Diagonal Ramsey numbers and high dimensional geometry

Talk by Julian Sahasrabudhe

Notes by Sanjana Das

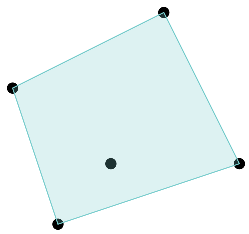
February 29, 2024

This is joint work with Marcelo Campos, Simon Griffiths, and Rob Morris.

§1 Introduction

§1.1 Ramsey's theorem

We'll start the story at the beginning, in a bit of an unusual place. The story goes back to an observation in the 1930s by Esther Klein about points in the plane — if we take any five points (with no three collinear), then there always exist four in convex position.



She then asked, does this hold for numbers greater than 4? For example, is there some number of points such that we can always find 25 in convex position?

> **Question 1.1** (Klein). For all $k \in \mathbb{N}$, does there exist some $n(k) \in \mathbb{N}$ such that for any set of points $X \subseteq \mathbb{R}^2$ with no three points on a line and of size $|X| \geq n(k)$, there exist k points in X in convex position?

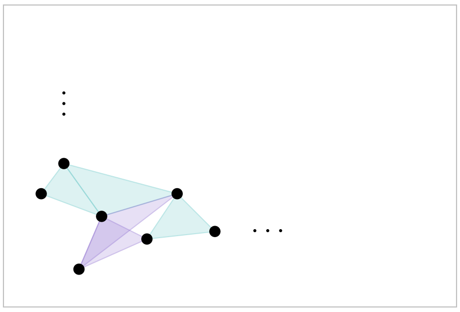
She asked this question to her two friends Paul Erdős and George Szekeres, who realized that this problem in a sense has nothing to do with geometry at all, and actually follows from a much more general principle.

Theorem 1.2 (Erdős–Szekeres 1935)

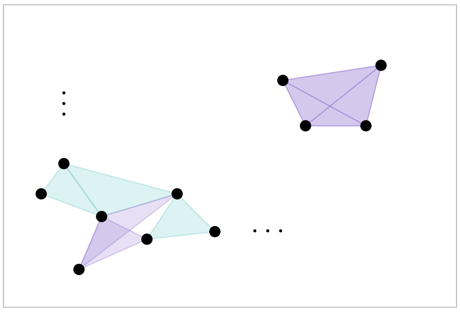
The answer is yes (i.e., for every k, there exists $n(k)$ such that for any set of at least $n(k)$ points with no three on a line, there exist k in convex position).

That more general principle is something discovered just a few years earlier, by Frank Ramsey.

Before we state the theorem formally, imagine that we have a world of abstract points. And what we're interested in is r-sets of these points (i.e., subsets of size r). Imagine that I hand you this world of points, and you color each of these r-sets with two colors — red and blue — in any way you want.



Then I (or rather, Ramsey) claim that for any k, I can choose the size of this world big enough such that no matter how you color, I can find a little sub-world of size k such that in that sub-world, all the subsets of size r are the same color.



So I have a very structured small pouch inside the much bigger, completely arbitrarily colored world.

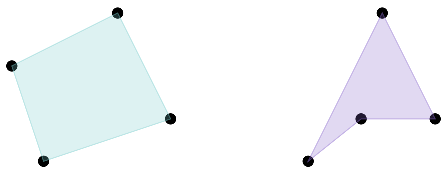
Here's a more formal statement of the theorem.

Theorem 1.3 (Ramsey 1930)
For all $k, r \geq 1$, there exists $R^{(r)}(k) \in \mathbb{N}$ such that for every $n \geq R^{(r)}(k)$ and every coloring $\chi \colon \binom{[n]}{r} \to \{\text{red}, \text{blue}\}$ (i.e., a coloring of the r-subsets of $[n]$ with red and blue), there exists some $X \subseteq [n]$ of size k such that all r-subsets of X are the same color.

We'll mainly be interested in the case $r = 2$, where we have an ordinary graph. But we'll use this theorem with $r = 4$ to answer the question of Klein.

Proof of Theorem 1.2. We're given k, and we want to choose $n(k)$ such that given k points, we can find k in convex position. Assume $k \geq 5$, and choose $n(k) = R^{(4)}(k)$ (the number coming from Ramsey's theorem).

Now imagine we have at least $n(k)$ points in the plane. We color the 4-tuples of points in this set depending on whether the four points are in convex position — we color a 4-tuple red if they're in convex position and blue otherwise.



And now we claim we're just done — by applying Ramsey's theorem, we can find k points such that all the sets among them are of the same color. But $k \geq 5$, so they can't all be blue — because there's always a convex quadruple among 5 points. So that means they all have to be red. But if every four points are in convex position, then the whole set has to be in convex position, giving us k points in convex position. $\square$

This is a really nice application of Ramsey theorem. But in this talk, we'll focus on *quantitative* questions about the Ramsey numbers.

Question 1.4. How *big* are the Ramsey numbers — i.e., how large does $R^{(r)}(k)$ need to be?

We'll only work with graphs (i.e., the case $r = 2$). We'll also define a slightly more general version, with *two* parameters (instead of a single parameter k).

Definition 1.5 (Ramsey numbers)**.** We define $R(\ell, k)$ as the minimum n such that for every red/blue colorings of the edges of K_n, there is either a blue K_ℓ or a red K_k.

This is a slightly asymmetric definition — we're only looking for something of size ℓ in blue and k in red. We'll always assume that $\ell \leq k$. We let $R(k) = R(k, k)$; we refer to these as the *diagonal* Ramsey numbers.

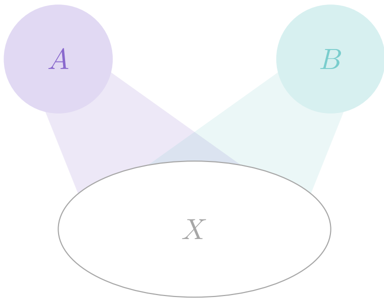
§1.2 A first upper bound

One of the first reasonable quantitative bounds comes from the same Erdős–Szekeres paper.

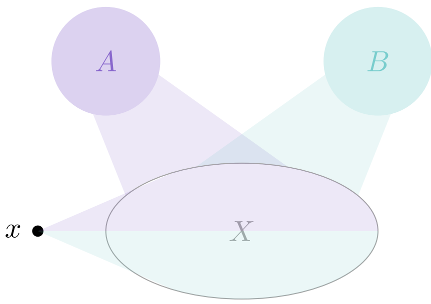
Theorem 1.6 (Erdős–Szekeres 1935)
We have $R(k) \leq 4^k$.

In fact, they do a bit better than this, but we'll get to that later.

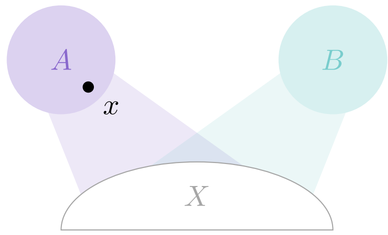
Proof. Suppose we have a graph with $n = 4^k$ vertices; we want to find a monochromatic K_k. Imagine the following picture: we'll have a 'world' X (which at the start is the entire graph), and as the algorithm chugs along, we'll hold in one fist a collection of 'red' vertices A, and in the other fist a collection of 'blue' vertices B. We'll insist that all edges within A and between A and X are red, and all edges within B and between B and X are blue. Initially A and B are empty, and X contains all vertices.



In a step of the algorithm, we'll grab a vertex $x \in X$ and pull it out. It'll be joined to a bunch of vertices in X with red edges, and a bunch of others with blue edges.



And there's two scenarios — one of these two colors constitutes at least 50% of our world X. If at least 50% of the vertices in X are *red* neighbors of x, then we shrink X to just consist of those red neighbors, and we put x in A. Similarly, if at least 50% are *blue* neighbors, then we shrink X to just consist of those blue neighbors and put x in B.



At each step, our world shrinks by a factor of $\frac{1}{2}$ (at worst). And if we take k red steps then we're done (we've found a red K_k), so we can take at most $k-1$ red steps; and in total, the red steps shrink our world by at most $(\frac{1}{2})^k$. Similarly, if we take k blue steps then we're done; so the blue steps shrink our world by at most $(\frac{1}{2})^k$. This means if we're not done, then we'll have at least $(\frac{1}{2})^k \cdot (\frac{1}{2})^k \cdot n$ vertices remaining in the world. And if $n \geq 4^k$ then this is at least 1, so we can keep going with the algorithm — and this means eventually we're going to succeed (i.e., to fill up one of the two fists with k vertices). $\square$

This is a very simple algorithm, but we'll see soon that it's rather hard to improve.

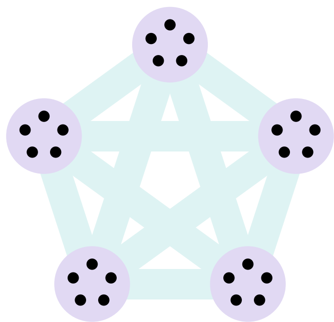
§1.3 A lower bound

For now, we've got an upper bound of 4^k; is it a *good* bound? To get some feel for this, let's try to prove a lower bound.

Here's the first lower bound you might get:

> **Claim 1.7** — We have $R(k) > (k-1)^2$.

Proof. We need to find a graph on $(k-1)^2$ vertices with no monochromatic clique on k vertices. To get such a graph, we take $k-1$ blobs of size $k-1$; inside each of these blobs we color everything red; and *between* these blobs we color everything blue.



Then there's no monochromatic K_k — you can't stick k vertices inside a red blob to get a red clique, and you don't have enough blobs to get a blue clique. $\square$

For some time, it wasn't understood whether the Ramsey numbers even grow exponentially or polynomially. So it came as somewhat of a shock when Erdős proved an exponential lower bound.

Theorem 1.8 (Erdős 1947)
We have $R(k) \geq 2^{k/2}$.

This means the exponential upper bound is closer to the truth. The proof of this was revolutionary, and one of the first uses of the probabilistic method in combinatorics. The idea is very different from the above construction — instead of giving an explicit coloring, we just color the edges of our graph *randomly*, and we show that among the whole universe of graph colorings that we could have made, there *exists* one with the property we want.

Proof. Let $n = 2^{k/2}$, and color each edge randomly and independently (e.g., by flipping a fair coin). We want to bound the probability that there is a monochromatic k-clique. And we can do this by a big union bound — we go to each clique of size k in our universe, and ask, is it monochromatic? There's $\binom{n}{k}$ of these k-cliques. And once we fix k points, what's the probability that they're monochromatic? We have $\binom{k}{2}$ edges and either all need to be red or all need to be blue, which means

$$\mathbb{P}[\text{a given } k\text{-clique is monochromatic}] = 2^{-\binom{k}{2}+1}.$$

Then the union bound gives

$$\mathbb{P}[\text{exists a monochromatic } k\text{-clique}] \leq \binom{n}{k} 2^{-\binom{k}{2}+1} \leq \left(\frac{en}{k} \cdot 2^{-\frac{k-1}{2}+\frac{1}{k}}\right)^k.$$

For large k, since $n = 2^{k/2}$ the expression inside the parentheses is less than 1, so this whole probability is less than 1; this means there exists a graph with no monochromatic k-clique. $\square$

If we're a bit more careful with the bounding, there's actually an extra k in the denominator of the above expression, and we can use this to improve the bound to

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

This bound is from 1947, but all that's been done in the years since is that we've moved the $\sqrt{2}$ from the denominator to the numerator — the best bound we know is

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

This is quite far from 4^k, but even improving this bound by a factor of $1 + \varepsilon$ seems to be a serious problem.

§1.4 Off-diagonal Ramsey numbers

Before we continue discussing the diagonal Ramsey numbers, we'll take a detour and talk about *off-diagonal* Ramsey numbers — first because they're one of our successes in this area, and second because there's a huge number of smart ideas that go into proving these bounds.

For $\ell = 3$, we understand $R(3, k)$ quite well. In fact, we know it up to a factor of 4 — we have

$$\left(\frac{1}{4} + o(1)\right) \frac{k^2}{\log k} \leq R(3, k) \leq (1 + o(1))\frac{k^2}{\log k}.$$

The history of this result touches on a lot of major contributions to probabilistic combinatorics. The upper bound (without the right constant) comes from a very important paper of Ajtai–Komlos–Szemerédi (1980). The right constant is due to Shearer (1980). It comes from the analysis of a random greedy process — after some initial reductions, we find our red clique by randomly selecting a vertex, removing all its blue neighbors and throwing it out of our graph, then randomly selecting the next vertex and doing the same there, and so on. And this actually works incredibly well.

Remark 1.9. As a plug for Julian's talk tomorrow, this circle of ideas was the major influence in his recent improvement — together with Campos, Jenssen, and Michelen — on sphere-packing lower bounds.

The lower bounds began with important ideas of Erdős from the 1960s and 1980s; then the lower bound was solved up to constants by Kim in the 1990s. The sharp constant comes from analyzing a random *triangle-free* process. The idea is that we can't just take a completely random graph — it'd have too many triangles. Instead, we start with an empty graph and then throw in edges randomly, one at a time, *conditioning* on not having any triangles. This condition starts steering the process quite hard, but you can actually control it; this was done by Bohman–Keevash and Fiz Pontiveros–Griffiths–Morris.

Meanwhile $R(4,k)$ has recently been determined up to log factors, by a breakthrough of Mattheus and Verstraete that came up with a totally new idea for a lower bound construction.

§1.5 Diagonal Ramsey numbers

Now we'll return to the diagonal case. So far, we've seen that $R(k) \leq 4^k$ by the Erdős–Szekeres argument from 1935. In fact, they did a bit better, using a different argument that gives a bound of

$$R(k) \leq \frac{4^k}{\sqrt{k}}. \qquad \text{(Erdős–Szekeres 1935)}$$

(The idea is to use a slightly better weight function than $\frac{1}{2}$ when deciding whether to put a vertex in the red or blue set.)

This was pretty much the state of the art until a couple of independent works in the 1980s. First, Rödl showed that

$$R(k) \leq \frac{a4^k}{\sqrt{k}(\log k)^c} \qquad \text{(Rödl 1988)}$$

(for some constants a and c), and then Thomason showed that

$$R(k) \leq \frac{c4^k}{k}. \qquad \text{(Thomason 1988)}$$

These might seem like meager improvements relative to the exponential factor, but they already require new and important ideas. This approach was taken farther by Conlon, who showed we can replace the k in the denominator with *any* polynomial factor — specifically, he showed that

$$R(k) \leq 4^{k-(\log k)^{2-o(1)}}. \qquad \text{(Conlon 2008)}$$

This was further improved by Ashwin Sah, who removed an extra $\log\log$ factor (the $o(1)$ term in the above bound) and proved that

$$R(k) \leq 4^{k-c(\log k)^2}. \qquad \text{(Sah 2023)}$$

If we look at the methods used, this last result is in some sense the limit of the approach. (We'll talk about these methods soon.) So we have to do something different to get beyond this barrier.

That brings us to the main topic of today, an *exponential* improvement.

Theorem 1.10 (Campos–Griffiths–Morris–Sahasrabudhe)
There exists $c > 0$ such that $R(k) \leq (4-c)^k$.

Remark 1.11. This means we've finally moved the constant of 4. However, we're still a long way from the $\sqrt{2}$ in the lower bound — c is pretty small, and the bound is basically $R(k) \geq 3.99^k$.

We haven't talked too much about the off-diagonal Ramsey numbers, but in some sense, this proof actually works *better* on the off-diagonal case — and the off-diagonal case was actually important for understanding the diagonal case.

Theorem 1.12 (Campos–Griffiths–Morris–Sahasrabudhe)
For all $\ell \leq k$, we have $R(\ell, k) \leq e^{-c\ell}\binom{k+\ell}{\ell}$.

Remark 1.13. The expression $\binom{k+\ell}{\ell}$ is one we haven't seen yet, but it's actually the bound that Erdős–Szekeres gives for all (k, ℓ). This bound for the off-diagonal case is new when $\ell \gg \log k$.

§2 Previous methods

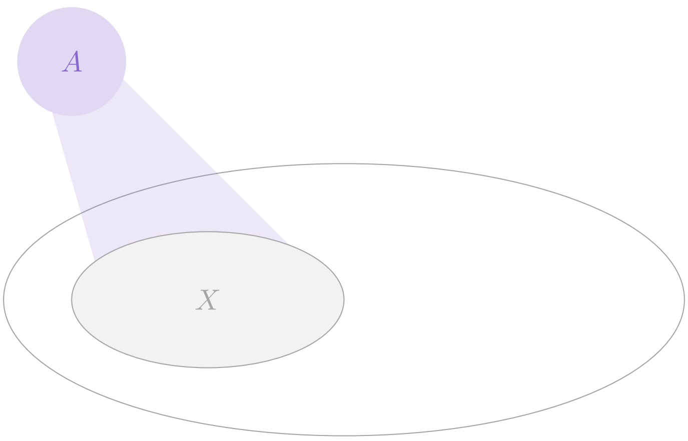
§2.1 The difficulty

First, what makes this problem hard? One way of starting to ask this question is to try to figure out what's lossy about the original Erdős–Szekeres argument, and why we can't improve it easily.

Imagine that we have one fist of red vertices and another of blue, and our algorithm is chugging along looking for a monochromatic K_k. In the algorithm, if we come up with a red K_k at the end of the day, in our accounting we're still paying for finding *almost* a blue K_k (corresponding to the fact that we pay one $(\frac{1}{2})^k$ for the red vertices, and another for the blue vertices).

So it seems that what we'd *like* to do is just look at the graph and say that it's clearly a graph where we want to take red steps — so let's just focus on red steps, and not waste vertices on blue steps.

But here's what's hard about this — imagine that I tell you the first ck steps to take. Once you've taken those steps, you're left with an *exponentially* small piece of the world. How can you possibly look at the original world and say what this exponential piece will look like?

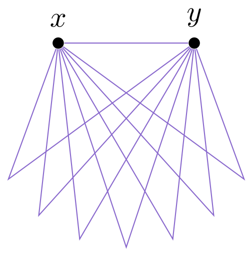


For example, imagine that an enemy gives you a graph where you can take many red steps at the start, but then you're forced to take lots of blue steps. So even if you can say all kinds of things about the *original* graph, you lose those things really quickly when you zoom in.

§2.2 Taking two red steps

Still, what if we try taking a bunch of red steps at once? Taking *two* red steps at once is actually already a very good idea, and the basis of Thomason's argument.

Thomason's idea is to take two red steps at once — this means we grab *two* vertices from X, and we pass to their common red neighborhood.



Trying to analyze this automatically leads us to consider triangles.

First, we want to control degrees. If in X some vertex x has large red degree, then we get an extra win (our world doesn't shrink as much when we move it to A). So we can assume that all red and blue degrees are roughly balanced.

And then we can use Goodman's formula, which states that if all degrees are roughly what they would be in a random graph, then the number of monochromatic triangles is *also* roughly what you'd expect in a random graph. Then by pigeonhole you can find some pair of vertices involved in lots of monochromatic triangles, which lets you take two steps in a row.

The improvement this gives at each step is very small, but we're taking k steps in total, so they accumulate; and this is what accounts for winning the $\sqrt{k}$ factor.

§2.3 Taking multiple red steps

Then Conlon takes this further — what if instead of taking *two* steps at once, we take r steps at once? This corresponds to controlling the number of monochromatic K_{r+1}'s. (We've been talking about taking multiple red steps at a time, but it also could be that we're forced to take a bunch of blue steps too.)

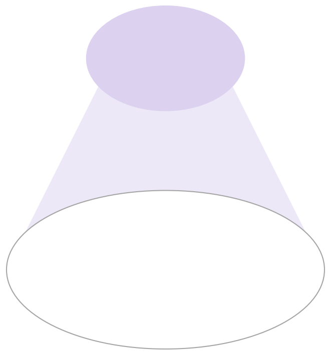
Here there's no analog of Goodman's formula (the analogous statement is false). But there's another set of tools we can use. Suppose that we can control the degrees in the graph, and *also* the number of monochromatic C_4's (cycles of length 4). Then we can use ideas going back to Thomason and Chung–Graham–Wilson — if the degrees are all roughly 50% and the number of monochromatic C_4's is roughly what you'd expect in a random graph, then the number of monochromatic K_{r+1}'s are *also* roughly what you'd expect. So you can get a count of these monochromatic K_{r+1}'s and use pigeonhole to find r good vertices to take a step with.

Then the work of Sah improves this quasirandomness step.

But the problem is this really seems to break down when $r \approx \log k$. (This accounts for one of the $\log k$ factors in the bound.) And to get an exponential improvement, we really want to take a *linear* number of red steps in a row (passing to an *exponentially* small piece of the graph.)

§2.4 A new perspective

For this, we'll change the perspective slightly. We're interested in taking a bunch of red steps at once; what does that look like as a structure? If we want to take t red steps, we essentially want a red clique of size t, and another set such that all edges between them are red (we don't know anything about what's inside that set). And we want $t \approx ck$.



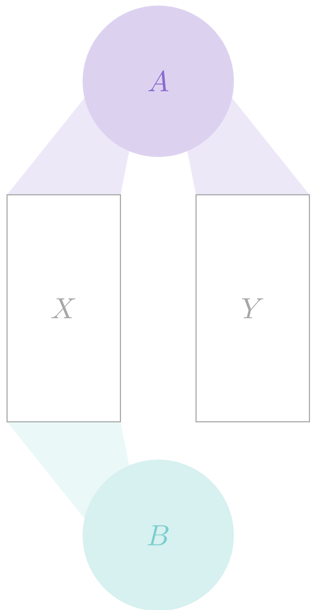
Suppose we've found this structure, and suppose that we can guarantee that the world inside the other set has size at least $R(k, k-t)$. Then we can actually finish — because either we have a blue K_k in there (in which case we're just done), or we have a red K_{k-t} in there, which we can complete into a red K_k by just adding back our t vertices.

So all this is to say that this is the kind of structure we want to look for — it's the same in spirit as taking red steps. So we'll try to focus on an algorithm that builds objects of this form.

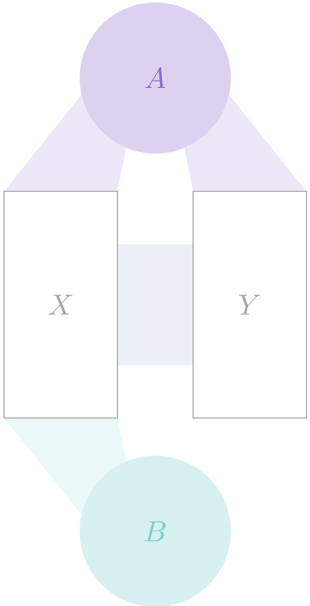
§3 The book algorithm

For the remainder of the talk, we won't prove anything, but we'll set up the algorithm for finding this object. We'll then see what the main difficulty for making it work is, and state a conjecture that in some sense sort of quantifies what's happening with this obstacle. The authors weren't able to prove this conjecture; instead, they proved a weaker version and pushed slack off into other parts of the argument.

As before, we'll maintain a certain picture as time goes on. Previously, we had three sets A, B, and X. Now we'll have sets A, B, X, and Y — where A is a handful of red vertices joined to both X *and* Y in red, and B is a handful of blue vertices joined to just X in blue.

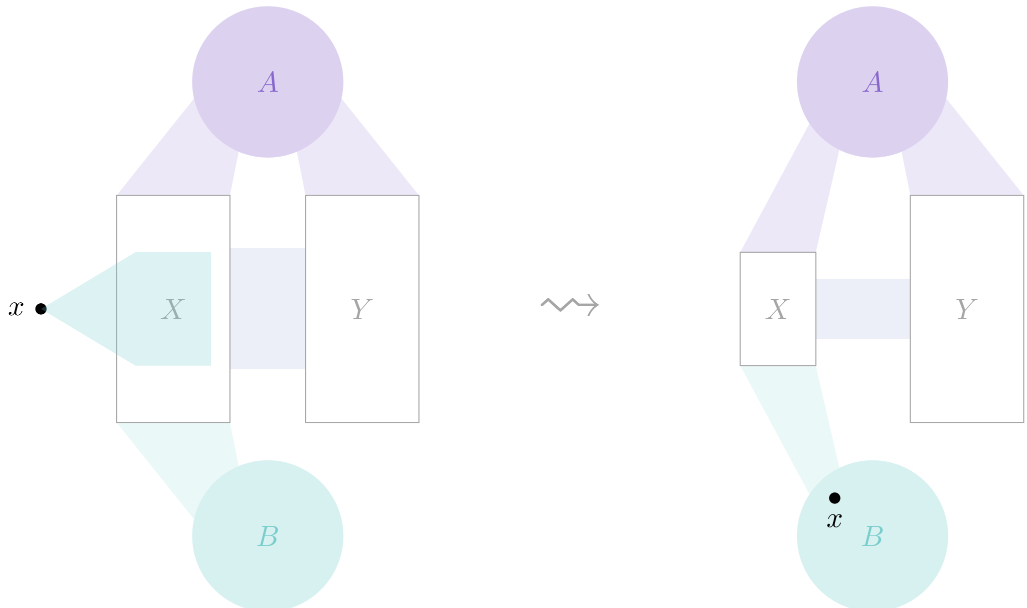


And as the algorithm zooms into a small portion of the graph, we'd additionally like to control the density of red edges between X and Y — so we'll track the evolution of this red density as well.

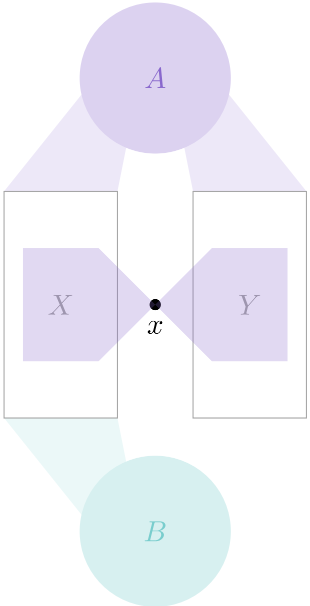


Our 'book' — the structure we're looking for — is going to consist of A on one side and the rest of the world (X and Y) on the other. We'll either find this or a blue clique in B. (So we're looking for a red book vs. a blue clique, rather than a clique vs. a clique.)

As before, we'll think of X as our reservoir (where we pull vertices out of). Taking a blue step will essentially be the same as before — we choose a vertex $x \in X$ with a bunch of blue neighbors in X, and we shrink X to just consist of those blue neighbors (and keep Y as it is). There's a bit of a wrinkle because we need to track the red density between X and Y, but this turns out to not be too much of a problem (and we won't focus on it here).



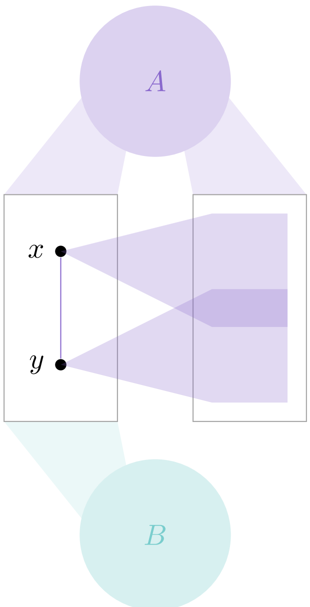
The meat of the argument is in the red steps. Suppose we have a vertex $x \in X$ with big red degree to X. Now for a red step, it's not enough to just consider what's happening in X — because we also need to maintain that A is complete in red to Y. So we also have to consider the red neighborhood of x in Y — and we'd like to shrink both X and Y to be those red neighborhoods, and put x into A.



There's a wrinkle — what if that red neighborhood is too small, so that Y shrinks too much? But we won't worry about that for now.

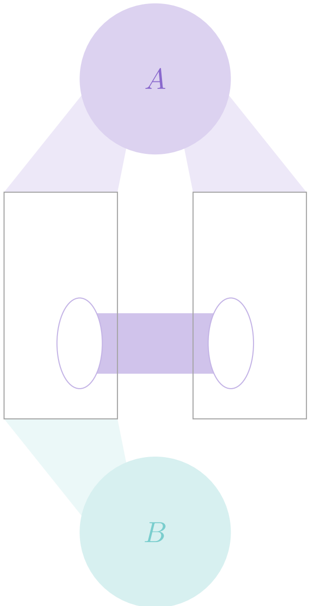
The real difficulty is in controlling the red density between X and Y when we do this. This is really subtle — for example, the regularity lemma is a kind of thing you can use to control densities, but it's not helpful here (the graph would have to be far too big for it to be useful).

And there's counterexamples to the sorts of things we'd hope to be true. For example, imagine your enemy hands you a setup where every time we have vertices $x, y \in X$ joined by a red edge, the red neighborhoods of x and y in Y are both $\frac{1}{2}$ of Y, but these neighborhoods — rather than being totally random (in which case they'd have an intersection of $\frac{1}{4}$) — slightly repulse each other, so that the intersection has size slightly smaller than $\frac{1}{4}$. Specifically, suppose that this intersection has size $\frac{1}{4}(1 - \frac{c}{k})$ for some constant c.



This would be a total disaster for us — then when we take a red step, all the neighborhoods of x are poking outside of what we want a bit too much. And so we lose a factor of $1 - c/k$, and this accumulates over k steps and really spins out of control.

You might hope that this sort of thing never exists — maybe this just can't happen. Unfortunately, it can. But you'll find that the first examples you think about where it does happen have some sort of 'macroscopic clustering' — there will be lots of edges xy in X whose neighborhoods push apart, but you'll also see clumps of the graph that have much more edges than you'd expect from a random graph.



And then you can zoom in on these clumps, and we get a huge bump in density that's good for other reasons.

So you might think, maybe this sort of bad situation *only* happens when we have this macroscopic clustering. But that's also false — there are also graphs where you get 'mezzoscopic clustering' (more than microscopic, but not really macroscopic).

But then you might hope that maybe when this situation happens, we always get this mezzoscopic clustering that we can still take advantage of (to get a density increment). The authors spent a long time trying to prove this, but couldn't (we'll state a toy version of this statement soon).

What they *did* prove is a weak version of this implication. And it turns out that this weak version is not good enough for the diagonal case, but it works much better on the off-diagonal case — so you can get exponential improvements for something like $R(k/10, k)$. And then for the diagonal case, you run an algorithm similar to this, but at a loss. If we pushed that algorithm all the way to the end, it'd be a really complicated way to prove $R(k) \leq 4.1^k$. But instead of pushing all the way to the end, we push down to $R(k/10, k)$ — the algorithm lets you force $R(k)$ to depend on $R(k/10, k)$, and the exponential win in $R(k/10, k)$ makes up for the loss in the algorithm.

§4 A conjecture in high dimensional geometry

That's all we'll say about the proof, but we'll finish with a sort of toy example that should shed light on the conjecture the authors tried hard to prove. (This is where geometry comes in.)

Consider a random variable $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$ (we think of d as going to ∞). For simplicity, we'll assume that X is centered and isotropic — so $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$, and $\mathbb{E}[X_iX_j] = 0$ for all $i \neq j$.

Here's two definitions (the first one is standard, and the second is particular to this problem).

Definition 4.1. We say X is *δ-subGaussian* if for all $\theta \in \mathbb{R}^d$ with $\|\theta\|_2 = 1$, we have

$$\mathbb{P}[\langle X, \theta \rangle > t] \leq e^{-\delta t^2}.$$

Think of θ as representing a direction on the d-dimensional unit sphere. Then $\langle X, \theta \rangle$ is some random variable; and this condition says that its tails decay at least as fast as a Gaussian's.

Definition 4.2. We say X is *ε-lopsided* if taking Y to be an independent copy of X, we have

$$\mathbb{P}_{X,Y}[\langle X, Y \rangle \leq -\varepsilon\sqrt{d}] \geq \frac{2}{3}.$$

Here Y is an independent copy of X, and $\langle X, Y \rangle$ varies on a scale of $\sqrt{d}$ and is expected to be 0. So we're looking at the probability that it's noticeably below this; and the point of that definition is that this probability is much bigger than you'd expect. (The value of $\frac{2}{3}$ doesn't matter; it's just some constant bigger than $\frac{1}{2}$.)

Conjecture 4.3 — For all $\varepsilon, \delta > 0$, for all large d, there does not exist a random variable $X \in \mathbb{R}^d$ that is both δ-subGaussian and ε-lopsided.

How does this correspond to the picture we had earlier? SubGaussianity is sort of like saying you're not clustered in any way (so it corresponds to the mezzoscopic clustering condition). And lopsidedness is capturing the repulsion of neighborhoods — in some sense, $\langle X, Y \rangle$ is precisely capturing the error term in the overlap of neighborhoods (which we need to not be $-c/k$).