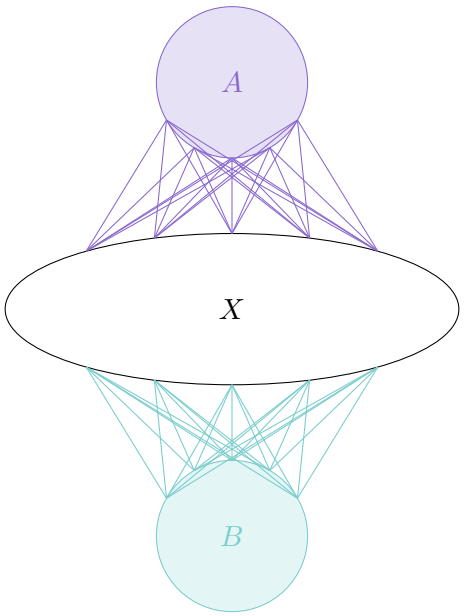
(this is roughly the same as the Erdős–Szekeres bound on $R(k,\ell)$, using Stirling's formula).

In order to prove this, we'll start with a complete graph G on n vertices with edges colored red and blue. Suppose that there is no red clique of size k or blue clique of size ℓ. Our goal is to show that then n is smaller than the above expression.

The idea of the algorithm is to start with our graph and perform certain steps, such that if the algorithm runs too long, then it finds a red clique of size k or a blue clique of size ℓ. This means the algorithm must terminate before this happens, and then from the quantitative behavior of the algorithm, this will give us a bound on n. (This is an unusual way to phrase the original bound, but it's the way of thinking that'll be helpful for the proof.)

We'll recursively build three disjoint subsets A, B, and X of $V(G)$, such that:

- A is a red clique, and all edges from A to X are red.
- B is a blue clique, and all edges from B to X are blue.



We start with $X = V(G)$ and $A = B = \emptyset$. We stop when $|X| \leq k + \ell$.

At every step, we pick a vertex $x \in X$ (since we haven't stopped yet, X has more than $k + \ell$ vertices, so it is not empty).

We want to move x to either A or B. If we move x to A, then we'll need to look at its set of red neighbors inside X. If this set is pretty big, then that's good — the new updated X will be pretty big as well. Meanwhile if this set is *not* pretty big, then we'll instead move it to B.

In other words, we look at the red and blue neighborhoods of x inside X, and compare their sizes. There should be a certain threshold for when we move x to A or B:

- If x has a lot of blue neighbors in X — more precisely, if

$$|N_B(x) \cap X| \geq \frac{\ell}{k+\ell}(|X| - 1),$$

then we update $A \mapsto A$, $B \mapsto B \cup \{x\}$, and $X \mapsto N_B(x) \cap X$. (In other words, we move x to B and update X accordingly.)

- Similarly, if
$$|N_B(x) \cap X| \leq \frac{\ell}{k+\ell}(|X|-1),$$
then since the two neighborhoods add up to $|X|-1$ we must have
$$|N_R(x) \cap X| \geq \frac{k}{k+\ell}(|X|-1).$$
Then we update $A \mapsto A \cup \{x\}$, $B \mapsto B$, and $X \mapsto N_R(x) \cap X$.

The threshold we use here is perhaps not that surprising — in the *diagonal* case we'd want the threshold to be $1/2$. If they're *not* the same — we want a clique of size k in A or ℓ in B — then it makes sense that the ratio of the threshold should be around the ratio between k and ℓ. (If we put a variable for the threshold and optimize it afterwards, we'll find that this is basically optimal.)

This algorithm is rather simple, so it's also rather easy to analyze — we need to track the sizes of A, B, and x throughout the process. We essentially chose the algorithm such that the size of X doesn't drop too much at each step, and we can prove this — inductively, we can prove that throughout the algorithm, we always have
$$|X| \geq \left(\frac{\ell}{k+\ell}\right)^{|B|} \left(\frac{k}{k+\ell}\right)^{|A|} \left(1-\frac{1}{k+\ell}\right)^{|A|+|B|} \cdot n.$$
The first factor comes from the fact that in the first case, we increase $|B|$ by 1 and shrink $|X|$ by around $\ell/(k+\ell)$; the second term si similar. The third term is somewhat artificial and just comes from the rounding with $|X|-1$ (it's a lower-order term that just comes to make the inequality correct adjusting for $|X|-1$ vs. $|X|$, but this term doesn't matter much in the end). Finally n is just the number of vertices we start with in the beginning.

Now let's analyze what happens in the end. Note that throughout the whole algorithm, A will never reach size k, and B will never reach size ℓ (since we assumed that there is no red clique of size k or blue clique of size ℓ) — so we always have $|A| \leq k$ and $|B| \leq \ell$. This means
$$|X| \geq \left(\frac{\ell}{k+\ell}\right)^{\ell} \left(\frac{k}{k+\ell}\right)^{k} \cdot \frac{1}{e} \cdot n.$$
This must always be true, so it's in particular true when the algorithm terminates. On the other hand, when the algorithm terminates, we have $|X| \leq k+\ell$. Combining the two, we obtain an inequality for n — this tells us
$$n \leq e \cdot (k+\ell) \left(\frac{(k+\ell)}{\ell}\right)^{\ell} \left(\frac{k+\ell}{k}\right)^{k},$$
which is the bound we want. (The $k+\ell$ and e are for convenience to deal with the -1 issue; they're lower-order terms, so they don't really matter — the point of this isn't so much to get a good bound, since the bound qualitatively agrees with $\binom{k+\ell-2}{k-1}$ but is worse by lower-order factors as well.)

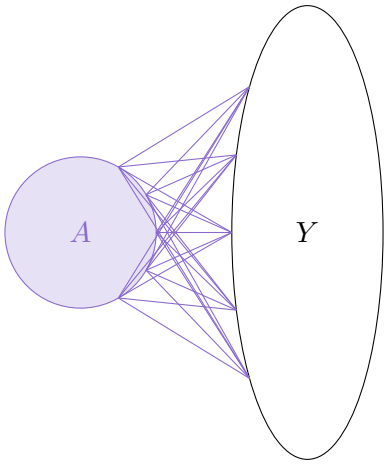
In some sense, this argument is the exact same as the Erdős–Szekeres argument we saw in class — we didn't phrase it in this way, but what the algorithm does is roughly the same.

§22.2 Books

Of course, this is not the algorithm that the new paper uses, since they get a better bound. Their algorithm is vaguely similar, but more complex.

The first idea they use is the following (which is from an earlier paper): in this algorithm, we have a few structures — we have a red clique, and a large set of vertices which all have red edges to A. This structure has a name — it's called a *book*. (This may be because if A has size 1, then this looks a bit like if you take a book held at its binding, and have a lot of pages fanning out.)

Definition 22.4. In a complete graph G with edges colored red and blue, for disjoint $A, Y \subseteq V(G)$, we say that (A, Y) is a *red book* if A is a red clique and all edges between A and Y are red.



The first idea going into the proof (which has been known before) is that rather than making an algorithm which tries to find a monochromatic *clique* of the desired size, it's good enough to find an algorithm which finds a large *book*. Here's a statement making this precise (it tells us that if we've found a book which is large in some sense, this automatically implies the existence of a large monochromatic clique).

Lemma 22.5

Let G be a complete graph with edges colored red and blue. If (A, Y) is a red book in G with $|Y| \geq R(k - |A|, \ell)$, then G contains a red clique of size k or a blue clique of size ℓ.

Note that the Ramsey number we have here depends on $|A|$ — if $|A|$ is bigger, then the $|Y|$ we need is smaller.

Proof. Since $|Y| \geq R(k - |A|, \ell)$, Y contains a blue clique of size ℓ or a red clique of size $k - |A|$. In the first case we're done, because we've found a blue clique of size ℓ. In the latter case, the red clique of size $k - |A|$, together with A, forms a red clique of size k. (Both pieces are red cliques, and all edges between them are red by the definition of a book.) □

This tells us that if we run an algorithm of this type, we're already happy once we've found a red book where Y is big enough.

§22.3 Idea of the Algorithm

The next thing we'll do is explain what the algorithm actually builds (it doesn't just consist of sets A, X, and B — there's actually four sets — and we'll see what they build). In order to keep things more understandable for now, we won't write down all the conditions with exact parameters, because they're very complicated to set up; we'll explain what the algorithm builds and what steps we have, and what we control and some rough intuition of how we control these things, and how that helps us finish. On Tuesday we'll precisely define the parameters and thresholds.

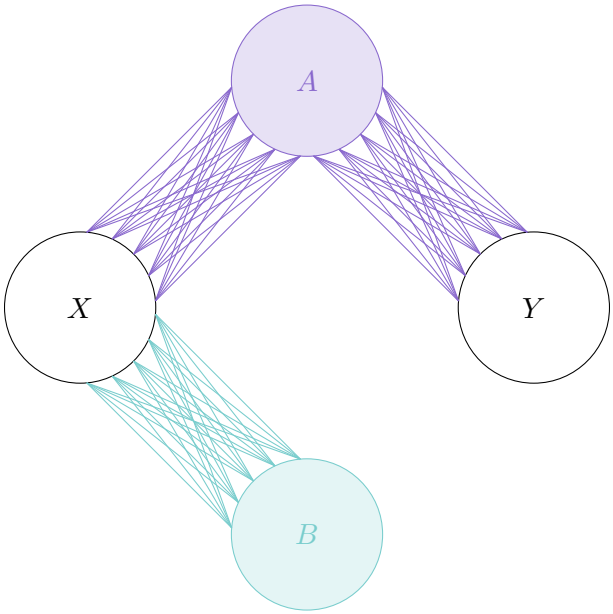
The theorem we've written is for diagonal Ramsey numbers, but we'll need to consider the off-diagonal case as well — in the end, the bound is good for all $\ell \leq k$ with ℓ linear in k.

Let $\ell \leq k$ and let G be a commplete graph with edges colored red and blue. Assume that G has no red clique of size k and no blue clique of size ℓ. Our goal is to upper-bound $|V(G)|$.

The algorithm will build disjoint subsets A, B, X, $Y \subseteq V(G)$ such that:

- A is a red clique, and all edges from A to X and to Y are red.
- B is a blue clique, and all edges from B to X are blue.

It might be surprising that B doesn't have all edges to X *and* Y blue, but that's the asymmetry — which is why we distinguish X and Y in the first place.



Throughout the algorithm, we know that A never gets up to size k, and B never gets up to size ℓ. In the end, we'll look at the red book (A, Y). The role of these sets will be that (A, Y) is the red book, while X is our pool of vertices that we'll use to move to A and B — so X and Y have very different roles. Vertices in Y don't ever get moved to A or to B — we'll move vertices from X to A and B. Of course we will have to update Y sometimes (when we move vertices from X to A).

Throughout the algorithm, there's a few things we need to track:

- The sizes of the sets — $|A|$, $|B|$, $|X|$, $|Y|$. This is unsurprising — we want A and B to grow (as we move things to them), and we want the sizes of X and Y to not drop too much.
- The density of red edges between X and Y. (This is perhaps more surprising.) We'll keep a variable p for this — we use $p := d_R(X, Y)$ to denote the density of red edges between X and Y.

(We'll see in a moment why this density is important.)

We first need to initialize the algorithm — we start with $A = \emptyset$ and $B = \emptyset$, and X and Y each being half of the vertices in an arbitrary manner — we partition $V(G) = X \cup Y$ arbitrarily into sets of size at least $\lfloor |V(G)|/2 \rfloor$ (so up to constant factors, X and Y start off the same size as the ground set).

We also want to define when the algorithm terminates — the algorithm terminates when $|X|$ becomes very small. (The threshold for 'very small' is $|X| \leq R(k, \ell^{3/4})$ — right now it's very unclear why this is the case, so we can ignore it.)

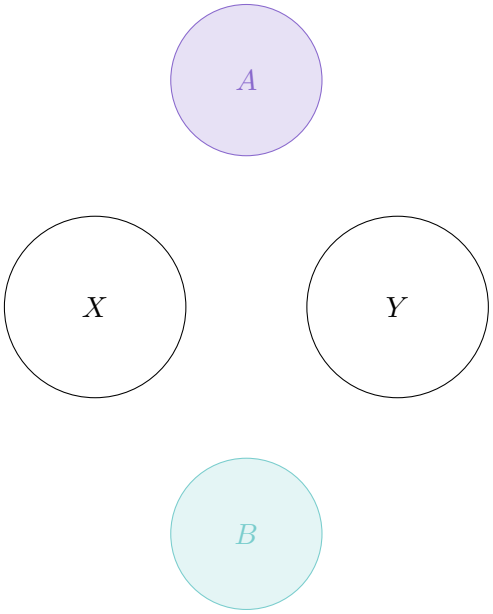
Roughly, the steps of the algorithm are as follows (later we'll add a fourth pre-processing step, but these are the three types of main steps).

- *Red step* (this is roughly what we already know from what we did before) — we take a 'suitable' vertex $x \in X$ and move it to A, and update the sets accordingly. As before, we update $X \mapsto N_R(x) \cap X$ and $Y \mapsto N_R(x) \cap Y$ (for the same reason), $A \mapsto A \cup \{x\}$, and $B \mapsto B$.

  We'll define what 'suitable' means later. But here's a comment on why we care about the red density between X and Y — as in the original algorithm, we need x to have many red neighbors in X, so that when we update X it doesn't change too much. But when we move x to A we need to update Y as well — this means we need x to have many red neighbors in Y. If all edges between X and Y were blue, then we'd have a big problem — doing such an update would cause Y to disappear. Similarly if there were only a few red edges, then Y would shrink a lot, and we wouldn't be able to afford to perform this update. So this is why we need reasonable red density between X and Y — so that we can make moves of this form without the size of Y dropping too much.

  We might guess that the next type of step is a blue step, but it's actually more subtle. The algorithm doesn't have just one blue step (the way it has one red step), but two types of blue steps with different features.

- *Big blue steps* — this is roughly what we'd think a blue step is (moving a vertex from X to B and updating), but moving lots of vertices at once. More precisely, we find a large blue book $(S, T) \subseteq X$.



  We now want to move all the vertices from S to B. So we update $X \mapsto T$ (we could replace it by the intersection of blue neighborhoods of vertices in S, but that's complicated, and the only thing we know about it is that it contains T). We don't need to change Y (because we didn't put anything into A). So we update $Y \mapsto Y$, $B \mapsto B \cup S$, and $A \mapsto A$. (This is a *big* blue step because all the vertices in S get moved to B.)

  Since our book is large, meaning that T is large, the size of X doesn't drop too much. Meanwhile Y doesn't change, and B grows quite a lot. But what about the red density? The new X is T, so the new density between X and Y is the density between T and Y. In theory, with what we've just written, we have no control between the density of X and Y. But this is ensured by a *pre-processing* step — before we do any of these steps, we first need to clean up X and Y by removing all vertices in X which have low density to Y. (Before we do one of our main steps, we do this cleanup step, called 'degree regularization,' where we kick out all vertices in X which have only a few neighbors in Y. When we do this, then we know the vertices in T survive this step, so they have enough neighbors in Y and therefore the density between T and Y is good enough.)

  There's one natural question — why do we need to do this with many vertices at a time, rather than just one? This is because every time we do these steps, we lose a bit of density — it's important to

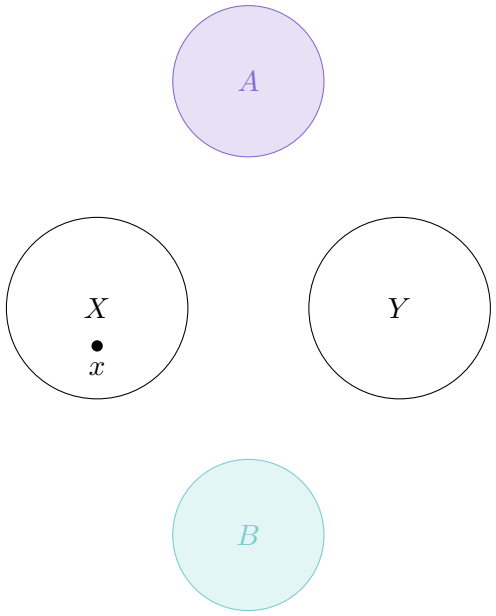
track the density p. We've just said that with the cleanup, we don't completely lose control of the density. But we lose a little bit of density during the cleanup (we can't kick out all vertices just a tiny bit below the average, because that could be almost everything). So on this step we lose a bit of density, which means we can't perform too many of these steps. The fact that we move many (it'll be around $\ell^{1/4}$) vertices at once gives us a bound as to how often these steps occur — we can have at most ℓ vertices in B, so this can happen at most $\ell^{3/4}$ times. This means we can take the density loss that we have in these steps without having to worry about it — it'll be a lower order term.

In the red steps, we also lose a bit of density; that will be part of a very delicate analysis.

- The third step is also a type of blue step, but we don't call it a blue step but rather a *density-boost step*, because its additional feature (besides moving a vertex from X to B) is that it boosts the density from X to Y. This step will counter the losses in density we have in the other two.

  The point is that this step will happen when neither of the first two steps can happen (it could be that we can't find a suitable vertex or a large book, so we can't perform the first two steps; but in those cases we'll be able to prove that we can always move a vertex from X to B while boosting density.)

  Take a suitable vertex $x \in X$; in particular, such that the red density between $N_B(x) \cap X$ and $N_R(x) \cap Y$ is relatively high. We then update $X \mapsto N_B(x) \cap X$, $Y \mapsto N_R(x) \cap Y$, $A \mapsto A$, and $B \mapsto B \cup \{x\}$. This is basically a small blue step, with a density boost between X nad Y.



  So we take a vertex $x \in X$, and we care about its blue neighborhood in X and its red neighborhood in Y. We care that the red density between these two sets should be quite high (quite a bit bigger than p). Then we will move x to B. The new set X will be what we needed it to be (the blue neighborhood of x in X). But the surprising thing is that we also update Y. When we move a vertex from X to B, there is *a priori* no need to update Y — there's no conditions between B and Y. But we still do update it — the point is that then we have a boosted density between our new X and our new Y.

  This density-boost step is quite costly in the sense that both X and Y shrink. But on the other hand, we get a good boost in the density; we have other steps which are good in terms of sizes but shrink the density, while this step is costly in terms of sizes but boosts the density. As you might expect, we'll then have to delicately analyze how the changes in these steps compare to each other. This is the reason our blue steps are big — that means that the density loss will be somewhat negligible since there aren't many such steps, and the main tradeoff will be what happens between the red steps and the density-boost steps.

Before doing any of these steps, we need to remove the vertices from X which have too small red degree to Y. This is the *degree regularization* step (it's interleaved with the other three main steps — we perform this pre-processing step, then a main step, then a pre-processing step, then a main step, and so on). Informally, this step removes vertices $x \in X$ from X for which $|N_R(x) \cap Y|$ is significantly smaller than $p\,|Y|$. (On average $|N_R(x) \cap Y|$ will be $p\,|Y|$, since p is the red density.) Of course, we'll need to analyze that we don't remove too many vertices.

If the first two steps fail, why should we hope to find a vertex such that the third step works? Here is the intuitive argument: first there is always a vertex $x \in X$ such that $d_R(X, N_R(x) \cap Y) \geq p$. This is by an averaging argument using convexity (we'll say more about it on Tuesday) — this isn't completely obvious, but it's not difficult, and it follows from an averaging argument counting red walks from X to Y back to X and using convexity.

The point is then sort of that if a blue step does not happen, then we can find a vertex $x \in X$ such taht $|N_R(x) \cap X|$ is large and $d_R(X, N_R(x) \cap Y)$ is not much smaller than p (we find a vertex with this property with a bit of loss, such that the additional condition holds that the red neighborhood of x in X is large, i.e., its blue neighborhood isn't).

Then there are two options. We know the red density between X and Y is pretty good. So either the red density between the *red* neighborhood of $x \in X$ and the red neighborhood in Y is good (with a threshold quite a bit below p — this means we'll need to take quite a bit of density loss in the red steps) — in which case we can do a red step — or if it's quite a bit smaller, then by the tradeoff the density between the blue neighborhood in X and the red neighborhood in Y must be quite a bit bigger than p (to make up for the loss).

To be more critical, we chose x such that the red ring is quite big, which means the blue bit might be tiny. But now we're replacing X by this tiny bit. So we're losing a lot in the size of X — that sounds super dangerous. But the point is that this boost of density from the averaging argument is increasingly higher the smaller this becomes. So in this step, there's a delicate tradeoff between how much X shrinks vs. how much we boost our density (as a corollary the set can't be empty). These tradeoffs are quite delicate, and it's sort of miraculous this all works out in the end.

The big blue steps don't affect the density too much, since there's only a few. The red steps make A grow but cause a loss in density. The density-boost steps don't make A grow, but they boost the density; but they make X and Y quite a bit smaller.

In the end, $|A|$ will be affected by the number of red steps, and $|Y|$ will be affected by the number of both. We'll want the inequality $|Y| \geq R(k - |A|\,, \ell)$.

There's a picture on page 8 of the paper where the x-axis shows the number of red steps we've taken, and the y-axis the number of density boost steps. The algorithm only terminates if X is small, so not all combinations of these numbers are possible; the blue region tells us which combinations are possible if we start with $n = (4 - o(1))^k$. In the end, we want

$$|Y| \geq R(k - |A|\,, \ell).$$

We already have the Erdős–Szekeres bound, so we're happy if this is satisfied — i.e., if

$$|Y| \geq \binom{2k - |A|}{k - |A|}.$$

In the picture, this is satisfied outside the red region (the red region is the region of problematic). So the only problem occurs if we're in both the red and blue regions simultaneously (because we have to be in blue, and if we're outside red then we're sad).

It looks like they don't intersect, but a zoomed in picture shows that they intersect on a tiny green thing. If they were disjoint, then the proof would be done; but if we're in the little green thing, then we're *not*

done. But the green thing is rather small — in particular, it only happens if $|A|$ is roughly between $0.75k$ and $0.85k$. This means we only need to focus on the case where $|A|$ is in this interval. In this case, we will prove a better bound on the Ramsey number than the Erdős–Szekeres one — when $0.75k \leq |A| \leq 0.85k$, then we'll want a Ramsey bound in the case $0.15k \leq \ell \leq k/5$. If we can get a better bound for this Ramsey number, then it'll be easier to prove Y is at least this Ramsey number.

So we do everything from scratch again, starting with a different ℓ. Why would you hope this succeeds now? The tradeoff we saw earlier was something like $\ell/(k+\ell)$. Now this isn't going to be close to $1/2$ — it'll be close to $1/4$. So the size tradeoff is different, which means the density boost is bigger. So if $\ell \leq k/4$, the density boost in the density-boost steps will be bigger, and we will get an exponential improvement over the Ramsey bound in the case $0.15k \leq \ell \leq 0.2k$.

Now using this better bound, the red region (the problematic region) becomes smaller. And now the regions really are disjoint, and you can improve 4^k to $(4-c)^k$.

§23 May 16, 2023

Last class, we began discussing the spectacular new upper bound on the diagonal Ramsey numbers from two months ago:

Theorem 23.1 (Campos, Griffiths, Morris, Sahasrabudhe 2023+)
There exists a constant $c > 0$ such that

$$R(k,k) \leq (4-c)^k$$

for all $k \geq 2$.

Last class, we saw an overview of the proof strategy, but everything was vague; today the plan is to make the algorithm precise, and maybe to see some of the tradeoffs and why you can expect certain things to happen at certain points (last time, we only gave very rough intuition on why if we can't do a big blue step or a red step then the density-boost step really boosts density). Of course, the paper is 50 pages, so we can't cover all the details; we'll try to be precise about what the algorithm is, but we can't discuss the 30 pages of analysis.

§23.1 The Overall Structure

The overall structure is to first show a better bound for *off-diagonal* Ramsey numbers of the form

$$R(k,\ell) \leq e^{-\ell/50+o(k)}\binom{k+\ell}{\ell}$$

for $k/10 \leq \ell \leq k/4$. (The 10 here is somewhat arbitrary, but the 4 is important — last time we saw a picture with a red and blue region which slightly overlap, but only in a region of the x-axis between $0.15k$ and $k/4$ — that precisely corresponds to the range here. The point is that the diagonal approach almost works, but it fails in this range; in that range, one needs a better bound on the off-diagonal Ramsey number.) This gives an exponential improvement on the Ramsey numbers in this case — we already know $\binom{\ell+k}{k}$ is an upper bound by Erdős–Szekeres, and the $e^{-\ell/50}$ is a tiny exponential improvement.

The next step is to use this to show that $R(k,k) \leq (4-c)^k$. (We saw a strategy for how to prove this, which fails in the place where two regions overlap; there we need a better upper bound on the Ramsey numbers in that region, because we eventually want to show that $|Y|$ is greater than a certain Ramsey number using the book lemma.)

Both of these parts follow the same 'algorithmic' proof that we discussed last time and will discuss in more detail this time, with slightly different parameters and analysis.

Today the plan is to give a precise version of the algorithm.

§23.2 Proof Setup

Assume that $\ell \leq k$ and that both are sufficiently large, and assume that $\ell \geq k/10$ (you can put any constant here, but we'll just take 10 to have an explicit constant — we just need ℓ to be at most k but linear in k). Our goal is to upper-bound $R(k, \ell)$; in the end we care particularly about the case $k = \ell$, but as mentioned earlier we need to go through the more general case to get this one.

Consider a complete graph on n vertices with edges colored red and blue, with no red clique of size k and no blue clique of size ℓ. Our goal is to prove an upper bound on n. (We already know from the classical Erdős–Szekeres argument that $n \leq \binom{k+\ell}{\ell}$, but we want a (small) exponential improvement.)

As stated last time, the overall strategy is to have an algorithm building disjoint subsets A, B, X, $Y \subseteq V(G)$ such that:

- A is a red clique, and all edges from A to X and Y are red;
- B is a blue clique, and all edges from B to X are blue.

(There's no condition between B and Y, or between X and Y or inside X and Y.)

Throughout the algorithm, we need to track the sizes of A, B, X, and Y (this is unsurprising — we want these sets to not shrink too much) and the red density between X and Y (we explained why we want to track this last time); we use $p = d_R(X, Y)$ to denote this density of red edges between X and Y.

We use the set X to move vertices to A and B. Meanwhile, we use the set Y such that (A, Y) is a big red book in the end — a *red book* is precisely a red clique and another set where all edges from the red clique to the other set are red. Last time, we saw a simple lemma that if we have a big book where $|Y|$ is large enough (in terms of $R(k - |A|, \ell)$), then we must be able to find a red clique of size k or a blue clique of size ℓ.

We know that A has size less than k and B has size less than ℓ; in the end we'll examine the size of the book (A, Y), and we'll be done if $|Y|$ is big enough.

Notation 23.2. We use $N_R(x)$ and $N_B(x)$ to denote the red and blue neighborhoods of a vertex x.

(The B in this subscript is not related to the set B we're constructing.)

As we run through the algorithm, the sets will change — at every point, we'll have a current configuration of sets, and a current density. In particular, p is not fixed — we always use it to denote the *current* red density between X and Y.

Last time, we saw that there are three different main types of steps — the red steps, big blue steps, and density-boost steps. In the red and density-boost steps, we pick a particular vertex in X and move it to A or B, and do something with the neighborhoods in X and Y. So the first thing we need to define is related to how we pick these vertices.

Definition 23.3. Given the current configuration of sets X and Y, we define the *weight* of a pair (x, y) of vertices in X as

$$\omega(x, y) = \frac{1}{|Y|}(|N_R(x) \cap N_R(y) \cap Y| - p \cdot |N_R(x) \cap Y|).$$

Note that (x, y) is an ordered pair, and the two vertices may be the same or different.

This term looks odd, but as some motivation, we're picking two vertices $x, y \in X$ and looking at the red neighborhood of x in Y. Then the above expression counts the number of red neighbors of Y in this set, minus p times the size of the set. Very heuristically, because the red density between X and Y is p, you might expect that the red density of y into this set $N_R(x) \cap Y$ is roughly p, so this counts how much higher or lower the red density from y into this set is from what we would expect.

Last time, we saw that when we decide whether we're doing a red step or a density-boost step (we want density-boost steps to boost p), we want to pick x in such a way that the red density between X and the set $N_R(x)$ is a bit bigger than p. This is captured by this — this talks about the red density for every vertex $y \in X$, and if you fix x and take this information for all y together, then you get exactly information about the red density from X to $N_R(x)$.

We care about what happens for a fixed x and all y, so we also use the following definition:

Definition 23.4. We define

$$\omega(x) = \sum_{y \in X \setminus \{x\}} \omega(x, y).$$

(We remove x because we'll be splitting X into the red and blue neighborhoods of x, and x isn't part of either.) This talks about the number of red edges from $X \setminus \{x\}$ and the red neighborhood of x. The factor of $|Y|$ is just a normalization factor to ensure that our numbers are between 0 and 1, and isn't really important.

Heuristically, the idea is that we want to pick a vertex x where $\omega(x)$ is at least 0, or maybe just a tiny bit less; this will lead to having a vertex x such that the red density of $X \setminus \{x\}$ to $N_R(y)$ is good enough (at least p or just a tiny bit below).

To summarize, if $\omega(x)$ is nonnegative or just a bit below 0, then $d_R(X \setminus \{x\}, N_R(x) \cap Y)$ is at least p or just a bit below p. This is good because it'll allow us to take that vertex x, and either take a red step or a density-boost step.

First, the question is, why can we find such a vertex x such that $\omega(x)$ is at least 0 or just a bit below? A priori, we could be in bad shape, where every $\omega(x)$ is very negative. But this is ensured by the following lemma:

Lemma 23.5

We have $\sum_{x,y \in X} \omega(x, y) \geq 0$.

If we know this lemma, then it's not hard to believe that we can pick a vertex x such that $\omega(x)$ is nonnegative or just below 0. (The 'just below 0' comes from the fact that we need to exclude $y = x$ — this straightforwardly implies that there will be x such that $\sum_{y \in X} \omega(x, y) \geq 0$, but we'll also have a small contribution from the case $y = x$ that we need to get rid of.)

However, we don't just need *any* vertex with this property; we need it to have an additional property, that $N_R(x) \cap X$ is not too small. (This is because our vertex was supposed to be good for taking a red step, and it wouldn't be if $N_R(x) \cap X$ is too small.)

But the point is that if a blue step doesn't happen, then there can only be a few vertices in X whose red neighborhood is too small for a red step; the contreibution to the above sum (in the lemma) from the ineligible vertices is then not too big, so among the eligible ones you can still find a vertex x which is good (i.e., where $\omega(x)$ is not too negative).

We won't prove this lemma for now; it's a nice and relatively simple proof based on a convexity argument (we'll hopefully have time to see it at the end).

This tells us, in the algorithm, how we choose the vertex x which we look at for the red step or the density-boost step. The other question is what should the threshold be at which we decide which type of step to

do. We only want to do a red step if the red step doesn't decrease p by too much, so now the question is what too much means. For this we need a few more definitions.

Definition 23.6. Let $\varepsilon = k^{-1/4}$. Let p_0 be the initial (red) density between X and Y at the start of the algorithm.

The $1/4$ in the definition of ε is not important. Note that k is very large, so ε is very small (as $k \to \infty$, we have $\varepsilon \to 0$). We define p_0 to be the value of p at the start; this will be an important comparison point. The idea is that if our density throughout the algorithm increases a lot beyond p_0, then we're pretty rich and we can afford to lose more; while if it's close to p_0 or even below it, then we can't afford to lose very much. So the terms we will write down will all be based on comparison with p_0. (It's maybe not clear why p_0 at the start should be that important for the algorithm — they probably did this because it works, but there might be more dynamic ways to run the proof where you don't fix everything at a particular p_0.)

We need to ensure at the start of the algorithm that p_0 is not too small (so when we start the algorithm, we have a reasonably large red density between X and Y — we'll define 'reasonably large' later).

Definition 23.7. For all $h = 0, 1, 2, \ldots$, define

$$q_h = p_0 + \frac{(1+\varepsilon)^h - 1}{k},$$

and define $\alpha_h = q_h - q_{h-1}$.

We can see that $q_0 = p_0$, q_1 is a bit bigger, and so on; they increase exponentially, and so the differences α_h increase with h (the bigger h is, the bigger these differences are). These will sort of be the losses that we allow.

Definition 23.8. For $0 < p < 1$, we define $h(p)$ to be the smallest $h \in \{1, 2, \ldots\}$ such that $p \le q_h$.

Here p is our current red density; the point of the q_h is to divide the possible p-values into certain intervals. (If $p \le q_1$ then $h = 1$, if $p \in (q_1, q_2]$ then $h = 2$, and so on.)

Now in a red step, we will allow the density to decrease from p to $p - \alpha_{h(p)}$. So the role of $h(p)$ is that $\alpha_{h(p)}$ is the threshold by which we allow the density to go down in a red step — in a red step we start with density p, and we allow ourselves a loss of up to $\alpha_{h(p)}$. If this isn't possible, then we want a density-boost (which will also correspond somehow to $\alpha_{h(p)}$ in size).

Remark 23.9. We'll discuss what to do if p_0 is tiny later.

Definition 23.10. Define the threshold μ as

$$\mu = \begin{cases} \frac{\ell}{k+\ell} & \text{if } \ell < k \\ \frac{2}{5} & \text{if } \ell = k \end{cases}.$$

The role of μ is that we want to do red steps if $N_R(x) \cap X$ is big enough, or equivalently if $N_B(x) \cap X$ is not too big. We want this blue neighborhood to be at most $\mu |X|$. This is maybe not that surprising because in the algorithmic version of the Erdős–Szekeres bound, where we just took a vertex in X and moved it to A or B (there was no set Y), where we moved it had to do with how big the blue neighborhood of x inside X was, and the threshold for this was exactly $\ell/(k+\ell)$ (if the blue neighborhood was at least this then we'd do a blue step of moving x to B, and otherwise we'd move it to A). The role here is similar — if the blue neighborhood is at least $\mu |X|$ then something blue happens, and if it's smaller then a red step happens.

What's maybe more surprising is the value $2/5$. This comes from the fact that we first consider the case where $\ell \leq k/4$ and then define μ in the first case (in particular, μ is quite a bit smaller than $1/2$). As we hinted at last time, in this case since our threshold is quite a bit smaller than $1/2$, the boost we get in the density-boost steps is amplified by that. Once we have done this (for the first part), we want to actually do the diagonal case. In that case, we have a problem because μ is going to be around $1/2$; it turns out that this doesn't really work for the algorithm, and so we artificially make it smaller — making it smaller means that in the red steps we lose less in $|X|$. (μ is a threshold for how big the blue neighborhood is allowed to be if we want to perform a red step; in the diagonal case this means the blue neighborhood is at most $2/5$ of $|X|$, so the red neighborhood is at least $3/5$; this means $|X|$ loses a factor of $3/5$ instead of $1/2$). This $2/5$ makes our red steps better *and* our density-boost steps better (we get a better density-boost as well); where we pay for it is in the big blue steps. But the way the analysis is set up, the big blue steps don't contribute much — they only contribute a lower-order term — which is why we can afford that.

In the algorithm, you should think of μ as a fixed constant, and ℓ, k, and n as large with respect to μ. (If you think of ℓ in the given range, then μ is at least around $1/10$.) All our asymptotics will be for μ fixed and k, ℓ large.

§23.3 The Algorithm

(Note that this isn't an algorithm in the sense that it's efficient — for example, when this tells us 'find a vertex x such that $\omega(x)$ is pretty big,' this can't necessarily be done efficiently. So you couldn't use this in practice if someone gave you a coding task to find the clique; it's an algorithm in the sense that you algorithmically do something, but it's not efficient.)

The first step of any algorithm is to initialize it: we start with $A = B = \emptyset$ (this is unsurprising since we want to build the cliques throughout). Meanwhile to choose X and Y, we want both to be fairly big — we want to partition the vertex set into X and Y such that both have size at least $\lfloor n/2 \rfloor$, and such that p_0 — the initial value of $d_R(X, Y)$ — is reasonably large. By 'reasonably large,' we in particular need it to be bounded away from 0, but we need even more (roughly $k/(k+\ell)$).

First, how can we even ensure that p_0 is large? By taking a random partition, we can always ensure that it's at least the *total* red density. So then the question on how we can ensure a large starting red densisty between X and Y is the same as how we can ensure that the entire graph has a large red density.

In the diagonal case, we can definitely assume the red density is at least $1/2$ (otherwise we can switch the roles of red and blue). Unfortunately, this symmetry argument doesn't work in the off-diagonal case; in that case, you basically want the density to be at least roughly $k/(k+\ell)$. So what happens if it's not? In that case, we momentarily forget about everything, and go back to the beginning of last class — at the beginning of last class, we had a simpler algorithm without the set Y, where we just had the sets A, X, and B and moved a vertex from X to A or B.

If the blue density is bigger than $\ell/(k+\ell)$ by some constant factor, then when we do the argument with the moving, we win by a little bit at every step. And if we win by a little bit on every step, two things might happen — either at some point our red density goes into the range where we wanted it to be (i.e., the red density is at least $1 - \ell/(k+\ell)$), at which point we apply this framework to the graph that's left — or this never happens and when we run through the entire argument, at every step we save a bit. But the number of steps is linar in k, so we get an exponential total saving.

So basically, if the red density at the start is too low to do what we want to do, then we do what we did at the start of last class, and we run it until we reach a graph where the red density is high enough to do this, or if this never happens, then we've saved a constant factor at each step and get an exponential saving overall.

There's also the question of what happens if the density is very close to $\ell/(k+\ell)$ but not exactly at least it. This is why you do the argument a third time — you also consider a case where ℓ is even smaller (something like $\ell \leq k/40$, at which point you can do the Erdős–Szekeres argument even if you're just a little bit away).

Now we'll assume that our red initial density *was* good enough, and we'll finally do the algorithm.

We run the following algorithm until $|X| \leq R(k, \lceil \ell^{3/4} \rceil)$ (our simple algorithm from Thursday stopped when X was too small; here we again stop when X is too small) or if $p \leq 1/k$ (this is a technicality — we will ensure that this never actually happens, so when we stop we're actually in the first case).

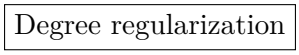
Before we do one of the main steps, we have to perform a *degree regularization* step — at this step, we kick out the vertices in X which have too small red density to Y. (This is in some sense just a technicality — we're cleaning up X and throwing out the vertices with abnormally low density.) This changes X a bit, so then we update p.

We then need to decide between our three main steps (red steps, big blue steps, and density-boost steps) which one to perform.

To do so, we first ask, *do there exist at least* $R(k, \lceil \ell^{2/3} \rceil)$ *vertices* $x \in X$ *with* $|N_B(x) \cap X| \geq \mu |X|$*?* (In other words, we're looking at how many vertices have a big blue neighborhood in X.) Note that this Ramsey number is quite a bit smaller than $R(k, \lceil \ell^{3/4} \rceil)$, our stopping condition on X. If the answer is yes, then we do a *big blue step*. (Then we go back and do degree regularization, and then another step, and so on.)

If the answer is no, then we choose a vertex $x \in X$ such that $\omega(x)$ maximal such that $|N_B(x) \cap X| \leq \mu |X|$. (In this case there only exist a few vertices $x \in X$ where the blue neighborhood is at least a μ-fraction, so there are many vertices $x \in X$ where the blue neighborhood is smaller; and of those, we choose one of them where $\omega(x)$ is maximal.) This is exactly what we were talking about earlier — we want to choose an eligible vertex where $\omega(x)$ is at least 0 or not too much below it.

Then we ask ourselves, *is* $d_R(N_R(x) \cap X, N_R(x) \cap Y) \geq p - \alpha_{h(p)}$*?* In other words, we look at the red density between the red neighborhood of x in X, and the red neighborhood of x in Y. This was our threshold for the density decrease in red steps; so if the answer is yes, then we do a red step, and if the answer is no, then we do a density-boost step (and in either case, after we do this, we go back to a degree-regularization step).



$R(k, \lceil \ell^{2/3} \rceil)$ vertices with $|N_B(x) \cap X| \geq \mu |X|$?

(We'll explain the degree regularization step in more detail later.)

Now we'll explicitly describe what happens in the red steps and the density-boost steps (as described last time).

In a *red* step, we replace $A \to A \cup \{x\}$, $B \to B$, $X \to N_R(x) \cap X$, and $Y \to N_R(x) \cap Y$. (These updates are in order to ensure that the correct edges between A and X or Y are red.)

In a density-boost step, as described last time, we replace $A \to A$, $B \to B \cup \{x\}$, $X \to N_B(x) \cap X$, and $Y \to N_R(x) \cap Y$. (We're taking the *complement* of x's red neighborhood in X, and the red neighborhood in Y. This will give a boost in the density because we chose x such that the red density between X and $N_R(x) \cap Y$ is at least p or a tiny bit below; meanwhile the red density between $N_R(x) \cap X$ and $N_R(x) \cap Y$ is quite a bit below this (at most $p - \alpha_{h(p)}$), so the red density between $N_B(x) \cap X$ and $N_R(x) \cap Y$ is quite a bit above it.)

In the degree regularization step, we define $h = h(p)$, and perform the updates $A \to A$, $B \to B$, $Y \to Y$, and

$$X \to \{x \in X \mid |N_R(x) \cap Y| \geq (p - \varepsilon^{-1/2}\alpha_h)\,|Y|\}.$$

In other words, we kick out all vertices in x which have too small red density to Y — we only keep the vertices with reasonably many red neighbors in Y. The average size would be $p\,|Y|$, and $\varepsilon^{-1/2}\alpha_h$ is the bit of slack we need t oensure we don't kick out too many vertices.

Doing this step never decreases p, since we're only kicking out vertices below average. Meanwhile, it ensures that every particular vertex x has reasonably large density to Y (where for each particular vertex, we lose a little bit compared to p). This is something we might lose in density when we do a blue step, because the only density guarantee we have afterwards is from this.

(Of course when you analyze the algorithm, you need to ensure that you don't lose too many vertices in X when you do this.)

Finally, in a *big blue step*, what we do is — we said last time that we want to add lots of vertices from X into B, so we want to find a large blue book. The big blue step happens when we have lots of vertices $x \in X$ with individually large blue neighborhoods, so we need to translate that into a blue book. The instructions are — find a blue book (S, T) in X where $|T| \geq \mu^{|S|} \cdot |X|/2$ with $|S|$ as large as possible (under this condition that $|T|$ has a reasonably big size — it's maybe not so strange that we get an additional μ-factor for every vertex in S, since we're moving all the vertices in S simultaneously to B, and each should correspond to losing a μ-factor). We then update $A \to A$, $B \to B \cup S$, $X \to T$, and $Y \to Y$. (In a big blue step we move our blue clique S to B, and our new set X becomes T — which by the definition of a blue book has all blue edges to S.)

The obvious question is, why does this actually make $|S|$ reasonably large (since we want big blue steps to really move many vertices to B)? This is the content of the following lemma:

Lemma 23.11

Whenever a big blue step happens, its book (S, T) satisfies $|S| \geq \ell^{1/4}$.

(This is indeed a lot of vertices, since ℓ is large.)

Corollary 23.12

There are at most $\ell^{3/4}$ big blue steps.

(This is because $|B|$ can never reach ℓ, since there's no blue clique of size ℓ. So if at each big blue step we add at least $\ell^{1/4}$ vertices, then we can do this at most $\ell^{3/4}$ times.)

This is important because we lose quite a bit on the value of p (namely the $\varepsilon^{-1/2}\alpha_h$ term), but there are not too many such steps (we'll get something like $k^{7/8}$ times the maximum values of the α_h).

Now we've seen what happens in each of the steps; we run this until either $|X|$ gets too small or p gets too small, and we can prove that p never gets too small. Then you analyze how these things change — how much density you lose and how much you lose in the sizes of X and Y — and you prove that in the end Y is big enough.

We have two lemmas on the board, and the one about big blue steps is perhaps the most surprising, so we'll now prove it. Before we do it, we'll quickly summarize the lemma $\sum_{x,y \in X} \omega(x, y) = 0$. This is basically a convexity argument counting red paths (or walks) from x to Y to y. If you analyze this with convexity or Cauchy–Schwarz, you get the desired inequality. (This is not obvious, but not particularly difficult.)

We'll now sketch the proof of the big blue step lemma.

Proof Sketch. We want to show that whenever our algorithm leads us to a big blue step — meaning there are many vertices $x \in X$ with reasonably large blue neighborhoods — we can find a blue book with $|S| \geq \ell^{1/4}$. It sufices to show that there exists a blue book (S, T) with $|T| \geq \mu^{|S|} \cdot |X| / 2$ (i.e., so that the book satisfies our condition) and $|S| = \left\lceil \ell^{1/4} \right\rceil$. (Then the one with maximal $|S|$ certainly has $|S|$ at least this quantity.)

Let $W \subseteq X$ be the set of vertices which satisfy our condition for a big blue step to happen — i.e., W is the set of vertices $x \in X$ with $|N_B(x) \cap X| \geq \mu |X|$. Then we know that $|W| \geq R(k, \left\lceil \ell^{2/3} \right\rceil)$. This tells us we can find a red clique of size k or a blue clique of size $\left\lceil \ell^{2/3} \right\rceil$ in W. But the first case is excluded, since the entire graph doesn't have a red clique of size k; this means we can find a blue clique U of size $|U| = \left\lceil \ell^{2/3} \right\rceil$ in W.

Now we consider $d_B(U, X \setminus U)$ (U is a blue clique living in W and therefore X, and we want to look at the blue edges between U and the rest of X). By definition, this is

$$d_B(U, X \setminus U) = \frac{e_B(U, X \setminus U)}{|U| \cdot |X \setminus U|} \geq \frac{|U| (\mu |X| - |U|)}{|U| \cdot |X \setminus U|}$$

(every vertex in U is in particular in W, so it has at least $\mu |X|$ edges into X; some of them might end up in U instead of $X \setminus U$, which is why we have the subtraction). This gives

$$d_B(U, X \setminus U) \geq \mu - \frac{|U|}{|X \setminus U|} \geq \mu - \frac{1}{k}$$

(the $1/k$ is very loose — $|X|$ is at least $R(k, \left\lfloor \ell^{3/4} \right\rfloor)$ and $|U|$ is $\ell^{2/3}$, so this error term is *tiny*).

We want S to be a blue clique of size $\ell^{1/4}$; we'll obtain this by looking inside U. So we take S to be a uniformly random subset of U of size $|S| = \left\lceil \ell^{1/4} \right\rceil$. Now we let T be the set of vertices $y \in X \setminus U$ such that $S \subseteq N_B(y)$ — i.e., all vertices such that S only has blue edges to y, so that S only has blue edges to T.

We want T to be reasonably large, so we're happy if $\mathbb{E}[T]$ is at least the bound we want.

To compute $\mathbb{E}[T]$, by linearity of expectation we can consider every vertex $y \in X \setminus U$ and consider the probability that we can include it. The probability that all edges from y to S are blue is

$$\frac{\binom{N_B(y) \cap U}{\lceil \ell^{1/4} \rceil}}{\binom{\lceil \ell^{2/3} \rceil}{\lceil \ell^{1/4} \rceil}}.$$

(This is the number of good choices for S over the total number.) Now we have

$$\mathbb{E}[T] = \sum_{y \in X \setminus U} \frac{\binom{N_B(y) \cap U}{\lceil \ell^{1/4} \rceil}}{\binom{\lceil \ell^{2/3} \rceil}{\lceil \ell^{1/4} \rceil}}.$$

The function $\binom{a}{b}$ is convex (if you define $\binom{a}{b}$ to be 0 for $b > a$, even for real numbers). So by convexity, this is at least $|X \setminus U|$ times the *average* of the quantity in the binomial coefficient; this gives

$$\mathbb{E}[T] \geq |X \setminus U| \cdot \frac{\binom{(\mu - 1/k)|U|}{\lceil \ell^{1/4} \rceil}}{\binom{\lceil \ell^{2/3} \rceil}{\lceil \ell^{1/4} \rceil}}.$$

After some calculations, this is at least

$$\frac{1}{2} \mu^{\lceil \ell^{1/4} \rceil} |X| \geq \mu^{|S|} \frac{|X|}{2}.$$

(To intuitively see where these terms come from, you can imagine estimating $\binom{a}{b}$ as a^b — in particular, this is why we have an exponent $\left\lceil \ell^{1/4} \right\rceil$ of $|U|$, while everything else will cancel.) □