# 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 Ramsey's theorem for graphs and hypergraphs

## §1.1 Introduction and 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, which essentially states the following.

**Theorem 1.1** (Ramsey's theorem, informally)

Every very large graph must contain a large clique or independent set.

In general, Ramsey theory is about taking a large combinatorial structure, and finding 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 states that no matter what the structure of the original graph looks like, as long as it is large enough, we can find one of these two very simple graphs inside it.

An example of Ramsey's theorem — the one usually used to explain Ramsey theory to non-mathematicians — is the following.

**Example 1.2**

Among any 6 people, there must be 3 who have pairwise met or 3 who have pairwise not met.

**Remark 1.3.** The connection is that you can interpret people as nodes of a graph, with edges between two people who have met. This example is typically phrased to be about people knowing each other (rather than having met), but the phrasing here makes it more clear that the relation is symmetric.

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 it to more colors later, and to make the symmetry clearer).

**Theorem 1.4** (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 with red and blue, we can find a red clique of size $k$ or a blue clique of size $\ell$.

This perspective (where we start with a complete graph, i.e., a graph with all edges, and color the edges with two colors) is equivalent to the one in the first statement (where we take an arbitrary graph and consider edges and non-edges) if we imagine taking the graph consisting of only the red edges.

**Remark 1.5.** Of course, it is possible that both options occur. But we really do need the two options — if we only had one, the graph where all edges were of the wrong color would be a counterexample.

**Remark 1.6.** If this statement holds for some number $R(k, \ell)$, then it automatically holds for all larger numbers as well (since we can simply look at a subset of vertices of size $R(k, \ell)$).

**Remark 1.7.** In Example 1.2, we had $k = \ell = 3$. In theory, allowing $k$ and $\ell$ to be different does not make the statement any stronger — if we only had the statement in the 'diagonal' case (where $k = \ell$), we could deduce the general case by applying the diagonal case to $\max(k, \ell)$. But even if we only care about the case $k = \ell$, writing the statement in this way is useful because we'll use induction. Furthermore, there are situations where it's natural to *not* have $k = \ell$. Although allowing $k \neq \ell$ doesn't make the theorem stronger as written, we generally care not just about the fact that $R(k, \ell)$ exists, but also about how large it needs to be. Then allowing $k \neq \ell$ does matter — the answer for $(1000, 5)$ may be very different from the answer for $(1000, 1000)$.

*Proof of Theorem 1.4.* We use induction on $k + \ell$.

The base case is when $k = 2$ or $\ell = 2$; in this case, the theorem is easy. If $k = 2$, we can take $R(2, \ell) = \ell$ — for any red-blue coloring of the edges of a complete graph with $\ell$ 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 $\ell$. (The case $\ell = 2$ is symmetric.)

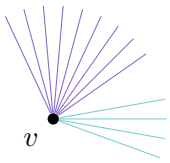
Now assume that $k, \ell > 2$, and that we have already proven the theorem for all $k', \ell' \geq 2$ with $k' + \ell' < k + \ell$. (Here we only need the inductive hypothesis for $k' + \ell' = k + \ell - 1$, but in general when doing induction, we might as well assume that we've proven the statement for all smaller values, as it doesn't lose anything and is sometimes helpful.)

Then we claim that taking

$$R(k, \ell) = R(k - 1, \ell) + R(k, \ell - 1)$$

works. To see this, 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 $\ell$.

First fix an arbitrary vertex $v$. We'll use $v$ as a starting point of exploration (but it won't necessarily be contained in our final clique).




Then $v$ has $R(k - 1, \ell) + R(k, \ell - 1) - 1$ neighbors, 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.

First assume that $v$ has at least $R(k - 1, \ell)$ red neighbors. Then among these red neighbors, we can find a red clique of size $k - 1$ or a blue clique of size $\ell$. If we've found a blue clique of size $\ell$, then we're done (as it's also a blue clique of size $\ell$ in the original graph). Meanwhile, if we've found a red clique of size $k - 1$, then since all its vertices are red neighbors of $v$, we can add $v$ to it to get a red clique of size $k$.


The second case, where $v$ has at least $R(k, \ell - 1)$ blue neighbors, is analogous. So in either case, we can find a red clique of size $k$ or a blue clique of size $\ell$ in the original graph. □

**Remark 1.8.** Note that even if we only care about the case $k = \ell$, the inductive proof requires us to consider cases with $k \neq \ell$ as well.

Of course, this isn't the end of the entire class — this theorem is just the starting point of Ramsey theory, and there's many directions to go from here. One direction is the following.

**Question 1.9.** Can we prove theorems of this flavor in other combinatorial structures?

For example, we could 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 setting of graphs completely, and prove statements about integers (which we'll see next week) or statements of a geometric flavor (starting with points in a plane and trying to find a subcollection with nice geometric properties).

Another direction is to consider the *size* of these numbers:

**Question 1.10.** How large does $R(k, \ell)$ have to be?

In this class, we will 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 for which the statement in Ramsey's theorem holds.

Note that we redefined $R(k, \ell)$ (in the statement of Theorem 1.4, we only required $R(k, \ell)$ to be *any* number such that the theorem statement holds); in particular, the statement $R(k, \ell) = R(k-1, \ell) + R(k, \ell-1)$ in our proof of Theorem 1.4 is no longer true.

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

This is an open problem. Determining the Ramsey numbers exactly is very difficult, even for very small values of $k$ and $\ell$ — for example, we know that $R(2, \ell) = \ell$, $R(3, 3) = 6$, and $R(4, 4) = 18$, but we don't even know $R(5, 5)$. But in extremal combinatorics, we don't generally care about the *exact* values — instead, we want to find bounds that determine how $R(k, \ell)$ grows as a function of $k$ and $\ell$. (In particular, most people in extremal combinatorics don't really care about what the exact Ramsey number $R(5, 5)$ is so much as the *growth rate* of the Ramsey numbers.) But even understanding the correct behavior of $R(k, \ell)$ is an open problem — we'll see some bounds on $R(k, \ell)$ shortly, but the correct ones aren't known.

First, we can obtain recursive bounds on $R(k, \ell)$ using the above proof of Theorem 1.4:

- We have $R(2, \ell) = \ell$ and $R(k, 2) = k$.
- We have $R(k, \ell) \leq R(k-1, \ell) + R(k, \ell-1)$ — our proof showed that the statement of Theorem 1.4 for $k$ and $\ell$ holds when we have $R(k-1, \ell) + R(k, \ell-1)$ vertices, and since $R(k, \ell)$ is defined to be the *smallest* number of vertices for which the theorem statement holds, it must be at most this quantity.

This gives us the following closed-form bound.

**Theorem 1.13** (Erdős–Szekeres 1935)
For all $k, \ell \geq 2$, we have $R(k, \ell) \leq \binom{k+\ell-2}{k-1}$.

**Remark 1.14.** Of course, the Ramsey numbers are symmetric in $k$ and $\ell$ — i.e., $R(k,\ell) = R(\ell,k)$ — as we can swap the two colors. The bound in Theorem 1.13 may not *look* symmetric, but it actually is — we have $k+\ell-2 = (k-1)+(\ell-1)$, so $\binom{k+\ell-2}{k-1} = \binom{k+\ell-2}{\ell-1}$.

*Proof.* We use induction, using the recursive bounds stated earlier. For the base cases, we have $R(2,\ell) = \ell = \binom{2+\ell-2}{2-1}$, and 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

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

## §1.3 Bounds on diagonal Ramsey numbers

The most natural case to consider is when $k = \ell$; the Ramsey numbers $R(k,k)$ are called *diagonal* Ramsey numbers. For this case, Theorem 1.13 gives the following upper bound.

**Corollary 1.15**

For all $k \geq 2$, we have $R(k,k) \leq \binom{2k-2}{k-1} \leq 4^k$.

**Remark 1.16.** More precisely, we have $\binom{2k-2}{k-1} = O(4^k k^{-1/2})$; however, $k^{-1/2}$ is much smaller than $4^k$, so we won't really care about it here.

Understanding the growth rate of $R(k,k)$ is one of the most famous problems in extremal combinatorics, and there's been a series of improvements over this upper bound. The best known bound is due to Ashwin Sah; it's still of the form $4^k$ divided by something sub-exponential. (*Note: This was true at the time of this lecture, but since then, there has been an exponential improvement (due to Campos, Griffiths, Morris, and Sahasrabudhe)! This is discussed in the final two lectures.*)

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

To answer this, we'll now discuss *lower* bounds — to prove a lower bound, we want to find some coloring of a large graph where we *don't* have a big red clique or blue clique.

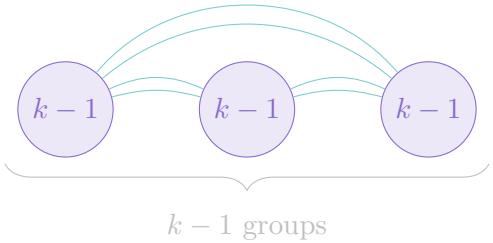
First, here are some examples of such constructions.

**Example 1.18**

We have $R(k,k) > k-1$ — *any* coloring of a graph on $k-1$ vertices won't have a red or blue $k$-clique.

**Example 1.19**

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




To see that this construction works, 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 of $(k-1)^2$ is much smaller than $4^k$. But in fact, it's extremely difficult to come up with explicit constructions — we don't know of any explicit construction that even gives an exponential lower bound. However, we *do* know the following lower bound.

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

Combined with the upper bound in Theorem 1.13, this tells us that the behavior of $R(k,k)$ is exponential in $k$, but we don't know the correct base of the exponent — $\sqrt{2}$ and 4 are still the best known bounds.

February 9, 2023

How can we have an exponential lower bound but no exponential construction? The reason is that the proof of this theorem is *probabilistic* — it takes a random coloring.

*Proof.* First we will eliminate the case where $k$ is small (since we'll later use bounds that only hold when $k$ is not very small). If $k \leq 4$, then we have $R(k,k) \geq (k-1)^2 \geq 2^{k/2}$ using the construction in Example 1.19. So we'll now assume that $k \geq 5$.

Let $n = \lfloor 2^{k/2} \rfloor$. Consider a complete graph on $n$ vertices, and color its edges red or blue *randomly* — we color each edge 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.) To do so, 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 forming our clique.

We will show that this quantity is smaller than 1; this means the *average* number of monochromatic cliques (over all colorings) is less than 1, so there must be some coloring where there are no monochromatic cliques. To do so, we 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

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

If $k \geq 5$ then this is less than 1, as desired.

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

**Remark 1.21.** 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

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

which is better by a factor linear in $k$.

**Remark 1.22.** We could also use the alteration method — we color randomly and show that the expected number of monochromatic cliques is *small*, and then delete one vertex from each to get rid of them. However, this only improves the bound by a factor of about $\sqrt{2}$.

The best-known lower bound is

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

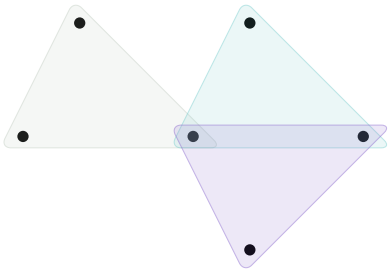
(which is a factor of 2 better than the bound here); it was proven by Spencer in 1977, by combining this argument with the Lovász local lemma.

## §1.4 Generalizations to hypergraphs and more colors

Next, we'll prove Ramsey's theorem for a generalization of graphs known as *hypergraphs*.

**Definition 1.23.** A $r$-uniform hypergraph is a collection of size-$r$ subsets of a ground set $V$.

Here we think of the ground set $V$ as our vertex set, and the size-$r$ subsets as edges in the hypergraph. 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 sizes of the subsets are not fixed.)



(This depicts a 3-uniform hypergraph, with each colored triangle representing one edge.)

**Definition 1.24.** The complete $r$-uniform hypergraph on $V$ is the hypergraph whose edges are *all* subsets of $V$ of size $r$. More generally, in any $r$-uniform hypergraph on vertex set $V$, a clique is a subset $S \subseteq V$ for which all possible subsets of $S$ are edges.

In other words, a clique is a copy of the complete $r$-uniform hypergraph inside some larger hypergraph. As in the setting of graphs, we say the *size* of such a clique is $|S|$, the number of vertices it contains.

We can now ask the exact same question from last class for hypergraphs.

**Question 1.25.** Given $k$ and $\ell$, if we color the edges of a $r$-uniform hypergraph on sufficiently many vertices with red and blue, can we always find a red clique of size $k$ (i.e., $k$ vertices such that all size-$r$ subsets are colored red) or a blue clique of size $\ell$?

The answer is yes, and the proof is essentially the same as last class; so we will make the question a bit more general by allowing $t$ colors instead of just two.

**Theorem 1.26** (Hypergraph Ramsey)

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: In any coloring of the edges of a complete $r$-uniform hypergraph with $R_r(k_1, \dots, k_t)$ vertices with colors $1, \dots, t$, there exists a clique of size $k_i$ in color $i$ for some $i$.

In other words, as long as we have enough vertices in our large coloring, there exists *some* color for which we can find a clique of its corresponding size. (It's possible that we can do so for more than one color.)

*Proof.* We'll again use induction — we will use double induction with the outer induction on $r$ and the inner induction on $k_1 + \cdots + k_t$. (We could also phrase this proof in terms of taking a minimal counterexample — then we'd consider a counterexample with minimal $r$, and among all such counterexamples, we'd take one with minimal $k_1 + \cdots + k_t$.)

Suppose first that $k_i = r$ for some $i$; without loss of generality, suppose that $k_t = r$. Then by induction, we know that the statement holds for $r$ and $k_1, \dots, k_{t-1}$ (as $k_1 + \cdots + k_{t-1} < k_1 + \cdots + k_t$). Now taking $R_r(k_1, \dots, k_{t-1})$ vertices suffices — consider any coloring of the complete $r$-uniform hypergraph on this many vertices with colors $1, \dots, t$. If there is an edge of color $t$, then we automatically get a clique of size $r$ in color $t$. Otherwise, color $t$ doesn't appear, which means we actually have a coloring using only colors $1, \dots, t-1$; then by the inductive hypothesis there must be a clique of size $k_i$ in color $i$ for some $i \in [t-1]$.

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

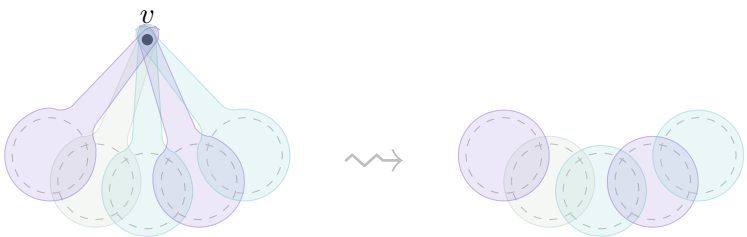
Now suppose $k_1, \dots, k_t > r$. The inductive hypothesis means that we have already proved the theorem for all smaller values of $r$ (regardless of $k_1 + \cdots + k_t$), and also for the same value of $r$ and all smaller values of $k_1 + \cdots + k_t$. Then we claim that taking

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

vertices suffices. Here the inner quantities all exist by the inner induction, and we can take $R_{r-1}$ of these quantities by the outer induction. (The numbers this produces are horrendously big.)

Consider some $r$-uniform hypergraph on a vertex set $V$ of at least this size, with its edges colored $1, \dots, t$.

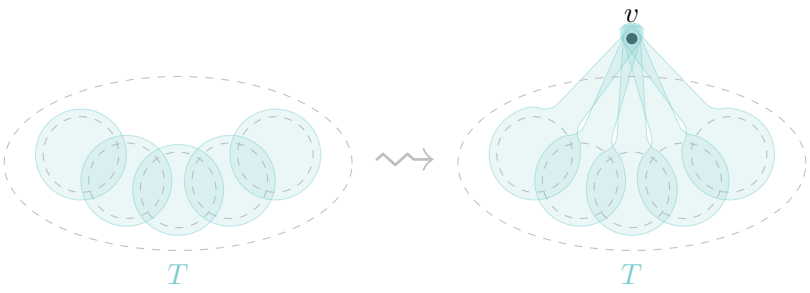
First, fix a vertex $v$. Then for every subset $S \subseteq V \setminus \{v\}$ of size $r-1$, we can look at the color of $S \subseteq \{v\}$. This defines a coloring of the complete $(r-1)$-uniform hypergraph on $V \setminus \{v\}$ with our $t$ colors.



Now we can apply the inductive hypothesis on this $(r-1)$-uniform hypergraph — for some $i \in [t]$, we can find a clique $T$ of size

$$|T| = R_r(k_1, \dots, k_i - 1, \dots, k_t)$$

of color $i$, with respect to the coloring of the complete $(r-1)$-uniform hypergraph on $V \setminus \{v\}$. This means $T$ has the property that any $(r-1)$-vertex subset of $T$ together with $v$ has color $i$ in the original coloring.




Now apply the induction hypothesis to the complete $r$-uniform hypergraph on $T$ (where we take the original coloring and restrict it to $T$ — we're *not* looking at the coloring of the $(r-1)$-uniform hypergraph obtained from removing $v$ anymore). This tells us there is a clique of one of the specified sizes and colors. There are two possibilities:

**Case 1** (For some $j \neq i$, there is a clique of size $k_j$ of color $j$). Then we're done, since this clique (in the restriction of the original coloring to $T$) is also a clique of size $k_j$ in the original graph.

**Case 2** (There 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 $i$. This is because every $r$-vertex subset *not* involving $v$ has color $i$ (by our choice of this subset as a $r$-uniform clique of color $i$), while every $r$-vertex subset which *does* involve $v$ has color $i$ by our construction of $T$.

Either way, we've found a clique of the desired color and size, so we're done. □

February 14, 2023

# §2 Ramsey-type results in geometric settings

This week, we'll see some other settings in which Ramsey-type questions are commonly studied. The first setting we'll look at is a geometric setting, where we look at points in a plane.

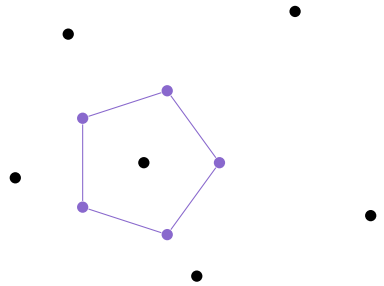
## §2.1 Points in convex position

The first result of this type was the Erdős–Szekeres theorem on point sets in convex position. (There is another theorem by Erdős and Szekeres in a different setting, about monotone subsequences; these are not the same.)

**Theorem 2.1** (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 such that no three are collinear, one can find $k$ points forming a convex $k$-gon.

In Ramsey-type results, we're always looking for a certain structure; and we want to say that we can find it as long as our original structure is big enough. Here there's no coloring, so the structure we're trying to find isn't some monochromatic thing; instead, it's $k$ 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 (Theorem 1.26) in two ways.

*Proof 1 of Theorem 2.1.* 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 take $r = 4$ and use two colors, *red* and *blue*. We'll 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 points (note that we're coloring sets of 4 points, not the points themselves).

Now we'll apply Ramsey's theorem for hypergraphs. We'll take the target clique sizes to be $k_{\text{blue}} = k$ and $k_{\text{red}} = 5$. The idea behind this is that we *can't* have a red 5-clique — we'll soon show that it's not possible to have 5 points such that all 4-tuples are not in convex position. Then Ramsey's theorem guarantees that we can either find a red 5-clique, which is impossible (so this definitely won't happen), or a blue $k$-clique, which gives the structure we want. (So we're essentially making a red clique impossible to find, which forces there to be a blue clique.)

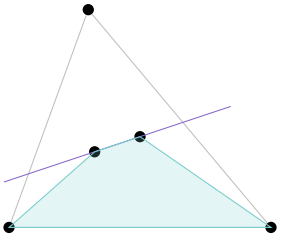
So we'll take $K = R_4(5, k)$. Now we'll give the details for why this works.

**Claim 2.2 —** 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 5 points looks like.

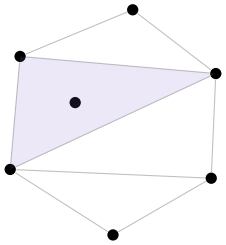
If the convex hull is a pentagon, then we're done (as *any* four points form a convex quadrilateral). If it's a quadrilateral, then we're also done (those four points form a convex quadrilateral). So the only case that requires an argument is the one where the convex hull is a triangle.

In this case, we can take the two points on the inside of the triangle, and draw the line through them. This line can't 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 our 4-uniform hypergrahp cannot have a red clique of size 5 (as this would cause 5 points among which no 4 form a convex quadrilateral). Then when we apply Ramsey's theorem to our hypergraph on $K = R_4(5, k)$ vertices, 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. We claim this means the $k$ points must themselves form a convex $k$-gon. Assume not. Then their convex hull doesn't contain all $k$ points, so at least one of the points is inside it. Then we can triangulate the convex hull in any way. The point on the inside must lie in the interior of one of these triangles (it can't lie on a boundary because no three points are on a line), giving four points which don't form a convex quadrilateral. □



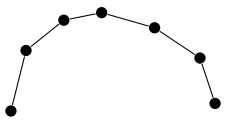
There's another proof which also applies Ramsey's theorem for hypergraphs, but in the 3-uniform setting (where we're coloring *triples* of points). Now it's not clear what our coloring should be — all triangles are convex, so we can't color triples based on convexity. Instead, the coloring we'll use will depend on the *direction* of the points. (Interestingly, this proof will depend on how the points are rotated — the theorem statement isn't affected if we rotate the points, but the proof will be.) This proof turns out to give a much better bound than the first.

*Proof 2 of Theorem 2.1.* First, we can assume that no line through two points is vertical (we can do this because there are finitely many lines, so we can rotate so that none are vertical).

For any three points, draw the lines between the leftmost and middle point, and the middle and rightmost point. If the graph this creates is concave, we call the three points a *cap*; if it is convex, we call the three points a *cup*.


Now take $K = R_3(k, k)$. Color a 3-tuple of points *red* if they form a cap, and *blue* if they form a cup. Then by Ramsey's theorem, we can find a red or blue $k$-clique — i.e., $k$ 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 (the fact that every consecutive triple forms a cap means that the slopes grow smaller at every step) — i.e., they 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, these $k$ points are in convex position. □

## §2.2 Caps and cups

The bounds we get from these proofs of Theorem 2.1 are not great, because the hypergraph Ramsey bounds are not great. But in fact, once we come up with the approach of looking for caps or cups of size $k$, we can take this idea further and forget about Ramsey's theorem, and prove a statement about caps and cups using induction. (It's very difficult to induct on Theorem 2.1 directly — it's difficult to induct on a convex $k$-gon. But caps and cups glue nicely together, which makes induction work much better.) This will give a much better bound for Theorem 2.1 (this is the primary motivation for studying caps and cups).

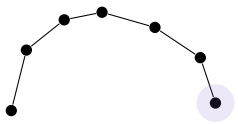
**Theorem 2.3** (Erdős–Szekeres 1935)
Let $k, \ell \geq 3$, and consider $\binom{k+\ell-4}{k-2} + 1$ points in the plane such that no three are collinear and no two are on a vertical line. Then we can find a cap of size $k$ or a cup of size $\ell$.

As with Ramsey's theorem for graphs, the induction works much better if we allow different sizes for the cap and the cup, even though the case we care about most is the one where they're the same size. Also, as with Ramsey's theorem, we really need the two cases (if the points we start with all form a cap, we can't find a cup of any size, and vice versa). The proof (and the bound we end up with) is also very similar to Ramsey's theorem.

*Proof.* We induct on $k + \ell$. First we'll deal with the case where one of the numbers is 3. If $k = 3$, 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, in which case the $\ell$ points form a cup of size $\ell$. The case $\ell = 3$ is analogous.

Now assume $k, \ell > 3$, and suppose we've proven the statement for all smaller $k+\ell$. Assume for contradiction that there is no cap of size $k$ or cup of size $\ell$. Then by applying the theorem to $(k-1, \ell)$, since we can't find a $\ell$-cap, we can find a $(k-1)$-cap. In fact, we can find *many* $(k-1)$-caps (we can find one and then remove one of its vertices, and there's still enough vertices to guarantee another).

Our goal is to extend our $(k-1)$-caps to $k$-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 that occur as endpoints (i.e., rightmost points) of a $(k-1)$-cap.

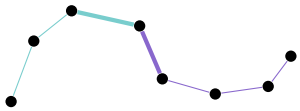


We claim that there are many points with this property: If we delete all these points, then there can't be any $(k-1)$-caps left (we've deleted the rightmost point form each). But there are also no cups of size $\ell$ left, because there were none to start with. This tells us the number of points left must be at most $\binom{k+\ell-5}{k-3}$ by the induction hypothesis on $(k-1, \ell)$ (if there were more points, then we could find a $(k-1)$-cap or $\ell$-cup). So there are 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 $\binom{k+(\ell-1)-4}{k-2}+1$, so we can apply the induction hypothesis for $(k, \ell-1)$ to these points; this tells us we can find either a cap of size $k$ (which is impossible by assumption) or a cup of size $\ell-1$.

Now focus on the first (i.e., leftmost) point of this cup of size $\ell-1$; we know this point is also the endpoint of a cap of size $k-1$. So we have 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 $k$.

Similarly, if it has lesser slope, then we can extend the cup of size $\ell-1$ to a cup of size $\ell$.

This produces a cap of size $k$ or a cup of size $\ell$, so we're done. □

In particular, applying Theorem 2.3 with $k=\ell$ shows that among any $\binom{2k-4}{k-2}+1$ points, there is a cap or 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 gives a bound for Theorem 2.1 of

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

This is much better than the bounds given by the hypergraph Ramsey numbers from the previous proofs.

In Theorem 2.3, the bound $\binom{k+\ell-4}{k-2}+1$ is tight — there is a recursive construction with $\binom{k+\ell-4}{k-2}$ points but no $k$-caps or $\ell$-cups. However, this doesn't necessarily imply that this is the correct number for Theorem 2.1, since there we're trying to find a weaker structure.

**Conjecture 2.4** (Erdős–Szekeres) — The precise best possible bound for Theorem 2.1 is $K = 2^{k-2}+1$.

There is a construction proving that we need $K \geq 2^{k-2}+1$. This conjecture is still open, but we do know that $K \leq 2^{k+o(k)}$ (Suk 2016) — so the bound is asymptotically almost known, but the precise number isn't.

**Remark 2.5.** The problem of determining the best $K$ is called the *happy ending problem*. This is because the people first thinking about the problem were Erdős, Szekeres, and Esther Klein; and Szekeres and Klein eventually got married.

# §3 Ramsey-type results in arithmetic settings

Last class, we saw Ramsey-type results in geometric settings. 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.

## §3.1 Schur's theorem

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

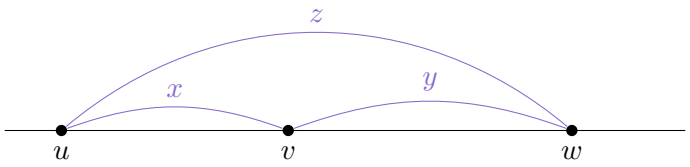
**Theorem 3.1** (Schur 1916)

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 $N = R(3, 3, \dots, 3)$ (with $t$ colors).

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

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 have the same color, and $z = x + y$, as desired. □

**Remark 3.2.** We're allowed to have $x = y$; if we required $x$ and $y$ to be distinct, this proof wouldn't work.

There's a lot of theory about which equations we can replace $x + y = z$ with in Theorem 3.1. This question was resolved by Rado, and we'll see this in a few weeks.

## §3.2 van der Waerden's theorem

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 can look for longer arithmetic progressions. This was answered by van der Waerden in 1927.

**Theorem 3.3** (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 $\{1, \dots, N\}$ with $t$ colors, there is a monochromatic arithmetic progression of length $k$ (i.e., distinct $x_1, \dots, x_k \in \{1, \dots, N\}$ 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. Note that in both Theorem 3.1 and Theorem 3.3, if the statement is true for some $N$, it's also true for all larger $N$ (as with Ramsey's theorem) — we can simply ignore all the extra numbers.

There's one significant difference between van der Waerden's theorem and Schur's theorem or Ramsey's theorem. In Schur's theorem, suppose that someone gave us a coloring with red, blue, and green, and we were looking for a solution to $x + y = z$. If half the points were red, a quarter were blue, and a quarter were green, we might start by trying to find one in red. But although Schur's theorem guarantees that we can find a solution in *some* color (as long as $N$ is large enough), this color might not be the most common one — for example, it could be that all the odd numbers are red (then red is the most common color, but it doesn't have any solutions).

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

So in both Schur's theorem and Ramsey's theorem, just looking at the biggest color class isn't enough. But interestingly, for van der Waerden's theorem, it *is* enough — 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 a $\frac{1}{t}$-fraction of all elements.

**Theorem 3.4** (Szemerédi 1975)

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

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 have at least a $\frac{1}{t}$-fraction of all points), and is much harder to prove.

**Remark 3.5.** We stated this for *all* sufficiently 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 specific $N$ to all larger ones.

**Remark 3.6.** We're not going to prove this. But it's 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. (Szemerédi proved it using regularity theory for hypergraphs; we won't see this in this class, but you can see *graph* regularity in **18.225**.)

## §3.3 The Hales–Jewett theorem

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

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 $k = 10$, so we're looking for a 10-term arithmetic progression).

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

**Example 3.7**

One arithmetic progression of this form is

$$10600108,\ 11601118,\ 12602128,\ 13603138,\ \ldots,\ 19609198.$$

We'll denote this progression by the 'code' 1∗60∗1∗8; then we get this progression by replacing the ∗'s with each of the digits 0 through 9.

Informally speaking, the Hales–Jewett theorem says that 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.

**Notation 3.8.** We write $[k]$ to denote $\{1, \ldots, k\}$, and $[k]^n$ for the set of $n$-tuples with elements in $[k]$.

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

**Definition 3.9.** A combinatorial line in $[k]^n$ is a subset of $[k]^n$ with size $k$ that can be obtained from some $n$-tuple $\lambda$, whose entries are all 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 $i$th point of the combinatorial line is obtained by replacing all $*$'s in $\lambda$ by $i$.

This is exactly what we did in Example 3.7 (with $\{0, \ldots, 9\}$ instead of $\{1, \ldots, 10\}$) — we started with $\lambda = (1, *, 6, 0, *, 1, *, 8)$ and obtained the first point by replacing all $*$'s by 0, the second by replacing all $*$'s by 1, and so on.

**Definition 3.10.** We say the endpoint of a combinatorial line is its $k$th point, i.e., the one where all $*$'s are replaced by $k$.

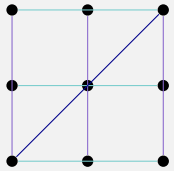
(The definition of a combinatorial line is standard. The definition of its endpoint is not, but we'll use it in the proof.)

**Theorem 3.11** (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 is a monochromatic combinatorial line.

This is again a case where if the statement is true for some $n$, then it's automatically true for all larger $n$.

**Remark 3.12.** The original reason Hales and Jewett came up with this was to generalize the game of Tic–Tac–Toe. If $k = 3$ and $n = 2$, then combinatorial lines describe some of the winning configurations in Tic–Tac–Toe (all but the northwest-southeast 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 Theorem 3.3 using Theorem 3.11.* Given $k$ and $t$ in van der Waerden's theorem, we'll apply Hales–

Jewett with the same $k$ and $t$; this gives some $n$. It's possible to take $N = k^n$ and use the argument from our motivation for Hales–Jewett, but we'll see a different proof that gets a better bound.

Take $N = kn$. Suppose we're given a coloring $\gamma : \{1, \dots, N\} \to \{1, \dots, t\}$; we wish to find an arithmetic progression whose $\gamma$-values are all equal. To apply the Hales–Jewett theorem, we first want to construct a coloring of $[k]^n$. From $\gamma$, we construct a coloring $\gamma' : [k]^n \to \{1, \dots, t\}$ where for all $x_1, \dots, x_t \in [k]$, we set

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

(Note that $x_1 + \cdots + x_n \le kn$, so this coloring is well-defined.)

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

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

*Proof of Theorem 3.11.* We'll use induction on $k$. The case $k = 1$ is easy — $[1]^n$ consists of a single point, and that single point is a combinatorial line. Now fix $k \ge 2$, and assume we've already proven the theorem for $k - 1$ (and *any* number of colors $t'$ — $t'$ can be much bigger than $t$, since we're only inducting on $k$).

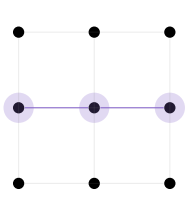
We know that the theorem holds for $k-1$, so we could restrict our coloring to $[k-1]^n$ and find a combinatorial line there. This would give us a combinatorial line of $k - 1$ points; in order to get a combinatorial line in $[k]^n$, we'd need to add the endpoint. This endpoint might not be the right 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.

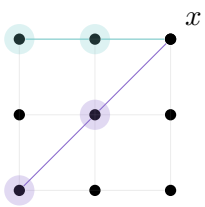
We'll prove this statement by induction as well. First, we'll now fix the number of colors $t$.

**Claim 3.13 —** 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 hoods:

- There exists a monochromatic combinatorial line in $[k]^n$.
- There 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.

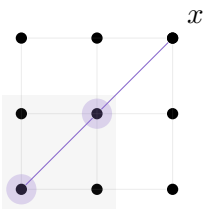
So for every coloring, there are two cases — either we're happy (we've found a complete combinatorial line), or we've found *many* combinatorial lines of distinct colors with the same endpoint, such that all are monochromatic if we omit this endpoint.





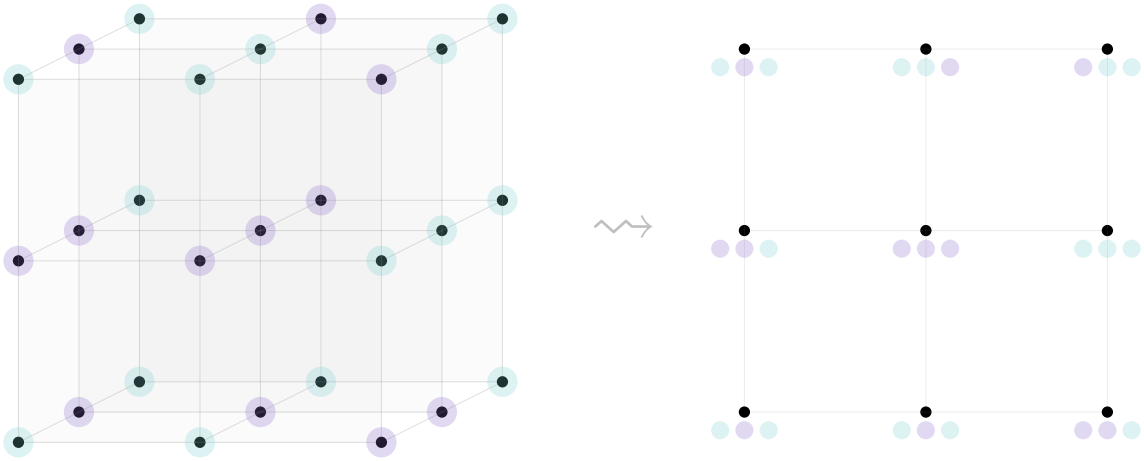

Note that this claim gets strictly stronger as we increase $j$; we state it for all $j$ because we'll induct on $j$. The reason this claim suffices is becaues taking $j = t$, either we're done or we have a line $\ell_i$ for every color; then one of these lines $\ell_i$ matches the color of $x$, giving a monochromatic line.

*Proof.* We use induction on $j$. If $j = 1$, 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 $\ell$ in $[k]^n$ with some endpoint $x$ (the point obtained by plugging $k$ into the code of $\ell$) such that $\ell \setminus \{x\}$ is monochromatic.



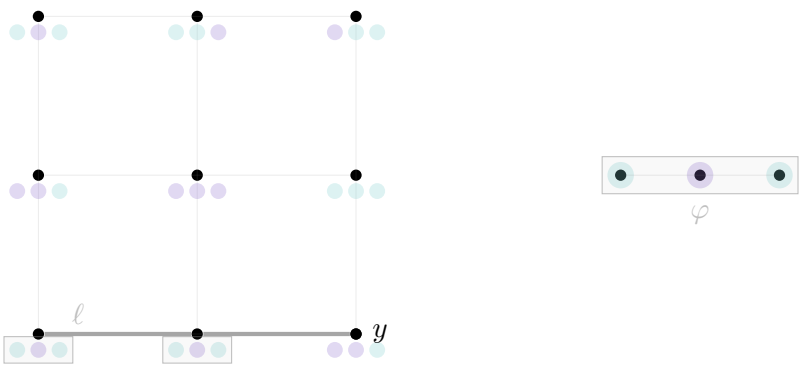

Now assume that $j \geq 2$, and that we've already proven the claim for $j-1$ (with the same $k$ and $t$). Let $n_1$ be as in the claim for $j-1$, and let $n_2$ be as in the Hales–Jewett theorem for $k-1$ with $t^{k^{n_1}}$ colors. (This is huge, but it's a well-defined finite number.) Let $n = n_1 + n_2$ (this is gigantic, but it's a number).

Consider a coloring on $[k]^n = [k]^{n_1} \times [k]^{n_2}$ with $t$ colors. Our first step is to view this coloring as a coloring on $[k]^{n_2}$ with $t^{k^{n_1}}$ colors — for every point $y \in [k]^{n_2}$, we get a list of $k^{n_1}$ colors coming from all the points $(x, y) \in [k]^n$ (where $x$ ranges over $[k]^{n_1}$). There are $t^{k^{n_1}}$ possible lists, so this gives a coloring of $[k]^{n_2}$ with this many colors. (In the below illustration, $k = 3$, $n_1 = 1$, and $n_2 = 2$.)



By our choice of $n_2$, there is a combinatorial line $\ell$ in $[k]^{n_2}$ in this coloring with some endpoint $y$ such that $\ell \setminus \{y\}$ is monochromatic in this coloring. Let $\lambda \in ([k] \cup \{*\})^{n_2}$ be its corresponding code.

The color of this monochromatic set $\ell \setminus \{y\}$ corresponds to a coloring $\varphi : [k]^{n_1} \to \{1, \ldots, t\}$, where for all $a \in [k]^{n_1}$ and $b \in \ell \setminus \{y\}$, the point $(a, b) \in [k]^n$ has color $\varphi(a)$ in the original coloring. (In other words, a set being monochromatic in the new coloring on $[k]^{n_2}$ means that for all $a \in [k]^{n_1}$ and $b$ in this monochromatic set, the color of $(a, b)$ in the original coloring only depends on $a$.)



Now apply the claim for $j-1$ to this coloring $\varphi : [k]^{n_1} \to \{1, \dots, t\}$. This gives combinatorial lines $\ell_1$, …, $\ell_{j-1}$ in $[k]^{n_1}$, all with the same endpoint $x$, such that $\ell_1 \setminus \{x\}$, …, $\ell_{j-1} \setminus \{x\}$ are monochromatic in $\varphi$ (and have different colors). Let their corresponding codes be $\lambda_1, \dots, \lambda_{j-1} \in ([k] \cup \{*\})^{n_1}$.

Now look at the $j$ lines corresponding to the codes $(\lambda_1, \lambda)$, …, $(\lambda_{j-1}, \lambda)$, and $(x, \lambda)$ in $([k] \cup \{*\})^n$. (So we're appending $\lambda$ to each of our $j-1$ lines in $[k]^{n_1}$, and we get a $j$th line by appending it to the common endpoint instead.) These combinatorial lines all have endpoint $(x, y)$.

By definition, all these combinatorial lines are monochromatic if we omit their endpoint $(x, y)$. Also, each of the lines $(\lambda_i, \lambda)$ has a distinct color. If $(x, \lambda)$ agrees with one of these colors, then we get a monochromatic line (take the line defined by $(\lambda_i, b)$ for any $b \in \ell \setminus \{y\}$). Otherwise, we get $j$ lines with distinct colors (and a shared common endpoint). □

Then by induction, we're done (we can take $j = t$ to get a monochromatic combinatorial line). □

March 2, 2023

## §4 Off-diagonal Ramsey numbers

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

**Definition 4.1.** For $k, \ell \geq 2$, the Ramsey number $R(k, \ell)$ is the smallest number such that the following holds: In any coloring of the edges 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 $\ell$.

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

### §4.1 Upper bounds

Earlier, we proved the following general upper bound for all $k$ and $\ell$.

**Theorem 1.13** (Erdős–Szekeres 1935)
For all $k$ and $\ell$, we have

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

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

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

(where the constant is roughly $\frac{1}{\ell!}$). In some sense, this is a good upper bound — we've seen that $R(k, k)$ grows exponentially with $k$, while this bound (for fixed $\ell$) is polynomial in $k$. But we might want a more precise bound. And in fact, one can improve this bound. For example, for $R(k, 3)$, 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 $\ell$ one can also improve upon Theorem 1.13 by logarithmic factors; in the case $\ell = 3$ we know this improvement is the right answer, but in the general case the right answer is not known.

**Theorem 4.2** (Ajtai–Komlós–Szemerédi 1980)

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

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

For $\ell = 3$, we more precisely have $R(k, 3) \leq \frac{k^2}{\log k - 1}$.

If we think of $\ell$ as fixed, then $C^{\ell-1}$ is a constant factor. The main behavior of this bound is still $k^{\ell-1}$, but we have a logarithmic improvement in the denominator (and the number of log factors grows with $\ell$ — this makes sense because when $\ell = 2$ you don't have any log factors, as $k$ is the *exact* answer). In some sense, logarithmic factors might not seem like a big improvement. But it's far from trivial, and it's actually the right answer when $\ell = 3$. (We don't know whether it's the right answer when $\ell > 3$ — we don't even know if the exponent is right.)

**Remark 4.3.** The special bound for the $\ell = 3$ case is a bit better because it removes the absolute constant $C$. This bound (and the proof we'll see) is due to Shearer. The constant factor doesn't matter too much, but it's nice because this is still the best constant factor we know.

### §4.1.1 The case $\ell = 3$

Today we'll prove Theorem 4.2 in the case $\ell = 3$; we'll prove the general case next class.

To prove an upper bound on $R(k, 3)$, 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 might 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 $k$ forms an independent set of size $k$ in the blue graph. So our goal is to bound the independence number of a triangle-free graph (we wish to show that it's at least $k$, if the number of vertices is big enough).

**Definition 4.4.** The independence number of $G$, denoted $\alpha(G)$, is the size of the largest independent set in $G$ (i.e., the size of the largest subset $U$ of vertices such that $U$ contains no edges).

We'll do this using the following lemma.

**Lemma 4.5**

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

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

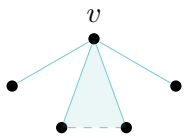
**Remark 4.6.** Without the triangle-free condition, there's a bound of $\alpha(G) \geq \frac{n}{d+1}$. You can get this bound when the *maximum* degree is $d$ by using a greedy algorithm, and it's also true for the *average* degree; this is called the Caro–Wei theorem (which can be proved probabilistically). Alternatively, if you don't care about constant factors, you can get a bound for the average-degree setting by running a greedy algorithm with only the vertices of degree at most $2d$; there are at least $\frac{n}{2}$ such vertices.

First we'll show why Lemma 4.5 implies the upper bound on $R(k, 3)$ in Theorem 4.2.

*Proof of Theorem 4.2 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. Let

$G$ be the graph consisting of the blue edges, which is triangle-free and satisfies $\alpha(G) \le k-1$ (an independent set in blue is a clique in red, so there can't be an independent set of size $k$).

Now we'd like to apply Lemma 4.5; to do so, we'd like to get a handle on the average degree of $G$. First, for every vertex $v$, its neighborhood in $G$ (i.e., its set of neighbors) must form an independent set — if there were an edge between two neighbors of $v$, we'd have a triangle consisting of those two vertices and $v$.




This means $\deg(v) \le \alpha(G) \le k-1$. So *every* vertex has degree at most $k-1$, which means $G$ has average degree $d \le k-1 \le k$.

For technical reasons, we'll also need a *lower* bound on the average degree. We can assume that $d \ge 10$ — otherwise, we can find an independent set of size at least $\frac{n}{40}$ by the greedy algorithm discussed in Remark 4.6 (or $\frac{n}{11}$ by Caro–Wei), which is greater than $k$.

Then by Lemma 4.5 and the observation that $d \le k$, we have

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

(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 that $d \ge 10$.)

Now plugging in our value of $n$ (we assumed that $n \ge \frac{k^2}{\log k - 1}$), we get

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

which is a contradiction. So there can't be a coloring with $n \ge \frac{k^2}{\log k - 1}$ vertices which contains no red clique of size $k$ or blue triangle. □

Now we want to prove Lemma 4.5. This isn't really about graph colorings anymore; instead, it's a statement about graphs in general. We'll see a proof from a few years after Theorem 4.2 (due to Shearer), which is basically magic. It proceeds by simple induction — you can't induct on the statement of Lemma 4.5, 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 4.7**

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

$$f(d) = \frac{d \log d - d + 1}{(d-1)^2} \quad \text{for } d \notin \{0, 1\},$$

and $f(0) = 1$ and $f(1) = \frac{1}{2}$.

(The definitions of $f(0)$ and $f(1)$ are what you'd get from taking 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 Lemma 4.7 implies Lemma 4.5. We can ignore the cases where $d = 0$ and $d = 1$; 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 by replacing Lemma 4.5 with Lemma 4.7 — this new function is more complicated, and the fact that it's bigger means that the statement is stronger (which seemingly makes it harder to prove). But the fact that this function is bigger also gives a better inductive hypothesis, which actually makes the statement easier to prove.

*Proof of Lemma 4.7.* First, $f$ has the following properties (which can be checked by relatively straightforward calculation):

- $f$ 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

$$(d+1)f(d) = 1 + (d - d^2)f'(d) \tag{4.1}$$

for all $d \in \mathbb{R}_{\geq 0}$. (Most likely, this 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 $f$.)

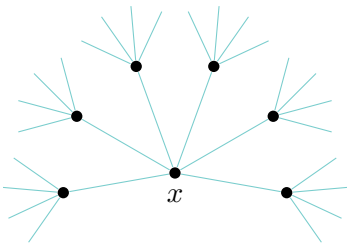
Now we'll induct on $n$. The base case $n = 0$ is trivial, as both $\alpha(G)$ and $n \cdot f(d)$ are 0.

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

We have $|N(x)| = \deg(x)$, so when we perform this process, the number of vertices we remove is $\deg(x) + 1$. Meanwhile, the number of *edges* we remove is

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

(This is because we remove all edges incident to either $x$ or any vertex in $N(x)$; and all edges incident to $x$ are also incident to some vertex in $N(x)$. Here 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 within $N(x)$.)



And now we know the number of edges and vertices we've removed, so we can calculate the average degree of the remaining graph. More formally, for every vertex $x$, let $G_x$ be the graph obtained from $G$ by removing the vertices in $\{x\} \cup N(x)$. Then we have

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

(where $d$ is the average degree of $G$), so the average degree of $G_x$ is

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

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.

By the induction hypothesis (since $G_x$ has fewer vertices than $G$), we know that $G_x$ contains an independent set of size at least

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

We'll now take such an independent set and add $x$ to it — this forms an independent set in $G$ because we removed all neighbors of $x$, so they're not part of the independent set in $G_x$. This shows

$$\alpha(G) \geq \alpha(G_x) + 1 \geq (n - 1 - \deg(x)) \cdot f(d_x) + 1. \tag{4.2}$$

(Converting this into the bound we want looks hopeless, since these expressions are rather horrendous; but it will actually work out.)

This inequality holds for all $x \in V(G)$, and now we want to find a good choice of $x$. In fact, we'll choose $x$ *randomly* and take the expectation of the right-hand side, and show that it's at least the desired bound $n \cdot f(d)$. In other words (to phrase this argument non-probabilistically), since $\alpha(G)$ is bounded by (4.2) for *all* vertices $x$, it's also bounded by the *average* of this expression over all $x$. Averaging (4.2) over all $x$ (and pulling the 1 out), we get that

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

And we want to show that this expression is at least $n \cdot f(d)$.

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 clear that $f(d_x)$ is the more ugly expression — $d_x$ is an ugly expression, and $f$ is itself also ugly. So it's hopeless to try to deal with $f(d_x)$ directly; instead, we'll try to bound it.

For this, we'll use the convexity of $f$ to relate $f(d_x)$ to $f(d)$ — by comparing $f(d_x)$ to the tangent line at $d$ (evaluated at $d_x$), 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 $x$ — so the only $x$-dependent term in this new expression is $d_x$, which we're no longer trying to plug into the function $f$. So we now have

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

Now, the latter expression has a lot of terms which don't depend on $x$. Separating out those terms, we can rewrite this as

$$\alpha(G) \geq 1 + \sum_{x \in V(G)} \frac{n - 1 - \deg(x)}{n} \cdot (f(d) - df'(d)) + \frac{1}{n} \sum_{x \in V(G)} (n - 1 - \deg(x)) \cdot d_x \cdot f'(d). \tag{4.3}$$

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 in the second summand of (4.3), the denominator cancels out and we simply get

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

Meanwhile, we can simplify the first summand of (4.3) to

$$\sum_{x \in V(G)} \frac{n - 1 - \deg(x)}{n} = n - 1 - d$$

(since $d$ is defined as the average degree). So then we have

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

Next, we'll evaluate this sum — we have

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

Since the summand only depends on $y$, we can swap the sum to rewrite this 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 $x$ of $y$ — so we get a sum of squares). Then by the AM–QM inequality, we get

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

(Note that $f'(d) < 0$, so we want a *lower* bound on this sum.) This tells us

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

Finally, by our differential equation (4.1) on $f$, we have

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

which makes this come out to exactly $n \cdot f(d)$. So by induction, we're done. □

### §4.1.2 The general case

Last class, we proved the $\ell = 3$ case of Theorem 4.2. Today we'll prove the $\ell > 3$ case.

*Proof of Theorem 4.2.* We'll use induction on $\ell$. Fix $C$ to be an extremely large constant (e.g., for concreteness we can imagine $C = 5000$). For the base case $\ell = 3$, we've already proven that

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

For technical reasons, we'll also need to allow $\ell = 2$ as a base case; in that case, Theorem 4.2 is trivially true because $R(k, 2) = k$.

Now assume that $\ell \geq 4$, and that we've already proven the statement for all $2 \leq \ell' \leq \ell - 1$. Let $k \geq 3$ and

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

and assume for contradiction that there is 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 $\ell$. (This is the right number of vertices because we're trying to show that having $(Ck)^{\ell-1}/(\log k)^{\ell-2}$ vertices 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'll rephrase the problem in terms of edges and non-edges (you can always translate back and forth between having red-blue edge-colorings and having the clique number and independence number of a graph, but in this case it's easier to think in terms of the latter).

In other words, let $G$ be the graph consisting of the blue edges. Then $G$ still has $n$ vertices, and it does not have a clique of size $\ell$. An independent set in $G$ is precisely the same as a red clique; we know there is no red clique of size $k$, so

$$\alpha(G) \leq k - 1 < k.$$

(We think of $k$ as large, so the difference between $k - 1$ and $k$ is not big, which means we don't have to be careful about the $-1$. Meanwhile, $\ell$ is small, so we *do* have to be careful with it — replacing $\ell$ with $\ell + 1$ would significantly change the bound in Theorem 4.2.)

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 proved Theorem 1.13, our first upper bound on the Ramsey numbers — we'll use a very similar idea. In that proof, we fixed a vertex and looked at its red neighbors and blue neighbors. We could bound the size of its red neighborhood by the inductive hypothesis, since it has no red clique of size $k - 1$ or blue clique of size $\ell$; and we could do the same with the blue neighborhood.

Here our problem is more asymmetric — we'll only look at the blue neighborhood, since we wouldn't get much out of replacing $k$ by $k - 1$ (which is what would happen if we looked at the red neighborhood). In the reformulation in terms of cliques and independent sets, this means we'll focus on the neighborhood of our vertex in $G$.

For every vertex $v \in V(G)$ its neighborhood $N(v)$ cannot have a clique of size $\ell - 1$ (if it did, we could add $v$ and get a clique of size $\ell$). And it also has no independent set of size $k$ (because the whole graph has no independent set of size $k$). So by the induction hypothesis applied to this neighborhood,

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

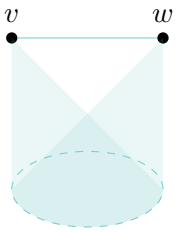
Now we have an upper bound on the size of the neighborhood of every vertex, or equivalently its degree — letting $d = (Ck)^{\ell-2}/(\log k)^{k-3}$, we have $\deg(v) \leq d$ for all $v \in V(G)$. In particular,

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

So 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'd like to apply Lemma 4.5, but we can't because our graph is not triangle-free. So instead, we'll go further with the induction hypothesis — we'll go one step deeper down. The reason for this is that it'll tell us something about the triangles of the graph; and this will be useful because then we can try to get rid of the triangles and apply Lemma 4.5.

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. So for any two adjacent vertices $v, w \in V(G)$, we look at their common neighborhood $N(v) \cap N(w)$ (the vertices adjacent to both $v$ and $w$). This common neighborhood has no clique of size $\ell - 2$ (if it did, we could add both $v$ and $w$ to get a clique of size $\ell$); and it also has no independent set of size $k$. So by the induction hypothesis, we have

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




(Note that we assumed $\ell \ge 4$, so $\ell - 2 \ge 2$; this is why we needed the technicality that Theorem 4.2 is also true for $\ell = 2$.)

Note that $(\log k)^4 < d$ (to verify this, note that $d$ is smaller when $\ell$ is smaller, so we can assume $\ell = 4$; then we're comparing some power of $\log k$ with some power of $Ck$, and as long as $C$ is large enough, this is true). Also let $\varepsilon = \frac{1}{\ell-2}$ (this quantity will appear a lot in the proof, so it's convenient to name it). Note that

$$(\ell - 2)(1 - \varepsilon) = \ell - 3 \quad \text{and} \quad (\ell - 3)(1 - \varepsilon) = \ell - 4 + \varepsilon.$$

(These are simple calculations, but it's easier to note them now rather than in the middle of a big calculation later, when we will use them.)

Returning to the actual argument, we now have a bound on the degree of every vertex, and a bound on $|N(v) \cap N(w)|$ for all adjacent vertices $v$ and $w$. We want to apply Lemma 4.5, but this lemma only applies to triangle-free graphs; so we'd like to make our graph triangle-free. For this, the next step is to upper-bound the *number* of triangles.

**Claim 4.8 —** 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

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

(In fact, this counts every triangle 6 times, once for each order, but this doesn't really matter.) So the total number of triangles is at most

$$n \cdot d \cdot \frac{(Ck)^{\ell-3}}{(\log k)^{\ell-4}} = nd \cdot \frac{(Ck)^{(\ell-2)(1-\varepsilon)}}{(\log k)^{(\ell-3)(1-\varepsilon)-\varepsilon}} = nd \cdot d^{1-\varepsilon} \cdot (\log k)^{\varepsilon}.$$

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

Now we have an upper bound on the number of triangles in $G$. We'd like to apply Lemma 4.5 to a 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 might be greater than the number of vertices, so this won't work. Instead, 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 tradeoff between the number of triangles and vertices, so that we'll actually be able to remove a vertex from every triangle.

First, why does it help us to pass to a random subgraph? 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 by more than vertices or edges. (It matters that our bound in Claim 4.8 is significantly less than $n^3$ — if it weren't, then we wouldn't be able to choose a good sampling probability.)

Let our sampling probability be $p = d^{-1+\varepsilon/4}$. (This is trivially between 0 and 1.)

**Remark 4.9.** As an exercise, you can see where this number comes from by going through the calculation keeping $p$ as a variable, and then seeing what value of $p$ makes it work. The 4 is a bit arbitrary, since $\frac{3}{4}$ was arbitrary in Claim 4.8; the point is that 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

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

Here we need to be a bit careful — we also want to say that most of the time, the actual number of vertices left isn't too much smaller than this. There's various ways to do this — for example, we can apply the Chernoff bound — but maybe the simplest is to use Chebyshev's inequality. We have

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

so Chebyshev's inequality shows that most of the time, $|V'|$ is at least half its expectation — more precisely,

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

(You can check that $pn \ge \frac{n}{d} \ge \frac{Ck}{\log k} > 16$. This bound is very weak, and in reality the probability is actually exponentially small; but we 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 $\frac{1}{2}pn$).

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 and triangles. First, we have

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

We also have

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

In the previous case (with the number of vertices), we had to say that with high likelihood we had *enough* vertices. But here we want the average degree and number of triangles to be *small* with high likelihood. This means we can just apply Markov (we don't have to go through the second moment) — we have

$$\mathbb{P}[\left|E(G')\right| > 4p^2\left|E(G)\right|] \le \frac{1}{4} \quad \text{and} \quad \mathbb{P}[\#(\text{triangles in } G') > 4p^3 \cdot \#(\text{triangles in } G)] \le \frac{1}{4}$$

by Markov's inequality. Now we have three unusual events, where each happens with probability at most $\frac{1}{4}$. So by the union bound, the probability that *any* of them happens is at most $\frac{3}{4}$; this means with probability at least $\frac{1}{4}$, none of them happen. In particular, there *exists* a choice of $V'$ for which none of these weird inequalities happen, meaning that there is some outcome for $V'$ and $G'$ satisfying the following inequalities:

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

Now we can forget about the randomness — fix an outcome for which these conditions hold. Our goal is to apply Lemma 4.5. Our graph $G'$ is still not triangle-free, but it has *few* triangles. So now we can finally perform the process of removing one vertex from every triangle in $G'$ — we're in better shape to do this with $G'$ than we were with $G$, because now the number of triangles is way smaller than the number of vertices.

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

$$\left|V''\right| \geq \left|V'\right| - \#(\text{triangles in } G') \geq \frac{1}{2}nd^{-1+\varepsilon/4} - 4nd^{-1}.$$

(This is an inequality because some vertices may be 'double-deleted' — if they're in multiple triangles — but that only helps us.) And the first quantity is at least as large as the second (this requires $d^{\varepsilon/4} \geq 16$, or $d^{\varepsilon} \geq 16^4$; but $d^{\varepsilon}$ is roughly $Ck$ (up to some logs), so if we make $C$ big enough then this is true). This means

$$\left|V''\right| \geq \frac{1}{4}nd^{-1+\varepsilon/4}$$

(so we haven't lost much in the number of vertices).

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

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

Now we're in good shape to apply Lemma 4.5 — we have a triangle-free graph on some number of vertices (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

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

So by Lemma 4.5, $G''$ has independence number

$$\alpha(G'') \geq \left|V''\right| \cdot \frac{\log(16d^{\varepsilon/4}) - 1}{16d^{\varepsilon/4}}.$$

**Remark 4.10.** Note that Lemma 4.7 only gives the bound of $n \cdot \frac{\log d - 1}{d}$ when we plug in the *actual* average degree (rather than just an upper bound). This is usually fine because $\frac{\log d - 1}{d}$ is monotone for $d \geq 10$, so if our average degree is at least 10 then this inequality is justified. However, if the average degree is less than 10 (which could be the case here), we can instead use the bound $\alpha(G) \geq \frac{n}{d+1}$, which is true in general; and we also have

$$\frac{|V''|}{11} \geq \left|V''\right| \cdot \frac{\log(16d^{\varepsilon/4}) - 1}{16d^{\varepsilon/4}},$$

since $d^{\varepsilon/4} > 100$ (for example).

Plugging in our bound on $|V''|$ (we replace $\frac{1}{4}$ with $\frac{1}{5}$ to eat up the $-1$), we get

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

Now plugging in the definition of $d$ and $n$, we get

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

(We change 320 to 325 to account for the rounding on $n$; then $d$ looks just like $n$, but with the exponents of $Ck$ and $\log k$ both lowered by 1.) The $\varepsilon$ cancels the $\ell-2$, and we can replace the $\ell-3$ with a $\ell-2$ (since we only want a lower bound), so that it also cancels; then we're left with

$$\alpha(G'') \geq \frac{1}{325} \cdot \frac{Ck}{\log k} \cdot (\log(Ck) - \log\log k).$$

We have $\log(Ck) - \log\log k = \log k + \log C - \log\log k \approx \log k$ (since $\log\log k$ is much smaller than $\log k$ unless $k$ is small, in which case $\log\log k$ is much smaller than $\log C$). So bounding it by $\frac{1}{2}\log k$, we get

$$\alpha(G'') \geq \frac{1}{325} \cdot \frac{Ck}{\log k} \cdot \frac{1}{2}\log k \geq \frac{1}{650} \cdot Ck > k.$$

But now we have $\alpha(G) \geq \alpha(G'') > k$, which is a contradiction (as $G''$ is an induced subgraph of $G$ — we obtained it by restricting to a vertex subset — so every independent set in $G''$ is also one in $G$). So this coloring can't exist, showing that $R(k,\ell) \leq n$. □

**Remark 4.11.** For some intuition on why increasing $\ell$ by 1 gives an additional log factor, this essentially comes from the last step. We're defining $d$ as our upper bound on $R(k,\ell-1)$, and in the end we get a bound of $\alpha(G) \gtrsim \frac{n}{d} \cdot \log k$ (where the $\log k$ essentially comes from Lemma 4.5); we want this to be greater than $k$, so we take $n \asymp d \cdot \frac{k}{\log k}$.



## §4.2 Lower bounds

Last class, we proved an upper bound for the off-diagonal Ramsey numbers, that

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

*Off-diagonal* means we're interested in the case where $\ell$ is fixed and $k$ is large; then $C^\ell$ is a constant, so this bound is of the form

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

Today we'll talk about lower bounds.

**Theorem 4.12** (Spencer 1977)
For every fixed $\ell \geq 3$, there is some constant $c_\ell > 0$ such that for all $k \geq 3$,

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

Comparing this to the upper bound in Theorem 4.2, when $\ell = 3$ the exponent of $k$ matches (it's $k^2$ in both cases), but the exponents of the $\log k$ 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 $\log k$ was an open problem for a long time. This was resolved by Kim in the mid-1990s; it turns out that 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

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

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 behavior is still wide open. (*Note: This was true at the time of this lecture. But sometime later in 2023, Mattheus and Verstraete determined the correct exponent for the case $\ell = 4$ — they proved that $R(k,4) \geq c \cdot \frac{k^3}{(\log k)^4}$, which matches the upper bound up to log factors. As of now, we still don't know the correct exponent for $\ell \geq 5$.*)

We won't discuss this improvement, which is very complicated, but we'll discuss the proof of Theorem 4.12. The approach of Spencer's proof is a probabilistic construction — this is unsurprising, given that our lower bounds for the *diagonal* Ramsey numbers (Theorem 1.20). But if you approach this probabilistically with the same approach as we had for diagonal Ramsey numbers, you 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 (which we'll soon discuss). Still, it's a relatively simple probabilistic proof.

The additional improvement (of Bohman and Keevash) is much more complicated. There, the idea is to not just 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 probability $p$, they consider the $K_\ell$-free process — we start with the correct number of vertices and no edges. Then we place the first edge randomly; then we place the second edge randomly among all available edges; but as more and more edges get added, we add the next new edge randomly while making sure not to create a $K_\ell$. In other words, if we have a pair of vertices where adding that edge would create a $K_\ell$, then we forbid it; at every step, among all edges we're allowed to add, we add one uniformly at random. As you can imagine, the edges are now not independent; these dependencies make the process much more complicated to analyze.

### §4.2.1 The Lovász local lemma

The proof of Theorem 4.12 relies on the Lovász 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 this is the form it's usually applied in, and it's probably the easiest to parse. (If you've seen the statement of the Lovász local lemma before, you might remember there's a notion of a dependency graph, which is kind of subtle — you have to care about *mutual* independence of non-neighbors, not just *pairwise* independence. But we'll state the lemma in a specialized form that avoids these subtleties.)

**Theorem 4.13** (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 $J_i \subseteq [m]$. Suppose there exist $x_1, \ldots, x_n \in [0,1)$ such that for each $i \in [n]$ we have

$$\mathbb{P}[\mathcal{E}_i] \leq x_i \cdot \prod_{k \sim i} (1 - x_k), \tag{4.4}$$

where we write $k \sim i$ to mean that $k \neq i$ and $J_i \cap J_k \neq \emptyset$. Then we have

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

In words, we have a collection of events $\mathcal{E}_i$, each of which depends on some subset $J_i$ of our random variables $Z_j$; and $k \sim i$ means that $\mathcal{E}_i$ and $\mathcal{E}_k$ have at least one variable in common that they depend on. When we apply this theorem, we'll think of the random variables $Z_j$ as the colors of the different edges. And the events $\mathcal{E}_i$ 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 the Lovász local lemma before, it might look a bit strange. But here's some motivation: We think of the events $\mathcal{E}_i$ as 'bad' events (events we want to avoid). The conclusion of the theorem says that with positive probability, we can avoid all the events. The precise probability bound typically doesn't matter — we only really care that it's positive, because that tells us there *exists* a good outcome (one that avoids all the bad events).

What does $x_i \prod_{k \sim i}(1-x_k)$ 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 interact at all. This means the events $\mathcal{E}_i$ are all independent. In this setting, the second product has no factors, so this condition tells us $\mathbb{P}[\mathcal{E}_i] \le x_i$. And in that case, Theorem 4.13 is trivial — we have *independent* events with $\mathbb{P}[\mathcal{E}_i] \le x_i$, so the probability that none hold is $\prod_{i=1}^n (1 - \mathbb{P}[\mathcal{E}_i]) \ge \prod_{i=1}^n (1 - x_i)$.

The point of Theorem 4.13 is that we can still draw a good conclusion if there's some dependencies — the theorem still works well when there's only a *few* intersections between the sets $J_i$. If there's too many intersections, then the product $\prod_{k \sim i}(1 - x_k)$ becomes really small, and we can't satisfy the condition. But if there's only a few, then this product doesn't have too many factors, which means it doesn't change $x_i$ by too much (and we can satisfy the condition by taking $x_i$ to be a bit bigger than $\mathbb{P}[\mathcal{E}_i]$).

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

(We'll 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 (the off-diagonal Ramsey numbers), but it'll be helpful for another homework problem. (This symmetric form 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 a reasonable amount of symmetry.)

**Theorem 4.14** (Symmetric Lovász local lemma)

Let $Z_1, \dots, Z_m$ be independent random variables and let $\mathcal{E}_1, \dots, \mathcal{E}_n$ 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 $J_i \subseteq [m]$). Suppose that $\mathbb{P}[\mathcal{E}_i] \le p$ 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 $ep(d+1) \le 1$, we have

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

In other words, we're assuming every set $J_i$ intersects at most $d$ other sets (so our product $\prod_{k \sim i}(1 - x_k)$ in Theorem 4.13 would have at most $d$ factors). Here $e = 2.718\dots$ is the usual constant.

*Proof.* First, if $d = 0$ then the statement is trivial (as all the events $\mathcal{E}_i$ are independent). Now assume $d \ge 1$, and apply the Lovász local lemma (Theorem 4.13) with $x_i = \frac{1}{d+1}$ for all $i \in [n]$. We have

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

since the number of factors in this product is at most $d$; so the condition (4.4) of the Lovász local lemma is met, and it tells us that $\mathbb{P}[\overline{\mathcal{E}_1} \cap \cdots \cap \overline{\mathcal{E}_n}] > 0$. □

We're not going to 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 $\ell$, which behave very differently (as $k \gg \ell$). But if you wanted to lower-bound $R(k,k)$ (which we will do on the homework), then everything *does* become symmetric, and it's easier to apply the symmetric version.

### §4.2.2 Proof of Theorem 4.12

Now we'll prove Spencer's lower bound on $R(k,\ell)$.

*Proof.* Fix $\ell \geq 3$ (whenever we use asymptotic notation, we think of $\ell$ as fixed and $k$ as large). We may assume that $k$ is sufficiently large with respect to $\ell$ — if we prove the bound for all sufficiently large $k$, then we can adjust the constant $c_\ell$ to satisfy the bound for the finitely many remaining values of $k$. In particular, we'll use $o(1)$ notation to refer to terms that converge to 0 as $k \to \infty$. (Then at some point, these terms will be smaller than certain thresholds we care about, so for large $k$ the inequalities we want on those terms will be true.)

Let $c_\ell$ be small enough that we have

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

(We'll see where this comes from later, but the logic is clearer if we fix $c_\ell$ now. If you were trying to come up with this proof, you'd go through this argument thinking in your head that $c_\ell$ is small enough, and later determine how small it needs to be for things to work.) (Note that the true $c_\ell$ in Theorem 4.12 might be even smaller, to account for the finitely many remaining $k$.) Now let

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

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 $n$-vertex complete graph with red and blue with no red clique of size $k$ or blue clique of size $\ell$. (This would show that $R(k,\ell) > n$.)

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 (Theorem 1.20), we again want to take a random coloring. In the diagonal case, we colored each edge red with probability $\frac{1}{2}$ and blue with probability $\frac{1}{2}$. In this setting, $\frac{1}{2}$ is not the right probability, since $k$ and $\ell$ are very different. In particular, we want edges to be more likely to be red (since we have much more space with red edges). So we color each edge *blue* with probability

$$p = 12\ell \cdot \frac{\log k}{k},$$

and red otherwise, independently for all edges. (This value of $p$ comes from running the whole argument with $p$ left as a parameter and then optimizing; we will not do this, because it's kind of messy. The 12 is not important, since we don't really care about the constant; the important thing is the $\ell \cdot \frac{\log k}{k}$ behavior.) As a reality check, as $k$ grows, this probability is indeed pretty small.

This defines independent random variables $Z_1, \ldots, Z_m$ indicating the colors of the edges, where $m = \binom{n}{2}$. (Technically, you need to number the edges from 1 to $\binom{n}{2}$. This can be done arbitrarily, because the ordering of the variables doesn't really matter; it only matters which events have variables in common.)

Our bad events $\mathcal{E}_i$ will be the occurrences of the cliques we want to avoid — we'll have $\binom{n}{k} + \binom{n}{\ell}$ events of two types (red $k$-cliques and blue $\ell$-cliques). (We won't name these events, in order to not blow up the notation.)

We can compute the probabilities of these bad events — for any set of $\ell$ vertices, the probability they form a blue clique of size $\ell$ is $p^{\binom{\ell}{2}}$ (since we have $\binom{\ell}{2}$ edges which all need to be blue). Similarly, for any set of $k$

vertices, the probability they form a red clique of size $k$ is $(1-p)^{\binom{k}{2}}$. Since these probabilities will come up quite often, let's name them

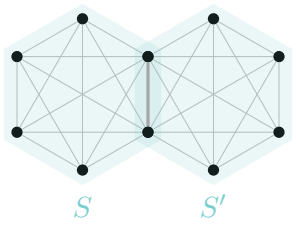
$$q = p^{\binom{\ell}{2}} \quad \text{and} \quad q' = (1-p)^{\binom{k}{2}}.$$

We want to avoid all these bad events, so we want to show that with positive probability none of these events occur. We'll do this using the Lovász local lemma; so our goal is to find numbers $x_i$ such that the condition (4.4) of the Lovász local lemma is satisfied.

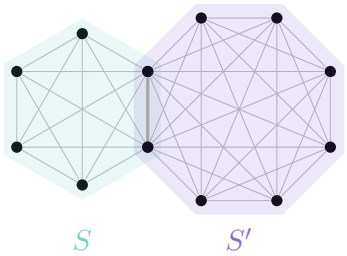
The Lovász local lemma has some flexibility — it lets us choose a different $x_i$ for each of the events. But we only want to choose two — the $\ell$-clique events are all symmetric, and the $k$-clique events are all symmetric, so we really only want to define two values $x$ and $x'$.

To figure out the relevant inequalities, let's figure out how many factors of each type we have in (4.4) — this means for every $i$, we want to figure out how many different $k$ there are and which types of events they correspond to.

For a subset $S$ of $\ell$ vertices (corresponding to one of the blue $\ell$-clique events), let $d_{\ell\ell}$ be the number of vertex subsets $S'$ with $|S'| = \ell$ and $|S \cap S'| \geq 2$ (meaning that $S$ and $S'$ contain a common edge). We care about this because this corresponds to the dependencies — for a set of $\ell$ vertices, its corresponding set $J$ consists of the $\binom{\ell}{2}$ edges between these vertices (since the corresponding event — whether these $\ell$ vertices form a blue clique — only depends on the edges between them). So for two events to have nonempty intersection, we need them to share an edge, which occurs if and only if they share at least two vertices. This means $d_{\ell\ell}$ counts the number of factors $(1-x_k)$ in (4.4) for $S$ corresponding to a blue-clique event.




Similarly, for $S$ of size $\ell$, we also need to consider events where $|S'| = k$ (corresponding to red $k$-cliques) — 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 our 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 (4.4). So to apply the Lovász local lemma, we want to find $x, x' \in [0, 1)$ such that the following two conditions hold. For the blue $\ell$-clique events, we need

$$q \leq x \cdot (1-x)^{d_{\ell\ell}} \cdot (1-x')^{d_{\ell k}} \tag{4.5}$$

(our definitions of $d_{\ell\ell}$ and $d_{\ell k}$ precisely track the number of $\ell$-sets and $k$-sets which interfere with our given $\ell$-sets). Similarly, for the red $k$-clique events, we need

$$q' \leq x'(1-x)^{d_{k\ell}}(1-x')^{d_{kk}}. \tag{4.6}$$

Now all we need to do is find $x$ and $x'$ satisfying these two conditions; then we've satisfied the conditions of the Lovász local lemma and we're done. To do this, let's first try to upper-bound these exponents $d_{\ell\ell}$, $d_{\ell k}$, and so on — the exact values are pretty ugly, but we can give reasonable upper bounds. First, we have

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

The $\binom{\ell}{2}$ term comes from choosing two vertices for $S$ and $S'$ to share; then we've chosen two things to place into $S'$, so we need to choose the remaining $\ell - 2$. (We use $n-2$ instead of $n-\ell$ to account for the fact that $|S \cap S'|$ can be larger than 2; we're overcounting these cases, but that's fine.) This bound is quite lossy — we have $(\ell-2)!$ in the denominator, which is usually much bigger than $\frac{1}{2}\ell^2$. But $\ell$ is fixed, so we don't really care about the extra factorial. Similarly, we have

$$d_{k\ell} \le \binom{\ell}{2}\binom{n-2}{k-2} \le \frac{1}{2}\ell^2 \cdot \frac{n^{k-2}}{(k-2)!} \le \frac{1}{2}n^{k-2}$$

(we're assuming $k$ is large with respect to $\ell$, so the $(k-2)!$ in the denominator eats the $\ell^2$). Similarly,

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

(this time we can't omit the $k^2$ since $(\ell-2)!$ is not big), and

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

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

**Remark 4.15.** The factors of $\frac{1}{2}$ are not important. But we'll keep them around because we'll pick up a factor of 2 later, so keeping them around to cancel it makes our numbers slightly nicer.

It's a bit annoying to have $(1-x)$ raised to some power. The easiest way to deal with such terms is to take logs, but $\log(1-x)$ is also kind of ugly. Often, the way we deal with it is by using the following fact:

**Fact 4.16 —** For all $x \in \mathbb{R}$, we have $1 - x \le e^{-x}$.

(This is because $e^x$ is convex, so it always lies above its tangent at $x = 0$.)

In this situation, we really want an inequality in the opposite direction. When $x$ is close to 0, the two sides actually pretty close — for example, for $0 \le x \le 0.1$ we have $1 - x \ge e^{-2x}$. (The 0.1 has a lot of room.) Equivalently, $\log(1-x) \ge -2x$. This gives a reverse inequality to Fact 4.16, if we lose a factor of 2.

Now we can rewrite (4.5) as

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

Using the above bound $\log(1-x) \ge -2x$ (which is valid as long as $0 \le x, x' \le 0.1$), it suffices to show that $\log \frac{q}{x} \le -2d_{\ell\ell}x - 2d_{\ell k}x'$; flipping signs, this is equivalent to

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

Now plugging in our bounds on $d_{\ell\ell}$ and $d_{k\ell}$, it suffices to have

$$\log \frac{x}{q} \ge \ell^2 n^{\ell-2}x + n^{k-2}x'. \tag{4.7}$$

By the same calculation, to get (4.6) it suffices to have

$$\log \frac{x'}{q'} \geq k^2 n^{\ell-2} x + n^{k-2} x'. \tag{4.8}$$

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 do this, we'll calculate our actual expressions for $q$ and $q'$. We have

$$q = p^{\binom{\ell}{2}} = (12\ell)^{\binom{\ell}{2}} \left(\frac{\log k}{k}\right)^{\binom{\ell}{2}};$$

in particular, $q < 0.05$ if $k$ is large. Meanwhile,

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

(The $k-1$ is a bit annoying, so we bound $k-1 \geq \frac{k}{2}$. The first inequality is by Fact 4.16 — in general, this bound is good to use whenever $x$ is small.)

Now it's time to define $x$ and $x'$. As a bit of motivation, on the left-hand side of (4.7) we have $\log \frac{x}{q}$, so it's natural to define $x$ as something times $q$. So in fact we'll take $x = 2q$ (the 2 is not important; the point is just that we want $x$ a bit larger than $q$, so that the left-hand side is a constant). (Note that $2q < 0.1$.)

Similarly, we'll also take $x'$ to be something times $q'$. But the right-hand side of (4.8) will be larger than a constant, so we'll actually take $x' = k^{\ell k} q'$. (Note that $x' \leq k^{-2\ell k} < 0.1$ for $k$ large as well.)

Now we need to check our two inequalities (4.7) and (4.8). We have

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

(the last two factors come from plugging in the definition of $n$; the first few factors come from $q$). The $\frac{\log k}{k}$ terms are very similar — in both cases the exponent is quadratic in $\ell$, and in fact these two exponents differ by exactly 1. So this simplifies to

$$n^{\ell-2} x \leq c_\ell^{\ell-2} \cdot 2(2\ell)^{\binom{\ell}{2}} \cdot \frac{\log k}{k}.$$

We defined $c_\ell$ so that the factor in front is less than $\frac{\ell}{2}$, giving

$$n^{\ell-2} x \leq \frac{\ell}{2} \cdot \frac{\log k}{k}.$$

Meanwhile, we can bound

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

(here we're using the very loose bound $n < k^\ell$). So the $n^{k-2}x'$ terms in (4.7) and (4.8) are both $o(1)$, while the first term (the one with $n^{\ell-2}x$) is more significant. Now for (4.7), we have

$$\ell^2 n^{\ell-2} x \leq \ell^2 \cdot \frac{\ell}{2} \cdot \frac{\log k}{k} = o(1)$$

as well (because $\ell$ is fixed and $k$ is large), so

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

as well. In particular, for large $k$ this is less than $\log 2 = \log \frac{x}{q}$.

Meanwhile, for (4.8), we have

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

This proves both the inequalities we needed, so we're done. □

# §5 Bounds for hypergraph Ramsey numbers

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

**Definition 5.1.** For 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 (we require $k_i \geq r$ 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 $r = 2$.)

## §5.1 Upper bounds

We already proved that $R_r(k_1, \dots, k_t)$ exists (Theorem 1.26). 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* upper 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 something *really* gigantic. So this recursion gives an extremely weak upper bound for $R_r(k_1, \dots, k_t)$.

**Example 5.2**

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

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

so $R_2(k, k)$ grows exponentially in $k$ (though we don't know the correct base of the exponent). If we plug this into the above recursion, the bound we get for $R_3(k, k)$ is roughly of the form

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

where the number of 2's is linear in $k$.

(This is because every time we increase $k$ by 1, we're plugging the previous value into $R_2$, which is exponential; this gives an iterated exponential.)

This is an extremely fast-growing function, called a *tower-type* function; for example, its first few values are

$$2,\ 2^2 = 4,\ 2^{2^2} = 2^4 = 16,\ 2^{2^{2^2}} = 2^{16} = 65536,\ 2^{2^{2^{2^2}}} = 2^{65536}, \dots.$$

(We'll define the tower function more formally next week and talk about it more.)

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

**Theorem 5.3** (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. When we recurse, we're decreasing the uniformity by 1 (and all of the clique sizes). The binomial coefficient looks scary, but it's bounded by $R_{r-1}(k_1-1,\dots,k_t-1)^{r-1}$. And though exponentials are large, at least there's only one. So this bound is actually pretty mild.

**Example 5.4**

Applying Theorem 5.3 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 (where the top is linear in $k$).

**Remark 5.5.** When we build these towers, we don't care that much about what's at the top (e.g., $k$ vs. $k^2$); what's most important is the number of times we exponentiate. Here we're only exponentiating 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 5.6**

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

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

(Here we're using Theorem 4.2 for the upper bound on $R(k-1,\ell-1)$; we probably don't care about the log factors, so we could even just use the Erdős–Szekeres theorem (Theorem 1.13).) This expression is *single*-exponential (where the exponent is a polynomial function of $k$).

**Remark 5.7.** In the case of fixed $\ell$, we know that this single-exponential behavior is correct (though we don't know what the exponent should be). In the diagonal case $k = \ell$, the correct answer isn't known.

More generally, the tower height in the bound we get from Theorem 5.3 is approximately the uniformity $r$. This is much better than a tower height of approximately $k$, since we generally think of the uniformity as fixed and $k$ as large.

Now let's prove Theorem 5.3.

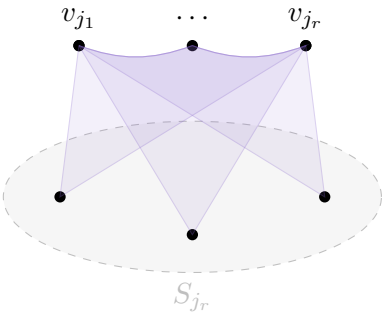
*Proof.* Let $R = R_{r-1}(k_1-1,\dots,k_t-1)$, and consider a complete $r$-uniform hypergraph on $t^{\binom{R}{r-1}}$ vertices where every edge is colored with one of the colors $1,\dots,t$. We need to show that for some $i \in \{1,\dots,t\}$, there exists a clique of size $k_i$ in color $i$.

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

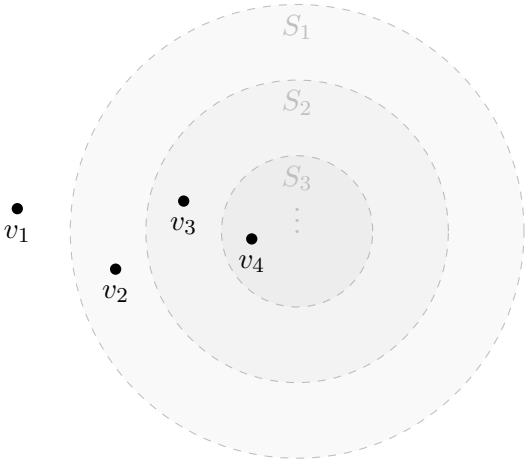
We'll construct a sequence of distinct vertices $v_1,\dots,v_R$ and a nested sequence of vertex sets $S_1 \supseteq S_2 \supseteq \cdots \supseteq S_R$, such that the following three conditions hold:

(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 all $v \in S_{j_{r-1}}$ have the same color.

In other words, we think of $S_j$ as a 'candidate set' associated with the vertices $v_1, \dots, v_j$. Then this condition states that if we choose any $r-1$ distinct vertices among $v_1, \dots, v_R$, then as long as we choose $v$ from the appropriate candidate set, the color of the corresponding edge (where we add $v$ to those $r-1$ vertices) doesn't depend on $v$.




(2) For every $1 \le j \le R$, we have $v_{j+1}, \dots, v_R \in S_j$, but $v_1, \dots, v_j \notin S_j$. (In particular, this means $v$ can't coincide with any of the previous vertices $v_{j_i}$ in (1), because then (1) wouldn't make sense.)




(3) We have $|S_j| \ge t^{\binom{R}{r-1}-\binom{j}{r-1}} - j$ for all $1 \le j \le R-1$, and $|S_R| \ge 1$.

(When $j < r-1$, this means $S_j$ contains nearly the entire graph — all but $j$ vertices. But this makes sense because $S_j$ can never appear as the candidate set in (1), so there are no constraints on it except that it shouldn't have $v_1, \dots, v_j$.)

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

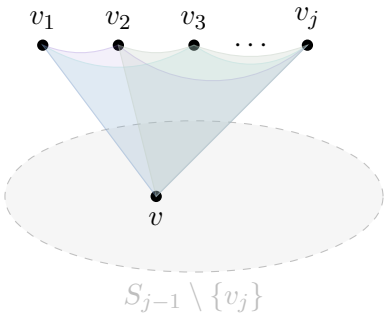
We'll construct these vertices and sets one at a time. First, let $S_0$ be the entire vertex set (for convenience); this satisfies (3) because

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

Now for $1 \le j \le R$, suppose that we've already constructed $c_1, \dots, v_{j-1}$ and $S_0 \supseteq S_1 \supseteq \cdots \supseteq S_{j-1}$. Then we construct $v_j$ and $S_j$ as follows.

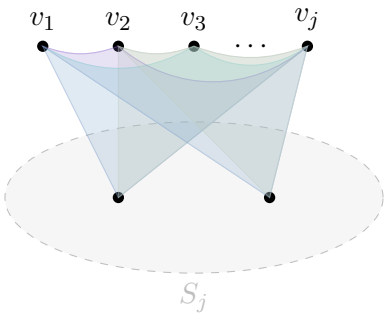
First, we construct $v_j$ by picking an arbitrary vertex $v_j \in S_{j-1}$. (To do this, we need $S_{j-1}$ to be nonempty; but the lower bound on its size in (3) is positive, so this is true.)

Now we need to construct $S_j \subseteq S_{j-1} \setminus \{v_j\}$ satisfying (1). So 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 $1 \le j_1 < \cdots < j_{r-2} < j$. These are the colors we need to analyze for (1). (We don't need to consider the case where we don't take $v_j$ — then there's no new condition, since all vertices in the previously chosen sets already satisfy (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. So this means there are $t^{\binom{j-1}{r-2}}$ possibilities for these colors. In other words, imagine that for every $v$, we record the color associated to each $(r-2)$-tuple. Then there are $t^{\binom{j-1}{r-2}}$ possibilities for this data, for any given $v$.

Looking at these different possibilities, we get a partition of $S_{j-1} \setminus \{v_j\}$ (where we split up our vertices $v$ based on their data, so each group of the partition consists of vertices with the same data). This partition has $t^{\binom{j-1}{r-2}}$ subsets (one for each possibility for the 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 it. So (1) holds by construction — for every $v \in S_j$, by definition we have the same assignment from $(r-2)$-tuples to colors. In other words, 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 $v \in S_j$. This means (1) holds when $j_{r-1} = j$ (and the cases when $j_{r-1} < j$ 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), we have $v_j \notin S_j$ by construction (we took $S_j$ as a subset of $S_{j-1} \setminus \{v_j\}$). But we do have $v_j \in S_{j-1}$, and therefore $v_j \in S_i$ for all $i < j$ (as $S_0 \supseteq S_1 \supseteq \cdots$). So we have $v_j \in S_i$ if and only if $i < j$, which gives exactly the condition in (2).

Finally, for (3), we need to check that $|S_j|$ is large enough. The important fact here is that we chose $|S_j|$ to be the *biggest* of the subsets in our partition, so it's at least their average — we have

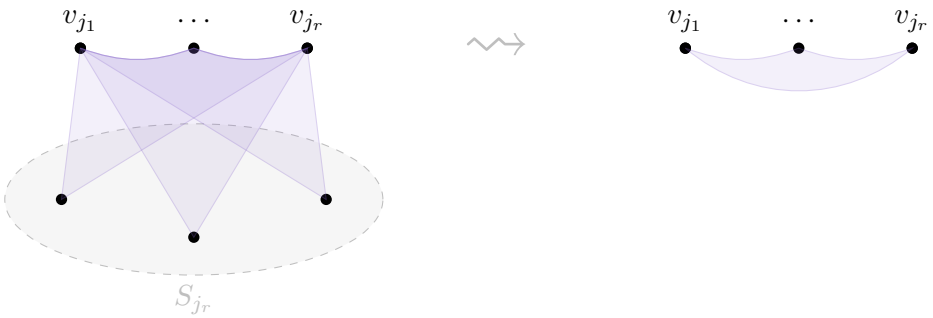
$$|S_j| \geq \frac{|S_{j-1} \setminus \{v_j\}|}{t^{\binom{j-1}{r-2}}} \geq \frac{t^{\binom{R}{r-1}-\binom{j-1}{r-1}} - (j-1) - 1}{t^{\binom{j-1}{r-2}}} \geq t^{\binom{R}{r-1}-\binom{j-1}{r-1}-\binom{j-1}{r-2}} - j = t^{\binom{R}{r-1}-\binom{j}{r-1}} - j$$

(the third step is because moving the subtractions out of the fraction can only make our expression smaller, and the last is by Pascal's identity). This proves the desired lower bound for $j \leq R-1$. For the special case $j = R$, we need to check that $|S_R| \geq 1$. We have

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

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 $|S_R| \geq 1$ (it's a set, so its size is an integer). So (3) is satisfied.

Now we've constructed our sequence of vertices $v_1, \dots, v_R$ and sets $S_1 \supseteq \cdots \supseteq S_R$, and we want to use this to construct a monochromatic clique. 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 (1) — i.e., with the common color of the edges $\{v_{j_1}, \dots, v_{j_{r-1}}, v\}$ in the original coloring over all $v \in S_{j_{r-1}}$. (Any $(r-1)$-edge in our $(r-1)$-uniform hypergraph is of this form for some indices $j_1 < \cdots < j_{r-1}$; then we can look at (1) for those indices, and it tells us that all $r$-edges of this form have a common color in the original coloring; we then assign that common color to this $(r-1)$-edge.)




Since we defined $R = R_{r-1}(k_1 - 1, \dots, k_t - 1)$, in this coloring of a $(r-1)$-uniform hypergraph, we can find a clique of size $k_i - 1$ in color $i$ for some $i \in \{1, \dots, t\}$. 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 our $r$-uniform hypergraph. This is good enough, as then we'll have 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 $T \cup \{v\}$, this $r$-set has color $i$. There's two cases, depending on whether or not this $r$-set involves $v$.

In the first case, for any vertices $v_{j_1}, \dots, v_{j_{r-1}}$ in $T$ (ordered with $j_1 < \cdots < j_{r-1}$), the edge $\{v_{j_1}, \dots, v_{j_{r-1}}, v\}$ has color $i$ because of (1) and the definition of the $(r-1)$-uniform hypergraph coloring — we have $v \in S_R \subseteq S_{j_{r-1}}$, so (1) requires that this $r$-edge has the color associated to $\{j_1, \dots, j_{r-1}\}$. And that color 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 $v_{j_1}, \dots, v_{j_r}$ in $T$ (ordered with $j_1 < \cdots < j_r$). Then the edge $\{v_{j_1}, \dots, v_{j_r}\}$ has color $i$ for essentially the same reason — we must have $v_{j_r} \in S_{j_{r-1}}$ by (2) (since $j_r > j_{r-1}$), so we can again apply (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. This means it forms a clique of size $k_i$ in color $i$, and we're done. □

The strategy we saw in this proof is useful more generally. What's going on is that we progressively add vertices $v_1$, $v_2$, ... 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). We make our sequence long enough so that at some point, we can look at the set of vertices we've chosen and find a 'good' subset. The idea of having a sequence of vertices and a candidate set comes up reasonably often, and we'll see it again later in the course.

## §5.2 Lower bounds

Now we've proven Theorem 5.3, and we've already drawn some conclusions which are much better than the giant bounds obtained from the original recursive proof. Now we'll talk about the best-known *lower* bounds.

### §5.2.1 An overview

**Theorem 5.8** (Erdős–Hajnal 1972)
There is an absolute constant $c > 0$ such that $R_3(k, 4) \geq 2^{ck}$ for all $k \geq 3$.

In Example 5.6, we saw a single-exponential *upper* bound for the off-diagonal Ramsey numbers $R_3(k, \ell)$ (i.e., for fixed $\ell$); Theorem 5.8 shows that this single-exponential behavior is correct. For *graph* Ramsey numbers, $R(k, \ell)$ is *polynomial* in $k$, so this isn't obvious. (Of course, we have $R_3(k, \ell) \geq R_3(k, 4)$ for any fixed $\ell \geq 4$, so Theorem 5.8 gives a lower bound for $R_3(k, \ell)$ for every $\ell \geq 4$; in comparison with the upper bound from Example 5.6, the exponent is linear in $k$ rather than a higher-degree polynomial.)

(We will prove Theorem 5.8 next class.)

The best-known bounds are a bit better, but have 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 improves the upper bound from roughly $k^4$ to $k^2 \log k$ in the exponent. Still, the fundamental behavior is the same — we still have single-exponential bounds with a polynomial gap in the exponent. These bounds are due to Conlon–Fox–Sudakov 2010. More generally, they improved the upper bound from Example 5.6 to

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

which essentially square-roots the polynomial in the exponent (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 what polynomial should be in the exponent; and the same is true for $R_3(k, \ell)$ for any fixed $\ell$.

For the *diagonal* case, we have the following theorem (which we will prove on the homework).

**Theorem 5.9**
There is an absolute constant $c > 0$ such that $R_3(k, k) \geq 2^{ck^2}$.

This lower bound is better than the one Theorem 5.8 gives for fixed $\ell$ (here the exponent is quadratic rather than linear in $k$), but it's still far from the double-exponential upper bound from Example 5.4. So we know

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

These are essentially the best-known upper and lower bounds — it's still unclear whether $R_3(k, k)$ is single-exponential or double-exponential.

**Conjecture 5.10** (Erdős) — The true behavior of $R_3(k, k)$ is double-exponential.

In other words, Erdős conjectured that there should also be a lower bound of the form $2^{2^{ck}}$. He offered \$500 for this conjecture; it is still open.

So far, all the bounds we've thought about are for $r = 3$. Next week, we'll see that in some sense, answering these questions for the 3-uniform case is enough. Theorem 5.3 gives a way to go from an upper bound for $r - 1$ to an upper bound for $r$. Next week we'll see a way to do the same with *lower* bounds (where both the upper and lower bounds increase by one exponential). But this will only work starting with uniformity 3 or higher (i.e., when $r \geq 4$). So the 3-uniform case is the most interesting one, since it would give all the others automatically.

### §5.2.2 The off-diagonal setting

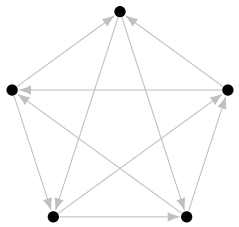
We'll first prove Theorem 5.8 from last class; here we're looking at off-diagonal 3-uniform Ramsey numbers, where we want either a red $k$-clique or a blue 4-clique.

*Proof of Theorem 5.8.* As usual, we can assume that $k$ is large — if we find some $c > 0$ such that Theorem 5.8 holds for all sufficiently large $k$ (e.g., $k \geq 100$), then we can deal with the finitely remaining values of $k$ (e.g., $k < 100$) by adapting the constant $c$. (For every *particular* $k$ there certainly exists a constant $c$ for which Theorem 5.8 is true, so by shrinking $c$ enough we can take care of these finitely many $k$.)

We'll 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 its edges without a red clique of size $k$ or a blue clique of size 4.

As usual, it's difficult to get good *explicit* Ramsey colorings, so we instead use a probabilistic approach, where we use *randomness* to get our coloring. So far, in our previous probabilistic constructions, we looked at every edge and colored it red or blue with a certain probability. Here, instead of directly coloring the edges randomly, we'll use an elegant trick — we'll construct another underlying random object, which we'll then use to define the coloring. This random object will be a random *tournament* on the same vertex set.

**Definition 5.11.** A tournament is a directed graph in which there is exactly one directed edge between every pair of vertices.

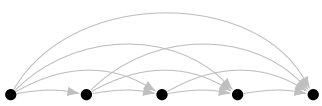


(For example, imagine we have $n$ soccer teams, and we organize a tournament where each team plays each other team exactly once. Then our graph records who won each game — an edge $a \to b$ means that team $a$ beat team $b$. Note that having the edges $a \to b$ and $b \to c$ doesn't tell you anything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

First consider a *random* tournament on the same vertex set — this means that 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 the edge between them — $u \to v$ or $v \to u$ — and we flip an independent coin for each edge.)

Now using this tournament, we want to define a coloring — we need to color every 3-tuple of vertices either red or blue. For any three vertices, there are two cases for what the tournament looks like between them — either it's a cycle, or it's an arrangement with a superior team, a middle team, and a weak team. In the latter case, we say the tournament between them is *transitive*:

**Definition 5.12.** A tournament is transitive if its vertices can be ordered in a line such that all edges go forwards (i.e., the stronger team always beats the weaker team).



Equivalently, a tournament is transitive if whenever $a$ beats $b$ and $b$ beats $c$, then $a$ also beats $c$ (i.e., there are no cycles of length 3).

So we color a 3-tuple of vertices *red* if they form a transitive tournament, and *blue* if they form a cycle.

We now want to check that with positive probability, this has the properties we want (i.e., it avoids red cliques of size $k$ and blue cliques of size 4).

**Claim 5.13 —** 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 (this is what a blue clique of size 4 would be).

*Proof.* Assume that there is such a tournament, and consider three of its vertices, which must form a cycle. Now consider the fourth vertex, which must form a cycle with each of the other edges. Each of these cycles fixes the orientation of two of its edges, and these orientations contradict each other. □

**Remark 5.14.** This is somewhat similar to our first proof of Theorem 2.1 (an upper bound for the Erdős–Szekeres problem) where we used a hypergraph Ramsey argument by taking a coloring of 4-tuples and using a lemma that among any 5 points, we can't have all 4-tuples be non-convex. Similarly, here the construction is based on there being a small configuration that we surely can't have.

Now we need to find the probability of having a red clique of size $k$. This is certainly *possible* (e.g., the entire tournament could be transitive — then all edges would be red, so the entire hypergraph would be a red clique). But we want to show it's *unlikely* (i.e., with high probability, there is no red clique of size $k$).

First, for any given $k$ vertices, we have a red clique on these $k$ vertices if and only if they form a transitive tournament (being a red clique is equivalent to having no blue triples, i.e., no cycles of length 3, and this is equivalent to being transitive). 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 have to be oriented correctly according to that ranking). We can bound this (for large $k$) by

$$k!\,2^{-\binom{k}{2}} \le k^k \cdot 2^{-k(k-1)/2} \le 2^{-k^2/4}$$

(we replace the denominator of $\frac{k^2}{2}$ with something a bit larger in order to eat all the lower-order terms).

This bounds the probability that any *given* $k$ vertices form a red clique. Now union-bounding over all $\binom{n}{k}$ possible choices of $k$ vertices, the total probability of having a red clique of size $k$ is at most

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

(Here we're using the bound $\binom{n}{k} \le n^k$. Depending on context, sometimes it makes sense to use this bound and ignore the $k!$, and sometimes we don't want to ignore the $k!$; here it's much smaller than everything else, so we don't lose much by ignoring it.)

So the probability of having a red clique of size $k$ is strictly less than 1, which means with positive probability there is no such red clique. Meanwhile, there is *necessarily* no blue clique of size 4. So there exists a coloring with no red clique of size $k$ or blue clique of size 4. □

### §5.2.3 The diagonal setting with multiple colors

Theorem 5.8 gives the lower bound $R_3(k,4) \ge 2^{ck}$, so the off-diagonal 3-uniform Ramsey numbers are at least exponential in $k$. We've also seen an upper bound which is exponential in some polynomial in $k$, so single-exponential is the correct behavior.

Theorem 5.8 also gives a lower bound for $R_3(k,k)$ (which is at least $R_3(k,4)$). But as seen last class, we can do better — there is an absolute constant $c > 0$ such that $R_3(k,k) \ge 2^{ck^2}$ (Theorem 5.9). (This can be proved by a relatively straightforward probabilistic argument, and we'll see it in the homework.) But this is still single-exponential in $k$, while 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).

We can also ask the same question if we take more than 2 colors. For any fixed $t$, our upper bound from last class (Theorem 5.3, which works for any number of colors) gives that

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

So the upper bound is still double-exponential in $k$, where the constant depends on the number of colors.

**Remark 5.15.** 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 later discuss what's known and conjectured in that case for *graphs*. Straightforwardly plugging in Theorem 5.3 should an upper bound of

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

which is roughly of the shape $t^{t^{kt}}$. But this probably isn't the right answer.

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

**Theorem 5.16** (Erdős–Hajnal)
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, using the lower bound for diagonal graph Ramsey numbers from Theorem 1.20, we get

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

for $k \ge 3$. This is double-exponential in $k$, so our upper bound is correct up to the constant $c_t$. (Note that if this lower bound holds for 4 colors, then it also holds for any number of colors at least 4.) 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, let $R = R(k-1, k-1) - 1$. This means that $R$ is *not* good enough for the graph Ramsey problem with 2 colors, so we can take an edge-coloring of the complete graph $G$ on vertex set $\{0, \dots, R-1\}$ such that there is no monochromatic red or blue clique of size $k-1$.

We want to find a coloring of a complete 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 for the 3-uniform 4-color problem). So consider the complete 3-uniform hypergraph on the vertex set $\{0,1\}^R$ (which has $2^R$ vertices, labelled by sequences of 0's and 1's), which we call $\mathcal{H}$. We need to find a coloring with 4 colors such that there is no monochromatic clique of size $k$.

To describe the coloring, we first want to think of our labels as binary numbers — every vertex $v \in \{0,1\}^R$ can be represented as a binary number $v = v_{R-1}v_{R-2}\dots v_0$ (i.e., the number $v_0 2^0 + v_1 2^1 + \cdots$). This gives a natural ordering of these vertices, where for $v, w \in \{0,1\}^R$, we say $v < w$ if $v_{R-1}\dots v_0$ is a smaller binary number than $w_{R-1}\dots w_0$, 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 start from the left. At each step, if the two digits are equal, we move one digit to the right. The first time we hit an entry where the two digits differ, if $v$ has a 0 and $w$ has a 1, then $v < w$ (and vice versa). So we'll name this index:

**Notation 5.17.** For $v, w \in \{0,1\}^R$ with $v \neq w$, we define $\delta(v,w) = \max\{j \mid v_j \neq w_j\}$.

In other words, $\delta(v,w)$ 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 5.18 —** The function $\delta$ has the following properties:

(i) For all $v, w, z \in \{0,1\}^R$ with $v < w < z$, we have $\delta(v,w) \neq \delta(w,z)$.

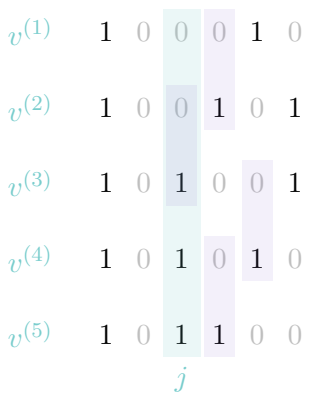
(ii) For $v^{(1)}, \dots, v^{(\ell)} \in \{0,1\}^R$ with $v^{(1)} < \cdots < v^{(\ell)}$, we have

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

(We use upper indices to refer to the $\ell$ distinct vectors, and lower indices to refer to coordinates.)

*Proof.* For (i), if both $\delta$-values were equal to $j$, then the fact that $v < w$ would mean $v_j = 0$ and $w_j = 1$, and the fact that $w < z$ would mean $w_j = 0$ and $z_j = 1$; these can't both be true.

For (ii), consider the first index $j$ where the $v^{(i)}$ are not all equal. Then at that digit $j$, some initial segment of the $v^{(i)}$ will be 0, and then the rest will all be 1. In particular, we'll have $\delta(v^{(1)}, v^{(\ell)}) = j$. Meanwhile, the terms $\delta(v^{(i)}, v^{(i+1)})$ will all be $j$ or greater; 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 the original coloring of $G$ and our function $\delta$ to define a coloring of our 3-uniform hypergraph $\mathcal{H}$ on vertex set $\{0,1\}^R$ with 4 colors. We need to color every triple of vertices. For every triple $v, w, z \in \{0,1\}^R$ ordered such that $v < w < z$, we define the color of $\{v, w, z\}$ 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 $\delta(v, w) < \delta(w, z)$.
- We color $\{v, w, z\}$ *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.

Note that by Claim 5.18(i), these two numbers $\delta(v, w)$ and $\delta(w, z)$ are different, so they define an edge in the original graph $G$ on $R$ vertices, which is either red or blue. Then the color we choose depends on which direction the inequality between these two values points, and the color of the edge they define.

We need to check that this coloring doesn't have a monochromatic clique of size $k$. The cases for the four colors are all analogous, so we'll check this just for dark red. Suppose there is a dark red clique of size $k$ in $\mathcal{H}$; name its vertices $v^{(1)} < \cdots < v^{(k)}$. Since this clique is dark red, every triple of three consecutive elements is dark red, so in particular

$$\delta(v^{(1)}, v^{(2)}) < \delta(v^{(2)}, v^{(3)}) < \cdots < \delta(v^{(k-1)}, v^{(k)}). \tag{5.1}$$

(Each individual inequality comes from looking at three consecutive vertices, which form a dark red edge; then we string those inequalities together.) Then by Claim 5.18(ii), for any $1 \le i < j \le k-1$, we have

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

Now, we know that the edge $\{v^{(i)}, v^{(i+1)}, v^{(j+1)}\}$ is dark red; this means that $\{\delta(v^{(i)}, v^{(i+1)}), \delta(v^{(i+1)}, v^{(j+1)})\}$ 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 $\delta(v^{(1)}, v^{(2)})$, $\ldots$, $\delta(v^{(k-1)}, v^{(k)})$ form a red clique of size $k-1$ in $G$ (the above argument shows that every pair forms a red edge in $G$, and they're all distinct by (5.1)). 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 the other three colors, so we're done. □

March 21, 2023

### §5.2.4 Stepping up

Last week, we saw the upper bound

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

(You can think of $t$ as 2; it's some fixed constant.) This gives a double-exponential bound for 3-uniform hypergraph Ramsey numbers, and by iterating it we get an upper bound for $r$-uniform ones for any fixed $r$.

Last class, we also started discussing lower bounds; but we only saw various lower bounds for the 3-uniform case. The reason we haven't yet talked about higher uniformity is that it turns out that from 3-uniform lower bounds, we can also get lower bounds for any higher uniformity. Theorem 5.3 (the above upper bound) lets us transfer an *upper* bound on Ramsey numbers to one uniformity higher, increasing the tower height by 1; we'll now see a way to do the same with *lower* bounds.

**Theorem 5.19** (Erdős–Hajnal, Stepping up lemma)

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

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

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

**Remark 5.20.** You can probably even omit the factors of 2 from the third term on. But it doesn't really matter (the factors of 2 don't play a major role).

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 could). 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 them for $r = 3$ — because to get to higher uniformities, we can transfer both bounds. We understand how $R_2(k, k)$ works fairly well, and if we could understand how $R_3(k, k)$ works, then we could also understand higher $r$. But $R_3(k, k)$ is not well-understood; we just know it's between single-exponential and double-exponential.

Now we'll prove this theorem.

*Proof.* The proof strategy is in some sense similar to the proof of Theorem 5.16 from last class. It'll be more complicated, though — in that proof, we got a lower bound on uniformity 3 with 4 colors in terms of a 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're not changing the number of colors in the bound, so we don't get to split red into dark red and light red.

Let $R = R_{r-1}(\lceil \frac{k_1-r+4}{2} \rceil, \dots, \lceil \frac{k_t-r+4}{2} \rceil) - 1$, and consider a coloring of the edges of the complete $(r-1)$-uniform hypergraph $H$ on vertex set $\{0, 1, \dots, R\}$ with the colors $1, \dots, t$ such that there is no clique of size $\lceil \frac{k_i-r+4}{2} \rceil$ in color $i$ for any $i$ (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 actually give a bound of $2^R + 1$, but the $+1$ is very unimportant, so we didn't write it down).

Consider the complete $r$-uniform hypergraph on the vertex set $\{0, 1\}^R$. We're going to find a coloring of its edges with colors $1, \dots, t$ without a clique of size $k_i$ in color $i$ for any $i \in \{1, \dots, t\}$.

As in the proof of Theorem 5.16, we interpret each vertex $v \in \{0, 1\}^R$ as a binary number $v = v_{R-1} v_{R-2} \dots v_0$. For $v, w \in \{0, 1\}^R$ 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). Last class, we saw $\delta$ has the following properties (Claim 5.18):

(i) For $v, w \in \{0, 1\}^R$ with $v < w$, we have $v_j = 0$ and $w_j = 1$.

(ii) For $v, w, z \in \{0, 1\}^R$ with $v < w < z$, we have $\delta(v, w) \neq \delta(w, z)$. This is because if they were the same index $j$, then (i) would tell us that $v_j = 0$ and $w_j = 1$ (from $v < w$), as well as that $w_j = 0$ and $z_j = 1$ (from $w < z$); but $w_j$ cannot be both 0 and 1.

(iii) If we have $\ell$ vectors $v^{(1)}, \dots, v^{(\ell)} \in \{0, 1\}^R$ with $v^{(1)} < \dots < v^{(\ell)}$, 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 these vectors aren't all equal, then because they're ordered in this way, some initial part of these vectors will have $j$th coordinate 0, and the rest will have $j$th coordinate 1. This means $j$ is the first index 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 values $\delta(v^{(i)}, v^{(i+1)})$ is $j$; and all the others are at most $j$, because the vectors are all equal at all indices before $j$.

Now we're ready to define our coloring. We're trying to color the edges of our hypergraph, where every edge consists of $r$ vertices; this means we want to define a color for every $r$-tuple of vertices. So for every $r$-tuple of vertices ordered so that $v^{(1)} < \cdots < v^{(r)}$, 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 $1 \le i \le r-1$ (i.e., these are the $\delta$-values between consecutive pairs). Note that $\delta_i \neq \delta_{i+1}$ for all $1 \le i \le r-2$ (by (ii), since $\delta_i$ and $\delta_{i+1}$ are defined as $\delta(v^{(i)}, v^{(i+1)})$ and $\delta(v^{(i+1)}, v^{(i+2)})$). We have four cases:

(1) If $\delta_1 < \delta_2 < \cdots < \delta_{r-1}$ or $\delta_1 > \delta_2 > \cdots > \delta_{r-1}$, then we color $\{v^{(1)}, \dots, v^{(r)}\}$ with the color of the edge $\{\delta_1, \dots, \delta_{r-1}\}$ in $H$ (recall that $H$ is our complete $(r-1)$-uniform hypergraph on vertex set $\{0, \dots, R-1\}$ — the $\delta_i$ are all elements of this set, and the monotonicity condition means they're all distinct, so $\{\delta_1, \dots, \delta_{r-1}\}$ is a set of $r-1$ distinct vertices in $H$, and has a well-defined color).

(2) If $\delta_1 < \delta_2$ and $\delta_2 > \delta_3$ (i.e., there's a local maximum at $\delta_2$), then we color $\{v^{(1)}, \dots, v^{(r)}\}$ in color 1. (In this case, we don't look at $H$ at all, and simply use color 1. Note that this is the same color 1 that appeared in $H$, since we can't use new colors.)

(3) If $\delta_1 > \delta_2$ and $\delta_2 < \delta_3$ (i.e., $\delta_2$ is a local minimum), then we color $\{v^{(1)}, \dots, v^{(r)}\}$ in color 2. (Again, this is some color that already appeared in $H$, different from color 1.)

(4) If none of these cases occur, we color $\{v^{(1)}, \dots, v^{(r)}\}$ with an arbitrary color (chosen from 1, …, $t$).

Now we've colored every edge (because every edge not colored in the first three steps is automatically colored in the last). This definition is strange (*a priori* it seems strange to zone in on $\delta_1$, $\delta_2$, and $\delta_3$ specifically — the proof is very asymmetric), and it's surprising that it works.

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

Suppose there is a clique in color 1 of size $k_1$, formed by vertices $w^{(1)} < \cdots < w^{(k_1)}$. Unsurprisingly, to analyze this clique we'll need to talk about the $\delta$-values between its vertices. So for each $1 \le i \le k_1 - 1$, let

$$\delta'_i = \delta(w^{(i)}, w^{(i+1)})$$

(we use the notation $\delta'_i$ to avoid notation clash with the $\delta_i$ we defined earlier when describing the color of an *edge*). Note that $\delta'_i \neq \delta'_{i+1}$ for all $1 \le i \le k_1 - 2$ (for the same reason as before — applying (ii) to $w^{(i)}$, $w^{(i+1)}$, and $w^{(i+2)}$).

Now we have a sequence $\delta'_1, \dots, \delta'_{k_1-1}$. So far, we only know that any two consecutive terms are distinct. But the next claim tells us that this sequence has much more structure — specifically, there's a reasonably long block which is nicely monotone.

**Claim 5.21 —** There are indices $j$ and $j'$ with $1 \le j < j' \le k_1 - r + 3$ and

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

such that either $\delta'_j < \delta'_{j+1} < \cdots < \delta'_{j'}$ or $\delta'_j > \delta'_{j+1} > \cdots > \delta'_{j'}$.

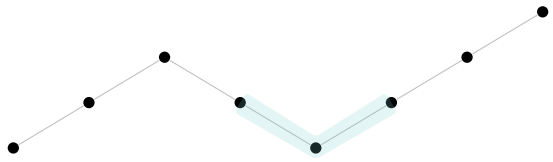
*Proof.* To prove this, we want to investigate the monotonicity behavior of the entire sequence $\delta'_1, \dots, \delta'_{k_1-1}$. *A priori*, it could do all kinds of crazy things. But recall that our vectors form a clique in color 1; so when we pick any $r$ of these vectors, we never arrive at the case (3), which would lead us to color 2 (we might arrive at any of the other cases, but not this one). 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 would force us to be in the case (3). Then because there are no local minima, the monotonicity behavior of the sequence will become much easier to describe.

More precisely, we claim that there does not exist any $1 \le i \le k_1 - r + 1$ such that $\delta'_i > \delta'_{i+1}$ and $\delta'_{i+1} < \delta'_{i+2}$ (i.e., such that $\delta'_{i+1}$ is a local minimum) — if there were such an $i$, then the edge

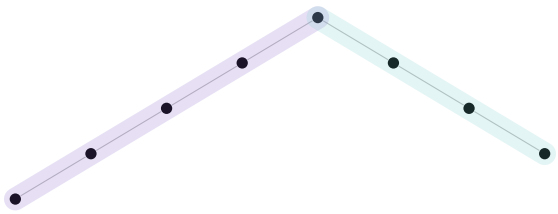
$$\{w^{(i)}, w^{(i+1)}, w^{(i+2)}, \dots, w^{(i+r-1)}\}$$

would have color 2 (which it doesn't, because it's part of our clique of color 1). To check this carefully, first note that this is a valid edge because $i + r - 1 \le k_1$. When we color this edge, we look at the consecutive $\delta$-values among its vertices, which are exactly $\delta'_i, \delta'_{i+1}, \dots, \delta'_{i+r-2}$. Then we're exactly in the case where the second is smaller than the first and third, which means we color it with color 2 by (3).

So the sequence $\delta'_1, \dots, \delta'_{k_1-r+3}$ cannot have a local minimum. This means that once the sequence starts decreasing, it must keep decreasing forever — if we went down and then up, then we'd have a local minimum.



This means the sequence can be split into an increasing part followed by a decreasing part (either part may be empty).



This sequence has $k_1 - r + 3$ terms, so it has $k_1 - r + 2$ steps. This means one of the two parts must consist of at least

$$\left\lceil \frac{k_1 - r + 2}{2} \right\rceil = \left\lceil \frac{k_1 - r + 4}{2} \right\rceil - 1$$

steps, which gives a monotone block of the desired length (note that the number of steps in $\delta_j, \delta_{j+1}, \dots, \delta_{j'}$ is $j' - j$). □

**Remark 5.22.** For the colors other than 1 or 2, the *entire* sequence up to index $k_1 - r + 3$ must be monotone for the same reason. So we can essentially omit the factors of 2 for the remaining colors in Theorem 5.19; but this doesn't really matter.

First suppose that $\delta'_j > \dots > \delta'_{j'}$ (in the output of Claim 5.21). The $j' - j + 1 \ge \lceil \frac{k_1-r+4}{2} \rceil$ vertices $\delta'_j, \dots, \delta'_{j'}$ in $H$ cannot form a clique of color 1 in $H$ (our $(r-1)$-uniform hypergraph, which we defined to have no clique of this size in color 1). This means among them, we can pick $r - 1$ vertices whose edge doesn't have color 1 in $H$ — i.e., there exist 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 $H$. We now want to use this information to contradict the fact that $w^{(1)}, \dots, w^{(k_1)}$ forms a clique of color 1 in our original hypergraph — we want to find an edge among these vertices with the same color as the above edge in $H$.

Let's consider the edge

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

(The reason it makes sense to choose $w^{(i_{r-1}+1)}$ is that we need to add one more vector to have $r$ of them, and the $\delta'_i$ store information about the $\delta$-values between consecutive $w^{(i)}$.) Plugging this edge into our color definition, these vectors are already ordered correctly (i.e., $w^{(i_1)} < w^{(i_2)} < \cdots$). When we look at their consecutive $\delta$-values, we have

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

(the first equality is by (iii), and the last is since the $\delta'_i$ 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 $\delta$-values are monotone decreasing, so we're in the case (1); this means the color of the edge $\{w^{(i_1)}, \dots, w^{(i_{r-1}+1)}\}$ is the same as the color of the edge $\{\delta'_{i_1}, \dots, \delta'_{i_{r-1}}\}$ in $H$. But the latter color is not 1 (by how we defined $\{\delta'_{i_1}, \dots, \delta'_{i_{r-1}}\}$). This is a contradiction — we've found an edge not of color 1 in what was supposed to be a clique of color 1.

The other case of Claim 5.21, where we instead have $\delta'_j < \cdots < \delta'_{j'}$, is very similar. We again pick indices such that the edge $\{\delta'_{i_1}, \dots, \delta'_{i_r}\}$ doesn't have color 1 in $H$. But this time, we look at the edge

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

The same analysis as in the previous case shows that this edge in our $r$-uniform hypergraph has the same color as the edge $\{\delta'_{i_1}, \dots, \delta'_{i_r}\}$ in $H$ (which we assumed is not color 1).

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 dealt with analogously. (For color 2, we run the same argument except that in the proof of Claim 5.21, we exclude a local *maximum* rather than a local minimum, so the sequence splits into a decreasing part followed by an increasing part. For the other colors, we can simply copy-paste the proof for color 1 — we get to exclude both a local maximum and a local minimum, so in particular we still get to exclude a local minimum.) □

Theorem 5.3 and 5.19 together mean that if we have both upper and lower bounds for Ramsey numbers of uniformity $r$, then we can boost them up to higher uniformity. To describe these bounds, we use the following definition.

**Definition 5.23.** The tower function is defined by $t_1(x) = x$ and $t_h(x) = 2^{t_{h-1}(x)}$.

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

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

$$t_{r-1}(c_{r,t}k^2) \leq R_r(k, \dots, k) \leq t_r(c'_{r,t}k).$$

So we have an upper bound and lower bound whose tower heights differ by 1; this is exactly because we have such bounds for $r = 3$ (the lower bound from Theorem 5.9 is single-exponential with $k^2$ in the exponent,

and the upper bound from Theorem 5.3 is double-exponential). In particular, to close this gap, it's enough to close it in uniformity 3; this is why uniformity 3 is the crucial open problem. Also, for $t \geq 4$ colors, the gap disappears because we've proven a double-exponential lower bound for the 3-uniform case (Theorem 5.16) — so we obtain

$$t_r(c_{r,t}k) \leq R_r(k, \dots, k) \leq t_r(c'_{r,t}k).$$



## §6 Ramsey numbers of graphs

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

In the ordinary Ramsey theorem (Theorem 1.4), we take $t$ colors and specify target clique sizes for each. But now, rather than cliques, we're looking for some other specific graphs.

**Definition 6.1.** For graphs $G_1, \dots, G_t$, 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.) The ordinary Ramsey numbers correspond to the case where the target graphs $G_1, \dots, G_t$ are cliques of the corresponding sizes.

**Example 6.2**

Let $G_1$ be the star on four vertices. Then the following is a copy of $G_1$ in purple in a 6-vertex graph (it doesn't matter what color the remaining edges in this 6-vertex graph are).

We wrote this as a definition rather than a theorem that such a number exists. But it does exist; in fact,

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

(where the right-hand side is the ordinary $t$-color Ramsey number where we're looking for cliques, which we know does exist) — 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 has also been some study of this more general setting.

## §6.1 A general lower bound

**Theorem 6.3** (Chvátal–Harary 1972)

For any two connected graphs $G$ and $H$, we have

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

If we ignore the $\pm 1$'s, this is roughly $\chi(G)\cdot|V(H)|$. To review what $\chi(G)$ means:

**Definition 6.4.** 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 6.5**

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

We'll prove Theorem 6.3 on the homework (there's a relatively simple construction).

**Question 6.6.** Is this a good bound?

In some cases, it's a terrible bound. For example, when $G$ and $H$ are both cliques on $k$ vertices, it gives the bound $R(k,k) \geq (k-1)^2+1$ which is quite terrible (the right answer is exponential in $k$). And in fact, the proof of Theorem 6.3 is similar to what we did in Example 1.19 in the first class, when we got this lower bound on $R(k,k)$ via an explicit construction. So it's natural to assume that maybe this bound is generally terrible. But that's not always true — there are cases where it's a pretty good bound, and in fact there are cases where it's *exactly* tight.

## §6.2 Ramsey numbers of trees vs. cliques

Now we'll state an example where we can determine the Ramsey number *exactly*, and it matches the lower bound of Theorem 6.3.

**Notation 6.7.** We use $K_s$ to denote a clique on $s$ vertices.

**Theorem 6.8** (Chvátal 1977)

Let $T$ be a tree with $t \geq 1$ vertices, and let $s \geq 1$. Then

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

(This matches the bound in Theorem 6.3 if we switch the places of $G$ and $H$ — of course $R(G,H)=R(H,G)$, since we can swap the names of the colors.)

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 from Theorem 6.3 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, we can apply Theorem 6.3 with $G = K_s$ and $H = T$ to get

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

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 it's easier to think about embedding a tree into this subgraph, and the condition that we have no blue clique of size $s$ still translates well). So let $H$ be the graph consisting of the red edges; then

$$|V(H)| = (s-1)(t-1) + 1.$$

We assumed that $H$ doesn't have a subgraph isomorphic to $T$; and since there's no blue clique of size $s$, the independence number of $H$ is

$$\alpha(H) \leq s - 1.$$

**Remark 6.9.** We saw this same translation — where we look at the subgraph formed by one color, and cliques of the other color become independent sets — when discussing off-diagonal Ramsey numbers (Theorem 4.2).

We'll work with the chromatic number of $H$. In general, 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 has to be an independent set, since any two neighbors must have different colors; then $\chi(H)$ color classes, each of size at most $\alpha(H)$, must cover all the vertices). Here, 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 $\chi(H) \geq t$.

The next trick is to consider a *minimal subgraph* of $H$ whose chromatic number is at least $t$ — let $H'$ be a minimal subgraph of $H$ with the property that $\chi(H') \geq t$. The fact that $\chi(H) \geq t$ means such a subgraph $H'$ does exist. (We don't require the subgraph to be induced — minimality means that we can't delete any vertices *or* edges while preserving this condition — but here it won't matter, and we'll only use the fact that we can't delete vertices.)

The reason we want to do this is that the minimality property gives us strong information about all the degrees in $H'$.

**Claim 6.10 —** 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 of its vertices have high degree; this will be pretty useful in being able to find a tree isomorphic to $T$. The point of doing all of this is not that we care about the chromatic number of $H$ or $H'$ — it's really to get this degree property, and the chromatic number is just a vehicle for getting us to a subgraph with this strong degree condition.

(Note that $H'$ is nonempty, since $\chi(H') \geq t > 0$.)

*Proof.* Assume for contradiction that $\deg_{H'}(v) \leq t-2$ for some vertex $v \in H'$. Now we'll 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 defined $H'$ to be the *smallest* graph with $\chi(H') \geq t$, so since $H''$ is smaller than $H'$, it can't also satisfy this condition).

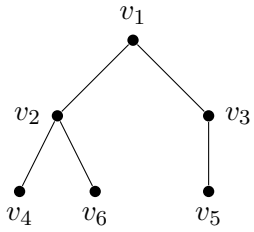
Now take a coloring of $H''$ with at most $t-1$ colors, where no two adjacent vertices have the same color. Then we can extend this coloring from $H''$ to $H'$ by assigning $v$ some color which is different from the color of all its neighbors. We can do this because it has $\deg_{H'}(v) \leq t-2$ neighbors, 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.

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. □

**Remark 6.11.** More generally, this argument shows that for any graph $H$, we can find a (nonempty) subgraph of $H$ where all degrees are at least $\chi(H)-1$ (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 a subgraph isomorphic to $T$ in $H'$. The main idea is that 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.

To more formally describe how we'll embed $T$ as a subgraph of $H'$, we first want to label the vertices of $T$ in some reasonable way, so that when we perform the greedy embedding procedure, we don't jump around the tree. So let the vertices of $T$ be $v_1, \ldots, v_t$, labelled in such a way that for every $2 \leq i \leq t$, the vertex $v_i$ has an edge to exactly one of $v_1, \ldots, v_{i-1}$. This condition intuitively means that we're having the tree grow naturally, rather than having two separate parts which join later. (For example, either depth-first or breadth-first search would give an ordering with this property.) We think of $v_1$ as the *root* of our tree.




We'll 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, and we need to preserve this edge in our embedding. For example, if we're trying to embed $v_6$, and its one edge is to $v_2$, then we need to select a new neighbor of $v_2$ at which we embed $v_6$. The number of neighbors of $v_2$ in $H'$ is at least $t-1$. We can't use a vertex we've already chosen, so some of these neighbors might be taken up by the previous images. But the number of previous images is at most $t-2$ (since we exclude the vertex we're currently trying to embed, and $v_2$ itself). So at most $t-2$ neighbors are blocked, which means among the $t-1$ neighbors, we can find one to use.

More formally, we can first embed $v_1$ because $H'$ is nonempty. For each $2 \le i \le t$, suppose we've already embedded $v_1, \dots, v_{i-1}$, and call their images $w_1, \dots, w_{i-1}$; we'll now try to embed $v_i$.

Let $v_j$ be the unique neighbor of $v_i$ among $v_1, \dots, v_{i-1}$ in $T$. Then its image $w_j$ in $H'$ has degree at least $t-1$ in $H'$; and among these $t-1$ neighbors of $w_j$ in $H'$, at most $t-2$ have already been used up by being the images of vertices in $\{v_1, \dots, v_{i-1}\} \setminus \{v_j\}$ (since $i \le t$ and $v_j$ is missing, there's at most $t-2$ vertices in this list). So there exists a vertex $w_i$ in $H'$ which is adjacent to $w_j$ and is not equal to any of the previous images $w_1, \dots, w_{i-1}$. Then $w_i$ is a suitable image for $v_i$ in the embedding, so we can map $v_i$ to $w_i$ (and continue the embedding process).

This shows that we can embed $T$ into $H'$ — so $T$ occurs as a subgraph of $H'$, and therefore of $H$. This is a contradiction, as we assumed $H$ doesn't have a subgraph isomorphic to $T$. □

**Remark 6.12.** As a philosophical remark, the reason we needed $H'$ is that $H$ might have some vertices with really low degree. In order to run this greedy embedding procedure, we need this minimum degree condition on *every* vertex. In $H$, *many* vertices may have high degree, but if we just embed greedily without thinking ahead, we might accidentally use a low-degree vertex where we'll get stuck later.

## §6.3 Ramsey numbers of bounded degree graphs

In some sense, Theorem 6.8 answers the question of how good our lower bound for $R(G, H)$ (from Theorem 6.3) is — for trees vs. cliques, it's exactly tight. However, there are other cases to consider as well.

As a thought experiment, suppose that $G$ has bounded chromatic number. Then Theorem 6.3 gives a lower bound which is linear in the number of vertices of $H$; this is not particularly interesting, because we definitely have $R(G, H) \ge |V(H)|$. So far, 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 at least quadratic. Even the bound in Theorem 6.8 looks quadratic (as a function of both $s$ and $t$). So all the Ramsey numbers we've encountered so far are at least quadratic; having a bound that's *linear* in the number of vertices seems totally off. But we'll now see a theorem (resolving a conjecture of Erdős and Burr) which tells us that some Ramsey numbers really *are* linear. This is really cool and surprising.

**Theorem 6.13** (Chvátal–Rödl–Szemerédi–Trotter 1983)
For any graph $H$ with $k$ vertices and maximum degree $d$, we have

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

where $C_d$ is a constant only depending on $d$.

This theorem is in the 'diagonal' regime, where we consider $R(H, H)$; we think of $d$ as fixed and $k$ as large. It tells us that this diagonal Ramsey number is linear in the number of vertices $k$, where the linearity factor depends on the maximum degree $d$ — in other words, for graphs of bounded maximum degree, we have a *linear* upper bound on $R(H, H)$. (Of course, we also have a linear *lower* bound, because we need at least $k$ vertices to find a copy of $H$.)



This is really surprising because most of the Ramsey numbers we've seen have been huge — we saw that Ramsey numbers for cliques are exponential, and hypergraph Ramsey numbers are terribly fast-growing. For trees and cliques we had roughly $st$-type behavior, but that's still quadratic (Theorem 6.8). Meanwhile, here we have something *linear*; that's really surprising, and why this is such an important result.

**Remark 6.14.** As a historical remark, the original proof of Theorem 6.13 used Szemerédi's regularity lemma. At that point, this was a rather new tool — it appeared in the 8 years between when this was conjectured (by Erdős and Burr in 1975) and when it was proven, and the proof was one of the first applications of Szemerédi's regularity lemma (beyond Szemerédi's original work introducing it).

But the proof we'll see today doesn't use the regularity lemma (not all of us know it, and it takes a while to discuss; you'll learn about it in **18.225**). Instead, we'll see an alternative proof due to Graham, Rödl, and Rucinski. In particular, because the proof doesn't use the regularity lemma, it gives better quantitative dependencies of $C_d$ on $d$, although we won't really care about that.

The proof of Theorem 6.13 is quite long (so it won't fit into today's class), but it uses several interesting ideas that also appear in other places.

As usual, we'll 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 red edges and the density of blue edges in several subsets. So we'll split the proof into several lemmas that cover different cases of how to find this red or blue copy of $H$.

First, we'll write down all the density definitions we need.

**Definition 6.15.** For a graph $G$ and disjoint vertex sets $X, Y \subseteq V(G)$, the density between $X$ and $Y$ is

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

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

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

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

In both cases, we're counting the number of edges which $G$ actually has (inside or between our vertex sets), and dividing by the number of edges which could possibly exist; this is what you'd intuitively expect density to mean. We also define the density of $G$ as

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

### §6.3.1 The very dense case

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

**Lemma 6.16**

Let $H$ be a graph 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 one color in some vertex *subset* of our original graph.

How should we interpret these two conditions? The first assumption, that $G$ has at least $4k$ vertices, 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 thing is the density assumption — note that $1 - \frac{1}{8d}$ is really close to 1, meaning that $G$ has

almost all the possible edges. (In the proof of Theorem 6.13, we'll see various density assumptions; but here we should emphasize that this density is really big.)

*Proof.* Our goal is to embed $H$ as a subgraph in $G$ — this means we want to find distinct images for all the vertices of $H$ in $G$ such that all edges in $H$ are mapped to edges in $G$. Most things in $G$ are edges, which is good; so we just need to be careful to not hit a non-edge. It's easier to think about this if we take the complement of $G$, where we interchange edges and non-edges — so this complement has few edges, and edges in this complement are what create conflicts when we embed $H$. So when we embed, we want to make sure to avoid high-degree vertices 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 we have

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

Let $n = |V(G)|$, 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 $\frac{n}{4d}$.

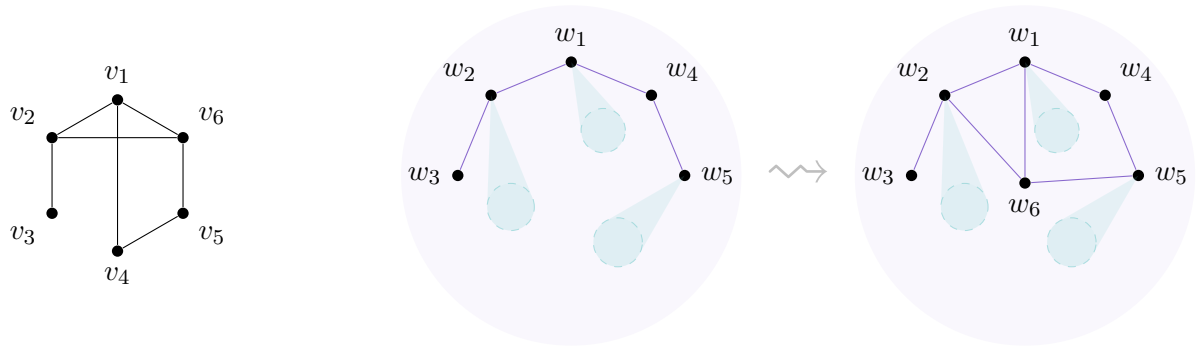
Now we kick these bad vertices out — let $G'$ be the graph obtained from $\overline{G}$ by deleting all vertices $v \in V(G)$ with $\deg_{\overline{G}}(v) > \frac{n}{4d}$. Then $|V(G')| \geq \frac{n}{2}$, and $\deg_{G'}(v) \leq \frac{n}{4d}$ for all $v \in V(G')$.

We now want to embed $H$ into $G$ using only the vertices in $G'$. And the idea is that we can do this greedily, since we've already kicked out all bad vertices (this is similar to the greedy embedding argument we saw when proving Theorem 6.8). Embedding greedily means that we embed one vertex of $H$ at a time (when embedding a tree in Theorem 6.8 we had to be careful about how we ordered the vertices in $H$; here we don't, and we can do it arbitrarily), being careful to map edges in $H$ to edges in $G$.

More precisely, let $v_1, \ldots, v_k$ be the vertices of $H$; we want to map these to vertices $w_1, \ldots, w_k$ of $G$ which are in $V(G')$. (It's important that we always stay inside this good set $V(G')$ — if we go outside it, we might create too many conflicts for later steps.) Suppose that for some $j$, we've already defined the images $w_1, \ldots, w_{j-1}$ in $V(G')$, and now we want to find $w_j \in V(G')$.

We need to respect all the edge relationships from the new vertex to the ones we've already defined — $w_j$ should have edges to all vertices $w_i$ (with $1 \leq i \leq j-1$) for which $v_j$ has an edge to $v_i$. First, there are at most $d$ such vertices $w_i$ (since $H$ has maximum degree $d$, so $v_j$ has at most $d$ neighbors $v_i$). 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 $\deg_{G'}(w_i) \leq \frac{n}{4d}$, since the edges in $G'$ are precisely the non-edges in $G$ (inside this 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 $d$ values 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

$$\left|V(G') \setminus \{w_1, \ldots, w_{j-1}\}\right| \geq \frac{n}{2} - (k-1) \geq \frac{n}{4} + 1$$

(since we're not allowed to choose any of the previously chosen vertices). So it is indeed possible to choose $w_j$. This tells us we can always choose the next vertex $w_j$ respecting the desired edge relationships. We can do this for each $j = 1, \ldots, k$, which means we can indeed embed all our vertices (i.e., find distinct images for all of $v_1, \ldots, v_k$ such that edges in $H$ are mapped to edges in $G$), so we've found an image of $H$ in $G$. (We've actually found an image of $H$ in $V(G')$, but we don't care about this in the end — we just needed it to maintain the condition in the greedy embedding that the number of forbidden vertices is small.) □

### §6.3.2 The bi-dense case

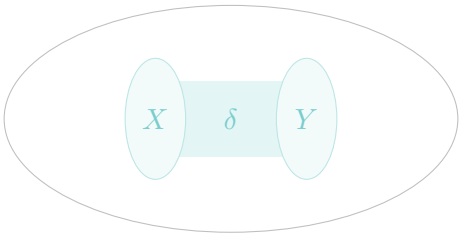
Lemma 6.16 tells us that if we have a graph which is very dense, then we can find $H$ as a subgraph. In particular, if either the red graph or the blue graph is very dense, then we're happy. But in most colorings, this won't happen (in most colorings, both colors will have density around $\frac{1}{2}$). But in fact, Lemma 6.16 tells us more, since we can also apply it to *subsets* of the red or blue graph. So if within some vertex 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 density 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 formally state this, we need a good notion of the blue edges being well-spread out. We'll use a definition which is a bit ad hoc for this proof (unlike the earlier density definitions, this isn't a standard term).

**Definition 6.17.** We say that a graph $G$ is bi-$(\rho, \delta)$-dense (for $0 \leq \rho \leq 1$ 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 $d(X, Y) \geq \delta$.

In other words, for any two vertex subsets which are not too small, the density between them should be at least $\delta$; and $\rho$ quantifies what 'not too small' means.




(This definition is only meaningful if $0 \leq \rho \leq \frac{1}{2}$ — if $\rho > \frac{1}{2}$, then *every* graph is bi-$(\rho, \delta)$-dense, as no such $X$ and $Y$ exist.)

The next lemma tells us that this assumption also lets us find a copy of $H$.

**Lemma 6.18**

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 $8\delta^{-d} d \cdot k$ vertices, then $G$ contains $H$ as a subgraph.

Here $H$ is a given graph with $k$ vertices and maximum degree $d$ (we think of $d$ as fixed and $k$ as large); this lemma gives another sufficient criterion for $G$ to contain $H$ as a subgraph (different from the one in Lemma 6.16). We want that criterion to be that $G$ satisfies a bi-$(\rho, \delta)$-density assumption, but what $\rho$ and $\delta$ should we use? This lemma says that we can take $\delta$ to be *anything*; and then if we choose $\rho$ appropriately (based on $\delta$), we do get such a criterion. Of course, we also need $G$ to have enough vertices — otherwise it would be hopeless to find a copy of $H$. This necessary number of vertices is again linear in $k$ (for fixed $d$ and $\delta$).

In the end, we'll take $\delta$ to be some function of $d$ (e.g., $\delta = \frac{1}{10d}$) to combine this with Lemma 6.16.

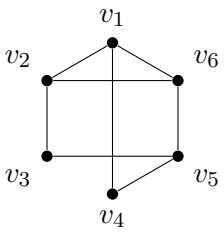
*Proof.* The idea is again to embed $H$ greedily into $G$, but we'll have to be more careful in how we do this than in the proof of Lemma 6.16. Again, we'll label the vertices of $H$ as $v_1, \dots, v_k$, and embed them one by one. But this time, we'll need to keep track of the allowed candidates for the images of all the remaining vertices (i.e., the vertices in $G$ which have the necessary edges to the vertices we've already embedded). We'll need to maintain the condition that these candidate sets are big enough.

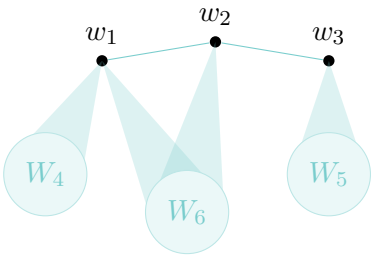
As a preprocessing step, we also want these candidate sets to be disjoint for any two vertices for which we need an edge, since the bi-$(\rho, \delta)$-density hypothesis only gives us a lower bound on edges between *disjoint* subsets. Obviously we can't make the candidate sets disjoint for *all* the vertices, since the number of vertices in $G$ is only linear in $k$ — this would make the candidate sets tiny, while we want them to have size linear in $k$. But we do want the candidate sets to be disjoint between all vertices for which we need edges. To do this, we'll 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.

More explicitly, first note that $\chi(H) \le d+1$, i.e., the vertices of $H$ can be colored with colors $1, \dots, d+1$ such that any two adjacent vertices have different colors. (This is because we can color greedily — whenever we want to assign a new vertex a color, it has at most $d$ neighbors, so there are at most $d$ forbidden colors, which means there's an available color we can choose.) We now use these color classes to get a partition of $G$ — partition $V(G) = U_1 \cup U_2 \cup \dots \cup U_{d+1}$ into parts of almost equal sizes (i.e., sizes differing by at most 1), so that

$$|U_i| \ge \left\lfloor \frac{|V(G)|}{d+1} \right\rfloor \ge \frac{|V(G)|}{4d}.$$

Now in the beginning, our candidate set for every vertex of $H$ is obtained by looking at its color, and taking the corresponding set $U_i$ — so this gives an initialization for the candidate sets. Then we'll greedily go on to embed the vertices of $H$ into $G$. We have to be careful that at each step, the newly chosen vertex has edges to all the vertices it's supposed to have edges to. We'll track this by adjusting the candidate sets — whenever we embed a new vertex, we'll adjust the candidate sets $W_i$ for the future vertices $v_i$ so that they only consist of vertices which have the correct edge relationships (to the vertices we've embedded so far) to possibly be an embedding of $v_i$. (In particular, vertices of the same color in $H$ start out having the same candidate sets; but their candidate sets may eventually become different, because those vertices have different edge relationships.)




So let $v_1, \dots, v_k$ be the vertices of $H$. Our goal is to find a copy of $H$ in $G$ — i.e., to map these to $w_1, \dots, w_k$ in $G$ such that $w_i \in U_h$ for every vertex $i$ of color $h$, respecting all edge relationships. The next claim says that we can run this greedy embedding procedure (where we update the candidate sets at every step) and maintain that these candidate sets have large enough size.

**Claim 6.19 —** For all $0 \le \ell \le k$, we can find distinct images $w_1, \dots, w_\ell$ of $v_1, \dots, v_\ell$ in $G$, such that:

(1) For each $1 \le i \le \ell$, if $v_i$ has color $h$, then $w_i \in U_h$.

(2) $w_i w_j$ is an edge whenever $v_i v_j$ is an edge.

(3) For each $\ell < i \le k$, the following holds: Let $h$ be 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 third condition tells us that the candidate sets for the future vertices are not too small. For each $i$, $N_\ell(i)$ records the neighbors of $v_i$ among the vertices $v_1, \dots, v_\ell$ that we've already embedded; when we embed $v_i$, we'll need to preserve all these edge relationships. Then we're looking at all the vertices in $U_h$ which are candidates for $v_i$ (meaning that they have edges to all vertices in $N_\ell(i)$). And this condition tells us that the number of such vertices is at least $\delta^{|N_\ell(i)|} |U_h|$. This is reasonable because for every edge relationship we have, we'd expect the number of candidate vertices to go down by a factor of $\delta$.

The proof of this claim is roughly as long as the statement — once we've set it up right, the claim basically inducts on its own.

*Proof.* We use induction on $\ell$. 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 adjacency conditions because nothing is embedded yet, so the candidate sets are just the $U_h$ themselves.

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 (such that these conditions hold). All we need to do is find a good image $w_\ell$ for $v_\ell$ in the correct set $U_h$ and with all the correct edges to the previously defined $w_i$ — both of these conditions will automatically be satisfied by choosing $w_\ell$ from the candidate set we currently have for it — and such that when we update the other candidate sets accordingly, they don't get too small.

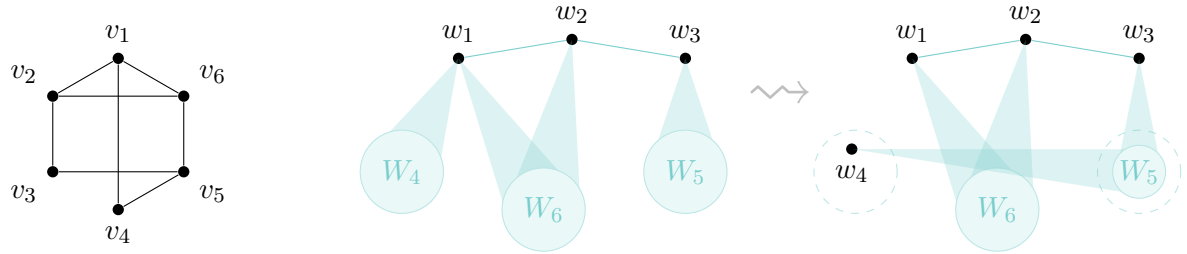
Let $W_1, \dots, W_k$ be the current candidate sets for the vertices $w_\ell, \dots, w_k$, as described in (3). So for all $i = \ell, \dots, k$, we have that

$$|W_i| \ge \delta^{|N_{\ell-1}(i)|} \cdot |U_h| \ge \delta^{|N_{\ell-1}(i)|} \cdot \frac{|V(G)|}{4d} \ge \frac{\delta^d}{4d} \cdot |V(G)|\,. \tag{6.1}$$

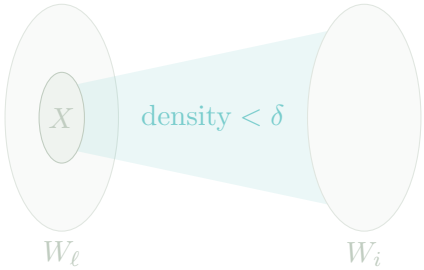
We now want to find a vertex $w_\ell \in W_\ell$ (this will automatically satisfy (1) and (2)) such that

$$|N(w_\ell) \cap W_i| \ge \delta\, |W_i| \tag{6.2}$$

for all $i = \ell + 1, \dots, k$ such that $v_i$ is adjacent to $v_\ell$. This is because the choice of $w_\ell$ doesn't influence any of the candidate sets for vertices $v_i$ which are not adjacent to $v_\ell$, so we don't need to update those; but for ones which are adjacent, we need to update the sets to only include neighbors of $w_\ell$. And in these cases, we've also increased $|N_\ell(i)|$ by 1 (compared to $|N_{\ell-1}(i)|$, since now we also have $\ell$). So we're allowed one extra factor of $\delta$, which means that if we can shrink $W_i$ by just a factor of $\delta$, then we're happy.



First, there are at most $d$ such indices $i$, since $v_\ell$ 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 (6.2). This is because if we take the set $X \subseteq W_\ell$ of all choices that fail the condition and think of it as a vertex set, the density between $X$ amd $W_i$ is less than $\delta$ (each failing vertex *individually* has density less than $\delta$ to $W_i$ — this is what it means to fail (6.2) — so this remains true when we combine them). And the lemma statement assumes that $G$ is bi-$(\frac{\delta^d}{8d^2}, \delta)$-dense, and $W_i$ itself has more than a $\frac{\delta^d}{8d^2}$-fraction of all vertices (it has at least a $\frac{\delta^d}{4d}$-fraction, as seen in (6.1)). So if $X$ was also at least a $\frac{\delta^d}{8d^2}$-fraction of all vertices, then these two sets would contradict the bi-density assumption; this means we have at most this many failing vertices.




There are at most $d$ such indices $i$, so the total number of forbidden choices for $w_\ell$ is at most

$$d \cdot \frac{\delta^d}{8d^2}|V(G)| = \frac{\delta^d}{8d}|V(G)| \le \frac{1}{2}|W_\ell|$$

(using the lower bound for $|W_\ell|$ from (6.1)). So at most half the choices for $w_\ell$ are forbidden, which means we can make some valid choice. (You also have to check that you can make a choice which is distinct from the previously chosen images; but there are at most $k$ of those, and we can check that $\frac{1}{2}|W_\ell| > k$.) □

Then the lemma trivially follows from taking $\ell = k$ in Claim 6.19 — then we've found distinct images for all the vertices in $G$ with the correct edge relationships, which means we've embedded $H$ into $G$. □



### §6.3.3 A dichotomy between the cases

So far, we've seen two lemmas which give a sufficient condition for a graph $G$ to contain $H$ as a subgraph:

- Lemma 6.16 says that if $G$ is extremely dense, then it contains $H$ as a subgraph.
- Lemma 6.18 says that if $G$ is bi-$(\rho, \delta)$-dense (where $\rho$ is an appropriate function of $\delta$), then it contains $H$ as a subgraph. Here, instead of requiring a density very close to 1, we're only requiring a *moderate* density $\delta$ (which we should think of as very small). But the point is that we don't require this density just on the graph as a whole, but rather between *any* two vertex subsets of reasonable size.

Now we'll show that these two lemmas together imply Theorem 6.13. The point is that we have two different sufficient criteria to find $H$ as a subgraph in some graph $G$. Then we want to take a complete graph on $C_d \cdot k$ vertices with edges colored red or blue, and try to find a red or blue copy of $H$ using Lemma 6.16 or 6.18. We'll try to apply Lemma 6.16 to red and Lemma 6.18 to blue. But so far, it's not obvious that one of these two lemmas always applies. To apply Lemma 6.16, 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 6.18 doesn't require reasonable density *within* a set, but rather *between* every pair of subsets.

The next lemma tells us that we *will* always be able to apply one of Lemma 6.16 and 6.18.

**Lemma 6.20**

Let $0 \le \rho \le \frac{1}{2}$ and $0 \le \delta \le \frac{1}{3}$, and let $s \ge \delta^{-1}$ be an integer. Then for every graph $G$, at least one of the following two statements holds:

(a) $G$ has a bi-$(\rho, \delta)$-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 $d(U) \le 3\delta$.

So in the first case (a), we can find a nice bi-$(\rho, \delta)$-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, also of linear size. (We lose a constant factor in the density — from $\delta$ to $3\delta$ — but this doesn't really matter.) It's tautological that you can find a sparse subset or a dense subset (i.e., a subset with density at most $\delta$ or at least $\delta$); the content of Lemma 6.20 is that you can find a sparse subset or a *bi*-dense subset.

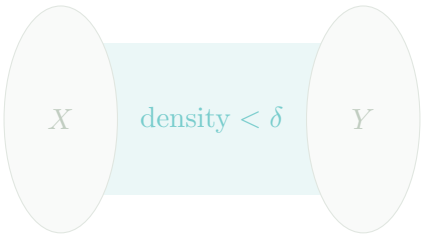
The upshot of Lemma 6.20 is that it tells us that in our coloring of the complete graph, the blue graph falls into one of these two cases. If it falls into (a), then we have a blue bi-$(\rho, \delta)$-dense subgraph, and we can apply Lemma 6.18. If it falls into (b), then we have a subset where the blue edges are very sparse, so the red edges are very dense, and we can apply Lemma 6.16 to the red graph on that subset.

Before we go through the details of this, we'll prove Lemma 6.20.

*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).

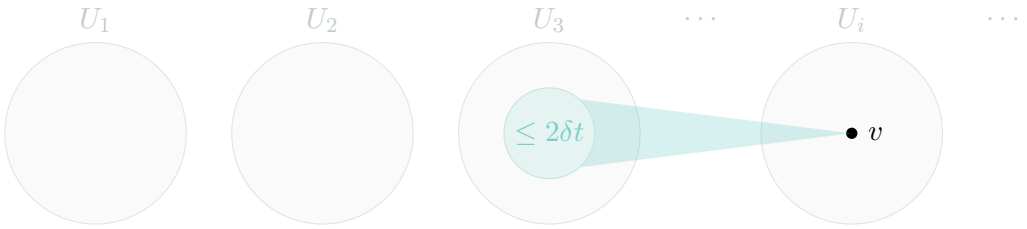
Assume that (a) does not hold; this means every subgraph of $G$ on at least $t$ vertices is *not* bi-$(\rho, \delta)$-dense. Our goal is to now show that (b) holds, i.e., to find a subset of at least $t$ vertices which is very sparse.

The idea for how to construct this subset is roughly as follows. We first look at $G$ as a whole; this graph is not bi-$(\rho, \delta)$-dense, so we can find disjoint sets $X$ and $Y$ of reasonable size with density less than $\delta$.




Our goal is to find a *single* subset in which there is low density. We might try to take $X \cup Y$, but this isn't good enough — this 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 this procedure in $Y$ — we repeatedly apply 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 will be good enough.

In other words, we'll construct disjoint vertex subsets $U_1, \dots, U_s$ of sizes $|U_1| = \cdots = |U_s| = t$ such that $U_1 \cup \cdots \cup U_s$ satisfies (b). The idea is to construct them in such a way that we have low density between any two. In fact, we'll construct them such that for all $1 \le i \le t$, *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 for example, if we take $U_3$ 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 the density between any two subsets is at most $2\delta$ (since we can compute this density by averaging over vertices in the later subset).




We'll construct these sets 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. So the claim we'll induct over is the following.

**Claim 6.21 —** For each $\ell = 0, \dots, s$, we can find disjoint vertex subsets $U_1, \dots, U_\ell, W \subseteq V(G)$ of sizes

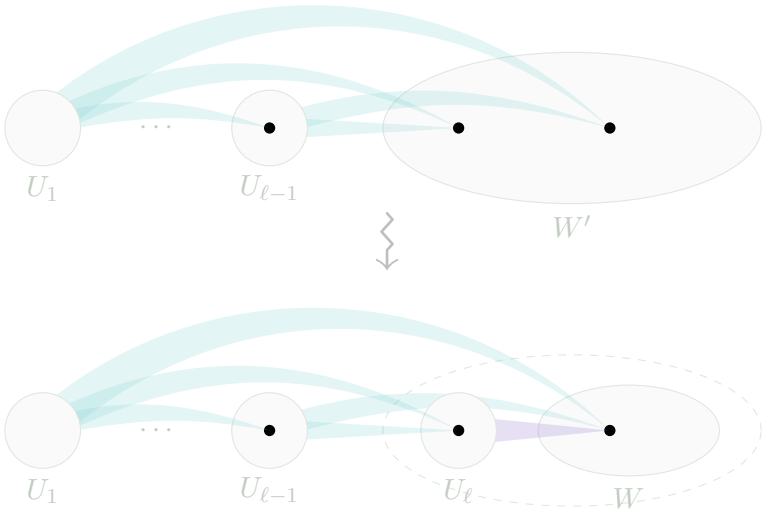
$$|U_1| = \cdots = |U_\ell| = t \quad \text{and} \quad |W| \ge \left(\frac{\rho}{2}\right)^\ell |V(G)|$$

such that for all $1 \le i \le \ell$, every vertex in $U_{i+1} \cup \cdots \cup U_\ell \cup W$ has at most $2\delta t$ edges to $U_i$.

This claim formalizes our goal in a recursive fashion, where $W$ is the set of candidates for the later vertex subsets. (The condition on degrees being at most $2\delta t$ shouldn't just hold for the current sets, but also for the pool of candidates, since this is where we'll draw the future sets from.)

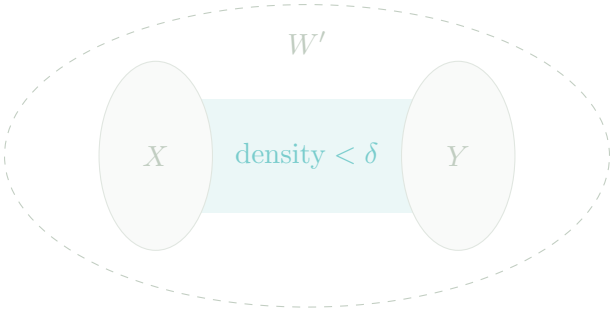
*Proof.* We use induction on $\ell$. The case $\ell = 0$ is vacuously true — we don't need to specify any sets $U_i$, and we can take $W$ to be the set of all vertices (since there are no conditions on it).

Now suppose $\ell \ge 1$ and we've already constructed $U_1, \dots, U_{\ell-1}, W' \subseteq V(G)$. We now want to find disjoint sets $U_\ell, W \subseteq W'$ such that $|U_\ell| = t$, $|W| \ge (\frac{\rho}{2})\,|W'|$, and every vertex in $W$ has at most $2\delta t$ edges to $U_\ell$.



The reason we have so few conditions here is that all others are automatically guaranteed if we take $U_\ell, W \subseteq W'$ — we know that for any set $U_i$ with $i \leq \ell - 1$ and any vertex in either a later set or $W'$, that vertex has at most $2\delta t$ edges to $U_i$. So if we pick $U_\ell$ and $W$ to be subsets of $W'$, then we automatically get all these conditions for free, which means the only pair of sets we need to check the degree condition between is $U_\ell$ and $W$. And 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 $U_\ell$ and $W$ is by using the fact that $W'$ is not bi-$(\rho, \delta)$-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 $t$ was defined as the ceiling of the right-hand side), so the induced subgraph $G[W']$ is not bi-$(\rho, \delta)$-dense (we assumed (a) doesn't hold, so there is no bi-$(\rho, \delta)$-dense subgraph on at least $t$ vertices). This tells us that there are two disjoint subsets $X, Y \subseteq W'$ of sizes $|X| \geq \rho |W'|$ and $|Y| \geq \rho |W'|$ such that $d(X, Y) \leq \delta$.

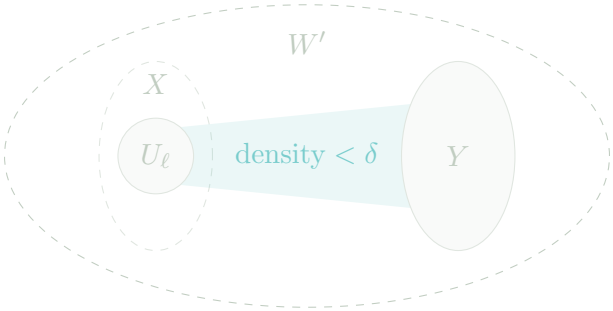



We want to take $U_\ell$ and $W$ to be subsets of $X$ and $Y$. But first, $X$ is way too big — $U_\ell$ should have size *exactly* $t$, while $X$ may be much bigger than this. So we'd like to pass to a subset of $X$ which still satisfies this density condition. And we can do this by an averaging argument.

More precisely, 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 $d(U_\ell, Y) \leq \delta$ (since the average density to $Y$ over all size-$t$ subsets of $X$ is just the density between $X$ and $Y$).




Now we need to choose $W \subseteq Y$. We can't just take $W$ to be $Y$ itself, since we need an upper bound on the number of edges from *every* vertex of $W$ to $U_\ell$. This density condition means that on *average* the vertices of $Y$ have less than $\delta t$ edges to $U_\ell$, but there could 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 $\delta t$ edges to $U_\ell$, so there are at most $\frac{1}{2} |Y|$ vertices $v \in Y$ with at least $2\delta t$ edges to $U_\ell$. We then kick out those vertices, and define $W \subseteq Y \subseteq W'$ as the set of vertices $v \in Y$ with at most $2\delta t$ edges to $U_\ell$. Since we kicked out at most half of $Y$, we have $|W| \geq \frac{1}{2} |Y| \geq (\frac{\rho}{2}) |W'|$ (since $|Y| \geq \rho |W'|$).

So we've found $U_\ell$ and $W$ (which are disjoint, since $X$ and $Y$ were disjoint), proving the claim. □

Now taking $\ell = s$ in Claim 6.21, we've found sets $U_1, \dots, U_s$ satisfying the conditions we wanted (we also have a leftover set of candidates, but we don't care about those anymore — the candidates were just for the purpose of the induction). Explicitly, we have disjoint subsets $U_1, \dots, U_s \subseteq V(G)$, each of size $t$, such that for all $1 \le i < j \le s$, 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 this satisfies (b). First, we have

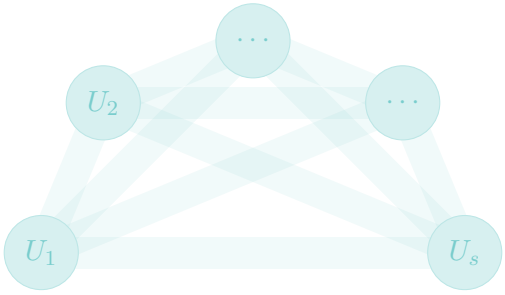
$$|U| = st \ge t \ge \left(\frac{\rho}{2}\right)^s |V(G)|,$$

so $U$ is indeed big enough. Now we need to check it has sufficiently low density: Its density is

$$d(U) = \frac{e(U)}{\binom{|U|}{2}} = \frac{e(U)}{\binom{st}{2}}.$$

To bound $e(U)$, there are two types of edges in $U$ — the edges *inside* the sets $U_1, \dots, U_s$, 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 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 $(U_i, U_j)$, each of which has at most $2\delta t^2$ edges). So we get

$$d(U) \le \frac{s\binom{t}{2} + \binom{s}{2}2\delta t^2}{\binom{st}{2}} = \frac{st(t-1) + s(s-1)2\delta t^2}{st(st-1)} = \frac{t-1}{st-1} + 2\delta \cdot \frac{(s-1)t}{st-1} \le \frac{1}{s} + 2\delta \le 3\delta.$$




Geometrically, what's happening here is that we have $s$ equally sized blobs, and we know that the density between any two blobs is at most $2\delta$; this in total contributes a density of at most $2\delta$. Then the question is how much the density increases if we add in the edges inside the blobs; this contributes at most $\frac{1}{s}$ (since at most a $\frac{1}{s}$-fraction of the possible edges live inside the blobs).

So then $U$ satisfies (b), and we're done. □

### §6.3.4 Conclusion and final remarks

We're now equipped with Lemmas 6.16, 6.18, and 6.20, and our goal is to deduce Theorem 6.13. All the work is basically done; now we just have to put the puzzle pieces together.

*Proof of Theorem 6.13.* The idea is that we want to apply Lemma 6.20 to the blue graph, so that we fall into either case (a) or (b). In (a), we'll get a blue bi-$(\rho, \delta)$-dense subgraph, and we'll apply Lemma 6.18 to that blue subgraph. In (b), we'll get a very dense *red* subgraph and apply Lemma 6.16. Now the question is what $\rho$ and $\delta$ to use. We want to have $3\delta \le \frac{1}{8d}$ (for Lemma 6.16), so we take $\delta = \frac{1}{24d}$. Then we take

$$\rho = \frac{\delta^d}{8d^2} = \frac{1}{8 \cdot 24^d \cdot d^{d+2}}$$

(for Lemma 6.18); this is a constant only depending on $d$. Finally, we take $s = \frac{1}{\delta} = 24d$.

We'll take the constant $C_d$ in the theorem statement to be

$$C_d = \left(\frac{2}{\rho}\right)^{25d} = \left(16 \cdot 24^d \cdot d^{d+2}\right)^{25d}$$

(so the main behavior of $C_d$ is $d^{d^2}$). Consider a complete graph on $C_d \cdot k$ vertices with edges colored red or blue; 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 6.20 is

$$\left(\frac{\rho}{2}\right)^s |V(G)| \geq \frac{2}{\rho} \cdot k \geq 8\delta^{-d} dk$$

(we could even write $(\frac{2}{\rho})^d$ instead of $\frac{2}{\rho}$, but this doesn't matter). Now applying Lemma 6.20 to $G$, we conclude that at least one of the following two cases holds:

(a) The blue graph $G$ has a bi-$(\rho, \delta)$-dense subgraph (for these particular values of $\rho$ and $\delta$) on at least $8\delta^{-d}dk$ vertices. Then we're in the situation where we can apply Lemma 6.18 (we chose these parameters so that the conditions of this lemma hold), so we can find a blue copy of $H$ in this bi-dense subgraph.

(b) There is a vertex subset $U \subseteq V(G)$ of size $|U| \geq 8\delta^{-d}dk \geq 4k$ with blue density $d(U) \leq 3\delta = \frac{1}{8d}$. Then the *red* density inside $U$ is at least $1 - \frac{1}{8d}$ (since the red and blue densities sum to 1). So we're in the situation to apply Lemma 6.16 to the red graph on $U$, which tells us we can find a red copy of $H$.

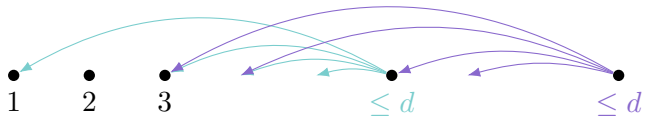
So in either case, we've found a monochromatic copy of $H$, and we're done. □

The value of $C_d$ we get from this proof is roughly $d^{25d^2}$, but this isn't the best-known bound. The original proof gave a much *worse* $C_d$ — it applied the regularity lemma, which gives a tower-type function (which is terrible as a quantitative dependence). The dependence from this proof is much better, but in fact better dependencies are known. A slightly more careful of this proof already improves the dependence a bit; and it was later improved by Conlon, Fox, and Sudakov to $C_d = \exp(cd \log d)$ (this removes one factor of $d$ from the exponent — we have $d^{cd}$ rather than $d^{cd^2}$). As of today, this is the best-known bound.

Theorem 6.13 can also be strengthened *qualitatively*. This theorem is amazing because it gives a *linear* dependence in the number of vertices $k$. The qualitative strengthening, which was proven relatively recently, still gets a linear dependence, but weakens the maximum degree assumption on $H$. Of course we need *some* assumption (since the behavior when $H$ is a clique is very different). But rather than requiring *maximum* degree $d$, this theorem widens the class of graphs $H$ to *$d$-degenerate graphs*. (This was proven by Lee, and also conjectured by Erdős and Burr.)

**Definition 6.22.** 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, such that for every vertex we add, we can draw at most $d$ edges to the previous vertices.

In other words, a graph is $d$-degenerate if its vertices can be labelled so that every vertex has degree at most $d$ to the preceding vertices.



**Theorem 6.23** (Lee 2017)

We have $R(H, H) \leq C_d \cdot k$ for every $d$-degenerate graph $H$ on $k$ vertices (where $C_d$ is a constant only depending on $d$).

In particular, every graph of maximum degree $d$ is degenerate (we can add the vertices in *any* order). So this is a strengthening of Theorem 6.13.

If we tried adapting our proof to this setting, Lemma 6.16 still works — in the proof of Lemma 6.16 we got to choose which order to embed vertices in, so here we can use the given ordering. However, for Lemma 6.18, our proof breaks down because we had to use the maximum degree condition going in *both* directions. Specifically, every time we chose a new vertex, we had to make sure that the candidate sets for the future vertices don't shrink too much; and there we used the fact that the current vertex has at most $d$ neighbors among the *not* yet embedded vertices (to say we had at most $d$ candidate sets to worry about).

**Remark 6.24.** Here's an unrelated mathematical announcement (*this is from the end of the March 23 lecture*): As we've 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 of roughly

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

where the upper bound (Theorem 1.13) came from a recursive bound giving a binomial coefficient of roughly $\binom{2k}{k}$, and the lower bound (Theorem 1.20) came from a probabilistic argument.

Three weeks ago, Prof. Sauermann told us that the best improvement on the lower bound is what we've seen on the homework (a probabilistic argument using the Lovász local lemma, which essentially gets a constant factor improvement); meanwhile, for the upper bound, there had been improvements on the lower order terms, but 4 was still the best-known exponent.

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

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

where $\varepsilon \approx \frac{1}{128}$ (the important thing is that $\varepsilon$ is a positive constant). There has been a paper by Campos, Griffiths, Morris, and Sahasrabudhe making this gigantic breakthrough, improving the base of the exponent from 4 to $4 - \varepsilon$. This is a gigantic breakthrough because the base of 4 came out of the simple inductive argument we saw in the first class; and there's been a lot of work on trying to improve it, but no one managed to push it beyond 4 until exactly a week ago.

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

April 13, 2023

## §7 Multi-color Ramsey numbers

In the last three classes, we've looked at Ramsey numbers of graphs, where we're looking for monochromatic graphs that aren't necessarily cliques; and we saw some surprising results in the bounded-degree setting. Today we'll go back to the classical Ramsey setting, where we're looking for cliques. We've already talked quite intensively about the known bounds for Ramsey numbers, especially diagonal ones, in the most classical 2-color setting. Today we'll talk about bounds for diagonal Ramsey numbers with *many* colors (where 'many' means more than 2).

**Definition 7.1.** 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, ..., $t$, there is a monochromatic clique of size $k_i$ in color $i$ for some $i \in \{1, \dots, k\}$.

So as usual, we have $t$ colors, and $k_1$, ..., $k_t$ 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'll start by recalling the bounds we already know. We already proved that $R(k_1, \dots, k_t)$ exists, and our proof (the simple proof of existence from the first week) gave a recursive upper bound of

$$R(k_1, \dots, k_t) \leq \sum_{j=1}^{t} R(k_1, \dots, k_{j-1}, k_j - 1, k_{j+1}, \dots, k_t)$$

(where we sum over all the different colors, and in each term, we reduce one target clique size by 1). (We may not have written this recursive formula down directly, since we did the multi-color proof in the $r$-uniform hypergraph setting. But we could get this recursion by either specializing that proof (of Theorem 1.26) to the case $r = 2$, or generalizing the 2-color proof (of Theorem 1.4) in the obvious way.)

**Remark 7.2.** In fact, the bound we get is a tiny bit better — we can say that

$$R(k_1, \dots, k_t) \leq \sum_{j=1}^{t} (R(k_1, \dots, k_j - 1, \dots, k_t) - 1) + 1$$

(where we subtract 1 from each term, and add an extra 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 defined as the number of ways to take the numbers 1, ..., $k_1 + \cdots + k_t - t$ and color $k_1 - 1$ of them with color 1, $k_2 - 1$ with color 2, and so on.)

We care most about the setting when all the target clique sizes are the same. For that case, this gives the relatively easy upper bound

$$R(k, \dots, k) \leq \binom{tk - t}{k - 1, \dots, k - 1} \leq t^{kt} \tag{7.1}$$

(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, and 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.)

## §7.1 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 (a 2-clique is just an edge). But even when $k = 3$, the answer is not understood.

**Conjecture 7.3** (Erdős, $100) — There is an absolute constant $C > 0$ such that $R(3, \dots, 3) \leq C^t$.

This is smaller than the above bound (7.1), which is exponential in $t \log t$. The best-known bounds are

$$C_1^t \leq R(3, \dots, 3) \leq C_2 \cdot t!$$

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 $t$ (Conjecture 7.3), but this isn't known. (Note that $t!$ is roughly like $t^t$, so in some sense it's not so much better than (7.1).)

## §7.2 The fixed-$t$ setting

Now we'll look at the setting where the number of colors $t$ is fixed, and $k$ is large. This setting 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 about the known bounds for $R(k, k)$).

For a fixed number of colors $t \geq 2$ and large $k$, the upper bound of

$$R(k, \dots, k) \leq t^{kt}$$

from (7.1) is exponential in $k$. We also have a lower bound of

$$R(k, \dots, k) \geq R(k, k) \geq \sqrt{2}^k$$

(from Theorem 1.20). So we know this quantity is exponential in $k$. But the next question is:

> **Question 7.4.** What is the correct base of the exponent?

It's in some sense unsurprising that this is where we are, since it's a similar state to what we know for $k = 2$. The best-known lower bound is essentially $2^{k/2}$, and at the time 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. The multi-color problem was also stuck for a long time without any movement, but in the last few years, more progress has been made.

### §7.2.1 A product construction

> **Question 7.5.** What's the best easy lower bound we can get for fixed $t$?

We can easily get a lower bound of $2^{k/2}$ from the 2-color 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 it gives a lower bound of roughly $t^{k/2}$ (this makes sense because it matches what we got when $t = 2$).

This is nice, but the base grows rather slowly with $t$ — if we double the number of colors, the base is only multiplied by a factor of $\sqrt{2}$. But having more colors should in some sense give us *way* more power — for example, in our proof of the stepping-up lemma for 3-uniform hypergraphs (Theorem 5.16), having 4 colors was much more powerful than just having 2.

And in fact, $t^{k/2}$ is *not* the best we can do. For $t \geq 4$, we can get a better bound via a *product construction* (which takes the construction from $t = 2$, but makes use of the additional colors in a better way).

**Lemma 7.6**

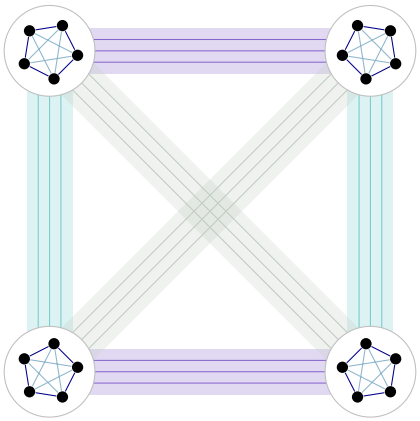
For any $t_1$ and $t_2$, we have

$$R(\underbrace{k, k, \dots, k}_{t_1+t_2 \text{ times}}) - 1 \geq (R(\underbrace{k, \dots, k}_{t_1 \text{ times}}) - 1)(R(\underbrace{k, \dots, k}_{t_2 \text{ times}}) - 1).$$

*Proof.* We want to show that given a construction with $t_1$ colors and no monochromatic $k$-clique, and another construction with $t_2$ colors, we can obtain a construction with $t_1 + t_2$ colors whose number of vertices is the product of the starting constructions' numbers of vertices.

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 (with $t_1$ colors) and blow up each vertex (turning it into a blob). Then inside each blob, we place a copy of the second construction (using $t_2$ new colors). For example, 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 — i.e., the size of the second construction), and we place a copy of the second construction in each blob.



Suppose that there is a monochromatic clique of size $k$ in some color; then it has to be in either one of the first $t_1$ colors, or one of the last $t_2$. If this clique is in one of the first $t_1$ colors, then it can 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 forms a monochromatic $k$-clique in the original graph (i.e., the first construction before we blew it up), which we assumed doesn't exist.

Similarly, if there is a clique in one of the last $t_2$ colors, then it has to live entirely in a blob (since no edges between blobs have that color). But our blobs don't have any $k$-cliques by construction, so this is again a contradiction. □

Together with $R(k, k) \geq 2^{k/2}$, iterating Lemma 7.6 now gives

$$R(\underbrace{k, \dots, k}_{t \text{ times}}) \geq 2^{kt/4}$$

for any even number of colors $t$ — we can split the $t$ colors into pairs and iterate this lemma; then we get a product of $\frac{t}{2}$ factors of $R(k, k) - 1$. This is significantly better than our previous bound of $t^{k/2}$. When $t$ is odd, we have to be a bit more careful. If we pair colors, then we get one color left; we could just waste it, but that's a bit wasteful. We can do a bit better by pairing all but three of the colors, and then grouping those three.

But in fact, pairs aren't the most efficient — we get something better if we use groups of 3 instead. So combining Lemma 7.6 with the bound $R(k, k, k) \geq 3^{k/2}$, we get

$$R(\underbrace{k, \dots, k}_{t \text{ times}}) \geq 3^{kt/6}$$

when $t$ is divisible by 3. (We can check that $3^{1/6}$ is a bit bigger than $2^{1/4}$.) Then if $t$ is not divisible by 3, we can combine the leftover colors in pairs (if $t \equiv 2 \pmod 3$ then we'd have two leftover colors, so we pair them; if $t \equiv 1 \pmod 3$ then we instead form *two* pairs).

### §7.2.2 A better lower bound

The above bounds were known for several decades, and this is where they had been stuck for a long time (similarly to the two-color case, where we had an exponential upper bound and an exponential lower bound with different bases, and both were stuck). But then there was movement a few years ago — there was a pretty surprising paper that improved this lower bound. The improvement doesn't look drastic, but it's significant because that 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 7.7** (Conlon–Ferber, Wigderson 2020)
For any fixed $t \geq 2$, we have

$$R(\underbrace{k, \dots, k}_{t \text{ times}}) \geq 2^{(\frac{3}{8}t - \frac{1}{4} - o(1))k}.$$

Again this bound is exponential in $k$, but the exponent (which depends on $t$) is now $2^{\frac{3}{8}t - \frac{1}{4}}$. This is better than the previous bound for all $t > 2$ — for $t = 2$ it's the same as the $2^{k/2}$ bound, but for all greater $t$ it's an improvement.

**Remark 7.8.** This theorem is a bit hard to attribute. First, a paper by Conlon and Ferber proved this 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 (Lemma 7.6). This gives an improvement from the previously known bound, but not one as strong as the one we stated above (in Theorem 7.7). Then the bound in Theorem 7.7 was established by Wigderson, who took the Conlon–Ferber construction but found a better way to generalize it to more 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 0.383796 we had $\frac{3}{8} = 0.375$, then we'd get the same expression as in Theorem 7.7; 0.383796 is a bit bigger than $\frac{3}{8}$, so this is a slight improvement. 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'll prove the bound in Theorem 7.7 (with $\frac{3}{8}$), but using Sawin's approach. 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, because it requires carrying around some ugly terms in the computations).

The proof of Theorem 7.7 has two parts (half of the argument is basically the same as the part 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 7.9.** For a graph $G$ and integer $k \geq 2$, we define

$$q_G(k) = \mathbb{P}[\{v_1, \dots, v_k\} \text{ is an independent set in } G],$$

where we take $v_1, \dots, v_k$ to be independent uniformly random vertices.

In other words, we choose $k$ independent random vertices in $G$ (which do not have to be distinct), and define $q_G(k)$ as the probability that these vertices form an independent set.

The idea of the following lemma is due to Wigderson, though it was phrased in this way by Sawin.

**Lemma 7.10** (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 it comes up naturally in the proof.

*Proof.* As usual, we'll let

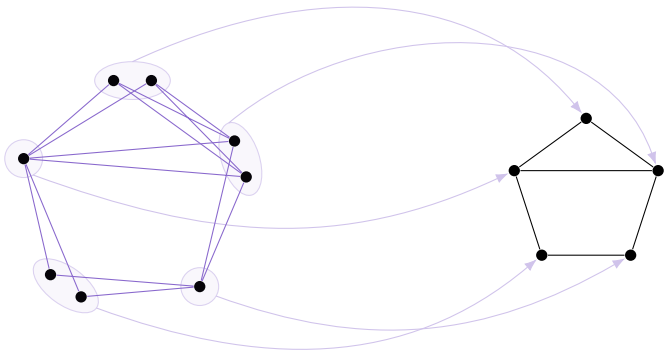
$$n = \left\lfloor \left(\frac{1}{q_G(k)}\right)^{(t-2)/k} \cdot 2^{(k-1)/2} \right\rfloor.$$

Our goal is to find a coloring of the edges of a complete graph on the vertex set $\{1, \dots, n\}$ with $t$ colors and no monochromatic clique of size $k$ (this will mean the corresponding Ramsey number is greater than $n$).

As you can imagine (since we're trying to prove a bound involving $q_G(k)$, which is defined by a random experiment), we should take a random coloring, but this random coloring should have something to do with $G$. The way we want to do this is 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 those of $G$. This is where randomness comes in — we'll use *random* functions for this.

We take $t-2$ independent and uniformly random functions $f_1, \dots, f_{t-2} : \{1, \dots, n\} \to V(G)$. (We'll see the reason for taking $t-2$ functions soon.) So 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, for any 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 cliques of size $k$.

We define color 2 the same way for $f_2$ (if this causes an edge to get both color 1 and color 2, we choose which one to use arbitrarily). We then define color 3 the same way for $f_3$, 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 with the first $t-1$ colors and give all the remaining edges the last color $t$. But it turns out that it's better to only do it with $t-2$ colors, and then for all the remaining edges, we color them randomly with one of the last two colors. (You could also do this with $t-3$ colors and color the remaining edges randomly with the last three colors, or so on; there isn't a clear reason to see why 2 is the best, but if you try the computations, you'll see that it is.)

More explicitly, for any $x, y \in \{1, \dots, n\}$ 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 $xy$ with color $i$. (If there are several such $i$, then we pick one arbitrarily, e.g., we can choose the smallest such $i$ or choose one randomly.)
- Otherwise, we color $xy$ randomly with either color $t-1$ or color $t$ (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 $f_i$, and the random choices we're making here for the edges not yet colored.

We want to show that with positive probability, there is no monochromatic clique of size $k$ — equivalently, we want to show that the probability this coloring has a monochromatic clique of size $k$ is strictly less than 1. We've already seen that 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 doesn't 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 in one of these colors. For any given subset of $\{1, \dots, n\}$ of size $k$, we can find the probability it forms a monochromatic clique of color $t-1$ or color $t$ as follows. First, none of the edges between these $k$ vertices can be assigned one of the first $t-2$ colors, so each of the 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 the colors are all independent, this has probability $q_G(k)^{t-2}$. Then we also need to make the same choice between the colors $t-1$ and $t$ for all $\binom{k}{2}$ edges. This gives

$$\mathbb{P}[k \text{ given vertices form a monochromatic clique}] = q_G(k)^{t-2} \cdot 2 \cdot 2^{-k(k-1)/2}.$$

Now we can union-bound over all possible subsets, so the total probability of having a monochromatic clique of size $k$ in color $t-1$ or $t$ is

$$\mathbb{P}[\text{exists a monochromatic clique}] \leq \binom{n}{k} \cdot q_G(k)^{t-2} \cdot 2 \cdot 2^{-k(k-1)/2}.$$

Now we can do some computations. First, we have $\binom{n}{k} < \frac{n^k}{k!}$. Since $k \geq 2$, we have $k! \geq 2$, so we can use the $k!$ to cancel out the 2, and we get

$$\mathbb{P}[\text{exists a monochromatic clique}] < n^k \cdot q_G(k)^{t-2} \cdot 2^{-k(k-1)/2}.$$

When we plug in our definition of $n$, both terms perfectly cancel out, and the right-hand side becomes 1. So the 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 7.11.** 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 $v_1, \dots, v_k$ to be distinct. But we're going to take $n$ to be much bigger than $|V(G)|$, so we can't try to force our maps $f_1, \dots, f_{t-2}$ to be injective; this is why we allow repetition in the definition of $q_G(k)$.

Now we'll see the second lemma. Our goal in the end is to get a lower bound for $R(k, \dots, k)$, and so far we have a lower bound in terms of $q_G(k)$. So we'd like to construct a graph $G$ where $q_G(k)^{-(t-2)/k}$ is pretty big, meaning that $q_G(k)$ is pretty small.

**Lemma 7.12** (Sawin 2021, Conlon–Ferber 2020)
For every $k \geq 2$, there is a graph $G$ not containing a clique of size $k$ such that

$$q_G(k) \leq 2^{-(\frac{3}{8}-o(1))k^2}.$$

**Remark 7.13.** A graph with this property was first established by Conlon–Ferber (in their proof of Theorem 7.7) with an explicit construction. But the lemma was stated in this form by Sawin, who replaced $\frac{3}{8}$ with the slightly better constant 0.383796. We'll go through Sawin's proof but only get $\frac{3}{8}$, and we'll briefly comment on what to adapt to get this larger constant.

We don't have enough time to prove Lemma 7.12 right now, so we'll first prove Theorem 7.7 from these two lemmas.

*Proof of Theorem 7.7.* Choosing $G$ as in Lemma 7.12, by Lemma 7.10 we get that

$$R(\underbrace{k,\dots,k}_{t \text{ times}}) \geq \left(\frac{1}{q_G(k)}\right)^{(t-2)/k} \cdot 2^{(k-1)/2} \geq 2^{(\frac{3}{8}-o(1))(t-2)k} \cdot 2^{(k-1)/2}.$$

Since we're thinking of $t$ as fixed and $k$ as large, our $o(1)$ quantities are allowed to depend on $t$ (it's just important that for fixed $t$, it goes to 0 as $k \to \infty$); so we can pull the factor of $t-2$ inside the $o(1)$. We can also absorb the $-\frac{1}{2}$ in the second factor into the $o(1) \cdot k$, giving a bound of

$$R(\underbrace{k,\dots,k}_{t \text{ times}}) \geq 2^{(\frac{3}{8}(t-2)+\frac{1}{2}-o(1))k} = 2^{(\frac{3}{8}t-\frac{1}{4}-o(1))k}.$$ □

**Remark 7.14.** In this calculation, we can see that if we improve Lemma 7.12 by replacing $\frac{3}{8}$ with 0.383796, then the same calculation will give Sawin's stronger bound (stated in Remark 7.8).



Finally, we'll prove Lemma 7.12. We can now forget about the context from multicolor Ramsey numbers — our new task is just to find a graph $G$ with no clique of size $k$, and which satisfies this upper bound on the probability of a $k$-vertex random sample forming an independent set. First 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 shouldn't be too many edges in $G$, while upper-bounding the probability of getting 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 that we can find a graph $G$ satisfying both at the same time.

*Proof.* We'll construct the graph $G$ randomly — for each edge, we'll independently decide whether or not to include it. To get $\frac{3}{8}$, we can include each edge with probability $\frac{1}{2}$ (to get the better constant, you need to adjust this probability).

Let $m = \lceil 2^{(k-1)/2} \rceil$; this will be our number of vertices. Consider a random graph $G$ on $m$ vertices, where every edge is present with probability $\frac{1}{2}$ (independently for all edges). We're interested in getting an outcome which doesn't have a clique of size $k$ and satisfies the desired upper bound on $q_G(k)$ — so we need to deal with two restrictions at the same time. The way we'll think about this is that we only have a certain number of 'allowed' outcomes — those without a $k$-clique — and from these, we want to find one for which $q_G(k)$ is small. So it's natural to consider the *expectation* of $q_G(k)$, but only over these allowed $G$ (i.e., 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 that this conditional expectation is at most $2^{-(3/8-o(1))k^2}$; then we can pick some allowed graph $G$ with small $q_G(k)$.

First, 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 denominators, so that this statement then says

$$\mathbb{E}[q_G(k)] \ge \mathbb{E}[q_G(k) \mid G \text{ has no clique of size } k] \cdot \mathbb{P}[G \text{ has no clique of size } k].$$

And the right-hand side is the contribution to $\mathbb{E}[q_G(k)]$ coming from graphs $G$ with no clique of size $k$, while the contribution coming from graphs with a clique of size $k$ is nonnegative.

So now it suffices to show that

$$\frac{\mathbb{E}[q_G(k)]}{\mathbb{P}[G \text{ has no clique of size } k]} \le 2^{-(3/8-o(1))k^2}. \tag{7.2}$$

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 that $G$ *does* have a clique of size $k$. We'll do this by a union bound — there are $\binom{m}{k}$ possible $k$-cliques, and for each, the probability it really forms a clique is $2^{-\binom{k}{2}}$, so

$$\mathbb{P}[G \text{ has a } k\text{-clique}] \le \binom{m}{k} \cdot 2^{-\binom{k}{2}} \le \frac{m^k}{k!} \cdot 2^{-k(k-1)/2} \le \frac{(2^{(k-1)/2}+1)^k}{k!} \cdot 2^{-k(k-1)/2}.$$

The main terms cancel out (the $+1$ doesn't contribute much — the $k!$ in the denominator more than makes up for it), so we get

$$\mathbb{P}[G \text{ has a } k\text{-clique}] \le \frac{1}{2}$$

if $k$ is large enough. (We can assume $k$ is large because we have an $o(1)$ in the desired bound.)

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 of (7.2) is at least $\frac{1}{2}$, so it contributes a factor of at most 2. This is eaten by the $o(1)$, so all we have to show is that

$$\mathbb{E}[q_G(k)] \le 2^{-(3/8-o(1))k^2}. \tag{7.3}$$

Recall that $q_G(k)$ is defined by taking a random experiment where we choose $k$ independent random vertices of $G$. Then we have an outer expectation over the randomly chosen $G$, and $q_G(k)$ is also a probability (or expectation) over the choice of $k$ random vertices. So we can put these two sources of randomness together to make this into one single probability — letting $v_1, \dots, v_k$ be independent uniform random vertices of $G$, we have

$$\mathbb{E}[q_G(k)] = \mathbb{P}_{G,v_1,\dots,v_k}[\{v_1, \dots, v_k\} \text{ is an independent set in } G].$$

Now we can first look at the outcomes of $v_1, \dots, v_k$. Suppose that we've chosen $\ell$ distinct vertices (e.g., $\ell = k$ if they're all distinct, and $\ell = 1$ if they're all the same). Then the probability they form an independent set (meaning that we don't draw any edges between these $\ell$ vertices) is $2^{-\binom{\ell}{2}}$. So summing over all possible values of $\ell$, we get

$$\mathbb{E}[q_G(k)] = \sum_{\ell=1}^{k} \mathbb{P}[\#\{v_1, \dots, v_k\} = \ell] \cdot 2^{-\binom{\ell}{2}}.$$

Now we'll upper-bound the right-hand side.

**Remark 7.15.** Note that heuristically, we can't assume that all the vertices are distinct — we're choosing $k$ vertices from a pool of size roughly $2^{k/2}$. So the probability for each given pair of the vertices to collide is roughly $2^{-k/2}$, and there are roughly $k^2$ pairs, so the probability that some two vertices collide is roughly $k^2 \cdot 2^{-k/2}$. So it's true that this happens *rarely*, but it's not good enough for the bound — this is much bigger than our target probability of $2^{-(3/8-o(1))k^2}$.

In order to bound this probability, we can look at all possible ways to choose $\ell$ vertices from our ground set of $m$ vertices, and then find the probability that each of $v_1, \dots, v_k$ is among these $\ell$ vertices; this gives a bound of

$$\mathbb{E}[q_G(k)] \le \sum_{\ell=1}^{k} \binom{m}{\ell} \cdot \left(\frac{\ell}{m}\right)^k \cdot 2^{-\binom{\ell}{2}} \le \sum_{\ell=1}^{k} m^\ell \cdot \frac{\ell^k}{m^k} \cdot 2^{-\ell(\ell-1)/2}.$$

(There's an explicit expression which counts the probability that $\{v_1, \dots, v_k\}$ is a specific set of $\ell$ vertices, called a *Stirling number*; you could use the asymptotics on that number to get a tighter bound, but we won't need that.)

Here we have terms of $m^\ell$ and $m^k$, which are nice. The term of $\ell^k$ is a bit annoying, but we can bound $\ell^k \le k^k$ and pull it out (a $k^k$ factor might sound gigantic, but it actually gets absorbed into the $o(1)$ in (7.3), 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

$$\mathbb{E}[q_G(k)] \le k^k \cdot \sum_{\ell.=1}^{k} m^{\ell-k} \cdot 2^{-\ell(\ell-1)/2}.$$

Plugging in our definition of $m$, we get an upper bound of

$$\mathbb{E}[q_G(k)] \le k^k \cdot \sum_{\ell=1}^{k} 2^{(k-1)(\ell-k)/2} \cdot 2^{-\ell(\ell-1)/2}.$$

Now we can break up the first exponent into

$$\frac{(k-1)(\ell-k)}{2} = \frac{\ell(k-1)}{2} - \frac{k(k-1)}{2}.$$

The latter term doesn't depend on $\ell$, so we can pull it out of the sum; then combining the exponents of the remaining terms gives

$$\mathbb{E}[q_G(k)] \le k^k \cdot 2^{-k(k-1)/2} \cdot \sum_{\ell=1}^{k} 2^{\ell(k-\ell)/2}.$$

Now we can bound this summation by $k$ times the maximum ter; we always have $\frac{\ell(k-\ell)}{2} \le \frac{k^2}{8}$, so we get

$$\mathbb{E}[q_G(k)] \le k^k \cdot 2^{-k(k-1)/2} \cdot k \cdot 2^{k^2/8} = 2^{-(3/8-o(1))k^2}$$

(the only terms which matter are ones with $k^2$ in the exponent, and $-\frac{1}{2} + \frac{1}{8} = -\frac{3}{8}$). □

**Remark 7.16.** Sawin proved a version of Lemma 7.12 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 $0.454997\dots$, obtained by leaving $p$ as a variable and then optimizing for $p$ in the end) and take $m = p^{-(k-1)/2}$, then you get this. The ideas are the same, but we'll have $p$ and $1-p$ instead of the 2's, so the calculations become more complicated.

**Remark 7.17.** Conlon–Ferber found an *explicit* graph with the property in Lemma 7.12. We'll describe their graph, but not prove that it has this 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 $\mathbb{F}_2^k$), where $vw$ is an edge 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 on algebraic methods last year), and that the graph has this property about $q_G(k)$ (Conlon–Ferber don't write it down this way, but the idea is the same). This gives an explicit construction for $\frac{3}{8}$, but for the improved bound, you probably have to use a probabilistic contradiction.

**Remark 7.18.** Sawin also showed 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 0.383796, 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 Turán's theorem). However, graphs which are this dense but don't have a clique of size $k$ have pretty strong structure — they have huge independent sets, which causes $q_G(k)$ to be large. So it's not a contradiction to have a graph with pretty high density but still large $q_G(k)$.

April 18, 2023

# §8 Induced Ramsey theorem for graphs

Today we'll talk about a new topic. We've known Ramsey's theorem for graphs since the first class; here we'll write it in a slightly different but equivalent form.

**Theorem 8.1**

For every graph $H$, there is a graph $G$ such that the following holds: For every coloring of the edges of $G$ with red and blue, there is a subgraph of $G$ that is isomorphic to $H$ and such that all of its edges have the same color.

Here $H$ is our target graph. This isn't exactly how we phrased Ramsey's theorem — we said that there's some number such that if $G$ is a *complete* graph with that number of vertices, then there is a monochromatic copy of $H$. This is equivalent because we don't gain anything by allowing $G$ to be a non-complete graph, so we might as well assume that $G$ is a complete graph (adding additional edges can't hurt us).

However, as the title suggests, today we'll talk about what happens if we're looking at *induced* subgraphs. Imagine that we replace 'subgraph' in Theorem 8.1 with 'induced subgraph.' Then it's important to allow ourselves the added flexibility of taking $G$ to be a non-complete graph (if $H$ is not a clique, then this won't be true if $G$ is a clique). It turns out that this is still true, but much harder — the statement that $R(k, k)$ exists took 15 minutes to prove, but the proof we'll see today for the induced version is more complicated. (Without 'induced,' we may as well assume that $H$ is a clique and $G$ is a clique, and this is equivalent to Ramsey's theorem (Theorem 1.4). In the induced setting, if $H$ is a clique, this is still equivalent. But if $H$ is not a clique, then this becomes harder, and it was proven several decades later.)

**Theorem 8.2** (Deuber, Erdős–Hajnal–Pesa, Rödl 1973)

For every graph $H$, there exists a graph $G$ such that the following holds: For every coloring 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 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 of $G$ with red and blue, we can still find an induced subgraph in one color.

The proof for the non-induced version (equivalently, the proof of Theorem 1.4, that $R(k,k)$ exists) was by some sort of inductive argument. For this, we needed the extra flexibility of having different target clique sizes in red and blue — the recursion also involved off-diagonal Ramsey numbers $R(k,\ell)$. We'll do something similar here — we'll prove a version of Theorem 8.2 where we're looking for *different* induced subgraphs in red and blue.

**Theorem 8.3**

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 $H_b$ is the graph we're looking for in blue (which is the reason for the names). It's pretty clear that this implies Theorem 8.2 (since we can take $H_r$ and $H_b$ to both be $H$). The converse is also true (i.e., the diagonal case implies the case with distinct graphs), but we won't need that.

## §8.1 The proof strategy

We'll prove Theorem 8.3 by induction (which is why it's important to work with the non-diagonal version) — we'll induct on $|V(H_r)| + |V(H_b)|$.

First we'll deal with some trivial cases. The theorem is trivial if $|V(H_r)| = 1$ or $|V(H_b)| = 1$ (we can just take $G$ to be a single vertex), so we'll now assume that both are at least 2.

The case where $H_r$ has an isolated vertex $v$ is also easy — we can apply the theorem for $H_r \setminus \{v\}$ and $H_b$ (using the induction hypothesis, since these graphs have a smaller total number of vertices) to obtain some 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; 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 might not be strictly necessary to separate, but they'll cause some sets to be empty, so it's cleanest 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 one vertex $v_r$ in $H_r$ and one vertex $v_b$ in $H_b$, and let

$$H_r' = H_r \setminus \{v_r\} \quad \text{and} \quad H_b' = H_b' \setminus \{v_b\}.$$

We also want to keep track of the neighborhoods of $v_r$ and $v_b$. (In the non-induced version, we obtained a clique of size $k$ by inductively finding a clique of size $k-1$ and adding an additional vertex connected to everything; here we'll again need to add one more vertex 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 $v_r$); similarly, define $N(v_b) \subseteq V(H_b')$ as the neighborhood of $v_b$ in $H_b$.

We now win if we 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 exactly the right connecting edges to this copy of $H_r'$, all of which are red as well (or the same with $H_b'$ and blue).

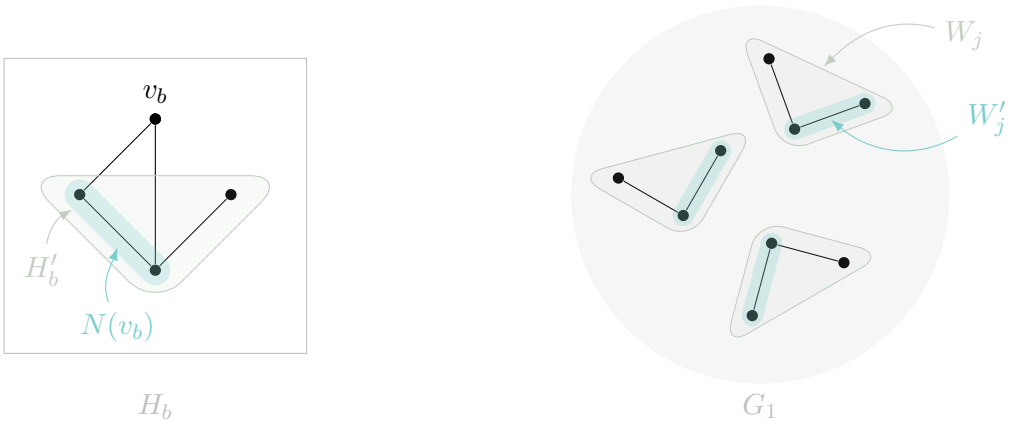
We want to use the induction hypothesis on $(H_r', H_b)$ and $(H_r, H_b')$. (We don't want to use it on $(H_r', H_b')$, since this would weaken the statement *twice*.) By the induction hypothesis, there exists a graph $G(H_r, H_b')$ satisfying the theorem for $H_r$ and $H_b'$, which we'll call $G_1$; 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 $G_1$ — this has the property that when we color it red and blue, either we find an induced copy of $H_r$ with all edges red (in which case we're done), or we find an induced copy of $H_b$ with all edges blue (which is not so good, since we're looking for $H_b$ and not $H_b'$).

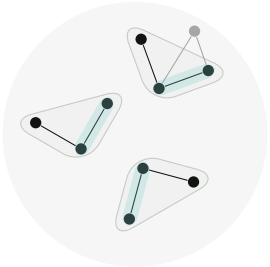
It's likely that $G_1$ actually has many different copies of $H_b'$, and the blue copy of $H_b'$ that we found might be any one of these copies. The strategy of the proof is to modify $G_1$ step by step (making it more complicated at each step) so that with every iteration, we exclude one more induced subgraph from being this blue copy of $H_b'$ (unless we find one of the two objects we're looking for).

## §8.2 Excluding possible blue copies

Consider our graph $G_1$. We want to look at all induced subgraphs of $G_1$ which are isomorphic to $H_b'$ (since these are the possible copies of $H_b'$). So let $W_1, \ldots, W_n$ 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 $j \in \{1, \ldots, n\}$, choose a particular isomorphism from $H_b'$ to $G_1[W_j]$, and let $W_j' \subseteq W_j$ be the image of $N(v_b) \subseteq V(H_b')$.




In other words, $W_j'$ records which vertices in $W_j$ correspond to the neighbors of our chosen vertex $v_b$ — we want to record this because we'd like to extend $W_j$ to a copy of $H_b$, and to do this, we need to attach a new vertex to $W_j'$.



The way we'll prove the theorem is by a $n$-step recursion through all these copies $W_j$. In order to avoid having too many indices, we'll summarize the properties we want to maintain through 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 give what we want (i.e., a graph satisfying the theorem for $H_r$ and $H_b$).

**Claim 8.4 —** For each $j \in \{1, \dots, n+1\}$, 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:

(1) For distinct vertices $v, v' \in V(G)$, for all $u \in U(v)$ and $u' \in U(v')$, we have that $uu'$ is an edge of $G_j$ if and only if $vv'$ is an edge of $G_1$.

(2) For any coloring of the edges of $G_j$ with red and blue, at least one of the following holds:

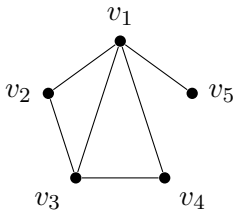
  (a) There is an induced subgraph of $G_j$ isomorphic to $H_r$ with all edges red.

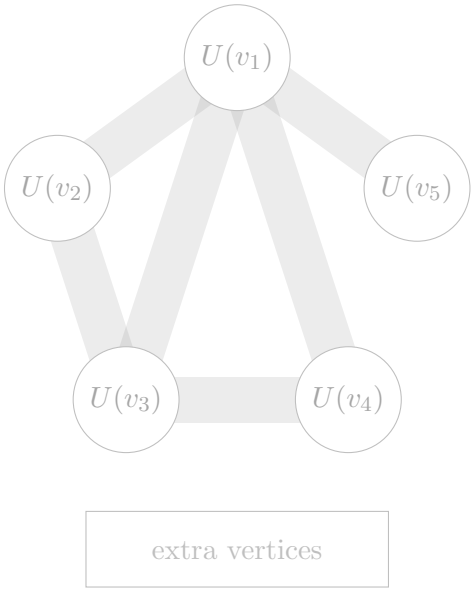
  (b) There is an induced subgraph of $G_j$ isomorphic to $H_b$ with all edges blue.

  (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 *starting* graph $G_1$, and for each of them, we should have a distinguished subset of $G_j$ corresponding to that vertex. We'll obtain this by starting with $G_1$ and then performing some operations where we blow up certain vertices in certain ways; and the set $U(v)$ will be everything that resulted from blowing up $v$.

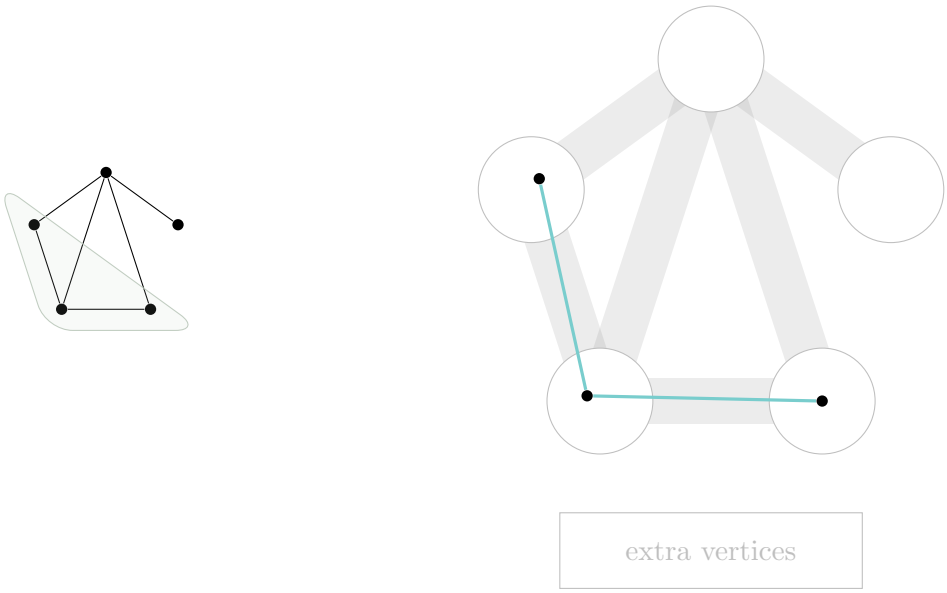
The first condition (1) says that if we look at two different vertices $v$ and $v'$ in $V(G_1)$ and their associated vertex sets $U(v)$ and $U(v')$, the edge relationships between these vertex sets are the exact same as the edge relationships between $v$ and $v'$ — so we have a sort of blow-up structure. In particular, between different sets $U(v)$ and $U(v')$, we either have all edges or no edges (corresponding to whether $vv'$ was an edge or not in $G_1$); we haven't yet specified what happens inside the sets $U(v)$. (There may also be additional vertices that aren't part of any of these sets $U(v)$.)




In the second condition (2), the first two options are the two things we're looking for, and the third is an option depending on $j$ which will be trivially satisfied for $j = 1$, and gets more restrictive as $j$ increases. It says that we can find an induced subgraph formed by choosing exactly one vertex from each of the subsets corresponding to $W_k$, such that all edges of this induced subgraph are blue. Any such subgraph has to be isomorphic to $H_b'$ (because we have the same edge relationships between the sets $U(v)$ as between their

corresponding vertices in $G_1$, and in $G_1$ the vertices in $W_k$ form an induced subgraph isomorphic to $H_b'$). So in particular, the condition (c) means that we have an induced copy of $H_b'$ in blue. But the condition is actually much stronger than this — we also have constraints on where this subgraph lies and what shape it has (it has to correspond to $W_k$ for some $k \geq j$).




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 (2) in the end. There, the first two options mean we've won (we've found an induced subgraph isomorphic to $H_r$ in red or $H_b$ in blue). And the third option cannot hold for $j = n + 1$ because there are no indices $k \in \{j, \dots, n\}$; so we're forced to be in (a) or (b), and we're happy. (One way to think about this claim is that we start with all possible options $W_k$, and then exclude them one by one at each step of the induction. In the end all the options are gone, so option (c) is not viable anymore.)

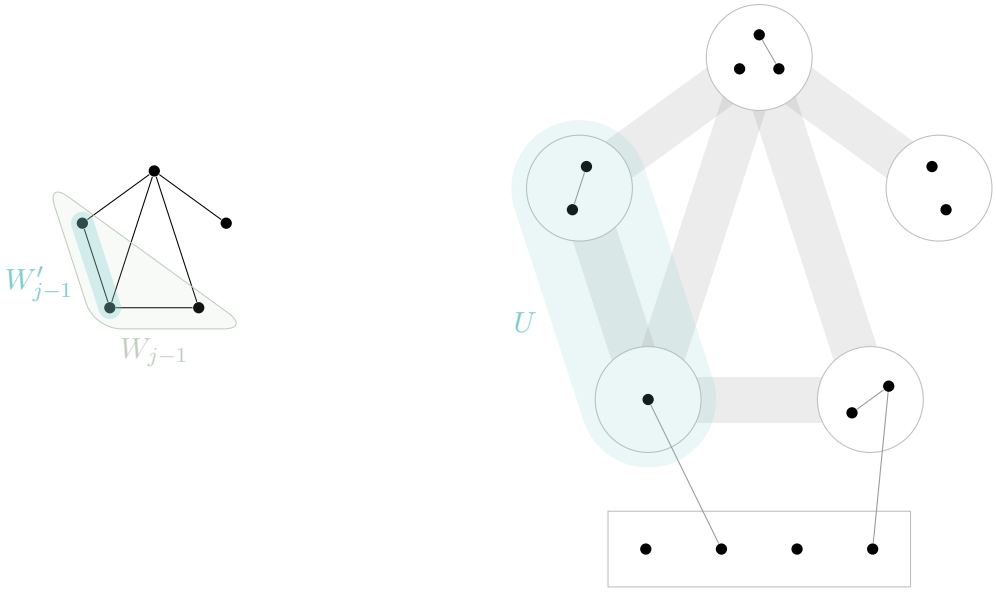
So our task is now to prove this claim, which we'll do by induction on $j$.

*Proof of Claim 8.4.* 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 $v \in V(G_1)$ (i.e., we start out with each vertex as its own individual blob). Condition (1) is tautologically true. For condition (2), we know that $G_1$ satisfies Theorem 8.3 for $H_r$ and $H_b'$. So in any coloring we'll find either an induced copy of $H_r$ with all edges red, in which case we're in case (a) and we're happy, or an induced copy of $H_b'$ with all edges blue. This must be one of $W_1, \dots, W_n$ (since this is a list of all possible subgrahps isomorphic to $H_b'$), so we're in case (c).

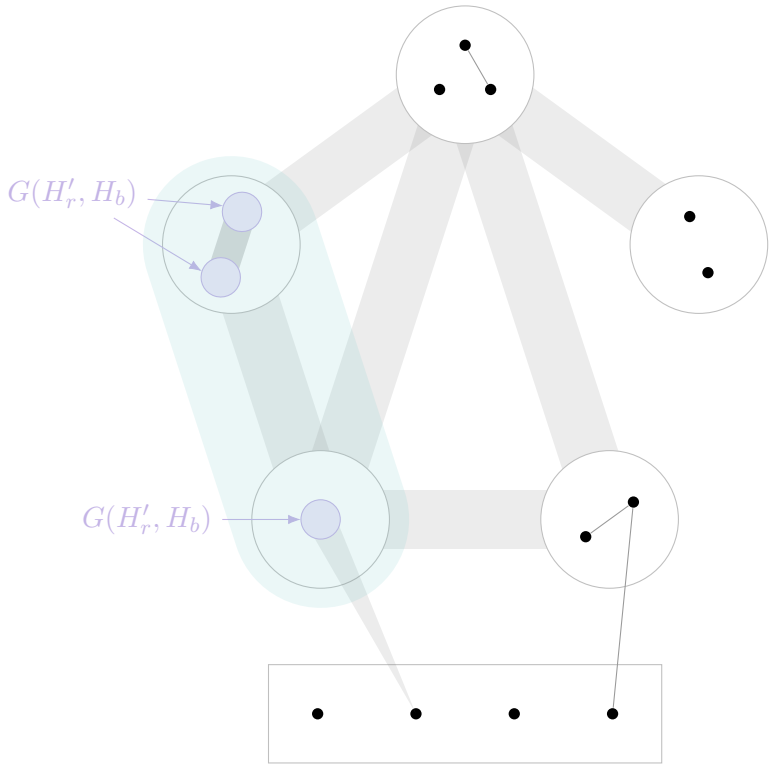
Now suppose that $2 \leq j \leq n + 1$, and that we already have such a graph $G_{j-1}$ for $j - 1$. Our goal is now to construct $G_j$ from $G_{j-1}$. Note that $G_{j-1}$ already has all the properties we want except that it's possible the vertex set in (c) is $W_{j-1}$, so the goal of the induction is to exclude $W_{j-1}$ as a possibility. We'll do this in two steps, as follows. First let

$$U = \bigcup_{v \in W_{j-1}'} U(v) \subseteq V(G_{j-1}).$$

(The reason we're looking at $W_{j-1}'$ is that we want the new stuff we do to only be connected to $W_{j-1}'$ — we essentially want to attach a vertex to $W_{j-1}'$ in order to form a copy of $H_b$.)

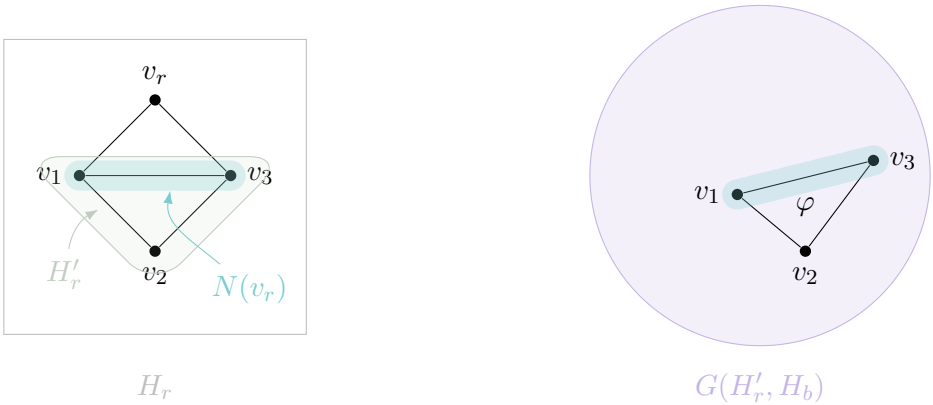



For the first step, we replace every vertex in $U$ by a copy of the graph $G(H'_r, H_b)$ (so we're taking our graph $G_{j-1}$ and blowing up each vertex in $U$ to something, namely the other graph from the induction hypothesis that we haven't used yet); we keep every vertex in $V(G_{j-1}) \setminus U$ as it is. This defines a new graph, which we call $G^*_{j-1}$. We define the sets $U(v)$ for this graph in the way you'd expect — whenever we blow up a vertex in $G_{j-1}$, the new vertices it becomes are still part of the same set $U(v)$. (In other words, for every vertex in a set $U(v)$ with $v \in W_{j-1}$, the entire associated copy of $G(H'_r, H_b)$ now 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 gardening plots, where each of us students has a plot and grows vegetables in that plot. The seeds (the original vertices) inside the plot grow into bigger graphs, but they never leave the plot.)

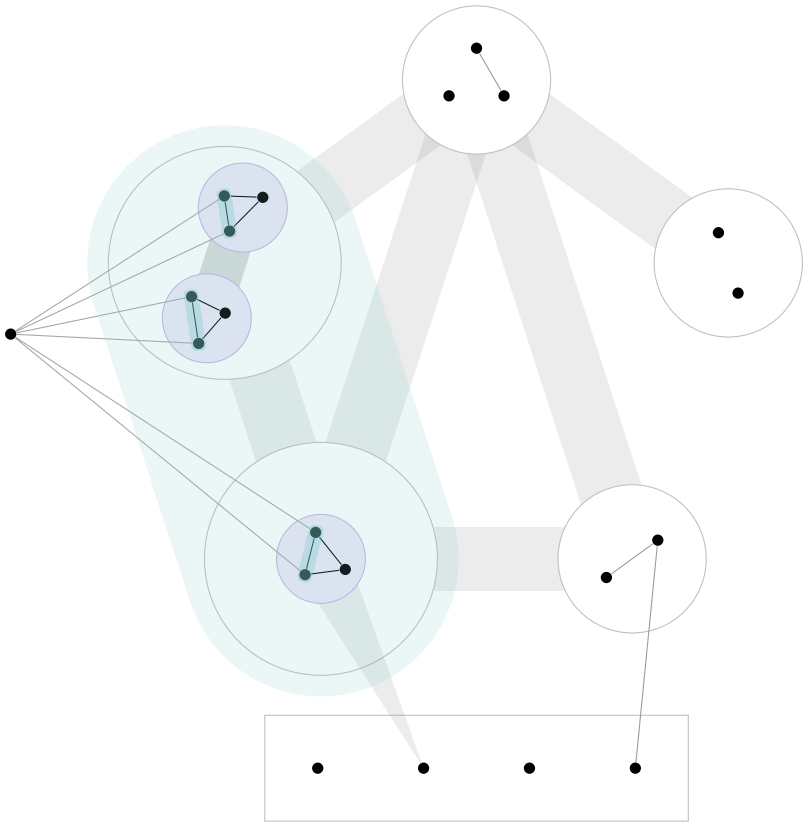



Now we'll do the second step. Let $m$ be the number of labelled induced subgraphs of $G(H'_r, H_b)$ (the graph we put into each of the blown-up vertices) which are isomorphic to $H'_r$. ('Labelled' means we're also fixing

an isomorphism. This definition isn't strictly necessary, but it'll give us some intuition.) For every $u \in U$, let $\Phi_u$ be the collection of all embeddings $\varphi_u : H_r' \to G_{j-1}^*$ whose image is an induced subgraph of $G_{j-1}^*$ which is isomorphic to $H_r'$ and lies in the copy of $G(H_r', H_b)$ corresponding to $u$ in $G_{j-1}$. This sounds complicated, but it's really just saying that for each of the vertices $u$ which we blew up in the previous step, we look at all the different ways to embed $H_r'$ into the thing we blew it up into (i.e., the small plant that grew from the seed $u$) as an induced subgraph. The number of such mappings for every $u$ is precisely $m$ (we won't need this, but it may help to keep things organized in your head) — there's $m$ different mappings from $H_r'$ into $G(H_r', H_b)$, and each $u$ corresponds to a copy of $G(H_r', H_b)$, so there's also $m$ mappings for each $u$.




Now we'll attach some additional vertices. For every choice of a tuple of maps $(\varphi_u)_{u \in U}$ where $\varphi_u \in \Phi_u$ for all $u$ (i.e., over all ways to choose one map $\varphi_u$ for each vertex $u \in U$ — so there are $m^{|U|}$ such tuples), we add an additional vertex to $G_{j-1}^*$ with edges to all vertices in $\bigcup_{u \in U} \varphi_u(N(v_r))$. In other words, for each of these maps $\varphi_u$, we look at the image of our set $N(v_r) \subseteq V(H_r')$ (this is the set of vertices we want to attach something at to turn $H_r'$ into $H_r$). We look at these images over all $u$, and we connect an additional vertex to all these images.



In particular, we add $m^{|U|}$ vertices in total in this step, because there are this many ways of choosing the tuple $(\varphi_u)_{u\in U}$. These newly added vertices won't be any parts of our sets $U(v)$.

We take $G_j$ to be the resulting graph, with the sets $U(v)$ defined as above. Now we'll check that $G_j$ satisfies the necessary conditions.

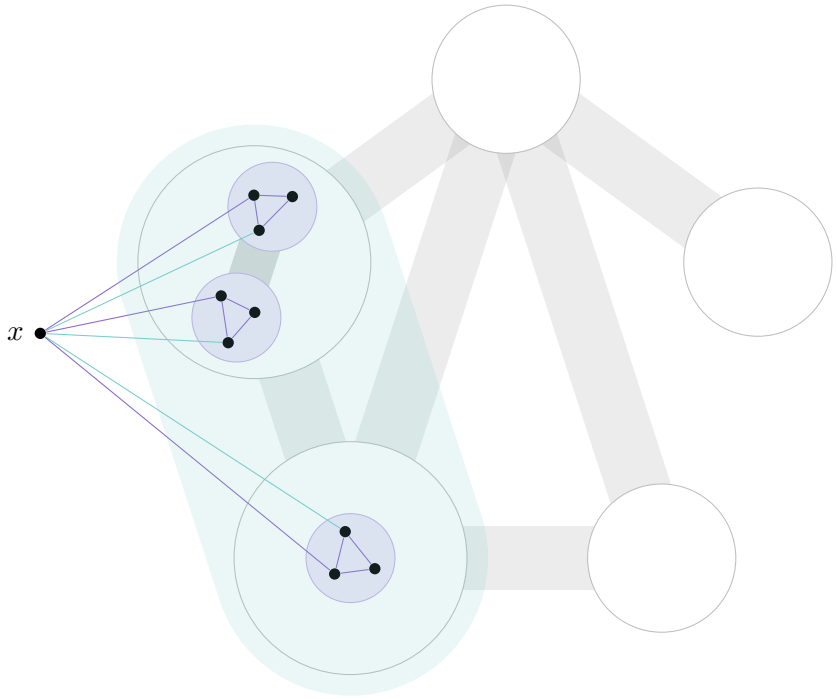
First, condition (1), on the blowup-like structure between the sets $U(v)$, is true by construction — in the first step (where we created $G^*_{j-1}$) we blew up certain vertices without changing the structure between our sets $U(v)$, and the newly added vertices in the second step don't affect this condition.

The more interesting part is condition (2). For this, we need to consider a coloring of the edges of $G_j$ with red and blue, and we need to show that this coloring satisfies (a), (b), or (c). Let's assume it doesn't satisfy (a) or (b), so we need to show that it satisfies (c).

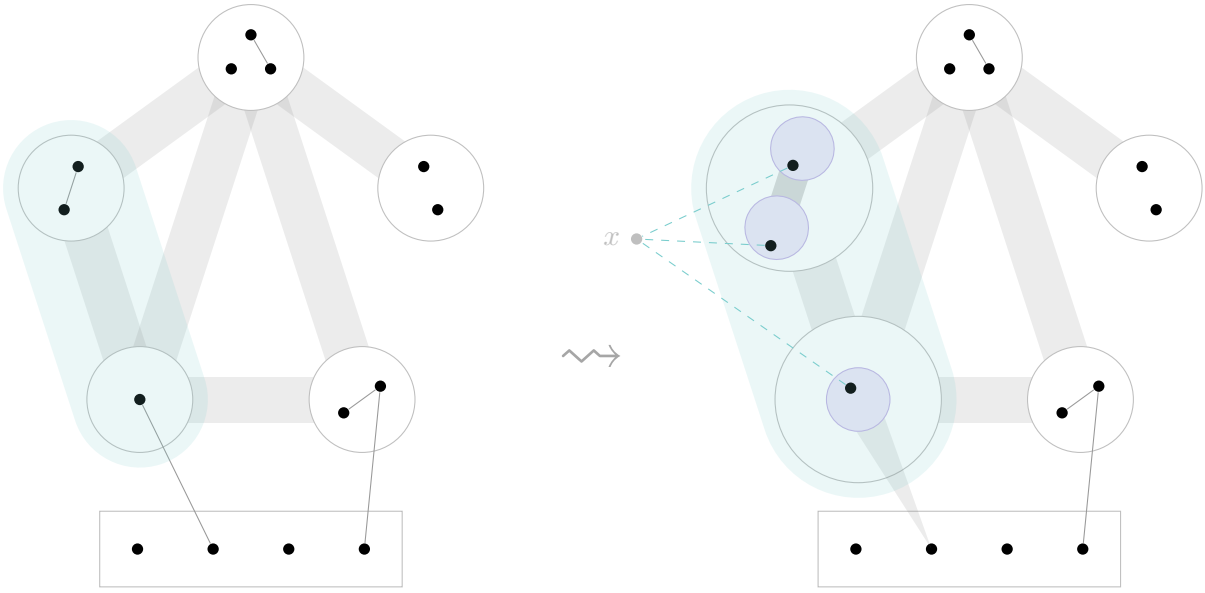
In the first step of the construction, we took $G_{j-1}$, and for every vertex $u \in U$, we replaced $u$ with a copy of $G(H'_r, H_b)$. We know the coloring of this copy has a red $H'_r$ (by definition it must have an induced copy of $H_b$ with all edges blue or an induced copy of $H'_r$ with all edges red; and we assumed that the first doesn't hold, so the second must hold). And this induced subgraph must be of the form $\varphi_u(H'_r)$ for some $\varphi_u \in \Phi_u$ (this is by the definition of $\Phi_u$, which just lists all the possible ways to find an induced subgraph in this graph isomorphic to $H'_r$).

So for every $u \in U$, we've chosen some $\varphi_u \in \Phi_u$; together, these form a tuple $(\varphi_u)_{u\in U}$. So now we have one special tuple associated to our coloring. And for this tuple, we added a particular vertex in the second step; call this vertex $x$.

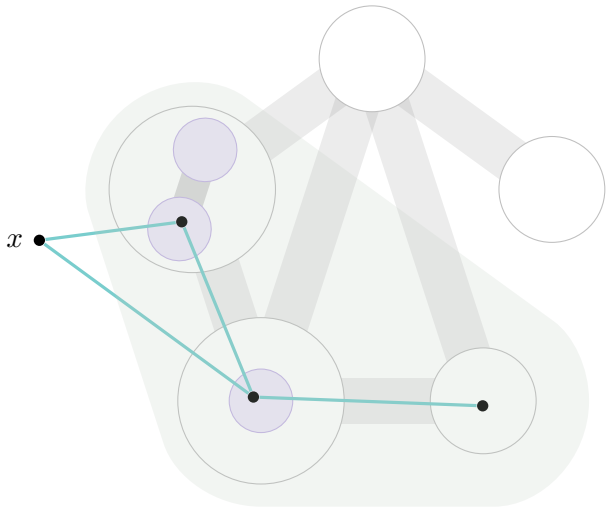
Now for every $u \in U$, by construction $x$ together with $\varphi_u(H'_r)$ forms an induced subgraph of $G_j$ isomorphic to $H_r$ (since $x$ has exactly the correct adjacencies to be the image of $v_r$, namely the ones corresponding to $N(v_r)$). And within $\varphi_u(H'_r)$, all the edges are already red. So if all the edges from $x$ to $\varphi_u(N(v_r))$ were red as well, then we'd get a red induced $H_r$. This means 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 them, the vertices $f(u)$ over all $u \in V(G_{j-1})$ together form an induced subgraph of $G_j$ isomorphic to $G_{j-1}$. This subgraph has also been colored, so we can apply condition (2) for $G_{j-1}$ — if this coloring of $G_{j-1}$ 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're done (since the coloring of $G_j$ then satisfies (c) for that same $k$). So our only problem is if $k = j-1$.




If $k = j-1$, then we can look at the induced copy of $H_b'$ from (c), and add $x$ to it — this forms a blue copy of $H_b$ (since $x$ has the right edges to it, i.e., the ones corresponding to $N(v_b)$, in blue), so we're done. □

# §9 Ramsey graphs

Our next topic is Ramsey graphs. (We've 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.)

## §9.1 Motivation and definition

In the first week of class, we saw the following theorem (in slightly different phrasing):

**Theorem 9.1** (Erdős–Szekeres)
Every graph on $n$ vertices contains a clique or independent set of size at least $\frac{1}{2}\log_2 n$.

We didn't state it in this form, but it's equivalent (to Theorem 1.13). We can color the edges red and 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 size $k$. We also stated the theorem in terms of an upper bound on Ramsey numbers, but these are equivalent — if $k = \lceil \frac{1}{2}\log_2 n \rceil$, then we saw that

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

which means the given coloring must have a monochromatic clique of size $k$.

**Remark 9.2.** The constant $\frac{1}{2}$ stood for about 80 years, until it was improved about a month ago by a recent result of Campos–Griffiths–Morris–Sahasrabudhe; we'll talk about this more in two weeks.

In the other direction, we also discussed the following theorem:

**Theorem 9.3** (Erdős 1947)
For $n \ge 3$, there exist $n$-vertex graphs with no clique or independent set of size at least $2\log_2 n$.

This is a restatement of Theorem 1.20 — taking $k = \lfloor 2\log_2 n \rfloor$, we saw that $R(k,k) > 2^{k/2} \ge n$, which means $n$ isn't good enough for Ramsey's theorem for $k$. This means there must be a red-blue coloring with no monochromatic $k$-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 class — given a graph on $n$ vertices, we're guaranteed to find a clique or independent set of size $\frac{1}{2}\log_2 n$, but we can't guarantee one of size $2\log_2 n$.

The proof of Theorem 9.3 relied on the probabilistic method (it was one of the first uses of probabilistic methods, and it very much popularized them). Even after many decades and a lot of attention, we still don't know any non-probabilistic proofs — in other words, we donn't know of 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 problem 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) — no explicit constructions of graphs as in Theorem 9.3 are known, even if we replace 2 by some other constant $C > 2$.

**Definition 9.4.** 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 the property in Theorem 9.3 where we replace 2 by $C$. The property gets weaker and weaker as we increase $C$. We can restate Theorem 9.3 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 can also restate Theorem 9.1 as saying $C$-Ramsey graphs don't exist for $C < \frac{1}{2}$, but we usually think of $C$ as large.)

**Remark 9.5.** What do we mean by 'explicit construction'? Obviously if we can write down the graph on a paper then it's explicit, but that doesn't make sense here (we're trying to get a construction for large $n$). Most definitions of 'explicit construction' say that there is an algorithm which can describe the graph in a reasonable amount of time. There are also potential number-theoretic constructions — for example, you can take the vertices of the graph to be $\mathbb{F}_p$, and draw an edge if and only if some number-theoretic property holds; this number-theoretic property might 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 know of any.

So it's known for $C \geq 2$ that 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 they must have — the heuristic is that they have to be sort of 'random-looking,' which is why it's hard to find them (since something we write down will have lots of structure).

## §9.2 Density bounds on Ramsey graphs

People have proven various results about $C$-Ramsey graphs, and there have been lots of conjectures about them, many of which have been resolved. In particular, here's a very nice theorem about $C$-Ramsey graphs.

**Theorem 9.6** (Erdős–Szemerédi 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 $\varepsilon$ and at most $1 - \varepsilon$.

Roughly speaking, this theorem says that $C$-Ramsey graphs can't be extremely sparse or extremely dense — their density is bounded away from 0 or 1. *A priori* this sounds like an extremely weak statement, since it sounds like it tells us nothing about the graph except its 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 size $C \log_2 n = 2C \log_2 \sqrt{n}$, and so we get a $2C$-Ramsey graph on those vertices; this means we can apply Theorem 9.6 on those $\sqrt{n}$ vertices. So we can apply this theorem to not just our graph itself, but also to all induced subgraphs with a reasonable number of vertices. This gives 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 the 'at least $\varepsilon$' and 'at most $1-\varepsilon$' statements are equivalent, since the complement of a $C$-Ramsey graph is also $C$-Ramsey. So to prove Theorem 9.6, it suffices to prove the 'at least $\varepsilon$' bound (and the bound in the other direction follows automatically).

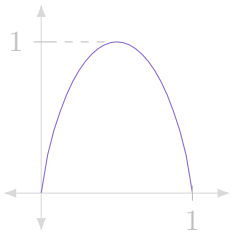
First, as some preparation for the proof, we'll use the following notion.

**Definition 9.7.** For $0 \leq x \leq 1$, we define the binary entropy of $x$, denoted $\mathsf{H}(x)$, by

$$\mathsf{H}(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}$$

if $0 < x < 1$, and $\mathsf{H}(x) = 0$ if $x \in \{0, 1\}$.

This function $\mathsf{H}$ is continuous 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's monotone increasing on $[0, \frac{1}{2}]$ and monotone decreasing on $[\frac{1}{2}, 1]$. When we plug in $x = \frac{1}{2}$, we obtain $\mathsf{H}(\frac{1}{2}) = 1$; this means $\mathsf{H}(x) \in [0,1]$. We can also see that $\mathsf{H}$ is symmetric about $\frac{1}{2}$, and that it's concave (we won't need this for our argument, but it's sometimes useful elsewhere).




The relevance of this is that we can use binary entropy to bound binomial coefficients:

**Fact 9.8 —** For any $0 \le k \le n$, we have $\binom{n}{k} \le 2^{\mathsf{H}(k/n)\cdot n}$.

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 this bound gives us, as $\mathsf{H}(\frac{1}{2}) = 1$). But what about $\binom{n}{n/3}$ or $\binom{n}{n/4}$ or so on — what's the right base of the exponent? That's exactly what the entropy funtion tells us (this bound is asymptotically tight — the same lower bound holds up to an $o(n)$ in the exponent). So if we ever have a combinatorial argument where we run into $\binom{n}{n/10}$ (for example), bounding it by $2^n$ is quite wasteful; the entropy gives the correct bound.

*Proof of Fact 9.8.* First, Stirling's formula tells us that

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

(we generally don't care about the constant factors). Then we can write

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \le \frac{e\sqrt{n} \cdot (\frac{n}{e})^n}{\sqrt{2\pi} \cdot \sqrt{k} \cdot (\frac{k}{e})^k \cdot \sqrt{2\pi} \cdot \sqrt{n-k} \cdot (\frac{n-k}{e})^{n-k}}.$$

Conveniently, we have $e < 2\pi$ and $\sqrt{n} \le \sqrt{k(n-k)}$ (unless $k$ is 0, 1, $n-1$, or $n$, and we can check these cases by hand), and the exponentials of $e$ cancel out; so we get

$$\binom{n}{k} \le \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} = 2^{\mathsf{H}(k/n)\cdot n}. \qquad \square$$

April 25, 2023

*Proof of Theorem 9.6.* Let $0 < \varepsilon < \frac{1}{16}$ be such that $\mathsf{H}(8\varepsilon) < \frac{1}{4C}$. (Given some $C$, we can always find such an $\varepsilon$ — this is because $\frac{1}{4C}$ is some small positive number, so we can make $\varepsilon$ small enough that $\mathsf{H}(8\varepsilon)$ is even smaller.)

We need to show that in any $C$-Ramsey graph on $n$ vertices, the density is at least $\varepsilon$ and at most $1-\varepsilon$. First, it suffices to show that the density is at least $\varepsilon$; then we automatically get that it's at most $1-\varepsilon$ by applying this to the complement (the complement of a $C$-Ramsey graph is also $C$-Ramsey; so the density of this complement must be at least $\varepsilon$, which means the density of the original graph is at most $1-\varepsilon$).

Suppose that $G$ is a $C$-Ramsey graph on $n$ vertices, where $n$ is sufficiently large — specifically, such that $C \log_2 n \leq \frac{n}{8}$ and $n \geq 2^{12}$ (we'll see later in the proof where we need these size conditions) — and assume for contradiction that the density of $G$ is less than $\varepsilon$.

First, since the density of $G$ is less than $\varepsilon$, its average degree is less than $\varepsilon n$ (since the density can be calculated by averaging all degrees and dividing by $n-1$). This means at most half the vertices can have degree at least $2\varepsilon n$. We think of these abnormally high-degree vertices as bad, and as usual, we'll just kick them out. So we 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, which means 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 we'll try to find a large independent set in $U$. *A priori* it's not clear how to do this. But the idea is to pick a maximum-size 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 $C \log_2 n$). 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

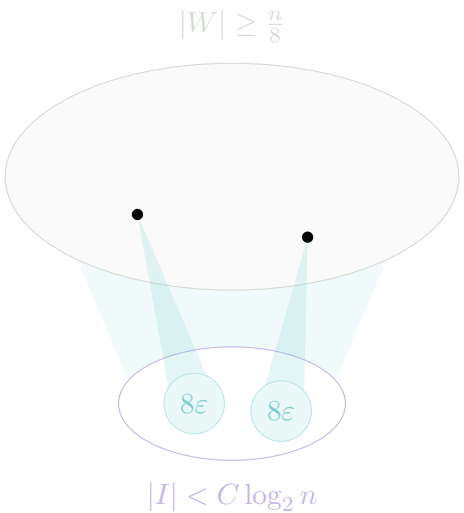
$$e(I, U \setminus I) \leq \sum_{v \in I} \deg(v) \leq 2\varepsilon n \cdot |I|$$

(because every vertex $v \in I$ contributes at most $\deg(v)$ edges, and since $I \subseteq U$, by the definition of $U$, 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 in $I$, 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 $I$ — if there were more, they would contribute a total of more than $\frac{n}{4} \cdot 8\varepsilon |I| = 2\varepsilon |I|$ edges between $I$ and $U \setminus I$. As before, we'll kick these vertices 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

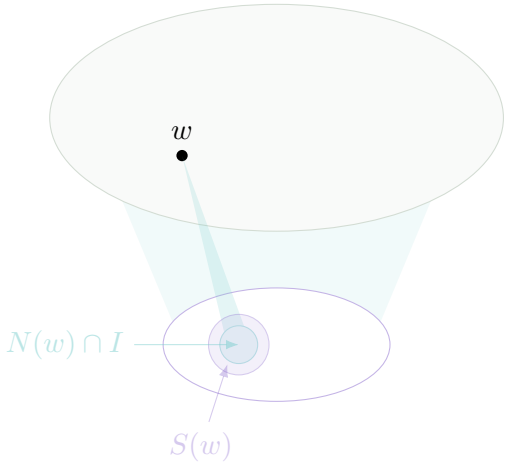
$$|W| \geq |U| - |I| - \frac{n}{4} \geq \frac{n}{2} - \frac{n}{8} - \frac{n}{4} = \frac{n}{8}$$

(this is what we needed the loose bound $|I| \leq \frac{n}{8}$ for).

So far, we've pruned out the high-degree vertices of $G$; then we looked at a maximum-size independent set $I$ and pruned away the vertices outside $I$ that had abnormally many edges into $I$. Now we have a set $W$ of $\frac{n}{8}$ vertices where each only has a few edges to $I$ — this means it has a neighborhood in $I$ which is at most an $8\varepsilon$-fraction of $I$. We'll analyze these neighborhoods using Fact 9.8 (our bound on binomial coefficients), which bounds the *number* of possible neighborhoods.



For each vertex $w \in W$, we choose a subset $S(w) \subseteq I$ of size *exactly* $s = \lfloor 8\varepsilon\,|I| \rfloor$ such that $N(w) \cap I \subseteq S(w)$. (This is possible because every $w \in W$ has at most $s$ neighbors in $I$. So if $w$ has *exactly* $s$ neighbors, then we define $S(w)$ to consist of those neighbors; if it has fewer, then we take those neighbors and add extra vertices for padding to get a set of size exactly $s$; note that $s < |I|$, so we can do this.)




The point is that we want to say that there aren't too many possibilities for $S(w)$, so there are mahy vertices in $W$ with the *same* set $S(w)$. First, the total number of subsets of $I$ of size $s$ is

$$\binom{|I|}{s} \leq 2^{\mathsf{H}(|I|/s)\cdot|I|} \leq 2^{\mathsf{H}(8\varepsilon)\cdot|I|}$$

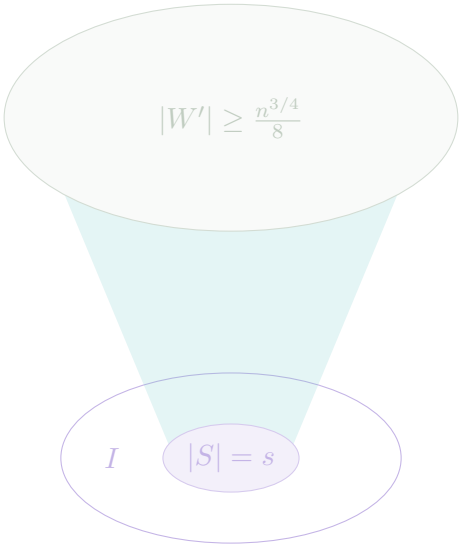
(we defined $s$ so that $s \leq 8\varepsilon\,|I|$; and since $8\varepsilon < \frac{1}{2}$, the entropy function is monotone up to $8\varepsilon$). We defined $\varepsilon$ so that $\mathsf{H}(8\varepsilon) < \frac{1}{4C}$, so this gives

$$\binom{|I|}{s} \leq 2^{\frac{1}{4C}\cdot C\log_2 n} = n^{1/4}.$$

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. Specifically, there must be a subset $W' \subseteq W$ and a subset $S \subseteq I$ of size $s$ such that $S(w) = S$ for all $w \in W'$, and

$$|W'| \geq \frac{|W|}{n^{1/4}} \geq \frac{n}{8n^{1/4}} = \frac{n^{3/4}}{8}.$$

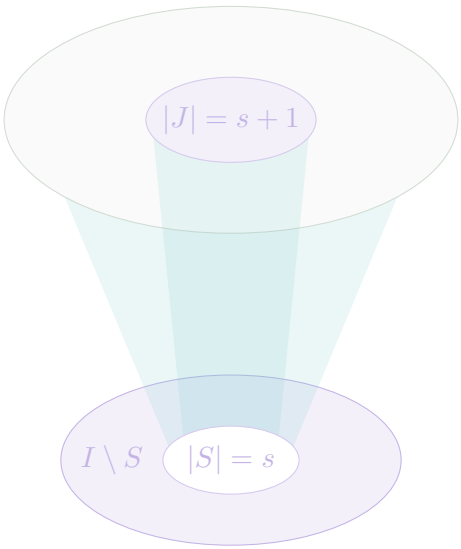
In particular, all edges from $W'$ to $I$ have their endpoints in $S$, i.e., there are no edges from $W'$ to $I \setminus S$.



(The point is that there are at least $\frac{n}{8}$ vertices $w \in W$ and each gets associated with a set $S(w)$; but there's at most $n^{1/4}$ choices for $S(w)$, so at least $\frac{n}{8n^{1/4}}$ vertices $w$ must have the same set $S(w)$. We call that set $S$, and let $W'$ be the set of these vertices $w$ associated to that set.)

**Claim 9.9 —** The set $W'$ 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 $(I \setminus S) \cup J$ (the set where we take $I$, remove $S$, and replace it with $J$) 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's also a subset of $U$ (all the relevant sets that occur here are subsets of $U$ — the vertices we kicked out at the beginning are gone from the picture); this contradicts the choice of $I$ as an independent set in $U$ of *maximum* size. □

Now we have an upper bound on the independence number of $W'$. Meanwhile, we also know that $W'$ does not contain a clique of size at least $C \log_2 n$ (because $G$ is a $C$-Ramsey graph, so it can't have a clique of this size). 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 $|I| \le t$ (since $|I| < C \log_2 n$).

So now we have an upper bound for $|I|$ as well as the sizes of cliques and independent sets in $W'$. We'll first combine the latter two bounds and use Ramsey's theorem to obtain an upper bound on $|W'|$ — since $W'$ forms a graph with no independent set of size $s+1$ or clique of size $t+1$, Ramsey's theorem (or more specifically, Theorem 1.13) gives

$$\left|W'\right| \le R(s+1, t+1) \le \binom{s+t}{s} \le 2^{\mathsf{H}(s/(s+t))\cdot(s+t)}.$$

Now using the fact that $|I| \le t$, we have

$$\frac{s}{s+t} \le \frac{s}{t} \le \frac{8\varepsilon\,|I|}{t} \le 8\varepsilon.$$

The above bound also shows that $s \le t$, so $s + t \le 2t$; plugging these into our bound for $|W'|$ gives

$$|W'| \le 2^{\mathsf{H}(8\varepsilon)\cdot 2t}.$$

Finally, plugging in our bounds on $\varepsilon$, $t$, and $n$, we obtain

$$|W'| \le 2^{\frac{1}{4C}\cdot 2C\log_2 n} \le n^{1/2} < \frac{n^{3/4}}{8} \le |W'|,$$

which is a contradiction. □

**Remark 9.10.** Note that in Theorem 9.6, we had two assumptions — that $G$ doesn't have a large clique or independent set — and we wanted to prove that its density is at least $\varepsilon$, i.e., we wanted to exclude the case that $G$ is very sparse. You might first think that it only matters that there is no large independent set (i.e., that just having no large independent set would imply a density of at least $\varepsilon$), since you'd naively expect a very sparse graph to have a large independent set. 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 Turán's theorem, in order to show that the density is at least $\varepsilon$ using only independent sets, you'd have to exclude independent sets of a much smaller (*constant* rather than logarithmic) size.

## §9.3 Induced subgraphs of Ramsey graphs

The next thing we'll discuss is the following theorem, which proves a conjecture of Erdős and Hajnal (which was open for around a decade).

**Theorem 9.11** (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 $\lambda \log_2 n$ vertices as an induced subgraph.

The intuition is that it's hard to find explicit constructions of $C$-Ramsey graphs; heuristically, we think this is because they should be random-like or unstructured. This theorem is sort of a concrete instance of this unstructuredness — such graphs must have induced subgraphs of all shapes (up to a certain size).

In some sense, this should be very surprising. The definition of a $C$-Ramsey graph says that it does not contain a clique or independent set of size at least $C \log_2 n$. But the conclusion of Theorem 9.11 is about subsets of the same order of magnitude (just changing the big constant $C$ to a small constant $\lambda$), and it says that looking at those subsets, we can find *everything*. In particular, we can find cliques and independent sets of size $\lambda \log_2 n$. This is maybe not that surprising, but it's definitely *somewhat* surprising — Ramsey's theorem tells us that we can find *either* a clique or an independent set, but this theorem tells us that we can find both, as well as anything else we might wish for.

We'll now prove Theorem 9.11. The first ingredient in the proof is a modification of a lemma that we saw before (in the proof of Theorem 6.13 — that if $H$ has bounded degree, then $R(H, H)$ is linear in the number of vertices of $H$). First, here's the statement of the lemma that we saw before.

**Lemma 6.20**

Suppose that $0 \le \rho \le \frac{1}{2}$ and $0 \le \delta \le \frac{1}{4}$, and let $s \ge \frac{1}{\delta}$ 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 $|U| \ge (\frac{\rho}{2})^s |V(G)|$, such that for all pairs of disjoint subsets $X, Y \subseteq U$ of sizes $|X| \ge \rho |U|$ and $|Y| \ge \rho |U|$, we have $d(X, Y) \ge \delta$.

(b) There is a vertex subset $U \subseteq V(G)$ of size $|U| \ge (\frac{\rho}{2})^s |V(G)|$ with density $d(U) \le 3\delta$.

(When we originally stated Lemma 6.20, for (a) we just said that there's a vertex subset $U$ such that the induced subgraph on $U$ is bi-$(\rho, \delta)$-dense. The condition written here is just the definition of what it means to be bi-$(\rho, \delta)$-dense, which we won't repeat since we'll modify it shortly.)

This isn't exactly the statement we need to prove Theorem 9.11 — (a) only tells us that there aren't too *few* edges between pairs of subsets (it says the density between pairs is *at least* something). But if we want to find induced subgraphs, we also want it to not be too *many* edges — we need both non-edges and edges to behave correctly. So we'll modify it as follows:

**Lemma 9.12**

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 $d(U) \le 3\delta$ or $d(U) \ge 1 - 3\delta$.

Compared to the previous lemma, in (a) we added an upper bound of $1 - \delta$. We have to pay for this in two ways. First, we change the bound on $s$ by a constant factor; this is not important. We also get another exit case where the density of $U$ is extremely *high* (instead of extremely *low*); this gives the second case in (b). (Note that (a) gives both the bounds of $\delta$ and $1 - \delta$ at the same time, while (b) has an 'or.')

So Lemma 9.12 is essentially a more symmetric version of Lemma 6.20 in terms of edges and non-edges — we basically took all edge conditions and added in the analogous non-edge conditions.

As you'd expect, we can prove Lemma 9.12 similarly to how we proved Lemma 6.20; this will be on the homework. (We have to tweak the proof a bit to get this modification — Prof. Sauermann doesn't known of a way to black-box Lemma 6.20 and apply it first on edges and then non-edges, because the parameters don't work out if we apply it twice.)

We'll now use this to prove Theorem 9.11. (The proof will make use of Theorem 9.6.)

*Proof of Theorem 9.11.* First note that we can assume $n$ is sufficiently large with respect to $C$ — if we're able to prove the theorem for all $n$ above some threshold depending only on $C$, then we get some $\lambda$ which works for all $n$ above that threshold. There are only finitely many $n$ below that threshold, so we can just artifically lower $\lambda$ so that $\lambda \log_2 n \le 1$ for all such $n$; then the statement trivially holds for these values of $n$ as well.

Now let $G$ be a $C$-Ramsey graph on $n$ vertices; we'll assume that $n$ is large enough (in order to satisfy the threshold for Theorem 9.6, as well as a few other inequalities we'll want).

Theorem 9.6 tells us something about the density of $G$, but that's rather weak — it only tells us something about the total number of edges. But as mentioned last class, the true power of Theorem 9.6 lies in applying it to smaller induced subgraphs. We'll apply it to 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 we have $C \log_2 n = 2C \log_2 \sqrt{n}$, so if it has $m \ge \sqrt{n}$ vertices then it can't have a clique or independent set of size $2C \log_2 m$).

So let $\varepsilon > 0$ be such that Theorem 9.6 holds for $2C$-Ramsey graphs (with sufficiently many vertices). Then it tells us that every induced subgraph of $G$ on at least $\sqrt{n}$ vertices must have density at least $\varepsilon$ and at most $1 - \varepsilon$. Now let

$$\delta = \frac{\varepsilon}{4} \quad \text{and} \quad s = \left\lceil \frac{2}{\delta} \right\rceil$$

(the relevant inequalities here are that $3\delta < \varepsilon$ and $s \geq \frac{2}{\delta}$; note that $\varepsilon$, $\delta$, and $s$ only depend on $C$, and not $n$). Now we're in a position to apply Lemma 9.12. Let $\rho = 2n^{-1/2s}$, so that

$$\left(\frac{\rho}{2}\right)^s = n^{-1/2}.$$

Then applying Lemma 9.12 (with these parameters) 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 $|U| \geq \sqrt{n} = (\frac{\rho}{2})^s n$ satisfy

$$3\delta < \varepsilon \leq d(U) \leq 1 - \varepsilon < 1 - 3\delta.$$

Since option (b) is impossible, this means option (a) must happen. So we obtain a vertex subset $U \subseteq V(G)$ of size $|U| \geq \sqrt{n}$ 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 $\delta \leq d(X,Y) \leq 1 - \delta$. In other words, whenever we find disjoint $X$ and $Y$ which are not too small (specifically, such that each is at least a $\rho$-fraction of $U$), the density between them can't be too extreme. You can imagine that such a property is very helpful in order to embed things as induced subgraphs — to embed an induced subgraph 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, then it seems reasonable that we can succeed at this embedding task.

April 27, 2023

Now we'll use this nice property to embed all graphs of size at most $\lambda \log_2 n$ in our subset $U$. First we'll choose $\lambda$ — let $\lambda > 0$ be chosen such that

$$\lambda \log_2 \frac{1}{\delta} < \frac{1}{3s}$$

(both $\frac{1}{\delta}$ and $\frac{1}{3s}$ are positive, so we can make $\lambda$ small enough that this inequality holds). Since $\delta$ and $s$ only depends on $C$, we can define $\lambda$ such that it only depends on $C$ as well. 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 of $G$, so let $H$ be a graph with $m \leq \lambda \log_2 n$ vertices, named $v_1, \ldots, v_m$. Our goal is to embed these vertices $v_1, \ldots, v_m$ 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 $v_1, \ldots, v_m$ to distinct vertices $w_1, \ldots, w_m$ in $U$ (such that in the end, their images form an induced subgraph isomorphic to $H$).

We'll embed vertices one by one. As we do so, we'll look at the candidates we have left for the remaining vertices; we want to make sure that at every step, the sizes of these candidate sets don't go down too much. (This is similar to the proof of Lemma 6.18 from earlier.)

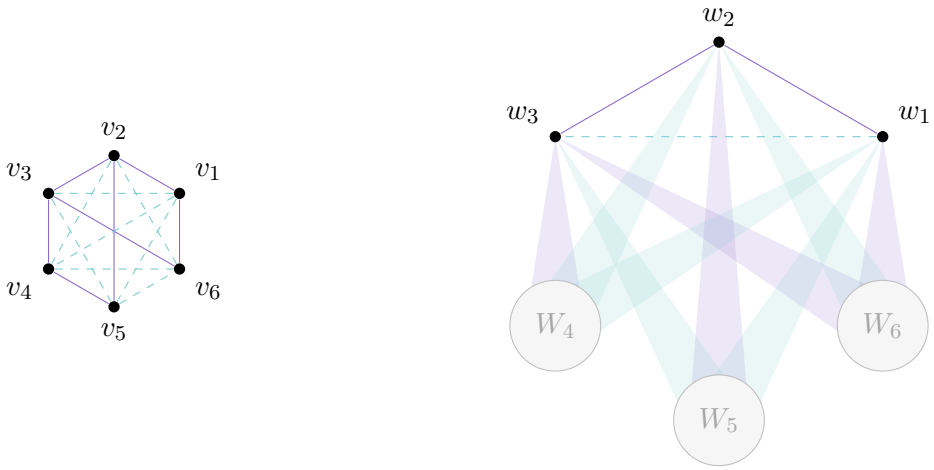
The following claim will let us do this — we'll inductively embed vertices one by one, and the claim will capture the properties we'll keep track of during the process. Then taking $j = m$ (which corresponds to the last step of the process) will give the desired embedding.

**Claim 9.13 —** For every $0 \leq j \leq m$, there exist distinct vertices $w_1, \ldots, w_j$ in $U$ and disjoint subsets $W_{j+1}, \ldots, W_m$ of $U \setminus \{w_1, \ldots, w_j\}$, each of size at least $\frac{\delta^j}{2m} \cdot |U|$, such that:

(1) The adjacencies between $w_1, \ldots, w_j$ are the same as between $v_1, \ldots, v_j$ in $H$ (i.e., $w_k w_\ell$ is an edge if and only if $v_k v_\ell$ is an edge).

(2) For every $i = j+1, \ldots, m$, every vertex in $W_i$ has the same adjacencies to $w_1, \ldots, w_j$ as $v_i$ has to $v_1, \ldots, v_j$ 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 is big enough. The second condition captures what it means to be a suitable candidate set — for each candidate set, all its vertices should have the right adjacencies to the already chosen vertices. (We don't have any assumptions about the edges between two candidate sets.) For example, if $v_i$ has an edge to $v_2$, then every vertex in $W_i$ should have an edge to $w_2$.




(In this picture, purple represents edges and blue represents non-edges.)

If we prove this claim, then we're done — if we take $j = m$, it tells us that there are vertices $w_1, \dots, w_m$ with the same adjacencies as $v_1, \dots, v_m$, giving an induced copy of $H$.

Before we prove the claim, we'll do the following side calculation (because it will come up repeatedly in the proof) — note that

$$\frac{\delta^m}{(2m)^2} \geq \frac{\delta^{\lambda \log_2 n}}{(2\lambda \log_2 n)^2} \geq \frac{n^{-1/3s}}{(2\lambda \log_2 n)^2}. \tag{9.1}$$

(The left-hand side comes from bounding how small the candidate sets $W_i$ become (as a fraction of $U$), with an extra factor of $2m$ thrown in.) Recall that $\lambda$ only depends on $C$ and not $n$, so the $n$-dependence of the right-hand side is $n^{-1/3s}/(\log n)^2$. In particular, if $n$ is sufficiently large, then this is greater than $\rho = n^{-1/2s}$ (since $\frac{1}{3s} < \frac{1}{2s}$). So the point is that these candidate sets are at least a $\rho$-fraction of $U$, even if we introduce an extra factor of $2m$ in the denominator. This is important because our property on $U$ (that the density between any pair of reasonably sized subsets is not too extreme) only applies to sets of size at least $\rho |U|$; this calculation will ensure that we can apply it to the sets that will come up.

*Proof of Claim 9.13.* We'll use induction on $j$. The base case is when $j = 0$. Here we don't need to specify any vertices $w_i$, and we just need to specify $m$ disjoint subsets $W_1, \dots, W_m$ of $U$. So we can just obtain these subsets by taking a partition of $U$ into sets of size at least

$$\left\lfloor \frac{|U|}{m} \right\rfloor \geq \frac{|U|}{2m}.$$

(Note that $|U| \geq \sqrt{n}$ and $m$ is only logarithmic in $n$, so $|U|$ is much bigger than $m$, which means rounding down only loses a factor of at most 2.)

**Remark 9.14.** *A priori*, we might think of all vertices in $U$ as candidates for all of $w_1, \dots, w_m$, since we don't have any restrictions on the candidate sets yet. But the reason we're doing this partition is that we want to ensure our candidate sets are *disjoint* (the condition on $U$ is only for disjoint subsets $X$ and $Y$, so this makes it easier to use). And we can easily afford this because losing a factor of $2m$ doesn't really hurt us.

Now assume that $1 \leq j \leq m$, and that we've already proven the claim for $j-1$. This means we've found distinct vertices $w_1, \ldots, w_{j-1}$ in $U$, and disjoint subsets $W'_j, W'_{j+1}, \ldots, W'_m$ of $U \setminus \{w_1, \ldots, w_{j-1}\}$ (the primes are there to distinguish these from the sets we'll obtain for $j$ — we'll have to make these sets smaller), which satisfy the conditions in the claim for $j-1$.

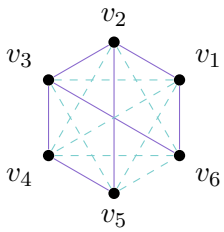
Now we need to prove the claim for $j$. This means we need to produce the next vertex $w_j$, which should come from $W'_j$. This automatically ensures that it's distinct from the previous vertices $w_1, \ldots, w_{j-1}$, and that it has the right adjacencies to those vertices (because of (2) for $W'_j$). But picking $w_j \in W'_j$ will restrict the remaining candidate sets (since we need to ensure they have the correct edges to $w_j$); so we need to ensure that this doesn't make the candidate sets too small.

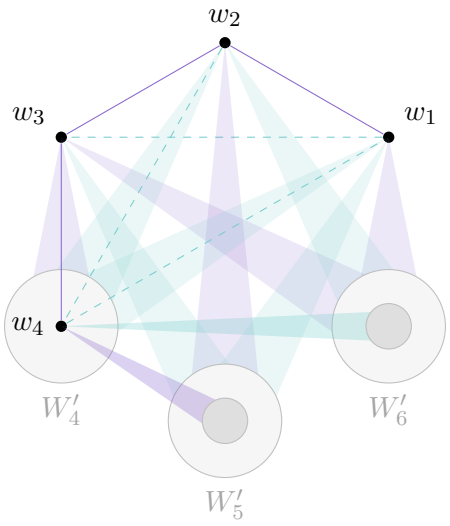
In other words, we need to find a vertex $w_j \in W'_j$ such that:

- For all $i \in \{j+1, \ldots, m\}$ with $v_j v_i \in E(H)$, the vertex $w_j$ has at least $\delta |W'_i|$ neighbors in $W'_i$.
- For all $i \in \{j+1, \ldots, m\}$ with $v_j v_i \notin E(H)$, the vertex $w_j$ 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 (meaning that for each $i$, we take the subset of $W'_i$ with the correct adjacencies to the newly picked vertex $w_j$), each shrinks by a factor of at most $\delta$. (If $v_j v_i$ is an edge in $H$, then we need to update $W'_i$ to the set consisting of the neighbors of $w_j$ in $W'_i$, so we need this set of neighbors to have size at least $\delta |W'_i|$. If $v_j v_i$ is not an edge in $H$, then we need to update $W'_i$ to the set of *non-neighbors* of $w_j$; 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'll automatically have the right adjacencies (taking $W_i$ to be these subsets of either neighbors or non-neighbors of $w_j$ in $W'_i$), and these new candidate sets $W_i$ will still be disjoint and have the right sizes.



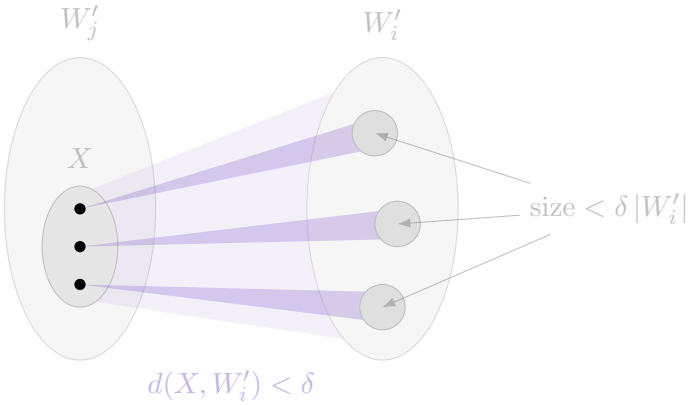




In order to show that there is such a vertex $w_j$ in $W'_j$, for every $i$ we'll upper-bound the number of choices for $w_j$ which *fail* this condition, using the nice property of $U$ (that the density between any two reasonably large sets $X$ and $Y$ is not too extreme). Specifically, we claim that for each $i \in \{j+1, \ldots, m\}$, there are fewer than $\rho |U|$ vertices $w_j \in W'_j$ which violate this condition for $i$. To see this, fix $i$, and take $X$ to be the set of vertices in $W'_j$ which violate this condition, and $Y = W'_i$. If $v_j v_i$ is an edge in $H$, then every violating vertex $w_j$ has density less than $\delta$ to $W'_i$, so the density between the set of *all* violating vertices and $W'_i$ is also less than $\delta$. Similarly, if $v_j v_i$ is a non-edge, then this density is more than $1-\delta$. So we've found two disjoint subsets of $U$ such that the density between them is less than $\delta$ 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 that $W'_i$ has size at

least $\rho |U|$ by our earlier calculation from (9.1), as

$$\left|W_i'\right| \geq \frac{\delta^{j-1}}{2m} |U| \geq \frac{\delta^m}{(2m)^2} |U| \geq \rho |U| .$$

So the only way this is possible is if $|X| < \rho |U|$ (otherwise we'd have two reasonably large sets of vertices with density at most $\delta$ or at least $1 - \delta$, which isn't allowed).




Now we can just count — for every $i$ there are less than $\rho |U|$ vertices in $W_j'$ which fail (meaning that they violate the condition for $i$), so in total, there are at most

$$(m-j)\rho |U| \leq m\rho |U| \leq \frac{\delta^m}{4m} |U| \leq \frac{1}{2} \cdot \frac{\delta^{j-1}}{2m} |U| \leq \frac{1}{2} |W_j'|$$

vertices which violate one of these conditions (using our calculation (9.1) from earlier). So at most half the vertices in $W_j'$ are bad, which means there must be at least one good vertex; this means we can find $w_j \in W_j'$ with the desired properties. □

This finishes the proof — the point is that our strong condition on $U$ was good enough to let us run this greedy embedding. □

# §10 Rado's theorem

We've talked about graphs for the last eight weeks. Now we'll return to a topic which is *not* about graphs, and is instead about coloring numbers and solving equations.

## §10.1 Motivation and setup

We've seen several results along these lines earlier, which we'll now recall.

**Theorem 3.1** (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 for $N$ large enough, whenever we take the ground set $\{1, \dots, N\}$ and color these 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) numbers with the same color.

Here are two other statements of this type, which we've seen on the homework.

**Notation 10.1.** We write $\mathbb{N} = \{1, 2, \dots\}$ to denote the set of positive integers.

**Proposition 10.2** (Homework 1, Problem 2a)

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 exist $x, y_1, \dots, y_d \in \mathbb{N}$ 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 asked for the $2^d$ numbers in the above set to be distinct — but it's more natural to omit that condition for today's topic.)

**Proposition 10.3** (Homework 2, Problem 1)

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 $x_1, \dots, x_k, y \in \{1, \dots, N\}$ of the same color such that $x_1, \dots, x_k$ form a $k$-term arithmetic progression with common difference $y$, i.e.,

$$x_2 - x_1 = x_3 - x_2 = \cdots = x_k - x_{k-1} = y.$$

This is a strengthening of van der Waerden's theorem, since in particular it means $x_1, \dots, x_k$ is a monochromatic arithmetic progression (which would be van der Waerden's theorem); and it gives the additional guarantee that we can also make the common difference $y$ have the same color.

These three statements have a common pattern — we fix some number of colors $t$ and some system of equations (which are linear and homogeneous), and we say there's some large enough $N$ such that for every coloring of $\{1, \dots, N\}$ with our $t$ colors, we can solve this system of equations monochromatically.

**Remark 10.4.** This is clear in Theorem 3.1 and Proposition 10.3. For Proposition 10.2, we can introduce $2^d$ additional variables corresponding to each of the sums, and enforce the equations $x_{00\dots0} = x$, $x_{10\dots0} = x + y_1$, $x_{010\dots0} = x + y_2$, $x_{110\dots0} = x + y_1 + y_2$, and so on. The statement that this system has a monochromatic solution is a bit stronger than Proposition 10.2 — it also tells us that $x$ and $y_1$, $\dots$, $y_d$ have the same color as this affine cube — but it's still true.

To set up some notation, in each of these statements, we're given a homogeneous system of equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1k}x_k &= 0 \\ a_{21}x_1 + \cdots + a_{2k}x_k &= 0 \\ \vdots \quad & \quad \vdots \quad \quad \vdots \quad \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mk}x_k &= 0. \end{aligned}$$

(*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.) And we're looking for a solution $(x_1, \dots, x_n) \in \mathbb{N}^k$ with $x_1, \dots, x_k$ of the same color. We can write this 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, $x$ is a $k \times 1$ vector, and 0 denotes the $m \times 1$ zero vector), where $x = (x_1, \dots, x_k)^\intercal \in \mathbb{N}^k$ is such that $x_1, \dots, x_k$ are all the same color.

(So far we haven't really said anything; we've just compressed all three statements into this more abstract framework.)

**Remark 10.5.** All the equations we're considering are homogeneous, meaning that we have 0 on the right-hand side (rather than some other constant). The results start getting weird if we put other constants instead. For example, if we look at 3-term arithmetic progressions, we have the equation $x - 2y + z = 0$. But if we consider $x - 2y + z = 1$ instead, it's easy to define a coloring with no solutions — we can take a parity coloring (where odds are red and evens are blue). Similarly, if we take $x - 2y + z = 2$, then we can take a mod 3 coloring, and so on. So when we put constants on the right-hand side, we start running into weird issues with modular constraints. We don't run into these issues with 0's on the right-hand side, so in that sense it's more natural to just have 0's.

**Question 10.6.** For which systems of equations $Ax = 0$ does a statement of this form (as in Theorem 3.1 or Propositions 10.2 or 10.3) hold?

(The answer isn't all of them.)

### §10.1.1 Finite vs. infinite colorings and partition regularity

Before we start discussing this question, we'll make an observation that reduces the complexity of our theorem statements.

So far, our theorem statements (for Theorem 3.1 and Propositions 10.2 and 10.3) have been written in terms of being given $t$ and saying there exists $N$ such that the statement holds on $\{1, \dots, N\}$. But the role of $N$ is kind of obsolete, in the following sense — we can also ask the same question for colorings of *all* positive integers (so instead of stopping at $N$, we consider a coloring of all of $\mathbb{N}$). If we know there exists some $N$ such that the statement holds if we stop at $N$, then it also holds for colorings of all of $\mathbb{N}$. The converse direction is less obvious, but it's also true. This means the 'there exists $N$' part of our theorem statements is not really necessary. (One could also consider quantitative aspects — what's the best possible $N$? — but if we just qualitatively care about finding *some* solution, then $N$ is unnecessary.) To state this more precisely:

**Lemma 10.7**

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 $N$ such that for every coloring of the elements of $\{1, \dots, N\}$ with $t$ colors, there is a solution to $Ax = 0$ for $x = (x_1, \dots, x_k)^\intercal \in \{1, \dots, N\}^k$ such that $x_1, \dots, x_k$ have the same color.

(ii) For every coloring of $\mathbb{N}$ with $t$ colors, there is a solution to $Ax = 0$ for $x = (x_1, \dots, x_k)^\intercal \in \mathbb{N}^k$ 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 a given coloring of $\mathbb{N}$. The fact that (ii) implies (i) is not tautological — this implication says that if we know the statement we want holds for colorings of all of $\mathbb{N}$, then we only need to look up to a certain point $N$. 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. But it is true, and we'll prove it on the homework. (It depends on the axiom of choice, but so does a large fraction of mathematics.)

This means that when we study Question 10.6, we might as well look at the simpler statement of coloring *all* positive integers. (But it still doesn't tell us for which equations $Ax = 0$ we actually succeed — it just tells us that the task where we only color the integers up to some point is equivalent to the task where we color all of them.)

**Definition 10.8.** We say the system $Ax = 0$ is partition regular if (i) and (ii) hold for every $t$ (i.e., for any coloring of $\mathbb{N}$ with finitely many colors, we can find a monochromatic solution to $Ax = 0$).

**Remark 10.9.** There are examples of systems which satisfy (i) and (ii) for $t = 2$ colors, but not for all $t$. One such example is the single 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 every number of colors (so this equation is not partition regular).

**Example 10.10**

- Schur's theorem (Theorem 3.1) tells us that the equation $z = x + y$ is partition regular.
- Proposition 10.2 tells us that the homogeneous system corresponding to an affine cube (as described in Remark 10.4) is partition regular.
- Proposition 10.3 tells us that the system $x_2 - x_1 = \cdots = x_k - x_{k-1} = y$ is partition regular.

**Remark 10.11.** The reason for the name 'partition regular' is that you can think of the coloring as a partition of $\mathbb{N}$ into finitely many color classes, and then we're trying to solve the given equation in one of the sets of this partition.

**Remark 10.12.** It's important that $\mathbb{N}$ starts with 1 and not 0 — otherwise every system would be trivially solvable (we could take $x = 0$).

## §10.2 Rado's theorem

This still doesn't answer Question 10.6 — so far, we've just given a name to the systems which work. But a beautiful theorem of Rado *exactly* characterizes these systems (i.e., it answers which homogeneous systems of equations are partition regular).

**Theorem 10.13** (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 $\{1, \dots, k\} = I_0 \cup \cdots \cup I_\ell$ (for some $\ell$) such that we have

$$\sum_{i\in I_0} a^{(i)} = 0, \tag{10.1}$$

and for every $j = 1, \dots, \ell$, we have

$$\sum_{i\in I_j} a^{(i)} \in \operatorname{Span}_{\mathbb{Q}}\{a^{(h)} \mid h \in I_0 \cup \cdots \cup I_{j-1}\}. \tag{10.2}$$

**Remark 10.14.** This is from Rado's PhD thesis, which had the name *Studies of Combinatorics*. His supervisor was Schur, and this result generalizes Schur's theorem from seventeen years ago.

**Remark 10.15.** The partition $I_0 \cup \cdots \cup I_\ell$ in Rado's theorem does not have anything to do with the name 'partition regular,' which makes the name a bit unfortunate.

In words, the first condition in Theorem 10.13 tells us that the sum of the columns in $I_0$ (the first set of the

partition) is 0 (note that both sides of this condition are vectors in $\mathbb{Z}^m$); 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 from the previous sets. The first condition is essentially just the second condition for $j = 0$, since $\text{Span}(\emptyset) = \{0\}$. (We only wrote it separately for clarity, because the span of the empty set might be a bit confusing.)

To get a bit more familiar with what this condition means, we'll try to understand it for our examples.

**Example 10.16**

Consider the equation $x + y = z$ (from Schur's theorem); this corresponds to the matrix

$$A = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}.$$

Then we can take $I_0 = \{2, 3\}$ and $I_1 = \{1\}$. The condition on $I_0$ is satisfied because $1 + (-1) = 0$, and the condition on $I_1$ is satisfied because $1 \in \text{Span}\{1, -1\}$.

**Example 10.17**

In Proposition 10.3, we have the system $x_2 - x_1 = \cdots = x_k - x_{k-1} = y$, which we can write as the matrix

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 & \cdots & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 1 & 1 \end{bmatrix}$$

(where the $i$th row represents the equation $x_i - x_{i+1} + y = 0$).

Then we can take $I_0$ to consist of all the columns except the last one — each row has one 1 and one $-1$ among these columns, so these columns sum to 0. And these columns span everything, so we can take $I_1$ to consist of the last column.

**Example 10.18**

In Proposition 10.2 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 $x$, $y_1$, $y_2$, $z_{00}$, $z_{01}$, $z_{10}$, $z_{11}$ in that order, we get the matrix

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}.$$

Then we can take $I_0 = \{1, 4, 5, 6, 7\}$. And columns 4, 5, 6, 7 span everything, so we can take $I_1 = \{2, 3\}$.

**Remark 10.19.** In these examples, we only needed $I_0$ and $I_1$. But there are examples where you need to take more sets (i.e., it's not always possible to move everything to $I_1$). In general, taking more sets gives you more power as you go along: For $I_1$, we have the condition that the sum of columns in $I_1$ needs to be in the span of the columns in $I_0$. But then when we go to $I_2$, for the corresponding span we have both $I_0$ and $I_1$ at our disposal (the sum of columns in $I_2$ needs to be in the span of the columns in $I_0 \cup I_1$). So we gain more power as we go on, and that's sometimes necessary — on the homework, there's a concrete example where a partition exists, but we can't find one with just $I_0$ and $I_1$.

Now we'll prove Rado's theorem (we probably won't finish today, so we'll do the proof over today and Thursday). Rado's theorem is an 'if and only if' statement, so there are two directions. We'll first prove the forwards direction — that if the system is partition-regular, then this funny condition holds. Then we'll start talking about the reverse direction, which we'll finish on Thursday.

## §10.3 Proof of necessity

We'll first consider the forwards direction of Rado's theorem: Suppose that our system $Ax = 0$ is partition regular; we want to show that it has the property described in Theorem 10.13.

For this, we'll need the following lemma (we'll soon see how it relates to our problem).

**Lemma 10.20**

Let $v_1, \dots, v_\ell \in \mathbb{Z}^m$ and $v \in \mathbb{Z}^m$ be vectors such that $v \notin \operatorname{Span}_{\mathbb{Q}}\{v_1, \dots, v_\ell\}$. Then there are only finitely many primes $p$ such that the following statement holds: We have

$$p^z v \equiv \lambda_1 v_1 + \cdots + \lambda_\ell v_\ell \pmod{p^{z+1}} \tag{10.3}$$

for some integer $z \geq 0$ and integer coefficients $\lambda_1, \dots, \lambda_\ell$.

(The congruence means coordinate-wise congruence — both sides are 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 of both vectors have the same remainder mod $p^{z+1}$, and so on.)

In other words, this lemma says that whenever we know a relation of the form (10.3) holds (for any $z$ and $\lambda_i$), this constrains $p$ to a finite list of primes (which does not depend on $z$ or the $\lambda_i$).

*Proof.* We're given that $v \notin \operatorname{Span}_{\mathbb{Q}}\{v_1, \dots, v_\ell\}$, so we can find a vector which witnesses this, in the sense that it's orthogonal to this 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 can assume that $w \in \mathbb{Z}^m$.

Now we want to show that whenever a relationship of the form (10.3) holds, the prime $p$ must be 'special.' Assume that for some prime $p$, (10.3) holds for some $z, \lambda_i \in \mathbb{Z}$ with $z \geq 0$. Then we can take the dot product of both sides with $w$ (this is why it's important that $w$ is in $\mathbb{Z}^m$, because otherwise the calculation with mods could be weird). This gives

$$p^z(w \cdot v) \equiv \lambda_1(w \cdot v_1) + \cdots + \lambda_\ell(w \cdot v_\ell) \pmod{p^{z+1}}.$$

But the right-hand side is 0, since we assumed $w \cdot v_i = 0$ for all $i$; so we get

$$p^z(w \cdot v) \equiv 0 \pmod{p^{z+1}}.$$

This means $p^{z+1} \mid p^z(w \cdot v)$, or equivalently

$$p \mid w \cdot v.$$

But $w$ is fixed (we chose it only depending on $v$ and the $v_i$, so it doesn't depend on $z$ or the $\lambda_i$), 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 $p$ for which a relation of the form (10.3) can hold. □

Now we'll return to Rado's theorem and prove the first direction, that any partition regular system $Ax = 0$ must have the property described by Rado's theorem.

*Proof of necessity in Theorem 10.13.* Let $A \in \mathbb{Z}^{m \times k}$ be an integer matrix with columns $a^{(1)}, \dots, a^{(k)} \in \mathbb{Z}^m$ such that the system $Ax = 0$ is partition regular, meaning that for every coloring of $\mathbb{N}$ with finitely many colors, we can find a monochromatic solution to this system. We want to show that we can then partition its column set as described by the theorem.

We'll use Lemma 10.20 in the following way: We'll use the fact that $Ax = 0$ is partition regular by considering a *particular* coloring; partition regularity means that we must be able to find a monochromatic solution in this coloring. This particular coloring will depend on a well-chosen prime $p$, and Lemma 10.20 will ensure that we can choose a prime with the properties we want (as captured by the following claim).

**Claim 10.21 —** We can find a prime $p$ satisfying the following condition: For any two subsets $I, J \subseteq \{1, \dots, k\}$ such that $\sum_{j \in J} a^{(j)} \notin \operatorname{Span}_{\mathbb{Q}}\{a^{(i)} \mid i \in I\}$, we do *not* have

$$p^z \sum_{j \in J} a^{(j)} \equiv \sum_{i \in I} \lambda_i a^{(i)} \pmod{p^{z+1}} \tag{10.4}$$

for any integer $z \geq 0$ and integers $\lambda_i$ for $i \in I$.

In other words, whenever we have two subsets $I$ and $J$ (not necessarily nonempty or disjoint) such that the sum of the columns $a^{(j)}$ in $J$ isn't in the rational span of the columns $a^{(i)}$ in $I$, we're *not* able to write a congruence where $p^z$ times this sum of the $a^{(j)}$ is congruent to some integer linear combination of the $a^{(i)}$ mod $p^{z+1}$. (Of course, it could be the case that for some choices of $I$ and $J$ this sum *is* in the span, but we don't have a condition from those $I$ and $J$.)

*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 for $z$ and the $\lambda_i$.) But this is implied by Lemma 10.20 — once we've fixed $I$ and $J$, if we take $v = \sum_{j \in J} a^{(j)}$ in Lemma 10.20, it tells us precisely that there are only finitely many primes $p$ excluded by $I$ and $J$.

So by Lemma 10.20, any *fixed* choice of $I, J \subseteq \{1, \dots, k\}$ only excludes finitely many primes. 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). □

We now have a prime $p$ with the (maybe weird-looking) property in Claim 10.21. 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 $r(x)$ 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 defined the coloring explicitly in this way because we'll need these quantities in a future calculation.)

**Remark 10.22.** Note that it's important we excluded 0 (i.e., $0 \notin \mathbb{N}$), since 0 can't be written 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 $x_1$, …, $x_k$ 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 $z(x_1)$, …, $z(x_k)$. (Some of these values may be equal and some may not be, but we can certainly make a finite list of all the values that appear among them; and we sort this list 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're using the same index set $\{0, \dots, \ell\}$): For each $j = 0$, …, $\ell$, we let

$$I_j = \{h \in \{1, \dots, k\} \mid z(x_h) = z_j\}.$$

In words, for each $h \in \{1, \dots, k\}$, its $z(x_h)$ is equal to exactly one of the values $z_j$, and we put $h$ into the corresponding set $I_j$. This gives a partition $\{1, \dots, k\} = I_0 \cup \cdots \cup I_\ell$ (and the subsets $I_j$ are all nonempty, since we only chose values $z_j$ that appear among $z(x_1)$, …, $z(x_k)$).

We now need to check that this partition satisfies the conditions in Theorem 10.13. The first condition (10.1) is really the same as the second condition (10.2) for $j = 0$, so to avoid repeating the same argument twice, we'll just check (10.2) for all $j = 0$, …, $\ell$.

In order to check the condition (10.2) for $j$, first recall that $x$ is a solution to $Ax = 0$ (both sides are vectors in $\mathbb{Z}^m$). We want to relate this to the column vectors, so we'll expand out the system as

$$x_1 a^{(1)} + \cdots + x_k a^{(k)} = 0. \tag{10.5}$$

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 $r(x_h) = r$ for all $h$. And whenever $h \in I_j$, we have $z(x_h) = z_j$ (this is how we defined $I_j$), so

$$x_h = p^{z_j}(p \cdot m(x_h) + r).$$

Now fix $j \in \{0, \dots, \ell\}$, and consider (10.5) mod $p^{z_j+1}$. Then this 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 $h$ in all the index sets coming after $I_j$), as their corresponding values of $z_{j'}$ are at least $z_j + 1$. So all the future terms in (10.5) disappear. Furthermore, for $h \in I_j$, we have

$$x_h = p^{z_j}(p \cdot m(x) + r) \equiv p^{z_j} \cdot r \pmod{p^{z_j+1}}$$

(i.e., the first term disappears). So now taking (10.5) 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 the remaining summands with $h \in I_{j+1} \cup \cdots \cup I_\ell$ all disappear).

Now we're almost done — we almost have a condition of the form (10.4) (as in Claim 10.21), and the only slightly annoying thing is $r$. But we can remove $r$ by multiplying by its inverse. Explicitly, since $r \in \{1, \dots, p-1\}$, there is some $s \in \{1, \dots, p-1\}$ 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

$$\sum_{h \in I_0 \cup \dots \cup I_{j-1}} s x_h a^{(h)} + \sum_{h \in I_j} (mp+1) p^{z_j} a^{(h)} \equiv 0 \pmod{p^{z_j+1}},$$

and since $mp \cdot p^{z_j} \equiv 0 \pmod{p^{z_j+1}}$, we can omit this term and rewrite this equation as

$$p^{z_h} \sum_{h \in I_j} a^{(h)} = \sum_{h \in I_0 \cup \dots \cup I_{j-1}} (-s x_h) a^{(h)} \pmod{p^{z_j+1}}. \tag{10.6}$$

So now we have a statement of the form (10.4) — that $p^{z_j}$ times the sum of certain columns $a^{(h)}$ can be written as a linear combination of certain other columns $a^{(h)}$ mod $p^{z_j+1}$.

And we chose $p$ (using Claim 10.21) with the property that whenever we have sets $I$ and $J$ and integers $z$ and $\lambda_i$ for which the relationship

$$p^z \sum_{j \in J} a^{(j)} \equiv \sum_{i \in I} \lambda_i a^{(i)} \pmod{p^{z+1}}$$

holds, $\sum_{j \in J} a^{(j)}$ must be in the rational span of the $a^{(i)}$ (since we chose $p$ such that for any $I$ and $J$ where this sum is *not* in the span, there do *not* exist any $z$ and $\lambda_i$ for which such a relationship holds). Here (10.6) is such a relationship, so by our choice of $p$, we must have

$$\sum_{i \in I_j} a^{(i)} \in \operatorname{Span}_{\mathbb{Q}}\{a^{(h)} \mid h \in I_0 \cup \dots \cup I_{j-1}\}.$$

This is exactly the condition (10.2) that we wanted to prove. (In particular, applying this for $j = 0$, we get that $\sum_{i \in I_0} a^{(i)} \in \operatorname{Span}_{\mathbb{Q}}(\emptyset) = \{0\}$, which means $\sum_{i \in I_0} a^{(i)} = 0$; this gives the first condition (10.1) as well.) So we have produced a partition $I_0 \cup \dots \cup I_\ell$ with the desired property. □

## §10.4 Proof of sufficiency

We've now proven one direction of Rado's theorem — that if a system is partition regular, then the condition in Rado's theorem holds. We now need to prove the converse — that if this condition holds, then the system is partition regular. This direction is maybe more interesting — we've seen examples where we worked rather hard to show partition regularity for a particular system (e.g., Schur's theorem and Propositions 10.2 and 10.3), and this direction of Rado's theorem gives all those results at once for free.

Rado's original proof used van der Waerden's theorem. We'll see a different argument; this proof will also give stronger side results, which we'll comment on next class.

### §10.4.1 Deuber sets

The proof will rely on the notion of *Deuber sets*.

**Definition 10.23.** For integers $\ell, d, c > 0$, an $(\ell, d, c)$-Deuber set is a subset $M \subseteq \mathbb{N}$ such that there exist $y_0, \dots, y_\ell \in \mathbb{N}$ such that $M$ consists of precisely all integers of the form

$$c y_j + \lambda_{j+1} y_{j+1} + \lambda_{j+2} y_{j+2} + \dots + \lambda_\ell y_\ell$$

for some $j \in \{0, \dots, \ell\}$ and $\lambda_{j+1}, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]$.

So the notion of a Deuber set depends on three parameters $\ell$, $d$, and $c$ (which are positive integers): $\ell+1$ is the number of $y_i$ we have. Then $M$ should consist of all 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 linear combinations of $y_0, \dots, y_\ell$ of this form; if all these combinations are positive integers (this won't always be the case) and $M$ is precisely the set of all those integers, then we call $M$ a Deuber set.

In other words, $M \subseteq \mathbb{N}$ is an $(\ell, d, c)$-Deuber set if $M = M_{d,c}(y_0, \dots, y_\ell)$ for some $y_0, \dots, y_\ell \in \mathbb{N}$, where

$$M_{d,c}(y_0, \dots, y_\ell) = \{cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_\ell y_\ell \mid j \in \{0, \dots, \ell\},\ \lambda_i \in \mathbb{Z} \cap [-d, d]\}$$

for some $y_0, \dots, y_\ell \in \mathbb{N}$. (In other words, for every choice of $y_0, \dots, y_\ell$ such that all these linear combinations are positive, we get an $(\ell, d, c)$-Deuber set, which we denote by $M_{d,c}(y_0, \dots, y_\ell)$.)

These sets are nice because we can find all linear combinations of this specific shape inside them; and this will help us solve equations of the form $Ax = 0$.

**Remark 10.24.** There exist $(\ell, d, c)$-Deuber sets for all $\ell, d, c > 0$. This is not entirely obvious, because if you choose $y_0, \dots, y_\ell$ badly, these linear combinations might not all be positive integers. But we can ensure that they *are* all positive integers by taking $y_{\ell-1}$ to be way bigger than $y_\ell$, then taking $y_{\ell-2}$ to be way bigger than $y_{\ell-1}$, and so on — for example, we can take $y_\ell = 1$, $y_{\ell-1} = d+1$, $y_{\ell-2} = (d+1)^2$, and so on. (Note that $c \geq 1$, so this will work.)

**Remark 10.25.** If you want, by making slightly better choices you can ensure all these linear combinations are *distinct*. Specifically, if $y_0, \dots, y_n \in \mathbb{N}$ are such that $y_i > (2d+1)y_{i+1}$ for all $i = 0, \dots, \ell-1$, then $M_{d,c}(y_0, \dots, y_\ell)$ will indeed be a Deuber set (meaning that it's a subset of $\mathbb{N}$, i.e., all these linear combinations are positive), 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; and it's not really relevant unless you care about finding solutions with *distinct* variables.

In fact, if $M_{(2d+1)c,c}(y_0, \dots, y_\ell)$ is a Deuber set (i.e., if it's a subset of $\mathbb{N}$), then we automatically get the condition $y_i > (2d+1)y_{i+1}$ for all $i$. So by forcing the value of $d$ to be larger, we can obtain additional properties on the $y_i$ (such as this distinctness condition).

### §10.4.2 Overview

There isn't enough time to finish the proof in the remaining seven minutes of class, but we'll now see a roadmap of how the proof works and how Deuber sets come in. The proof will have two pieces.

**Lemma 10.26**

Suppose that $A \subseteq \mathbb{Z}^{m \times k}$ is a matrix whose columns satisfy the condition in Rado's theorem. Then there exist integers $\ell, d, c > 0$ such that every $(\ell, d, c)$-Deuber set $M \subseteq \mathbb{N}$ contains a solution to $Ax = 0$ with $x = (x_1, \dots, x_k)^\intercal \in M^k$.

So we comsider a fixed system $Ax = 0$ satisfying the condition in Rado's theorem, and our goal in the end is to show that it's partition regular. This first lemma tells us that whenever we have a system satisfying this condition, we can find a solution to the system in *every* $(\ell, d, c)$-Deuber set, where $\ell$, $d$, and $c$ are chosen appropriately (depending on $A$).

This is nice because it already gives us some indication of why $(\ell, d, c)$-Deuber sets are useful — we can solve our system $Ax = 0$ in them. The proof of Lemma 10.26 isn't that deep — the definition of these Deuber sets means that they contain lots of linear combinations, and we'll choose particular linear combinations for $x_1, \dots, x_k$ such that our equations get solved.

Once we prove this, 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 $(\ell, d, c)$-Deuber set in any coloring.

**Theorem 10.27** (Deuber 1973)

Let $\ell, d, c, t > 0$ be integers. Then for every coloring of $\mathbb{N}$ with $t$ colors, we can find a monochromatic $(\ell, d, c)$-Deuber set.

Then putting these two pieces together will give the desired implication — if our system $Ax = 0$ satisfies the condition in Rado's theorem, then Lemma 10.26 lets us find fixed $\ell$, $d$, and $c$ such that every $(\ell, d, c)$-Deuber set contains a solution. Then applying Theorem 10.27 to these $\ell$, $d$, and $c$, we know that for every coloring we can find a monochromatic $(\ell, d, c)$-Deuber set, and this Deuber set will contain a solution to our system.



### §10.4.3 Proof of Lemma 10.26: Finding solutions in Deuber sets

Our remaining tasks are now to prove Lemma 10.26 and Theorem 10.27. We'll start by proving Lemma 10.26 (which is the easier part).

*Proof of Lemma 10.26.* By the assumption on $A$ (that it satisfies the condition in Rado's theorem), there is a partition $\{1, \dots, k\} = I_0 \cup \cdots \cup I_\ell$ for nonempty sets $I_j$ satisfying

$$\sum_{i \in I_j} a^{(i)} \in \operatorname{Span}_{\mathbb{Q}}\{a^{(h)} \mid h \in I_0 \cup \cdots \cup I_{j-1}\}.$$

In words, this says that we can write each $\sum_{i \in I_j} a^{(i)}$ as a linear combination of the previous vectors $a^{(h)}$ with rational coefficients. First, rational coefficients are a bit annoying (since we're working with integers), so we'll clear denominators: By taking a common multiple of *all* denominators that occur (over all values of $j$ and all coefficients in these linear combinations), we can find an integer $c > 0$ such that for all $j = 1$, $\dots$, $\ell$, we can write

$$c \sum_{i \in I_j} a^{(i)} = \sum_{h \in I_0 \cup \cdots \cup I_{j-1}} \lambda_{j,h} a^{(h)}, \tag{10.7}$$

where the coefficients $\lambda_{j,h}$ are now *integers*.

Our goal was to find $\ell$, $d$, and $c$ so that every $(\ell, d, c)$-Deuber set contains a solution to our system. We'll use this value of $c$ (this common multiple of all the denominators), and we'll also use the same value of $\ell$ as from the partition; so it just remains to define $d$. We let

$$d = \max\{|\lambda_{j,h}| \mid j \in \{1, \dots, \ell\},\ h \in I_0 \cup \cdots \cup I_{j-1}\}.$$

In words, we look at all the coefficients $\lambda_{j,h}$ appearing in our linear combinations (10.7), and we take $d$ to be (weakly) bigger than all of them.

We've now defined $\ell$, $d$, and $c$, so our goal is to show that every $(\ell, d, c)$-Deuber set contains a solution to our system. Let $M$ be an $(\ell, d, c)$-Deuber set, so that

$$M = \{cy_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_\ell y_\ell \mid j \in \{0, \dots, \ell\},\ \lambda_i \in \mathbb{Z} \cap [-d, d]\} \tag{10.8}$$

(for some $y_0, \dots, y_\ell \in \mathbb{N}$). We want to show that $M$ contains a solution to $Ax = 0$. We'll do this by writing down a solution explicitly — we want to define each of $x_1$, $\dots$, $x_k$ to be some linear combination of the form

$cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_\ell y_\ell$. The way we define $x_i$ will depend on which set in the partition it corresponds to: For each $j = 0, \dots, \ell$ and every $i \in I_j$, we define

$$x_i = cy_j - \lambda_{j+1,i}y_{j+1} - \cdots - \lambda_{\ell,i}y_\ell.$$

First note that the coefficients $\lambda_{j,h}$ are defined whenever $h \in I_{j'}$ for some $j' < j$ (i.e., $h$ belongs to a set with index strictly smaller than $j$). Here we have $i \in I_j$, and $j$ is strictly smaller than all of $j+1, \dots, \ell$ (the first indices of the above coefficients $\lambda_{-,i}$), so these coefficients are all well-defined. We can then see that $x_i$ is in $M$, meaning that this is a linear combination of the necessary form in (10.8) — each of the coefficients $\lambda_{-,i}$ is at most $d$ in absolute value (by how we chose $d$).

So $x_1, \dots, x_k$ are valid elements of $M$. Now we need to check that $Ax = 0$, i.e., that $\sum_{i=1}^k x_i a^{(i)} = 0$. This is a direct calculation — first, plugging in the definitions of the $x_i$, we have

$$\sum_{i=1}^k x_i a^{(i)} = \sum_{j=0}^{\ell} \sum_{i \in I_j} x_i a^{(i)} = \sum_{j=0}^{\ell} \sum_{i \in I_j} (cy_j - \lambda_{j+1,i}y_{j+1} - \cdots - \lambda_{\ell,i}y_\ell) a^{(i)}.$$

We'll reorder this 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 (the $j$th term), but it'll also appear with other coefficients from the other terms, and we want to group them. The reason we want to do this is that we know nothing about the $y_j$ (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 each $y_j$.

So we'll rewrite this as

$$\sum_{i=1}^k x_i a^{(i)} = \sum_{j=0}^{\ell} \left( \sum_{i \in I_j} cy_j a^{(i)} - \sum_{i \in I_0 \cup \cdots \cup I_{j-1}} \lambda_{j,i} y_j a^{(i)} \right) = \sum_{j=0}^{\ell} y_j \left( c \sum_{i \in I_j} a^{(i)} - \sum_{i \in I_0 \cup \cdots \cup I_{j-1}} \lambda_{j,i} a^{(i)} \right)$$

(the point is that $y_j$ appears for every $i$ in a set $I_{j'}$ with $j' < j$). And the coefficient of each $y_j$ is 0 — this is exactly our definition of $c$ and the $\lambda_{j,h}$ from (10.7). (Of course, when we were defining $x_1, \dots, x_k$, we chose the coefficients in order to make this step work out.) So this sum is indeed 0.

So we've shown that 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 10.28.** This 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 just using linear combinations of the shape in (10.8), you can form a solution to the system. The much harder part is to prove Theorem 10.27 (that's why the first part is called a lemma and the second a theorem).

**Remark 10.29.** We'll now comment on finding *distinct* variable solutions: In the definition of partition regularity, we want to find a solution with all variables of the same color; but we can also 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 — 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 the condition in Rado's theorem *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 it explicitly, but here's a sketch of how you'd do so. All you need to do is prove a slightly different version of Lemma 10.26 where we assume both these conditions, and then prove that in a Deuber set with appropriately chosen parameters, we can find a solution with distinct $x_1, \dots, x_k$. It's not hard to modify our proof to get this additional version: We define $x_1, \dots,$

$x_k$ by taking the same linear combinations of $y_0, \dots, y_\ell$ as in the above proof. Then we take $\ell$ to be one bigger, i.e., we take one more variable $y_{\ell+1}$. By assumption $Ax = 0$ has some fixed solution where all variables are distinct; and we'll use the integers in this solution to define the coefficients of $y_{\ell+1}$. Then when we perform the same calculation as above, again everything will cancel (because the coefficients on $y_{\ell+1}$ are from that solution).

So far, we've guaranteed that the *linear combinations* used to define the $x_i$ are all distinct, but we still need to guarantee that the $x_i$ themselves are distinct. To do this, we can use Remark 10.25 — by making $d$ artificially larger (replacing it with $(2d+1)c$), we can ensure that all these different linear combinations have distinct values.

### §10.4.4 A strengthening of Theorem 10.27

We've now proved Lemma 10.26, so all that remains (in order to prove Rado's theorem) is to prove Theorem 10.27. Before we prove the theorem, we'll first generalize it — in fact, Deuber actually proved something stronger than Rado's theorem.

**Remark 10.30.** Deuber gave this new approach to Rado's theorem in his PhD thesis; this is funny because Rado himself proved the theorem in his PhD thesis.

**Theorem 10.31** (Deuber 1973)

Let $\ell$, $d$, $c$, and $t$ be positive integers. Then there are positive integers $\ell^*$, $d^*$, and $c^*$ such that for every $(\ell^*, d^*, c^*)$-Deuber set $M^* \subseteq \mathbb{N}$ and every coloring of $M^*$ with $t$ colors, there exists a monochromatic $(\ell, d, c)$-Deuber set $M \subseteq M^*$.

Theorem 10.27 told us that we can find a monochromatic $(\ell, d, c)$-Deuber set if we color *all* the positive integers. This new theorem says that it's not necessary to look at *all* the positive integers — it suffices to just color an $(\ell^*, d^*, c^*)$-Deuber set (for some suitable choice of $\ell^*$, $d^*$, and $c^*$ depending on $\ell$, $d$, $c$, and $t$), i.e., for every such Deuber set $M^*$, coloring just $M^*$ with $t$ colors is enough to find a monochromatic $(\ell, d, c)$-Deuber set. This is stronger than Theorem 10.27, since we can find a $(\ell^*, d^*, c^*)$-Deuber set $M^* \subseteq \mathbb{N}$, and then take our coloring of $\mathbb{N}$ and only look at it on $M^*$.

Before we prove this new theorem, we'll talk about some additional consequences it has — in particular, it answers a question that Rado left open. To motivate this question, we'll state another result, which follows from Rado's theorem.

**Theorem 10.32**

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 $Ax = 0$.

The definition of partition regularity says that for each *individual* partition regular system, we can find a solution to that system in some color class. But this theorem says that we can choose one color class that *simultaneously* contains a solution to *all* these systems. This may be surprising at first, but it can actually be deduced by Rado's theorem. (To show this, suppose for contradiction that for every color class, there is a particular system that is not solvable in that color class (and which is partition regular). Now we have $t$ different systems, and we can combine all of them into one gigantic system (where we use disjoint variables for each, so that these different pieces don't interact). This will still be a linear system, and it will still be partition regular, so we must be able to solve this gigantic system in some color class. But this is a contradiction, since by assumption we can't solve the piece of it corresponding to that color class.)

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. Then we can ask the following question:

**Question 10.33.** Suppose we start with an amazing set (which solves every partition regular 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 Theorem 10.32. Rado conjectured but couldn't prove that the answer is yes; and Deuber proved this in his PhD thesis using Theorem 10.31.

**Theorem 10.34** (Deuber 1973)

Let $M \subseteq \mathbb{N}$ be a set containing a solution to every partition regular system $Ax = 0$. Then for every coloring of $M$ with finitely many colors, there exists a color class that also contains a solution to every partition regular system.

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 (i.e., color classes), then one of those pieces is still amazing. This is quite a cool statement.

To see how this follows from Theorem 10.31, the key idea is the following claim (we won't write down all the details for the sake of time).

**Claim 10.35 —** There must be a color class containing $(z, z \cdot z!, z!)$-Deuber sets for infinitely many $z$.

*Proof sketch.* For any particular $z$, we can consider Deuber sets with these parameters, and take $\ell^*$, $d^*$, and $c^*$ as in Theorem 10.31. Now, $(\ell^*, d^*, c^*)$-Deuber sets can be described by some complicated system of equations. This system will be partition regular (e.g., this is implied by Theorem 10.27, which says that we can find $(\ell^*, d^*, c^*)$-Deuber sets in any coloring of $\mathbb{N}$), so there will be a solution to it in $M$. Then that $(\ell^*, d^*, c^*)$-Deuber set in $M$ has been colored with $t$ colors, so by Theorem 10.31, one color class must contain a $(z, z \cdot z!, z!)$-Deuber set.

Finally, there are only finitely many color classes, so there must be some color class containing such a Deuber set for infinitely many $z$. □

Then the color class given by Claim 10.35 will be the one we want (for Theorem 10.34) — this color class will contain an $(\ell, d, c)$-Deuber set for all $\ell$, $d$, and $c$, since if $z > \ell, d, c$ then every $(z, z \cdot z!, z!)$-Deuber set will contain an $(\ell, d, c)$-Deuber set. So now we have a color class containing $(\ell, d, c)$-Deuber sets for all $\ell$, $d$, and $c$, which (by Lemma 10.26) means that it contains a solution to all partition regular systems.

**Remark 10.36.** The reason we're using the somewhat weird-looking parameters $z$, $z \cdot z!$, and $z!$ here is that $(\ell, d, c)$-Deuber sets don't behave as monotonously as you might hope for — it's not true in general that just making $\ell$, $d$, and $c$ bigger gives you containment. If you take a $(\ell, d, c)$-Deuber set, it *is* true that it'll contain an $(\ell', d', c)$-Deuber set for all $\ell' \le \ell$ and $d' \le d$; but for $c$, the monotonicity behavior is weird. (Reducing $\ell$ just means omitting some variables, which is fine; and reducing $d$ just means allowing fewer possibilities for the coefficients $\lambda_i$, 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 that's why we have $z!$. If $c = c'f$, then we need to multiply all the $y_i$ by $f$, which affects the threshold for $d$; this is the reason why we have another $z!$ term in $d$ as well. The formal statement we're using should be that every $(\ell, df, c)$-Deuber set should contain a $(\ell, d, c/f)$-Deuber set for $f \mid c$.

### §10.4.5 Proof of Theorem 10.31 — finding monochromatic Deuber sets

We'll now discuss the proof of Theorem 10.31. At this point, we can forget basically everything about linear systems, partition regularity, and so on — we just need to think about Deuber sets and colorings.

The terms in a Deuber set naturally fall into 'layers,' with one layer for each $j$ (consisting of the linear combinations that start with $y_j$). So as a first step towards Theorem 10.31, we'll prove something weaker — rather than finding an $(\ell, d, c)$-Deuber set where the *entire* set is monochromatic (i.e., *everything* has the same color), we'll find an $(\ell, d, c)$-Deuber set where each *layer* is monochromatic.

We'll do this inductively — we'll first take $j = 0$ and make its layer monochromatic; then we'll take $j = 1$ and make its layer monochromatic; and so on. So we'll show the following statement, by induction on $h$.

**Proposition 10.37**

Let $\ell$, $d$, $c$, and $t$ be positive integers, and let $h \in \{0, \dots, \ell\}$. Then there exist positive integers $\ell^*$, $d^*$, and $c^*$ (depending on $\ell$, $d$, $c$, $t$, and $h$) such that for every $(\ell^*, d^*, c^*)$-Deuber set $M^*$ and every coloring of $M^*$ with $t$ colors, there exists an $(\ell, d, c)$-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 which does have the property that every *layer* (with each layer corresponding to a value of $j$) is monochromatic. The role of $h$ is how far we can go; of course this statement becomes stronger as we increase $h$, and we'll eventually apply it for $h = \ell$. The point of writing the statement for general $h$ is that we'll prove it using induction on $h$ (but in the end, we'll only apply it with $h = \ell$).

Once we have Proposition 10.37 for $h = \ell$, it's still not exactly what we wanted (for Theorem 10.31) — we wanted an $(\ell, d, c)$-Deuber set which is altogether monochromatic, not just one where every layer is monochromatic. But deducing Theorem 10.31 from Proposition 10.37 is a pretty easy pigeonhole argument, which we'll see first (proving Proposition 10.37 is the main work of proving Theorem 10.31).

*Proof of Theorem 10.31 assuming Proposition 10.37.* The idea is that applying Proposition 10.37 with $h = \ell$, we can guarantee that each layer of our Deuber set is monochromatic, but these layers may have different colors. We only want to keep layers with the same color. By pigeonhole, at least a $\frac{1}{t}$-fraction of these layers will have the same color. So if we apply Proposition 10.37 with $\ell t$ instead of $\ell$, then we'll have $\ell t + 1$ layers, and we'll be able to find $\ell + 1$ with the same color; and then we'll define our $(\ell, d, c)$-Deuber set by only using the variables $y_j$ corresponding to these layers.



Now we'll fill in the details. First, in the statement of Theorem 10.31, we're given certain $\ell$, $d$, $c$, and $t$. We then apply Proposition 10.37 with $\ell t$ (in place of $\ell$), $d$, $c$, $t$, and $h = \ell t$. Then we obtain certain $\ell^*$, $d^*$, and $c^*$ such that Proposition 10.37 holds — i.e., for every $(\ell^*, d^*, c^*)$-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

$$\{cy_j + \lambda_{j+1}y_{j+1} + \cdots + \lambda_{\ell t}y_{\ell t} \mid \lambda_{j+1}, \dots, \lambda_{\ell t} \in \mathbb{Z} \cap [-d, d]\} \tag{10.9}$$

is monochromatic (so this set is monochromatic for each individual $j$, but the colors may be different for different values of $j$). Now we'll use pigeonhole — there are $\ell t + 1$ possible values of $j$ 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 obtain a monochromatic $(\ell, d, c)$-Deuber set by taking

$$M = M_{d,c}(y_{j_0}, \dots, y_{j_\ell}) \subseteq M_{d,c}(y_0, \dots, y_{\ell t}) \subseteq M^*.$$

More explicitly, we're starting with some Deuber set and restricting to a subset of the original variables; then the resulting set $M_{d,c}(y_{j_0}, \dots, y_{j_\ell})$ is a subset of the original set $M_{d,c}(y_0, \dots, y_{\ell t})$, since we can simply set the extra coefficients $\lambda_i$ to 0. (In particular, this means $M_{d,c}(y_{j_0}, \dots, y_{j_\ell})$ is a subset of $M^*$ and therefore $\mathbb{N}$, so it really is a Deuber set.) To see that it's monochromatic, we chose $j_0, \dots, j_\ell$ such that all of the layers corresponding to these indices in $M_{d,c}(y_0, \dots, y_{\ell t})$ (as in (10.9)) were monochromatic and had the same color. And all the linear combinations in the layers of our new Deuber set $M_{d,c}(y_{j_0}, \dots, y_{j_\ell})$ are subsets of these layers (again, we can set the coefficients of the variables we don't use to 0), so they're also monochromatic and have the same color. This means $M$ is indeed monochromatic. □

Now the only missing piece of the proof of Rado's theorem is Proposition 10.37, which we'll now prove. To prove it, we'll use the multidimensional version of the Hales–Jewett theorem, which we proved on the homework (we deduced it from the ordinary version of the Hales–Jewett theorem, Theorem 3.11).

**Theorem 10.38** (Multidimensional Hales–Jewett)

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 $S^n$.

This is a typical Ramsey-type statement — that 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 it's a bit annoying to write down.

**Definition 10.39.** An *$m$-parameter subset* of $S^n$ is a set 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 $n$-tuple obtained from $w$ by making these replacements. For fixed $w$, we call the set of all $|S|^m$ $n$-tuples obtained this way (over all possible replacements) a $m$-parameter subset.

As usual, $t$ is our number of colors, and we're coloring all $n$-tuples of elements of $S$. The ordinary Hales–Jewett theorem (Theorem 3.11) was about finding monochromatic combinatorial lines, which are precisely 1-parameter subsets — there, we just had one special symbol $*$, and we took a code $w$ (with some fixed entries and some $*$'s) and looked at all the ways to replace $*$ with an element of $S$. Here, we have multiple different variables $*_1, \dots, *_m$; it's called a 'multidimensional' version because rather than a monochromatic line, we're finding a monochromatic subset of greater dimension.

**Remark 10.40.** When we stated Theorem 3.11 (and the homework problem), we took the 'alphabet' $S$ to just be $\{1, \dots, q\}$ in place of $S$. But the changed notation will fit better into the framework we have here.

*Proof of Proposition 10.37.* First, we'll make an easy but useful observation. Last time (in Remark 10.36), we mentioned that $(\ell, d, c)$-Deuber sets are not entirely monotone because of issues with $c$. However, they *are* monotone with respect to $d$. This means Proposition 10.37 is monotone in $d$, in the sense that the bigger we make $d$, the harder the proposition gets — i.e., if we can show it for larger $d$, then we automatically get it for smaller $d$. (This is because if we take $M_{d,c}(y_0, \dots, y_\ell)$ and shrink $d$, then we only have fewer choices for the coefficients $\lambda_i$, so the new set is a subset of the original one; and having fewer linear combinations we need to check makes our job easier.) So we may assume that $d \geq c$.

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. But 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 want to apply the multidimensional Hales–Jewett theorem (Theorem 10.38). In the $h = 0$ case, the only condition we have is for $j = 0$ — we want to find an $(\ell, d, c)$-Deuber set such that the set

$$\{cy_0 + \lambda_1 y_1 + \cdots + \lambda_\ell y_\ell \mid \lambda_i \in \mathbb{Z} \cap [-d, d]\}$$

is monochromatic. Here, the first coefficient $c$ is fixed, while the $\lambda_i$ vary. So the 'alphabet' we want to use is $S = \mathbb{Z} \cap [-d, d]$. We'll use the same number of colors $t$ as the one we're given. Finally, we'll take $m = \ell$. (The reason for this is that looking at our statement for $j = 0$, there are $\ell$ different $\lambda_i$ which vary, so we want $\ell$ different parameters $*_i$ that we can vary.) This gives some $n \in \mathbb{N}$ such that the statement in the multidimensional Hales–Jewett theorem holds.

Now we take $\ell^* = n$. (This is because $\ell^*$ is basically the number of variables we have at our disposal in the *original* Deuber set $M^*$ — we'll only look at the layer where $j = 0$ in $M^*$, where we have $\ell^*$ different variables to vary (the coefficients $\lambda_1, \ldots, \lambda_{\ell^*}$); 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 $M^* = M_{d^*,c^*}(y_0, \ldots, y_n)$ to be an $(\ell^*, d^*, c^*)$-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 $S^n = (\mathbb{Z} \cap [-d, d])^n$ with $t$ colors, which should come from the given coloring of $M^*$. We do this as follows: We color each $(\lambda_1, \ldots, \lambda_n) \in S^n$ 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 $c^2$, $\lambda_1 c$, $\ldots$, $\lambda_n c$. This linear combination is in $M^*$ because $M^*$ was an $(n, d^*, c^*)$-Deuber set with $c^* = c^2$ and $d^* = cd$; so the first coefficient $c^2$ is correct, and since $|\lambda_i c| \le cd = d^*$, the other coefficients are also allowed. This means it has a well-defined color in $M^*$, and we use that color for $(\lambda_1, \ldots, \lambda_n)$ in our coloring of $S^n$.

> **Remark 10.41.** Right now, it's unclear why we have these additional factors of $c$; this will become clear later. But as a preview, the point is that when we apply the multidimensional Hales–Jewett theorem, we'll get a monochromatic $m$-parameter subset described by some word $w$, which will have a fixed part as well as parts that vary. We'll need the entire fixed part to go into the first term $cy_0$ in our new Deuber set $M$, which means we need to force it to be divisible by $c$.

This defines a coloring on $S^n$, so we'll now apply multidimensional Hales–Jewett to this coloring. We chose $n$ so that the statement of Hales–Jewett holds (for our specified parameters); this means there exists a monochromatic $\ell$-parameter subset of $S^n$ (recall that $m = \ell$) described by some word

$$w = (w(1), \ldots, w(n)) \in ((\mathbb{Z} \cap [-d, d]) \cup \{*_1, \ldots, *_\ell\})^n$$

such that each $*_i$ appears at least once. We don't *a priori* know in which order they appear; but the definition of a $m$-parameter subset is symmetric in the roles of $*_1, \ldots, *_m$, which means we can interchange them. 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 $w$ such that the $\ell$-parameter subset of $S^n$ it defines (using our replacement operation) is monochromatic. Our aim was to find an $(\ell, d, c)$-Deuber set $M = M_{d,c}(x_0, \ldots, x_\ell)$ such that its 0th level $\{cx_0 + \lambda_1 x_1 + \cdots + \lambda_\ell x_\ell\}$ is monochromatic. (We're using the letters $x_i$ instead of $y_i$ because we're already using $y_i$ for the variables generating the *original* Deuber set $M^*$.)

We'll define our variables $x_0, \dots, x_n$ as follows. First, 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 indices $i$ where $w(i)$ is not one of the special symbols $*_j$, and putting their contributions into $x_0$. This is why we needed the funny factors of $c$ — the point is that we want $x_0$ to basically be $\frac{1}{c}$ of the fixed term, which means we need the term to be divisible by $c$.)

Meanwhile, for each $j = 1, \dots, \ell$, we want to define $x_j$ by taking all the contributions of $*_j$ (so $x_1$ is associated with $*_1$, $x_2$ with $*_2$, and so on). So we take

$$x_j = \sum_{\substack{i \in \{1,\dots,n\} \\ w(i) = *_j}} cy_i.$$

(So we're looking at all indices $i$ where $w$ has a $*_j$ in that position, and taking the associated $y_i$.)

We'll now check that this works.

**Claim 10.42 —** We have $M_{d,c}(x_0, \dots, x_\ell) \subseteq M^* = M_{d^*,c^*}(y_0, \dots, y_n)$.

*Proof sketch.* This is basically true by definition. We need to look at all the terms appearing in our definition of an $(\ell, d, c)$-Deuber set, i.e., $cx_j + \lambda_{j+1} x_{j+1} + \dots + \lambda_\ell x_\ell$. If we write out each $x_i$ in terms of the original variables $y_i$, we can see that we'll get something of the correct form (i.e., one that appears in the definition of an $(n, d^*, c^*)$-Deuber set): The first $y_i$ that appears (which comes from the first term of $x_j$) will have a coefficient of $c^2$, since we have one factor of $c$ from the definition of $x_j$, and another from the fact that we're multiplying $x_j$ by $c$. Meanwhile, each $y_i$ coming from the $\lambda_{j'} x_{j'}$ terms will have a coefficient of $\lambda_{j'} c$ (note that $j'$ can't be 0 in these terms), which has absolute value at most $dc$; and the remaining $y_i$ coming from the $cx_j$ term will also have coefficients at most $dc$. □

Note that this in particular implies that $M_{d,c}(x_0, \dots, x_\ell)$ is a subset of $\mathbb{N}$, so it really is a Deuber set.

**Claim 10.43 —** 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.

*Proof.* This will come from the same calculation we sketched earlier — plugging in the definition of the $x_j$, we can rewrite each such linear combination as

$$cx_0 + \sum_{j \in \{1,\dots,\ell\}} \lambda_j x_j = c^2 y_0 + \sum_{\substack{i \in \{1,\dots,n\} \\ w(i) \in \mathbb{Z} \cap [-d,d]}} w(i) c y_i + \sum_{\substack{i \in \{1,\dots,n\} \\ j \in \{1,\dots,\ell\} \\ w(i) = *_j}} \lambda_j c y_i$$

(for the last term, we've combined the sum over $j$ (in the linear combination) with the sum over $i$ (in the definition of $x_j$) — whenever $w(i) = *_j$ we have a term of $cy_i$ from the definition of $x_j$, and we're multiplying it by $\lambda_j$ in the linear combination). We can rewrite this by thinking of it as a sum of the $y_i$ with various coefficients. Each term comes with a factor of $c$. In the first sum, we're looking at all $i$ for which $w(i)$ is an element of $S$, and taking $w(i)$ itself as our coefficient. Meanwhile, the second sum says that whenever $w$ has a $*_j$, we put a coefficient of $\lambda_j$. This means the coefficients of $y_1, \dots, y_n$ are precisely what we get from $w$ by replacing $*_j$ with $\lambda_j$ for all $j$. So we can rewrite our set as

$$\left\{ c^2 y_0 + \sum_{i \in \{1,\dots,n\}} w_{\lambda_1,\dots,\lambda_\ell}(i) c y_i \;\middle|\; \lambda_1, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d] \right\},$$

where $w_{\lambda_1,\dots,\lambda_\ell}$ is the word obtained from $w$ by replacing $*_1, \dots, *_\ell$ with $\lambda_1, \dots, \lambda_\ell$. And by our choice of $w$, this is monochromatic — we chose $w$ so that its corresponding $\ell$-parameter subset of $S^n$ corresponding to $w$ was monochromatic, and the colors in that subset (in our coloring of $S^n$) correspond to the colors of these terms (in the original coloring). □

We're now done with the base case of the induction. The inductive step is essentially the same; we'll sketch it, but we won't check all the details (for the sake of time).

For the inductive step, let $h \geq 1$, and assume that we've proven the statement for $h-1$. This means in $M^*$, we can always find some $M$ such that its $j$th layer is monochromatic for all $j = 0, \dots, h-1$; and our goal is to make the $h$th layer monochromatic as well.

To do this, we'll again apply the multidimensional version of Hales–Jewett. We'll again take $S = \mathbb{Z} \cap [-d, d]$, and this time we'll take $m = \ell - h$. (The reason we have $\ell - h$ is that we only need to work hard to get the condition for $j = h$ — smaller $j$ will come from the induction hypothesis — and for $j = h$, the varying $\lambda_i$ are $\lambda_{h+1}, \dots, \lambda_\ell$, so we only have $\ell - h$ parameters.) This gives some $n \in \mathbb{N}$ for which the statement of Hales–Jewett holds.

Meanwhile, we apply the induction hypothesis for $h-1$, replacing $\ell$ with $n+h$ (the reason for this is that on the last step, we want to have $n$ variables left after we put away the first $h$), $d$ with $cd^2$, and $c$ with $c^2$ (and use the same value of $t$). This gives some $\ell^*, d^*, c^* > 0$ (these are the values of $\ell^*$, $d^*$, and $c^*$ for which we'll show Proposition 10.37 holds for $h$).

Now let $M^* \subseteq \mathbb{N}$ be an $(\ell^*, d^*, c^*)$-Deuber set, and consider a coloring of $M^*$ with $t$ colors. The first thing we'll do is apply the induction hypothesis; this tells us 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

$$\{c^2 y_j + \lambda_{j+1} y_{j+1} + \cdots + \lambda_{n+h} y_{n+h} \mid \lambda_{j+1}, \dots, \lambda_{n+h} \in \mathbb{Z} \cap [-cd^2, cd^2]\}$$

is monochromatic. As before, we consider the coloring of $S^n = (\mathbb{Z} \cap [-d, d])^n$ with $t$ colors where we color each $(\lambda_{h+1}, \dots, \lambda_{h+n}) \in S^n$ with the color of

$$c^2 y_h + \lambda_{h+1} c y_{h+1} + \cdots + \lambda_{h+n} c y_{h+n},$$

which is in $M' \subseteq M^*$. (This is the same idea as before — we take coefficients of $c^2$ for the first term, and $\lambda_i c$ for the remaining ones; but here we're starting at $h$ instead of 0 because we care about proving 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 color.)

The strategy is the same as before — from our original coloring we've obtained a coloring of $S^n$, and we apply the multidimensional Hales–Jewett theorem to it: By our choice of $n$, there exists a monochromatic $(\ell - h)$-parameter subset (since $m = \ell - h$), described by some word

$$w = (w(h+1), \dots, w(h+n)) \in ((\mathbb{Z} \cap [-d, d]) \cup \{*_{h+1}, \dots, *_\ell\})^n.$$

By interchanging the order of these symbols, we can again assume that the first occurrence of $*_{h+1}$ is before the first occurrence of $*_{h+2}$, and so on.

> **Remark 10.44.** Note that we're using indices that start at $h+1$ instead of 1 (both for our $n$-tuples in $S^n$ and for the $*_i$), for reasons we'll soon see.

Now we want to define the $x_0, \dots, x_\ell$ that we'll use to define our $(\ell, d, c)$-Deuber set $M$ (whose layers 0, $\dots$, $h$ should be monochromatic). First, we'll define $x_0 = cy_0, \dots, x_{h-1} = cy_{h-1}$. Then 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 $x_h$ to consist of $cy_h$ together with the contributions of the fixed part of our word $w$.) Finally, for the remaining terms, for each $j = h+1, \dots, \ell$, we set

$$x_j = \sum_{\substack{i \in \{h+1,\dots,n+h\} \\ w(i) = *_j}} cy_i.$$

(This definition 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 corresponding $y_i$.)

By the same reasoning as before (i.e., in the base case), it's still true that

$$M_{d,c}(x_0, \dots, x_\ell) \subseteq M_{cd^2, c^2}(y_0, \dots, y_{n+h}) = M' \subseteq M^*.$$

(We can again expand out all linear combinations on the left-hand side in terms of the $y_i$ and check that they're the correct shape.) This in particular means that $M_{d,c}(x_0, \dots, x_\ell)$ is a set of positive integers, so it really is an $(\ell, d, c)$-Deuber set.

Now we need to check that a bunch of layers are monochromatic. First, for each $j = 0, \dots, h-1$, the set

$$\{cx_j + \lambda_{j+1} x_{j+1} + \dots + \lambda_\ell x_\ell \mid \lambda_{j+1}, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]\}$$

(the $j$th layer of $M$) is a subset of

$$\{c^2 y_j + \lambda_{j+1} y_{j+1} + \dots + \lambda_{h+n} y_{h+n} \mid \lambda_{j+1}, \dots, \lambda_{h+n} \in \mathbb{Z} \cap [-cd^2, cd^2]\}$$

(the $j$th layer of $M'$) — you can check that every linear combination of the first form can be expanded to one of the second (the coefficients work out properly). And the latter sets are each monochromatic (for each $j = 0, \dots, h-1$) by our choice of $M'$, so the former sets are each monochromatic as well.

This means the only case we need to check is $j = h$. And this is the same calculation as in the base case — we're interested in the set

$$\{cx_h + \lambda_{h+1} x_{h+1} + \dots + \lambda_\ell x_\ell \mid \lambda_{h+1}, \dots, \lambda_\ell \in \mathbb{Z} \cap [-d, d]\}$$

(the $h$th level of $M$). We can rewrite each of these linear combinations as

$$c^2 y_h + \sum_{\substack{i \in \{h+1,\dots,n+h\} \\ w(i) \in \mathbb{Z} \cap [-d,d]}} w(i) cy_i + \sum_{\substack{i \in \{h+1,\dots,n+h\} \\ j \in \{h+1,\dots,\ell\} \\ w(i) = *_j}} \lambda_j cy_i,$$

and again the coefficient of each $y_i$ is what we'd get if we took $w$ and replaced each $*_j$ with $\lambda_j$ (together with a factor of $c$), so we can rewrite this $h$th level as

$$\left\{ c^2 y_h + \sum_{i \in \{h+1,\dots,h+n\}} w_{\lambda_{h+1} \dots \lambda_\ell}(i) cy_i \;\middle|\; \lambda_i \in \mathbb{Z} \cap [-d, d] \right\}.$$

And this is exactly the set we chose to be monochromatic (when defining $w$). So we've checked the condition for $j = h$, and we're done. □

# §11 A 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 see 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 some 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 this number (Theorems 1.13 and 1.20).

**Theorem 11.1** (Erdős–Szekeres 1935 and Erdős 1947)
For all $k \geq 2$, we have $2^{k/2} \leq R(k,k) \leq 4^k$.

More generally, the Erdős–Szekeres argument for the upper bound gives $R(k,\ell) \leq \binom{k+\ell-2}{k-1}$.

Both bounds are exponentials, but with different bases ($\sqrt{2}$ and 4, respectively). Closing this gap has been maybe the most important problem in Ramsey theory since then. And in over 75 years, there has been no progress to improve either $\sqrt{2}$ or 4. There has been a lot of work which improved lower order terms — the lower bound was improved by a factor of 2 (as mentioned in Remarks 1.21 and 1.22), and the upper bound was improved by subexponential (but better than constant) terms. 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 11.2** (Campos–Griffiths–Morris–Sahasrabudhe 2023+)
There exists a constant $c > 0$ such that $R(k,k) \leq (4-c)^k$ for all $k \geq 2$.

**Remark 11.3.** Their proof also gives an exponential improvement for off-diagonal Ramsey numbers $R(k,\ell)$ when the ratio between $k$ and $\ell$ is bounded (compared to Erdős–Szekeres), but for time reasons we won't state exactly what that improvement is.

Today we'll discuss the proof outline. Our goal is to show that for any edge-coloring of a 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 want to talk about which vertices are connected in which colors):

**Notation 11.4.** In a graph with edges colored red and blue, for a vertex $x$, we write $N_R(x)$ to denote the red neighborhood of $x$, and $N_B(x)$ to denote the blue neighborhood of $x$.

## §11.1 Warm-up: An 'algorithmic' proof of the original upper bound

As a warm-up, and to put the approach of this new work into context, we'll first redo a proof of roughly the Erdős–Szekeres bound $R(k,\ell) \leq \binom{k+\ell-2}{k-1}$ (we'll lose some lower-order factors), phrased in a somewhat unusual way that will motivate the new approach. This new approach is in some sense to write down an algorithm that will lead us to find a monochromatic clique of size $k$ (it's not an algorithm in the sense of being efficient, which is why we have 'algorithm' in quotes). So along these lines, we'll first see a basic 'algorithmic' proof of the original upper bound. Specifically, we'll prove that

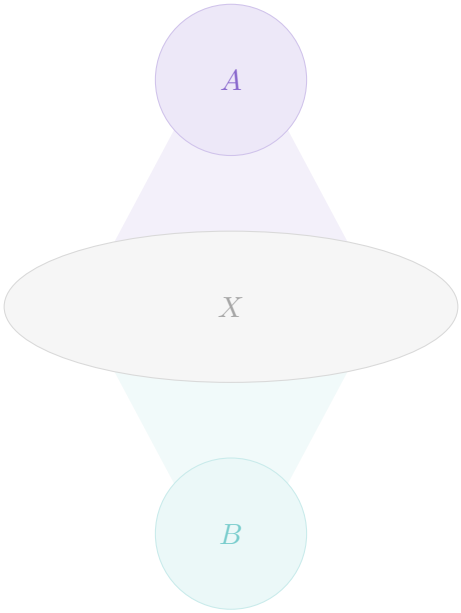
$$R(k,\ell) \leq e(k+\ell) \cdot \frac{(k+\ell)^{k+\ell}}{k^k \ell^\ell} \tag{11.1}$$

(this is roughly the same as the Erdős–Szekeres bound (Theorem 1.13), by Stirling's formula).

To prove this, we'll start with a complete graph $G$ on $n$ vertices with edges colored red and blue, and suppose that there is no red clique of size $k$ or blue clique of size $\ell$. Our goal is to show that then $n$ is smaller than the right-hand side of (11.1). The idea of the algorithm is to start with $G$ and perform certain steps, such that if the algorithm runs too long, then it either finds a red clique of size $k$ or a blue clique of size $\ell$. This means the algorithm must terminate before this happens; and then by looking at the quantitative behavior of the algorithm, we'll get a bound on $n$.

We'll recursively build three disjoint subsets $A, B, X \subseteq V(G)$ 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 $|X| \geq k + \ell$.

At every step, we pick a vertex $x \in X$ (note that since we haven't stopped yet, $X$ has more than $k + \ell$ vertices, so in particular it's nonempty).

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 $X$ (since this will be the set we update $X$ to). If this set is pretty big, then this is a good move (as $X$ will remain pretty big after the update). Meanwhile, if this set is *not* pretty big, then we'll instead move $x$ to $B$.

So in other words, we look at the red and blue neighborhoods of $x$ inside $X$, and compare their sizes; and we have a certain threshold for whether to move $x$ to $A$ or to $B$.

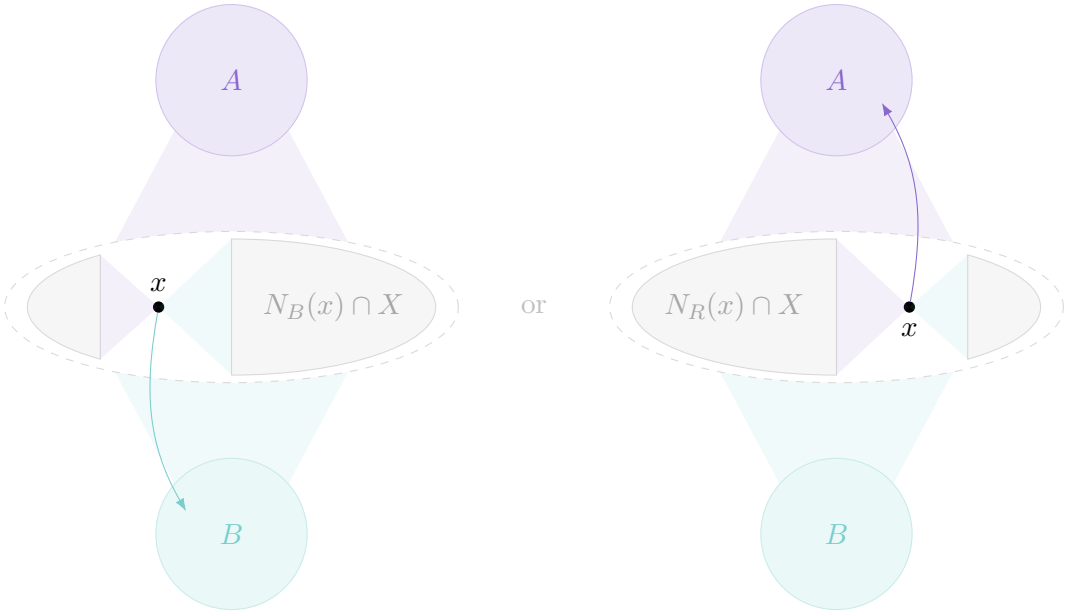
- If $x$ has a lot of blue neighbors in $X$, specifically

$$|N_B(x) \cap X| \geq \frac{\ell}{k+\ell}(|X| - 1),$$

  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.)
- Otherwise, since the red and blue 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 $X \mapsto N_R(x) \cap X$ (i.e., we move $x$ to $A$ and update $X$).




**Remark 11.5.** The threshold we use here ($\frac{k}{k+\ell}$ vs. $\frac{\ell}{k+\ell}$) is perhaps not that surprising. In the *diagonal* case $k = \ell$, we'd want the threshold to be $\frac{1}{2}$. When $k \neq \ell$, it makes sense that the ratio at the threshold should be around the same as the ratio between $k$ and $\ell$. And 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 want 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. More formally, we can inductively 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're increasing $|B|$ by 1 and shrinking $|X|$ by roughly $\frac{\ell}{k+\ell}$; the second term is similar (for the second case). The third term is somewhat artificial; it's a lower-order term that we just put in to adjust for having $|X| - 1$ rather than $|X|$ in the thresholds, and it doesn't matter too much in the end. Finally, $n$ is just the number of vertices we start with in the beginning.

Now let's analyze what happens at the end of the algorithm. Note that throughout the whole algorithm, $A$ can never reach size $k$ and $B$ can never reach size $\ell$ (since we assumed there is no red clique of size $k$ or blue clique of size $\ell$), so we always have $|A| < k$ and $|B| < \ell$. This means

$$|X| > \left(\frac{\ell}{k+\ell}\right)^{\ell} \left(\frac{k}{k+\ell}\right)^{k} \cdot \frac{1}{e} \cdot n.$$

This is *always* 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 these two bounds, we get that

$$n < e \cdot (k+\ell) \cdot \left(\frac{k+\ell}{\ell}\right)^{\ell} \left(\frac{k+\ell}{k}\right)^{k},$$

which is the bound we wanted (the right-hand side is the same as that of (11.1)).

**Remark 11.6.** The extra factors of $e$ and $k + \ell$ are there just for convenience to deal with the $|X| - 1$ vs. $|X|$ issue. They're lower-order terms, so we don't really care about them — the point of this wasn't so much to get a good bound, but rather to illustrate this algorithmic approach (even without these terms, $\frac{(k+\ell)^{k+\ell}}{k^k \ell^\ell}$ qualitatively matches $\binom{k+\ell-2}{k-1}$ but is still worse by lower-order factors).

**Remark 11.7.** In some sense, this argument is the exact same as the Erdős–Szekeres argument we saw in class when proving Theorem 1.13 — we didn't phrase the argument in that way, but what this algorithm does is roughly the same.

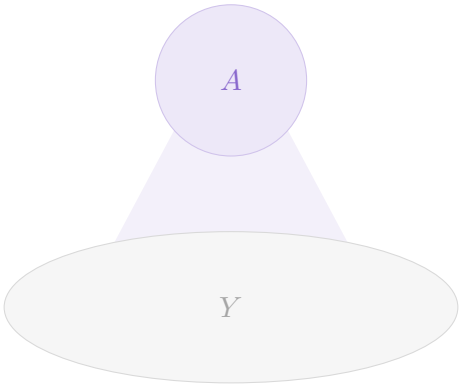
## §11.2 Outline of the new algorithm

Of course, this is not the algorithm that the new paper uses, since they get a better bound — their algorithm is vaguely similar, but it's more complex.

### §11.2.1 Books

The first idea they use is the following (this idea comes from an earlier paper). In our simple algorithm, we had a red clique $A$ and a large set of vertices which had all red edges to $A$. This structure has a name — it's called a *book*.

**Definition 11.8.** In a complete graph $G$ with edges colored red and blue, for disjoint $A, Y \subseteq V(G)$, we say $(A, Y)$ is a red book if $A$ is a red clique and all edges between $A$ and $Y$ are red.




**Remark 11.9.** The name 'book' may be because this structure looks a bit like if you take a book held at its binding and have lots of pages fanning out.

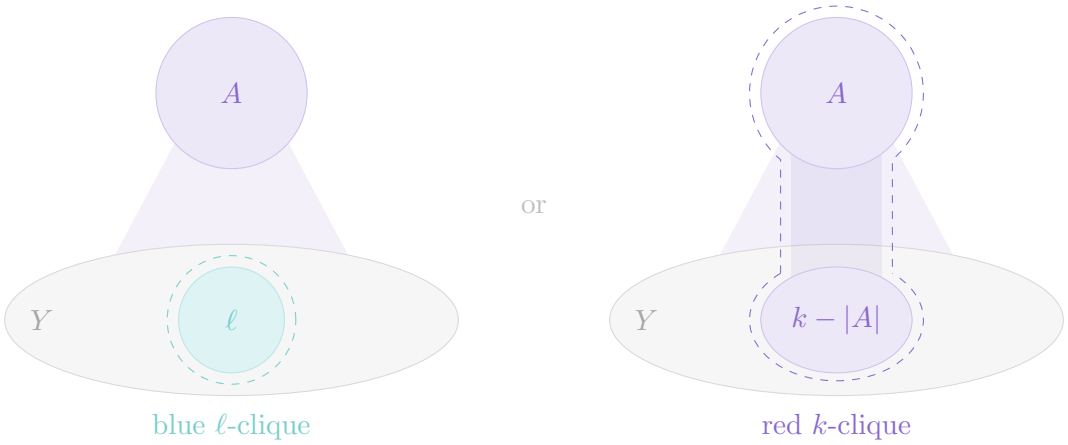
The first idea going into the proof (which had been known before) is that rather than making an algorithm which tries to find a monochromatic *clique* of the desired size, it's enough to make an algorithm which finds a large *book*. Here's a statement making this precise (it tells us that if we've found a book which is large in some sense, this automatically implies the existence of a large monochromatic clique).

**Lemma 11.10**

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 $\ell$.

Note that this Ramsey number $R(k - |A|, \ell)$ depends on $|A|$ — if $|A|$ is bigger, then the size $|Y|$ we need is smaller.

*Proof.* Since $|Y| \geq R(k - |A|, \ell)$, we know that $Y$ contains a blue clique of size $\ell$ or a red clique of size $k - |A|$. In the first case, we're done (we've found a blue clique of size $\ell$). In the second case, if we combine this red clique of size $k - |A|$ together with $A$, we get 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 (where we're trying to build a red or blue clique), we're already happy once we've found a red book where $Y$ is big enough.

### §11.2.2 The maintained structures

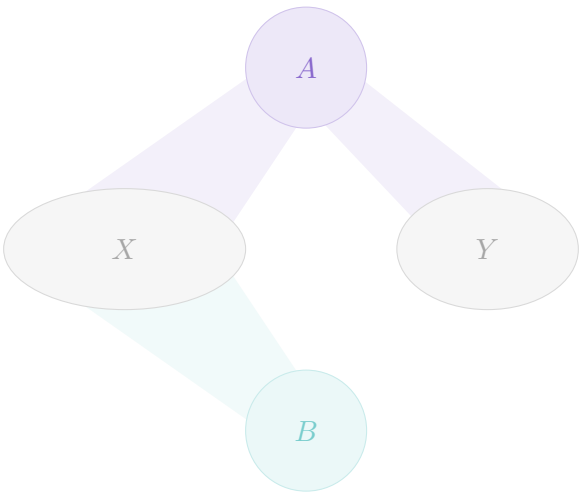
The next thing we'll do is explain what the algorithm actually builds. (It doesn't just consist of sets $A$, $X$, and $B$ as before — now there's actually *four* sets.) In order to keep things more understandable for now, we won't yet write down all the conditions with exact parameters, because they're complicated to set up. Today we'll just explain what the algorithm builds, what steps we have, what quantities we control and some rough intuition of how we control them, and how that helps us finish. On Tuesday we'll precisely define these parameters and thresholds.

We've written Theorem 11.2 for *diagonal* Ramsey numbers, but we'll need to consider the *off-diagonal* case as well; in the end, the proof gives a good bound (i.e., an exponential improvement) for all $\ell \le k$ with $\ell$ linear in $k$. So we'll let $\ell \le k$ and let $G$ be a complete graph with edges colored red and blue. We assume that $G$ has no red clique of size $k$ and no blue clique of size $\ell$, and 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 from $A$ to $Y$ are red.
- $B$ is a blue clique, and all edges from $B$ to $X$ are blue.

(It's maybe surprising that we don't also require the edges from $B$ to $Y$ to be blue; but this asymmetry is important.)



Throughout the algorithm, we know that $A$ never gets up to size $k$, and $B$ never gets up to size $\ell$. And in the end, we'll look at the red book $(A, Y)$ (for Lemma 11.10). So $X$ and $Y$ have very different roles — $(A, Y)$ is the red book we're building, while $X$ is the pool of vertices that we'll move to $A$ or $B$. (Vertices in $Y$ don't ever get moved to $A$ or $B$ — only vertices in $X$ do. But we will have to update $Y$ sometimes, e.g., when we move a new vertex to $A$.)

Throughout the algorithm, there's a few things we need to track:

- The sizes of the sets, i.e., $|A|$, $|B|$, $|X|$, and $|Y|$. This is unsurprising — we want $A$ and $B$ to grow (as we move things into them), and we want the sizes of $X$ and $Y$ to not drop by too much.
- The density of red edges between $X$ and $Y$. (This is perhaps more surprising, but we'll see in a moment why this density is important.) 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 first need to initialize the algorithm. We start with $A = \emptyset$ and $B = \emptyset$, and with $X$ and $Y$ each consisting of half of the vertices, chosen arbitrarily — so we partition $V(G) = X \cup Y$ arbitrarily into sets of size at least $\lfloor \frac{1}{2} |V(G)| \rfloor$. (So up to constant factors, $X$ and $Y$ start off the same size as the ground set.)

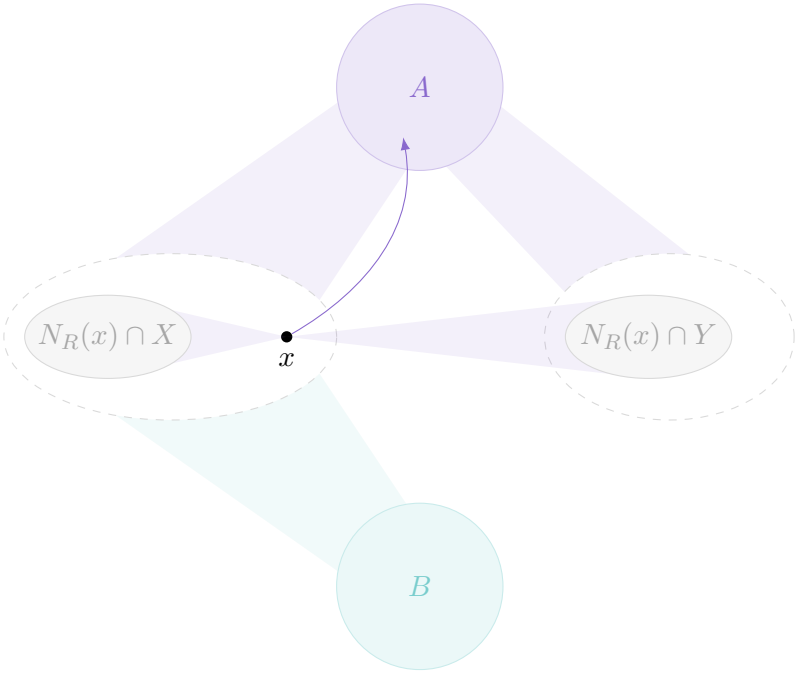
We also want to define when the algorithm terminates — we'll terminate when $X$ becomes very small. (The threshold for 'very small' is $|X| \leq R(k, \ell^{3/4})$. Right now it's very unclear why we use this threshold, so we can ignore it.)

### §11.2.3 Types of steps

We'll now discuss roughly what the steps of the algorithm are. There are three main types of steps (though later we'll add a fourth preprocessing step):

The first type of step is a **red step** — this is roughly what we did before (in the simple algorithm from Subsection 11.1), where we take a 'suitable' vertex $x \in X$ and move it to $A$, and update the sets accordingly. More precisely, we update

$$X \mapsto N_R(x) \cap X,\ Y \mapsto N_R(x) \cap Y,\ A \mapsto A \cup \{x\},\ B \mapsto B.$$



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 (from Subsection 11.1), we need $x$ to have many red neighbors in $X$, so that when we update $X$, it doesn't shrink by 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 between $X$ and $Y$, then $Y$ would shrink a lot, and we wouldn't be able to afford this update. So this is why we need a reasonable red density between $X$ and $Y$ — so that we can make moves of this form without the size of $Y$ dropping by too much.

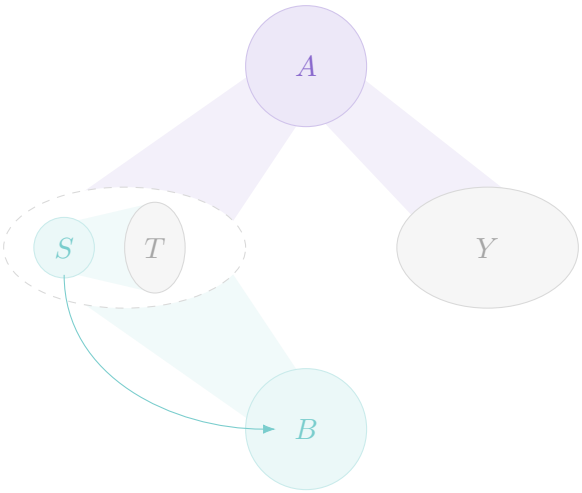
We might guess that the next type of step is a blue step, but it's actually more subtle. The algorithm doesn't have just one blue step (the way it has one red step); it actually has *two* types of blue steps with different features.

The next type of step is a **big blue step**. This is roughly what we'd think a blue step would be (moving a vertex from $X$ to $B$ and updating $X$ accordingly), but we move *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 $B$. When we do this, we'll update $X$ to $T$ (we *could* update $X$ to the intersection of the blue neighborhoods in $X$ 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

$$X \mapsto T,\ Y \mapsto Y,\ A \mapsto A,\ B \mapsto B \cup S.$$

(This is called 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 between $X$ and $Y$? We're replacing $X$ with $T$, so the new red density between $X$ and $Y$ is the current red density between $T$ and $Y$. In theory, with what we've just written, we have no control over the red density between $T$ and $Y$. But we will ensure this control using a *preprocessing* step. Before we do any of these main steps, we'll first clean up $X$ and $Y$ by kicking out all vertices in $X$ which have low red density to $Y$ (this cleanup step will be called *degree regularization*). Then we know the vertices in $T$ survive this step, which means they have enough neighbors in $Y$; so the red density between $T$ and $Y$ will be good enough.

There's one natural question — why do we need to do this step with many vertices at a time, rather than just one? The reason is that every time we do this, we lose a bit in our red density $p$. We've just said that with the cleanup step, we don't *completely* lose control of this density. But we do lose a little bit of density (we can't kick out all the vertices in $X$ just a *tiny* bit below the average density, because that could be almost everything), which means it would be bad if we performed too many of these steps. And the

fact that we're moving *many* vertices at once (specifically, the number we move will be around $\ell^{1/4}$) gives a bound as to how often these steps occur — we can have at most $\ell$ vertices in $B$, so these steps happen at most $\ell^{3/4}$ times. This means we won't have to worry about the density loss from these steps — it will be a lower-order term.

(We will also lose a bit of density in the red steps, for similar reasons; but that will be part of a very delicate analysis.)

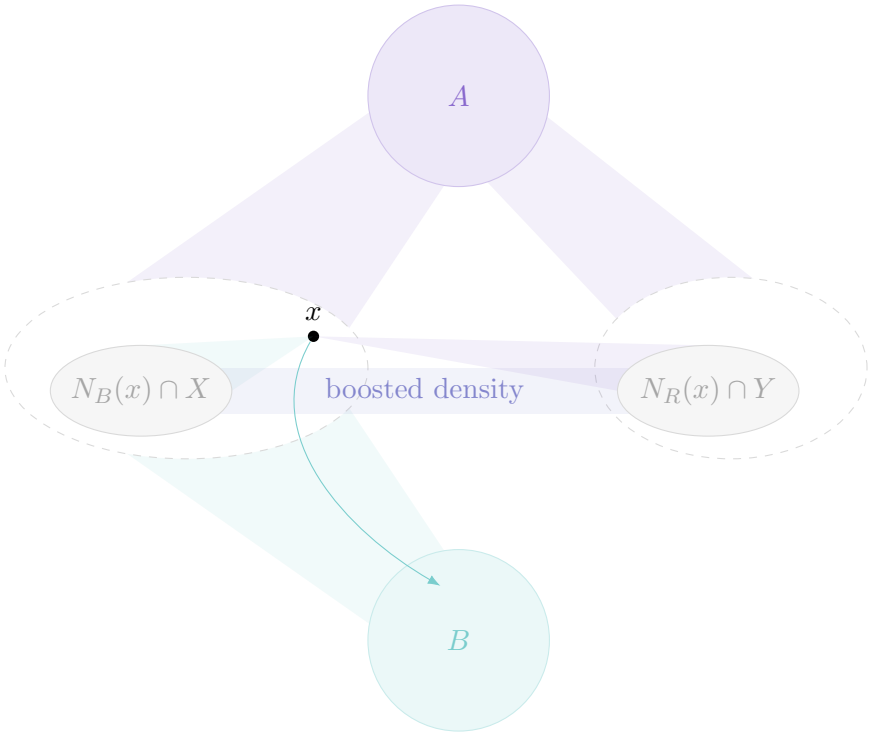
The third type of step is also a kind of blue step (in that it moves a vertex from $X$ to $B$), but we don't call it a blue step. Instead, we call it a **density-boost step**, because it has the additional feature that it boosts the red density between $X$ and $Y$. This step will counter the losses in density that we get from the other two steps.

This step will happen when neither of the first two steps can happen — it could be that we can't find a suitable vertex (for a red step) or a large blue book (for a big blue step), so we can't perform the first two types of steps. But in these cases, we'll be able to prove that we can always perform this third type of step, where we move a vertex from $X$ to $B$ while boosting the density.

In this type of step, we take a 'suitable' vertex $x \in X$; in particular, we require the red density between $N_B(x) \cap X$ and $N_R(x) \cap Y$ to be relatively high (quite a bit bigger than $p$). We then update

$$X \mapsto N_B(x) \cap X,\ Y \mapsto N_R(x) \cap Y,\ A \mapsto A,\ B \mapsto B \cup \{x\}.$$

This is a 'small blue step' in the sense that we're moving $x$ to $B$, and updating $X$ to what it needs to be (the blue neighborhood of $x$ in $X$). But the surprising thing is that we're also updating $Y$. When we move a vertex from $X$ to $B$, *a priori* there is no need to update $Y$ — there aren't any conditions between $B$ and $Y$. But we still do update it; the point of this is to get a boosted density between the new set $X$ and the new set $Y$.




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 imagine, we'll have

to delicately analyze how the changes in these steps compare to each other. This is the reason we want the big blue steps to be big — that means the density loss from those steps 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.

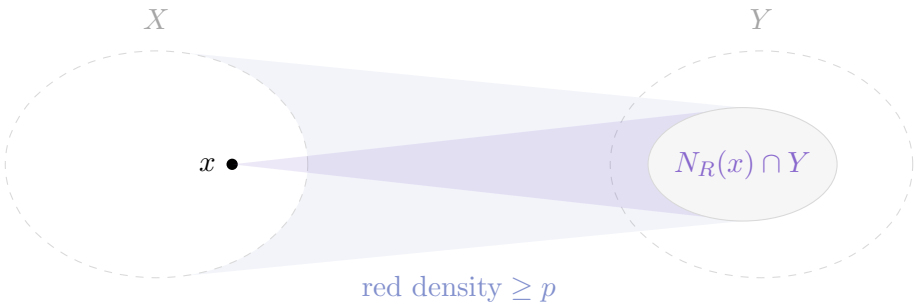
Finally (as mentioned earlier), before doing any of these steps, we need to remove vertices from $X$ which have too small red degree to $Y$. This is called the **degree regularization** step; it's interleaved with the three main steps (so we'll perform this preprocessing step, then a main step, then this preprocessing step, then a main step, and so on). Informally, this step removes vertices $x \in X$ for which $|N_R(x) \cap Y|$ is significantly smaller than $p|Y|$. (The *average* value of $|N_R(x) \cap Y|$ will be $p|Y|$, since $p$ is defined as the red density between $X$ and $Y$.) Of course, in the analysis we'll have to ensure that we don't remove too many vertices.

### §11.2.4 Choosing a step

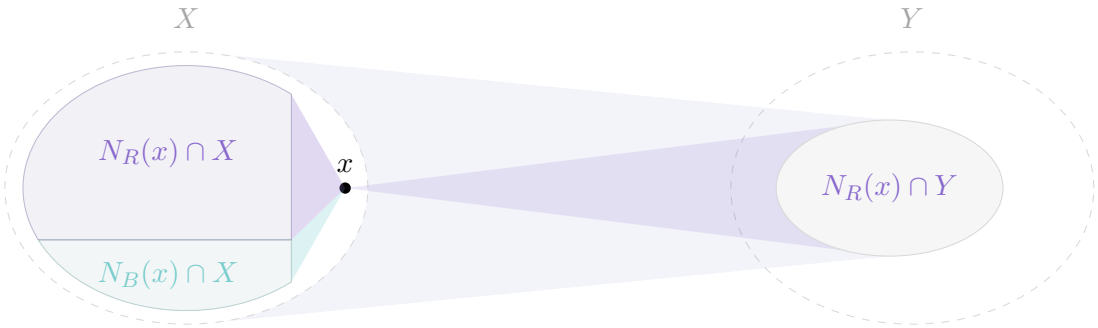
First, why is it always possible to perform one of the main steps — if the first two steps (a red step or big blue step) fail, why should we hope to be able to find a vertex $x$ such that the third step (a density-boost step) works? Here's the intuitive argument for this. First, we can show that there's always a vertex $x \in X$ such that

$$d_R(X, N_R(x) \cap Y) \geq p. \tag{11.2}$$

This is by an averaging argument using convexity, which we'll say more about on Tuesday (it's not obvious, but it's not difficult; the idea is to use an averaging argument where we count red walks from $X$ to $Y$ and back to $X$ using convexity).




Then the idea is sort of that if a big blue step can't happen, then we can find a vertex $x \in X$ such that $|N_R(x) \cap X|$ is reasonably large and $d_R(X, N_R(x) \cap Y)$ is not much smaller than $p$. In other words, we can find a vertex $x$ satisfying (11.2) with a bit of loss, with the additional condition that the red neighborhood of $x$ in $X$ is large (i.e., its blue neighborhood in $X$ isn't large).




Then there are two options. We know the red density between $X$ and $N_R(x) \cap Y$ (the red neighborhood of $x$ in $Y$) is pretty good. So one option is that the red density between $N_R(x) \cap X$ (the *red* neighborhood of

$x$ in $X$) and $N_R(x) \cap Y$ is pretty good; then we can take a red step. (The threshold here has to be quite a bit below $p$, which means we'll need to take quite a bit of density loss in the red steps.) Otherwise, the red density between $N_B(x) \cap X$ (the *blue* neighborhood of $x$ in $X$) and $N_R(x) \cap Y$ must be quite a bit bigger than $p$ to make up for this (since the red density from $X$ to $N_R(x) \cap Y$, which is nearly $p$, is basically a weighted average of the densities from these red and blue neighborhoods). And in this case, we can take a density-boost step.

There's something concerning here — we chose $x$ such that $N_R(x) \cap X$ is quite big, which means $N_B(x) \cap X$ might be tiny. But if we're taking a density-boost step, then we're replacing $X$ with this tiny bit. So we're losing a lot in the size of $X$, which sounds super dangerous. But the point is that the boost in density we get from the averaging argument gets higher as this bit gets smaller (as the smaller $N_B(x) \cap X$ gets, the higher its corresponding red density needs to be to make up for the red density from $N_R(x) \cap X$ being small). So in this step, there's a delicate tradeoff between how much $X$ shrinks and how much we boost our density (in particular, $X$ can't become empty). These tradeoffs are quite delicate, and it's sort of miraculous that this all works out in the end.

### §11.2.5 Overview of the final analysis

We've seen that the big blue steps don't affect the density $p$ too much, since there's only a few such steps. The red steps make $A$ grow but cause a loss in density. And the density-boost steps boost the density, but they make $X$ and $Y$ quite a bit smaller.

So in the end, $|A|$ will be affected by the number of red steps, and $|Y|$ will be affected by the number of red steps and density-boost steps. And we want the inequality

$$|Y| \geq R(k - |A|, \ell)$$

(so that we can use Lemma 11.10); in the diagonal case $k = \ell$, we want $|Y| \geq R(k - |A|, k)$. And we already have the Erdős–Szekeres bound (Theorem 1.13) for this Ramsey number, so we're happy if

$$|Y| \geq \binom{2k - |A|}{k - |A|}. \tag{11.3}$$

There's a picture on page 8 of the paper where the $x$-axis shows the number of red steps taken, and the $y$-axis shows the number of density-boost steps (at the end of the algorithm). The algorithm only terminates when $X$ is small, so not all combinations of these numbers are possible; the blue region in the picture shows which combinations are possible if we start with $n = (4 - o(1))^k$ vertices. And the red region in the picture shows where (11.3) is *not* satisfied (so the red region is the region that's problematic). Then the only case where a problem occurs is if we're in both the red and blue regions simultaneously (because we *have* to be in the blue region, and if we're not in the red region then we've succeeded).

At a first glance, it looks like the red and blue regions don't intersect; and if they really were disjoint, then the proof would be done. But a zoomed-in picture shows that they do intersect, on a tiny region shown in green; and if we're in that tiny green thing, then we're *not* done.

But the good news is that this green thing is rather small — in particular, it only occurs 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. And in this case, we'll actually try to prove a better bound on the Ramsey number $R(k - |A|, k)$ than the Erdős–Szekeres one. For the case $0.75k \leq |A| \leq 0.85k$, we want a better bound on $R(k, \ell)$ in the regime $0.15k \leq \ell \leq 0.25k$. (If we can get a better bound for this Ramsey number, then it'll be easier to prove the desired bound of $|Y| \geq R(k - |A|, k)$.)

So we do everything from scratch again, starting with a different $\ell$. Why would you hope this would succeed now? In the simple algorithm from Subsection 11.1, the size tradeoff was something like $\frac{\ell}{k+\ell}$ (we moved $x$

to $B$ if its blue neighborhood occupied at least this fraction of $X$, and to red if its blue neighborhood was smaller than this); we'll have a similar tradeoff here (in what it means for $N_B(x) \cap X$ to be small, im which case we perform a red or density-boost step). When $k = \ell$, this was roughly $\frac{1}{2}$; but here it's substantially smaller (roughly $\frac{1}{5}$). This means the density boost in the density-boost steps will be bigger (since this boost depends on how the red and blue neighborhoods of $x$ in $X$ compare). And this will be enough to get an exponential improvement over the Erdős–Szekeres bound in the case $0.15k \le \ell \le 0.25k$.

Now returning to the diagonal case and using this better bound for $R(k - |A|, k)$, the red region (the problematic region where $|Y| < R(k - |A|, k)$) becomes smaller. And it turns out that now the red and blue regions really are disjoint, which means we succeed at improving the bound from $4^k$ to $(4 - c)^k$.



## §11.3 Setup

Last class, we saw an overview of the proof strategy of Theorem 11.2, 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 (e.g., last time we only gave some 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 long, so we can't cover all the details. So we'll try to be precise about what the algorithm is, but we can't discuss the 30 pages of analysis.

### §11.3.1 The overall structure

The overall structure of the proof is to first show a better bound for *off-diagonal* Ramsey numbers:

**Theorem 11.11**
For $\frac{k}{10} \le \ell \le \frac{k}{4}$, we have $R(k, \ell) \le e^{-\ell/50 + o(k)} \binom{k+\ell}{\ell}$.

This gives an exponential improvement for the Ramsey numbers $R(k, \ell)$ in this range — we already know $R(k, \ell) \le \binom{k+\ell}{\ell}$ by Erdős–Szekeres, and the $e^{-\ell/50}$ factor is a tiny exponential improvement.

**Remark 11.12.** The 10 here is somewhat arbitrary, but the 4 is important — last time we saw that the diagonal approach almost works, but it fails in a certain range, and to fix it, we need a better bound on off-diagonal Ramsey numbers in this range (since in the end, we're trying to prove that $|Y| \ge R(k - |A|, k)$, so that we can use Lemma 11.10).

The next step is to use this to show that $R(k, k) \le (4 - c)^k$. Both parts follow the same 'algorithmic' proof that we discussed last time and will discuss in more detail this time, with slightly different parameters and analyses. Today the plan is to give a precise version of this algorithm.

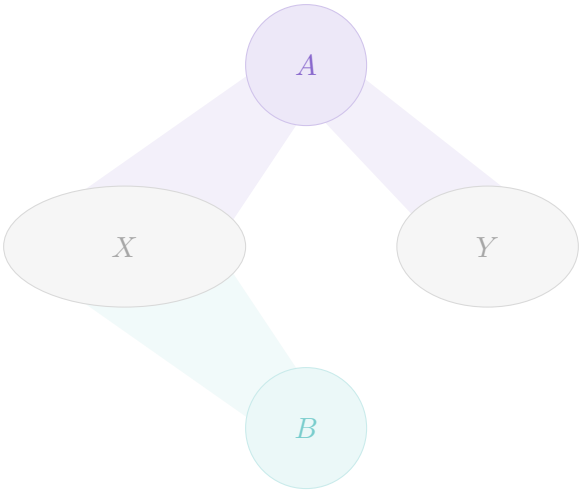
We'll assume that $\ell \le k$ and that both are sufficiently large, and that $\ell \ge \frac{k}{10}$. (You can put any constant here; we'll just take 10 for concreteness. We're ultimately interested in the case $k = \ell$, but as mentioned before, we have to go through the general case to get the diagonal case.)

Our goal is to upper-bound $R(k, \ell)$. So we 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 $\ell$; and we want to upper-bound $n$. (We already know that $n \le \binom{k+\ell}{\ell}$ by Erdős–Szekeres, but we want a (small) exponential improvement.)

As we saw last time, the overall strategy is to have an algorithm which builds 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 are no conditions between $B$ and $Y$, between $X$ and $Y$, or inside $X$ or $Y$.)




We use $X$ to move vertices to $A$ or $B$. Meanwhile, the role of $Y$ is that we want $(A, Y)$ to be a big red book in the end (specifically, if $|Y| \geq R(k - |A|, \ell)$, then by Lemma 11.10 we can find a red clique of size $k$ or blue clique of size $\ell$). We know $A$ has size less than $k$ and $B$ has size less than $\ell$; and in the end we'll examine the size of the book $(A, Y)$, and we'll be done if it's big enough.

Throughout the algorithm, we need to track the sizes of $A$, $B$, $X$, and $Y$ (this is unsurprising; we want $X$ and $Y$ to not shrink by too much). We also want to track the red density between $X$ and $Y$ (we explained why last time); we'll use $p = d_R(X, Y)$ to denote this red density.

(As we run through the algorithm, these sets $A$, $B$, $X$, and $Y$ will change. At every point in the algorithm, we'll have a current configuration of sets and a current density. In particular, $p$ is not fixed — we'll always use it to denote the *current* red density between $X$ and $Y$.)

### §11.3.2 Choosing a suitable vertex

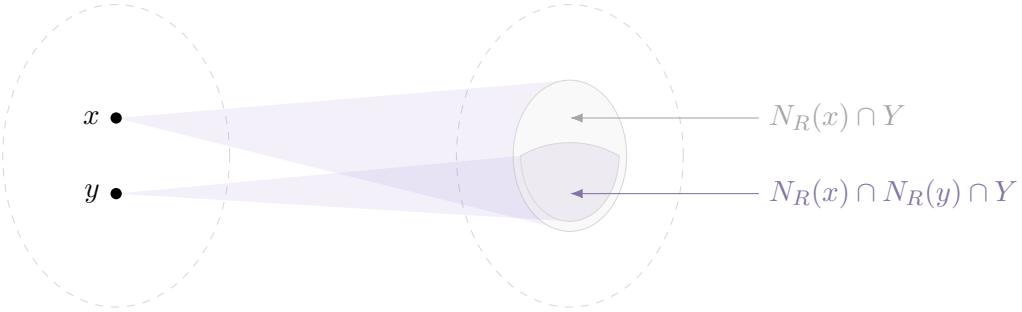
Last time, we saw that there are three different main types of steps — red steps, big blue steps, and density-boost steps. In red and density-boost steps, we pick a particular vertex in $X$ and move it to $A$ or $B$, and do something with its neighborhoods in $X$ and $Y$. So the first thing we'll define is related to how we pick these vertices.

**Definition 11.13.** Given the current configuration of sets $X$ and $Y$ and vertices $x, y \in X$, we define the weight of the pair $(x, y)$ as

$$\omega(x, y) = \frac{1}{|Y|}(|N_R(x) \cap N_R(y) \cap Y| - p \cdot |N_R(x) \cap Y|).$$

(We allow $x$ and $y$ to be the same vertex.)

This expression looks a bit odd, but as some motivation, we're first picking two vertices $x, y \in X$ and looking at the red neighborhood of $x$ in $Y$. Then the first term $|N_R(x) \cap N_R(y) \cap Y|$ counts the number of red neighbors of $y$ in this set, and we're subtracting $p$ times the size of that 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 $N_R(x) \cap Y$ would be roughly $p$; so $\omega(x, y)$ counts how much higher or lower the red density from $y$ into this set is compared to what we would expect. (The factor of $|Y|$ is just a normalization factor to ensure our numbers are at most 1; it's not super important.)




Last time, we saw that when we decide whether we're doing a red step or a density-boost step, we want to pick $x$ such that the red density between $X$ and $N_R(x) \cap Y$ is at least $p$ (or a tiny bit below $p$). This is captured by the above weights — these weights talk about the red density from *each* vertex $y \in X$ to $N_R(x) \cap Y$, and if we fix $x$ and combine this information over all $y$, then we get exactly the red density from $X$ to $N_R(x) \cap Y$. For this, we'll use the following definition:

**Definition 11.14.** We define $\omega(x) = \sum_{y \in X \setminus \{x\}} \omega(x, y)$.

This captures the number of red edges between $X \setminus \{x\}$ and the red neighborhood of $x$ in $Y$. (We remove $x$ because we'll eventually be splitting $X$ into the red and blue neighborhoods of $x$ (and looking at the density from each to $N_R(x) \cap Y$), and $x$ isn't part of either.)

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 below 0. This will mean $d_R(X \setminus \{x\}, N_R(x) \cap Y)$ is at least $p$, or just a bit below $p$. This is good because it's the condition we want on $x$ in order to take a red step or a density-boost step.

The first question is, why can we find such a vertex $x$? *A priori* it could be that every $\omega(x)$ is very negative, and then we'd be in bad shape. But the following lemma ensures this never occurs.

**Lemma 11.15**

We always have $\sum_{x,y \in X} \omega(x, y) \geq 0$.

Once we know this lemma, it's not hard to see that we can choose a vertex $x$ such that $\omega(x)$ is nonnegative or just a tiny bit below 0. (The reason we need the 'tiny bit below 0' case is that we need to exclude $y = x$ — this lemma immediately implies there is some $x$ such that $\sum_{y \in X} \omega(x, y) \geq 0$, but this sum has a tiny contribution from the case $y = x$ that we need to get rid of in order to get $\omega(x)$.)

However, we don't just need *any* vertex with this property — we need $x$ to have the additional property that $N_R(x) \cap X$ is not too small. (This is because $x$ is supposed to be good for taking a red step, and it won't be if $N_R(x) \cap X$ is too small.) But the idea is that if a big blue step doesn't happen, then there can only be a few vertices in $X$ whose red neighborhood in $X$ is too small for a red step. Then the contribution to this sum from ineligible vertices $x$ will be small; so among the eligible ones, we can still find a vertex $x$ where $\omega(x)$ is not too negative.

This lemma has a nice and relatively simple proof based on a convexity argument.

*Proof of Lemma 11.15.* We're interested in the sum

$$|Y| \cdot \sum_{x,y \in X} \omega(x, y) = \sum_{x,y \in X} |N_R(x) \cap N_R(y) \cap Y| - p\,|X| \sum_{x \in X} |N_R(x) \cap Y|\,.$$

For the second sum, we're essentially averaging the densities from each $x \in X$ to $Y$, so we get

$$p\,|X| \sum_{x \in X} |N_R(x) \cap Y| = p^2\,|X|^2\,.$$

Meanwhile, we can rewrite the first sum as a sum over $Y$ — each $z \in Y$ contributes to the first sum whenever $x, y \in N_R(z) \cap X$, so we can write

$$\sum_{x,y \in X} |N_R(x) \cap N_R(y) \cap Y| = \sum_{z \in Y} |N_R(z) \cap X|^2 .$$

And the *average* value of $|N_R(z) \cap X|$ over $z \in Y$ is $p|X|$ (we're essentially averaging the densities from each $y \in Y$ to $X$), so by convexity we get $\sum_{z \in Y} |N_R(z) \cap X|^2 \geq (p|X|)^2$. □

### §11.3.3 Density thresholds

So far we've seen how we'll choose the vertex $x$ which we look at for a red or density-boost step. The other question is, what should be the threshold at which we decide which kind of step to do? We only want to do a red step if it doesn't decrease $p$ by too much, so now we need to say what 'too much' means. For this, we'll need a few more definitions.

First, we let $\varepsilon = k^{-1/4}$. (The specific value $\frac{1}{4}$ is not important.) Note that $k$ is large, so $\varepsilon$ is small (i.e., as $k \to \infty$, we have $\varepsilon \to 0$).

We also let $p_0$ be the *initial* red density between $X$ and $Y$ (at the start of the algorithm). This will be an important comparison point: The idea is that if our density during the algorithm increases far beyond $p_0$, then we're pretty rich and can afford to lose more density. Meanwhile, if it's close to or below $p_0$, then we can't afford to lose very much. So all the terms we write down will be based on comparisons with $p_0$.

> **Remark 11.16.** It's maybe not clear why we'd want the density $p_0$ at the *start* of the algorithm to be so relevant for the intermediate behavior of the algorithm. The authors probably did this because it works, but there might also be more dynamic ways to run the proof where you don't fix everything based on a particular $p_0$.

First, we need to ensure at the start of the algorithm that $p_0$ is not too small (i.e., when we initialize the algorithm, we want to have a reasonably large red density between $X$ and $Y$; we'll define 'reasonably large' later, and discuss what to do if this isn't possible).

We'll now define some quantities that capture the losses we'll allow.

> **Definition 11.17.** For all $h = 0, 1, 2, \ldots$, we define
>
> $$q_h = p_0 + \frac{(1+\varepsilon)^h - 1}{k}.$$
>
> For $h = 1, 2, \ldots$, we also define $\alpha_h = q_h - q_{h-1}$.

We can see that $q_0 = p_0$; then $q_1$ is a bit bigger, and so on. They increase exponentially, which means the differences $\alpha_h$ increase with $h$ (i.e., the bigger $h$ is, the bigger $\alpha_h$ is).

> **Definition 11.18.** For $0 < p < 1$, we define $h(p)$ to be the smallest $h \in \{1, 2, \ldots\}$ such that $p \leq q_h$.

Here $p$ represents our current red density. So the point of the $q_h$ is to divide the possible $p$-values into certain intervals. (For example, if $p \leq q_1$ 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 amount by which we allow the density to go down in a red step. If this isn't possible, then we'll instead want a density boost (which will also somehow correspond to $\alpha_{h(p)}$ in size).

### §11.3.4 Red vs. blue neighborhood thresholds

We'll also define a threshold

$$\mu = \begin{cases} \frac{\ell}{k+\ell} & \text{if } \ell < k \\ \frac{2}{5} & \text{if } \ell = k. \end{cases}$$

The role of $\mu$ is that we only want to do a red step if $N_R(x) \cap X$ is big enough, or equivalently if $N_B(x) \cap X$ isn't too big. Specifically, we want this blue neighborhood to be at most $\mu |X|$.

The quantity $\frac{\ell}{k+\ell}$ is maybe not that surprising because in the algorithmic version of the Erdős–Szekeres bound (from Subsection 11.1), where we just took a vertex $x \in X$ and moved it to $A$ or $B$ (there was no set $Y$), which one we moved it to was decided based on how big the blue neighborhood of $x$ inside $X$ was, and our threshold for this was exactly $\frac{\ell}{k+\ell}$ (i.e., if $|N_B(x) \cap X| \geq \frac{\ell}{k+\ell} \cdot |X|$ then we'd do a 'blue step' where we move $x$ to $B$; otherwise we'd do a 'red step' where we move it to $A$). And the role of $\mu$ here is similar.

What's maybe more surprising is the value $\frac{2}{5}$. This comes from the fact that we first consider the case where $\ell \leq \frac{k}{4}$, where $\mu = \frac{\ell}{k+\ell}$; in particular, $\mu$ is quite a bit smaller than $\frac{1}{2}$. As we hinted at last class, this threshold $\mu$ being small means that the boost we get in the density-boost steps is amplified; and this allows the proof to work out.

Once we've done this (i.e., gotten an exponential improvement when $\ell \leq \frac{k}{4}$), we want to return to the diagonal case. In that case, if we defined $\mu = \frac{\ell}{k+\ell} = \frac{1}{2}$, we'd have a problem — it turns out that $\mu = \frac{1}{2}$ doesn't really work for the algorithm. So we artificially make $\mu$ smaller, which means that in the red steps, we lose less in $|X|$ ($\mu$ is the threshold for how big the blue neighborhood of $x$ is allowed to be if we want to perform a red step; in the diagonal case, this means the blue neighborhood is at most a $\frac{2}{5}$-fraction of $X$, so the *red* neighborhood is at least a $\frac{3}{5}$-fraction; this means $|X|$ loses a factor of $\frac{3}{5}$ instead of $\frac{1}{2}$). So replacing $\frac{1}{2}$ with $\frac{2}{5}$ makes our red steps better, and it *also* makes our density-boost steps better (we get a bigger boost). The place 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 this.

> **Remark 11.19.** You should think of $\mu$ as a fixed constant, and $\ell$, $k$, and $n$ as large with respect to $\mu$. (If you think of $\ell$ in the range $\frac{k}{10} \leq \ell \leq k$, then $\mu$ is at least $\frac{1}{11}$.) In particular, all asymptotics will be for $\mu$ fixed and $k$ and $\ell$ large.

## §11.4 The algorithm

We'll now describe the algorithm more precisely. (Note that this isn't an algorithm in the sense of being efficient — some steps of the algorithm might not be possible to do efficiently. So you couldn't use this algorithm 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.)

### §11.4.1 Initialization

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 the algorithm). Meanwhile, we want $X$ and $Y$ to be pretty big, so we want to define them by partitioning the vertex set into two sets of size at least $\lfloor \frac{1}{2} n \rfloor$. But we also want the initial red density $p_0 = d_R(X, Y)$ to be reasonably large. We need it to be bounded away from 0, but we need even more — specifically, we need it to be at least roughly $\frac{k}{k+\ell}$.

First, how can we ensure that $p_0$ is large? By taking a random partition $X \cup Y$, we can always ensure that it's at least the *total* red density; so it's enough to ensure that the *entire* graph has a large red density.

In the diagonal case, we can definitely assume the red density is at least $\frac{1}{2}$ — otherwise we can switch the roles of red and blue. Unfortunately, this symmetry argument doesn't work in the off-diagonal case.

In the off-diagonal case, we want the density to be at least roughly $\frac{k}{k+\ell}$. So what happens if it's not? In that case, we momentarily forget about everything we've said, and go back to the start of last class. At the beginning of last class (Subsection 11.1), we saw a simpler algorithm without the set $Y$, where we just had the sets $A$, $X$, and $B$ and we moved a vertex from $X$ to $A$ or $B$. If the blue density is bigger than $\frac{\ell}{k+\ell}$ by any constant factor, then when we do this argument, we'll win by a little bit at every step. And then one of two things might happen:

- Our red density might go into the range we wanted it to be (i.e., at least roughly $\frac{k}{k+\ell} = 1 - \frac{\ell}{k+\ell}$) at some point. If that happens, then at that point, we apply this framework (the more complicated algorithm) to the graph that's left.
- Otherwise, this never happens. Then when we run through the entire argument from Subsection 11.1, at every step we save a bit. And the number of steps is linear in $k$, so we get an exponential total saving.

**Remark 11.20.** There's also the question of what happens if the density is very *close* to $\frac{k}{k+\ell}$, but isn't exactly at least $\frac{\ell}{k+\ell}$. To handle this case, you do the argument a third time (with a range where $\ell$ is even smaller, i.e., something like $\ell \leq \frac{k}{40}$).

Now we'll assume that our initial red density *was* good enough, and we'll finally do the algorithm.

### §11.4.2 Choosing a step

We'll run the algorithm until $|X| \leq R(k, \lceil \ell^{3/4} \rceil)$ (our simple algorithm from Subsection 11.1 stopped when $X$ became too small; here we again stop when $X$ is too small, though with a different definition of 'too small') or if $p \leq \frac{1}{k}$ (this is a technicality — we'll ensure that this never actually happens, so that 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. In 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 $X$ by throwing out vertices with abnormally low density.) This changes $X$ a bit, and we update $p$ accordingly. (We'll explain this step in more detail later.)

Then we need to decide which of our three main steps (a red step, a big blue step, or a density-boost step) 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 $|N_B(x) \cap X| \geq \mu |X|$? (Intuitively, 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 lower bound on $|X|$ from the stopping condition.)

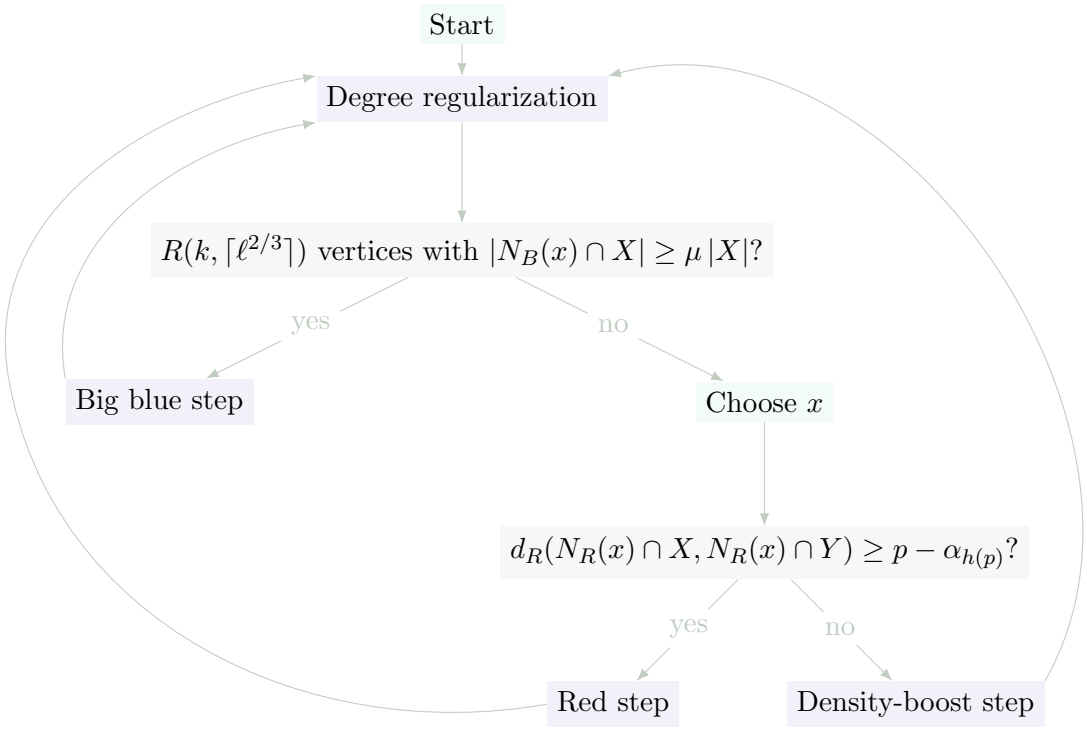
If the answer is yes, then we do a big blue step. (Then we go back and do another degree regularization step, then another main step, and so on.)

If the answer is no, then we choose a vertex $x \in X$ with maximal $\omega(x)$ such that $|N_B(x) \cap X| \leq \mu |X|$. (In this case, there are only a few vertices $x \in X$ whose blue neighborhood is at least a $\mu$-fraction of $X$, so there are many vertices whose blue neighborhood is smaller than this; and of those, we choose one where $\omega(x)$ is maximal.) (This is what we were talking about in Subsubsection 11.3.2 — we want to choose an eligible vertex $x$ where $\omega(x)$ is at least 0 or not too far 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're looking at the red density between the red neighborhood of $x$ in $X$ and the red neighborhood of $x$ in $Y$. The quantity $p - \alpha_{h(p)}$ was our threshold for how much density we're allowed to lose in a red step. 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, then another main step, and so on).




### §11.4.3 Details of the steps

Next, we'll explicitly describe what happens in the steps (as described last time). In a red step, we replace

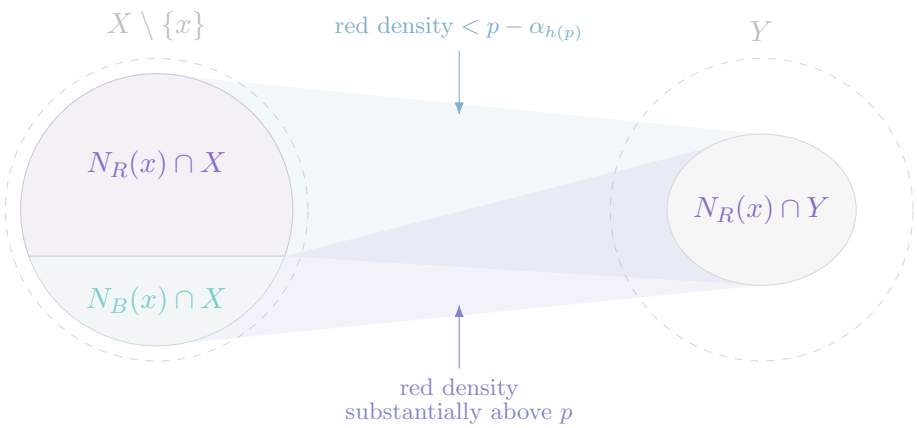
$$A \mapsto A \cup \{x\},\ B \mapsto B,\ X \mapsto N_R(x) \cap X,\ Y \mapsto N_R(x) \cap Y.$$

So we're moving $x$ to $A$, and updating $X$ and $Y$ to maintain the condition that the edges between $A$ and $X$ or $Y$ are red.

In a density-boost step, we replace

$$A \mapsto A,\ B \mapsto B \cup \{x\},\ X \mapsto N_B(x) \cap X,\ Y \mapsto N_R(x) \cap Y.$$

Note that we're replacing $X$ with the *complement* of the red neighborhood of $x$ in $X$, and $Y$ with the red neighborhood of $x$ 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 (it's at most $p - \alpha_{h(p)}$), so the red density between $N_B(x) \cap X$ and $N_R(x) \cap Y$ has to be quite a bit above it.



In a degree regularization step, we define $h = h(p)$, and update $A \mapsto A$, $B \mapsto B$, $Y \mapsto Y$, and

$$X \mapsto \{x \in X \mid |N_R(x) \cap Y| \geq (p - \varepsilon^{-1/2}\alpha_h)\,|Y|\}.$$

In words, we're kicking out all the 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 of this red neighborhood in $Y$ would be $p\,|Y|$, and $\varepsilon^{-1/2}\alpha_h$ is the bit of slack we need to ensure that we don't kick out too many vertices. (Of course when you analyze the algorithm, you need to ensure that you really don't lose too many vertices in $X$ from doing this.)

This step never decreases $p$, since we're only kicking out vertices below the average. Meanwhile, it ensures that every particular vertex $x \in X$ has a reasonably large density to $Y$, where we lose a little bit (specifically, $\varepsilon^{-1/2}\alpha_h$) compared to $p$. This is something we might lose in density when we do a big blue step, since the only density guarantee we have afterwards comes from this density guarantee on individual vertices.

Finally, we'll describe what we do in a big blue step. Last time, we said that in a big blue step, we want to add lots of vertices from $X$ to $B$; this means we want to find a large blue book $(S, T)$ (we'll move the blue clique $S$ to $B$; then $X$ becomes $T$, which by the definition of a blue book has all blue edges to $S$). A big blue step happens when we have lots of vertices $x \in X$ which *individually* have large blue neighborhoods in $X$ (each of size at least $\mu\,|X|$), and we want to translate that into a blue book.

We do this as follows: We find a blue book $(S, T)$ in $X$ where $|T| \geq \frac{1}{2}\mu^{|S|}\,|X|$, with $|S|$ as large as possible (subject to this condition). (It's maybe not so strange that we get an additional $\mu$-factor for every vertex in $S$ in this size condition — we're moving all the vertices in $S$ simultaneously to $B$, and each should cause us to lose a $\mu$-factor). We then update

$$A \mapsto A,\ B \mapsto B \cup S,\ X \mapsto T,\ Y \mapsto Y.$$

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 11.21**

Whenever a big blue step happens, its book $(S, T)$ satisfies $|S| \geq \ell^{1/4}$.

(This is indeed a lot of vertices, since we think of $\ell$ as large.)

**Corollary 11.22**

There are at most $\ell^{3/4}$ big blue steps.

*Proof.* Note that $|B|$ can never reach $\ell$, since there's no blue clique of size $\ell$. 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 from degree regularization) on each big blue step. But Corollary 11.22 says that there are not too many such steps; so we won't lose too much.

Now we've seen what happens in each of the steps. We keep running 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.

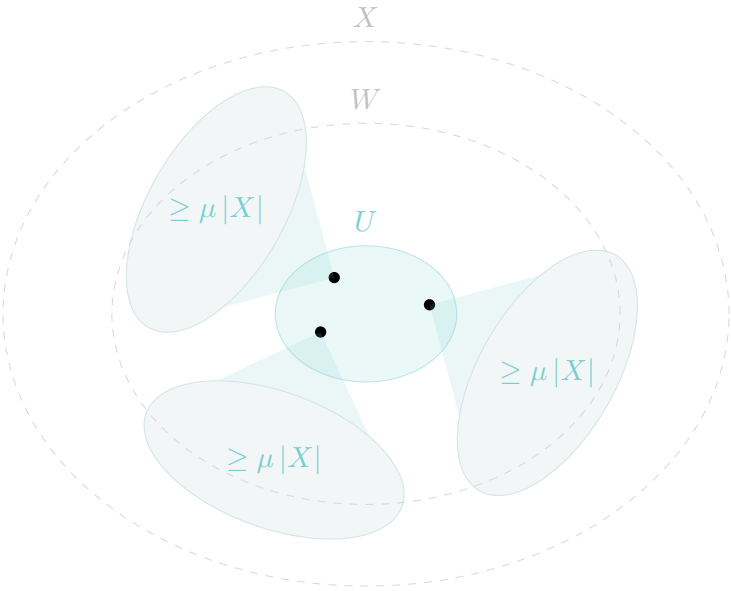
Finally, Lemma 11.21 is quite surprising, so we'll now sketch its proof.

*Proof sketch of Lemma 11.21.* We want to show that whenever our algorithm leads us to a big blue step — meaning that there are many vertices $x \in X$ with reasonably large blue neighborhoods — the blue book it produces satisfies $|S| \geq \ell^{1/4}$. It suffices to show that there *exists* a blue book $(S, T)$ with

$$|T| \geq \frac{1}{2}\mu^{|S|}\,|X| \quad \text{and} \quad |S| = \lceil \ell^{1/4} \rceil$$

(we're choosing a blue book satisfying the first condition with maximal $|S|$, so then that blue book will also satisfy $|S| \geq \lceil \ell^{1/4} \rceil$).

Let $W \subseteq X$ be the set of vertices $x \in X$ with $|N_B(x) \cap X| \geq \mu\,|X|$ (i.e., the set of vertices which satisfy the condition for a big blue step to happen). Then we know that $|W| \geq R(k, \lceil \ell^{2/3} \rceil)$, so we can find a red clique of size $k$ or a blue clique of size $\lceil \ell^{2/3} \rceil$ in $W$. But the first case is impossible because the *entire* graph doesn't have a red clique of size $k$; this means we can find a blue clique $U$ of size $|U| = \lceil \ell^{2/3} \rceil$ in $W$.




Now we'll consider $d_B(U, X \setminus U)$ (in words, $U$ is a blue clique living in $W$ and therefore $X$, and we're looking at the density of blue edges between $U$ and the rest of $X$). By definition, we have

$$d_B(U, X \setminus U) = \frac{e_B(U, X \setminus U)}{|U| \cdot |X \setminus U|} \geq \frac{|U|\,(\mu\,|X| - |U|)}{|U| \cdot |X \setminus U|}$$

(every vertex in $U$ is also in $W$, so it has at least $\mu\,|X|$ edges into $X$; some of those edges 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 $\frac{1}{k}$ is very loose — $|X|$ is at least $R(k, \lfloor \ell^{3/4} \rfloor)$ and $|U|$ is $\lceil \ell^{2/3} \rceil$, so this error term is *tiny*.)

We want $S$ to be a blue clique, so we'll obtain it by looking inside $U$ — we take $S$ to be a uniformly random subset of $U$ of size $|S| = \lceil \ell^{1/4} \rceil$. We then take $T$ to be the set of vertices $y \in X \setminus U$ such that $S \subseteq N_B(y)$ — i.e., all vertices $y$ such that $S$ only has blue edges to $y$ (this means $S$ 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 look at every vertex $y \in X \setminus U$ and consider the probability that it gets included in $T$, meaning that all edges from $y$ to $S$ are blue. We have

$$\mathbb{P}[\text{all edges from } y \text{ to } S \text{ are blue}] = \frac{\binom{|N_B(y) \cap U|}{\lceil \ell^{1/4} \rceil}}{\binom{|U|}{\lceil \ell^{1/4} \rceil}}$$

(the number of good choices for $S$ divided by the total number; note that $|U| = \lceil \ell^{2/3} \rceil$), so we have

$$\mathbb{E}[|T|] = \sum_{y \in X \setminus U} \frac{\binom{|N_B(y) \cap U|}{\lceil \ell^{1/4} \rceil}}{\binom{|U|}{\lceil \ell^{1/4} \rceil}}.$$

Now, the function $a \mapsto \binom{a}{b}$ is convex (if you define $\binom{a}{b}$ to be 0 whenever $b > a$, even for non-integers). So by convexity, we can bound this sum by replacing the quantities $|N_B(y) \cap U|$ in these binomial coefficients with their *average*, which is

$$d_B(U, X \setminus U) \cdot |U| \geq \left(\mu - \frac{1}{k}\right) |U|.$$

This gives a bound of

$$\mathbb{E}[T] \geq |X \setminus U| \cdot \frac{\binom{(\mu - 1/k)|U|}{\lceil \ell^{1/4} \rceil}}{\binom{|U|}{\lceil \ell^{1/4} \rceil}},$$

and after some calculations (using the fact that $|U| = \lceil \ell^{2/3} \rceil$), this gives

$$\mathbb{E}[|T|] \geq \frac{1}{4} \mu^{\lceil \ell^{1/4} \rceil} |X| = \frac{1}{2} \mu^{|S|} |X|.$$

(To intuitively see where this comes from, you can imagine estimating $\binom{a}{b} \approx \frac{a^b}{b!}$. Then the factorials and the powers of $|U|$ cancel out, and we're left with $\lceil \ell^{1/4} \rceil$ factors of roughly $\mu$.) $\square$