1 Introduction

The goal of this talk is to discuss the recent breakthrough of Campos, Griffiths, Morris, and Sahasrabudhe on upper bounds for diagonal Ramsey numbers (as well as a few of the follow-up results). I will not go into their proof in any level of detail; my hope is simply to explain what sorts of ideas they use, as well as where their new insights are most important.

Ramsey theory is a branch of combinatorics that studies order and disorder. The underlying mantra of the field, as articulated by Theodore Motzkin, is that "complete disorder is impossible"—any sufficiently large system must have a large, highly structured subsystem. The prototypical example of a Ramsey-theoretic statement is *Ramsey's theorem*, from which the field derives its name.

Theorem 1 (Ramsey, 1929). *For every integer $k \geqslant 2$, there exists some positive integer N such that any two-coloring of the edges of the complete graph*[1] *K_N contains a monochromatic K_k.*

In other words, no matter how we assign the edges of K_N a color, say red or blue, we can always find k vertices such that all edges between them receive the same color. That is, any such coloring, no matter how unstructured, contains a highly structured subcoloring.

Much of the modern research in Ramsey theory is concerned with the *quantitative* aspects of such statements: how large is the integer N in Theorem 1 as a function of k? Formally, we make the following definition.

Definition 2. The *Ramsey number* $r(k)$ is the least integer N such that every two-coloring of the edges of K_N contains a monochromatic K_k.

With that said, let us turn to the quantitative aspects of Theorem 1, that is, to the determination of the function $r(k)$ from Definition 2. The exact value of $r(k)$ is only known for $k \leqslant 4$, and it currently seems completely hopeless to obtain an exact formula for $r(k)$, so let us content ourselves with asymptotic bounds as $k \to \infty$. Essentially every proof of Theorem 1 yields (at least implicitly) an upper bound on $r(k)$, by proving the existence of *some* integer N. The original proof of Ramsey gave a bound of $r(k) \leqslant k!$, but Ramsey wrote "I have little doubt that [this upper bound is] far larger than is necessary". Indeed, a few years later, Erdős and Szekeres proved the following stronger bound.

Theorem 3 (Erdős and Szekeres, 1935). *$r(k) \leqslant 4^k$ for every $k \geqslant 2$.*

For about a decade, it was believed that this bound was also far larger than is necessary, namely that $r(k)$ should grow subexponentially as a function of k. However, Erdős dispelled this belief by proving an exponential lower bound.

Theorem 4 (Erdős, 1947). *$r(k) \geqslant \sqrt{2}^k$ for every $k \geqslant 2$.*

[1]Recall that the complete graph K_N has N vertices, and all of the $\binom{N}{2}$ possible edges are present.

After this breakthrough, progress stalled for 75 years. There were a number of improvements to these bounds over the years, including important results of Graham–Rödl, Thomason, Conlon, and Sah, but all of these improvements only affected the lower-order terms, and did not improve either of the exponential constants $\sqrt{2}$ and 4. This impasse finally ended with a breakthrough of Campos, Griffiths, Morris, and Sahasrabudhe.

Theorem 5 (Campos, Griffiths, Morris, and Sahasrabudhe, 2023)**.** *There exists a constant $\delta > 0$ such that $r(k) \leqslant (4-\delta)^k$ for all $k \geqslant 2$. Concretely, $r(k) \leqslant 3.993^k$ for all sufficiently large k.*

In the year and a half since their paper appeared on the arXiv, there have been some two major follow-up results. Firstly, Gupta, Ndiaye, Norin, and Wei improved the constant appearing in Theorem 5, proving that $r(k) \leqslant 3.8^k$ for all sufficiently large k. Although their proof is closely related to that of Campos et al., they also introduced some important new ideas. In particular, they recast the entire analysis in a different language, which is both somewhat simpler conceptually and which lends itself to easier numerical optimization.

Secondly, and very recently, a new proof of Theorem 5 was given by Balister, Bollobás, Campos, Griffiths, Hurley, Morris, Sahasrabudhe, and Tiba. Their proof is much more conceptual, and also has the advantage of working for any number of colors[2]. Rather remarkably, the key lemma underlying their new proof is purely geometric, concerning the self-correlation properties of probability distributions on high-dimensional spheres.

In the rest of the talk, I will try to cover most of the key ideas that go into the proofs of Campos et al. and Balister et al. However, the easiest way to understand many of these ideas is to see how they arise naturally in the (much simpler) proof of Theorem 3, so we begin by discussing this proof in three different ways, introducing three of the key ideas: the use of off-diagonal Ramsey numbers, the use of book graphs, and the algorithmic perspective.

2 Three ways of looking at Theorem 3

2.1 Off-diagonal Ramsey numbers

Before proceeding with the proof, we generalize the notion of Ramsey numbers from Definition 2. Here and throughout, we denote by $V(K_N)$ and $E(K_N)$ the vertex set and edge set, respectively, of the complete graph K_N.

Definition 6. Given integers $k, \ell \geqslant 2$, the *off-diagonal Ramsey number* $r(k, \ell)$ is the least integer N such that every two-coloring of $E(K_N)$ with colors red and blue contains a red K_k or a blue K_ℓ.

Note that $r(k, \ell) = r(\ell, k)$ as the colors play symmetric roles, and that $r(k) = r(k, k)$. The quantity $r(k)$ is often called the *diagonal Ramsey number*.

[2]Theorem 1 remains true even if we color the edges of K_N by more than two colors, so it is natural to study the asymptotics of such multicolor Ramsey numbers. There is a great deal to be said on this topic, but for simplicity I focus on the case of two colors for the remainder of this talk.

With this terminology, we can prove Theorem 3. In fact, we will prove the following more precise result.

Theorem 7 (Erdős and Szekeres, 1935). *For all integers $k, \ell \geqslant 2$, we have*

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

In particular,

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

Proof. We proceed by induction on $k+\ell$, with the base case $\min\{k,\ell\} = 2$ being trivial. For the inductive step, the key claim is that the following inequality holds:

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

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

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

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

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

We can now complete the induction. By (1) and the inductive hypothesis, we find that

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

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

2.2 Book graphs

Definition 8. Let t, m be positive integers. The *book graph* $B_{t,m}$ consists of a copy of K_t, plus m additional vertices which are adjacent to all vertices of the K_t, but not adjacent to one another. Equivalently, $B_{t,m}$ is obtained from the complete bipartite graph $K_{t,m}$ by adding in all the $\binom{t}{2}$ possible edges in the side of size t. Equivalently, $B_{t,m}$ consists of m copies of K_{t+1} which are glued along a common K_t.

Note that two important special cases are $m = 1$, where $B_{t,1}$ is simply the complete graph K_{t+1}, and $t = 1$, where $B_{1,m}$ is simply the star graph $K_{1,m}$, consisting of one vertex joined to m others (and no other edges). The "book" terminology comes from the case $t = 2$, in which case $B_{2,m}$ consists of m triangles sharing an edge, which looks, to some extent, like a book with m triangular pages. Continuing this analogy, the K_t in $B_{t,m}$ is called the *spine*, and the m additional vertices of $B_{t,m}$ are called the *pages*. We will often denote a book as a pair of sets (A, Y), where A is the spine and Y comprises the pages.

The reason book graphs are important in the study of Ramsey numbers comes down to the following simple observation.

Lemma 9. *Suppose that a two-coloring of $E(K_N)$ contains a monochromatic red copy of $B_{t,m}$, where $m \geqslant r(k-t, \ell)$. Then this coloring contains a red K_k or a blue K_ℓ.*

Proof. Let A be the spine of the book, and let Y be its pages. By assumption, $|Y| = m \geqslant r(k-t, \ell)$, so Y contains a blue K_ℓ or a red K_{k-t}. In the former case we are done, and in the latter case, we may add A to the red K_{k-t} to obtain a red K_k. □

This proof should look familiar—we have already encountered the same idea in the proof of Theorem 7, where we implicitly used the $t = 1$ case of Lemma 9. Indeed, in that proof, we showed that if a coloring contains a red star with $r(k-1, \ell)$ leaves, then it contains a red K_k or a blue K_ℓ. The only new idea in Lemma 9 is that we don't need to consider a single vertex (i.e. the case $t = 1$), but may take an arbitrary book.

Although the idea of Lemma 9 basically goes back to the work of Erdős and Szekeres, it was first formulated in essentially this language by Thomason, who used Lemma 9 to propose a natural approach to improving the upper bounds on $r(k)$. Namely, if one can show that every two-coloring of $E(K_N)$ contains a monochromatic $B_{t,m}$, for some appropriate parameters t and $m \geqslant r(k-t, k)$, then one can plug this into Lemma 9 and conclude that $r(k) \leqslant N$. Again, this is essentially the approach we used in the proof of Theorem 7, where a simple argument based on the pigeonhole principle showed that any coloring of $E(K_N)$ contains a large monochromatic star, that is, a monochromatic book with many pages and a spine of size $t = 1$. The idea behind Thomason's program is that perhaps for larger values of t, more sophisticated arguments than the pigeonhole principle could yield stronger results, and improve the upper bounds on $r(k)$. This general framework has been quite successful, and the three prior improvements on Theorem 3, before the work of Campos et al., all used variants of this idea.

2.3 The algorithmic lens

One of the many new ingredients introduced by Campos et al. is the following simple idea: rather than searching for some *specific* book $B_{t,m}$, they define an exploration algorithm for finding *some* book, and then prove that regardless of which book is found, the parameters involved are good enough to plug into Lemma 9. Although this idea is almost a triviality, this change of perspective is crucial for the proof of Theorem 5.

Before we discuss this exploration algorithm—which they termed the *book algorithm*—let us first rephrase the proof of Theorem 3 as an exploration algorithm, the *Erdős–Szekeres algorithm*. Let us fix a two-coloring of $E(K_N)$. We assume that this coloring has no monochromatic K_k, and our goal is to eventually obtain a contradiction if N is sufficiently large. For the moment we only seek to get a contradiction if $N > 4^k$, and thus reprove Theorem 3.

For a vertex $v \in V(K_N)$, we write $N_R(v)$ for the *red neighborhood* of v, that is, the set of vertices $w \in V(K_N)$ such that the edge vw is colored red. Similarly, $N_B(v)$ denotes the *blue neighborhood* of v.

In the Erdős–Szekeres algorithm, we maintain three disjoint sets $A, B, X \subseteq V(K_N)$; the sets A and B will grow throughout the process, whereas X will shrink. The key property we maintain is that (A, X) is a red book, and (B, X) is a blue book; that is, A is completely red, B is completely blue, all edges between A and X are red, and all edges between B and X are blue. To initialize the process, we set $A = B = \varnothing$ and $X = V(K_N)$. We now repeatedly run the following steps.

1. If $|X| \leqslant 1$, $|A| \geqslant k$, or $|B| \geqslant k$, stop the process.

2. Pick a vertex $v \in X$, and check whether v has at least $\frac{1}{2}(|X| - 1)$ red neighbors in X.

3. If yes, move v to A and shrink X to the red neighborhood of v. That is, update $A \to A \cup \{v\}$ and $X \to X \cap N_R(v)$, and keep B the same. Call this a *red step*.

4. If not, then v has at least $\frac{1}{2}(|X| - 1)$ blue neighbors in X. We now move v to B, and shrink X to the blue neighborhood of v. That is, we update $B \to B \cup \{v\}$ and $X \to X \cap N_B(v)$, and keep A the same. Call this a *blue step*.

5. Return to step 1.

By the way we update the sets, we certainly maintain the key property that (A, X) and (B, X) are red and blue books, respectively, throughout the entire process, since every time we add a vertex v to A (resp. B), we shrink X to the red (resp. blue) neighborhood of v.

Using this algorithm, we can give an alternative proof of Theorem 3.

"Algorithmic" proof of Theorem 3. Let $N = 4^k$, and fix a two-coloring of $E(K_N)$. Assume for contradiction that this coloring contains no monochromatic K_k. We now run the Erdős–Szekeres algorithm until it terminates.

Suppose first that the algorithm terminated because $|A| \geqslant k$. Throughout the process, we maintain the property that all edges inside A are red. Therefore, if $|A| \geqslant k$ at the end of the process, we have found a monochromatic red K_k, a contradiction. Similarly, if $|B| \geqslant k$ at the end of the process, we have found a blue K_k, another contradiction. We may thus assume that at the end of the process, we have $|A| < k$ and $|B| < k$.

Therefore, the process can only end when $|X| \leqslant 1$. The key observation now is that at every step of the process, we have

$$|X| \geqslant 2^{-|A|-|B|} N. \tag{2}$$

Indeed, this certainly holds when the process begins, for then we have $|A| = |B| = 0$ and $|X| = N$. We can now check that it holds by induction: every time we do a red step, we increase $|A|$ by 1, and decrease $|X|$ to at least[3] $\frac{1}{2}|X|$, thus preserving the validity of (2). Similarly, in a blue step, we increase $|B|$ by 1 and decrease $|X|$ to at least $\frac{1}{2}|X|$, again preserving (2). By induction, we conclude that (2) also holds at the end of the process.

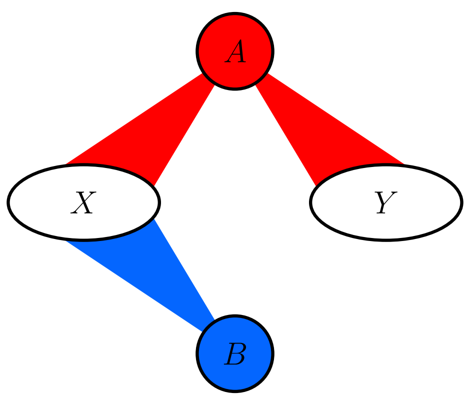
At the end of the process, we thus have

$$N \leqslant 2^{|A|+|B|}|X| < 2^{k+k} \cdot 1 = 4^k,$$

where we plug in our upper bounds $|A| < k, |B| < k, |X| \leqslant 1$. This contradiction completes the proof. □

3 The book algorithm

We now turn the book algorithm of Campos et al. As before, we fix a two-coloring of $E(K_N)$, and assume that there is no monochromatic K_k; our goal is to obtain a contradiction if N is sufficiently large. Throughout the process, we maintain four disjoint sets A, B, X, Y, with the following properties: (A, X) is a red book, (B, X) is a blue book, and (A, Y) is another red book[4]. Thus, the only difference from the Erdős–Szekeres algorithm is the presence of the new set Y. At the end of the process, our goal is to output the pair (A, Y), and to prove that $t \coloneqq |A|$ and $m \coloneqq |Y|$ satisfy $m \geqslant r(k-t, k)$, so that we can apply Lemma 9 to obtain a contradiction. We initialize the process with $A = B = \varnothing$, and $X \sqcup Y$ an arbitrary partition of $V(K_N)$ with $|X| = |Y|$. By permuting the colors if necessary, we may assume that at the beginning of the process, at least half the edges between X and Y are red.



As in the Erdős–Szekeres algorithm, we will iteratively build this picture by moving vertices from X to A or B, and then shrinking X and Y. A move from X to A will be called a red step, and a move from X to B will be called a blue step.

[3]Strictly speaking, we should write here $\frac{1}{2}(|X| - 1)$, although the claimed bound (2) can also be proved inductively by judicious use of ceiling signs. However, from now on, we will start ignoring such additive ± 1 terms.

[4]Equivalently, we could say that (B, X) is a blue book and $(A, X \cup Y)$ is a red book.

What is the advantage of maintaining such a picture? Recall that in the Erdős–Szekeres algorithm, $|X|$ shrinks by a factor of two whenever we do a red or a blue step, hence we end up with $|X| \geqslant 2^{-|A|-|B|}N$ as in (2), yielding the bound $r(k) < 4^k$. However, it is reasonable to hope that since we are imposing "half as many constraints" on Y as on X—that is, we are only maintaining that the edges between A and Y are red, and not that any edges incident to Y are blue—we may be able to obtain better control on $|Y|$. Indeed, we might hope that every blue step does not shrink Y at all, while every red step shrinks Y by only a factor of two, as before, yielding[5] a bound of $|Y| \gtrsim 2^{-|A|}N$.

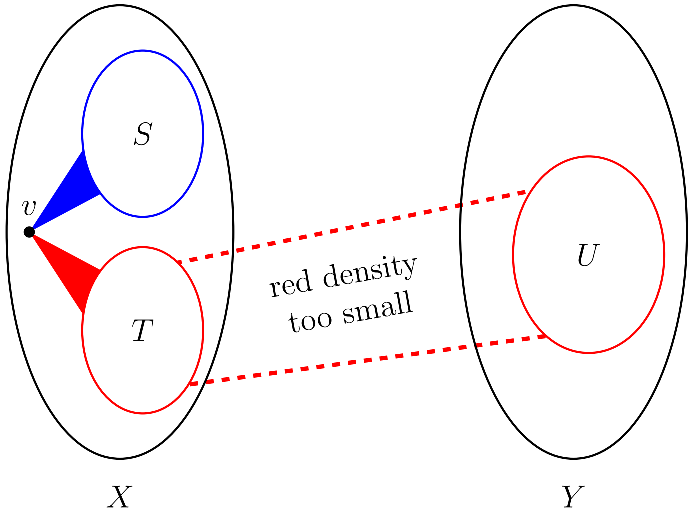
In other words, our goal will be to "sacrifice" the vertices in X, and use them as the fuel we use to build the large red book (A, Y). This approach comes with a fundamental asymmetry between the colors, in marked contrast to the Erdős–Szekeres proof. We will really insist on finding a *red* book (A, Y), and will do our best to build it. Only when doing so is really impossible will we take blue steps.

Because of this, our preferred move would be taking a red step. That is, we would like to pick a vertex $v \in X$, move v to A, and update $X \to X \cap N_R(v)$. Moreover, since we need to maintain that (A, Y) is a red book, we will also need to update $Y \to Y \cap N_R(v)$. In particular, when deciding whether to add a vertex $v \in X$ to A, we need to check not only that v has many red neighbors in X—so that X doesn't shrink too much—but also that v has many red neighbors in Y, so that Y doesn't shrink too much. In particular, we see that in addition to tracking the sizes of A, B, X, and Y, we will also need to track a fifth parameter, the red edge density between X and Y, that is, the fraction of edges in $X \times Y$ that are red. We denote this red edge density by p_R. Again, the reason we are tracking p_R is that if it ever gets too small, then the red steps become very costly, as they start shrinking Y by a larger factor. Hence we would like to ensure that p_R stays fairly large throughout the process.

Unfortunately, when we take a red or a blue step, p_R can change, since red and blue steps shrink X and Y. In order to deal with this, we will refuse to do a red step or a blue step if doing so decreases p_R by an unacceptable amount. This trivially ensures that p_R never becomes too small, but comes with its own major problem: if we are allowed to refuse to do a blue or a red step, there may be no steps available for us to take!

In order to deal with this issue, Campos et al. introduced another type of step, termed *density-boost steps*. The basic picture is as follows:

[5]If we could really obtain such strong control on $|Y|$, we would show that $r(k) \lesssim 2^k$, a dramatic improvement over Theorem 3. Unfortunately, and unsurprisingly, the devil is in the details, and a lot of work is needed to make such an approach work, and the extra complications yield a substantially weaker bound.



In the picture above, we consider whether to take a red step from v, which would entail shrinking X to T and Y to U. As in the picture above, suppose doing so would decrease p_R unacceptably, meaning that the red edge density between T and U is too small. However, if we recall that the global density of red edges between X and Y at the moment is p_R, this tells us something: the edges in $T \times U$ contain a *less than average* fraction of red edges. In particular, in the rest of the picture, there must be a *more than average* fraction of red edges. So by restricting X and Y to (roughly) the complements of T and U, we can boost the red edge density.

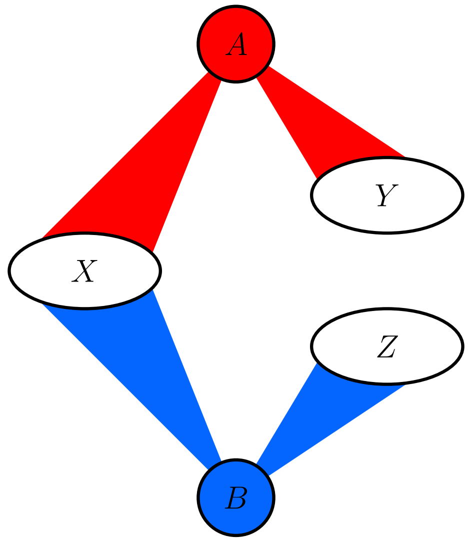
Thus, the three basic steps of the book algorithm are red steps, blue steps, and density-boost steps. The analysis then boils down to understanding how the sizes of the relevant parameters, namely the sizes of the sets A, B, X, Y, as well as the red density d_R, change during a single step. However, this analysis is extremely complicated. Moreover, naive applications of the ideas above actually *do not* work, and cannot prove any bound stronger than roughly $r(k) \leqslant 4.006^k$. To overcome this, Campos et al. begin by using the book algorithm to improve the upper bound on off-diagonal Ramsey numbers in a certain regime, which they can then plug into Lemma 9, which is now quantitatively strong enough to obtain an improved upper bound on $r(k)$. The details are far too involved to give in this talk, but see my exposé for a much more detailed, although still far from complete, proof sketch.

4 The second book algorithm and geometry

Now, I want to discuss the new approach of Balister et al., and to explain how high-dimensional geometry even enters the picture. In their modified book algorithm, they undo the asymmetry that was inherent to the argument above, and instead work with the following symmetric picture[6]. We will maintain five disjoint sets A, B, X, Y, Z, such that $(A, X \cup Y)$

[6]In some sense, the fact that the colors play symmetric roles is the key reason why this argument generalizes to more than two colors, whereas the earlier one did not.

is a red book and $(B, X \cup Z)$ is a blue book. Thus, the only difference from the setup above is the new set Z, which is the blue version of the set Y.



We initialize the process with $X \sqcup Y \sqcup Z$ being an arbitrary equitable partition of the vertices, and with $A = B = \varnothing$. As before, a step of the process consists of taking a vertex from X and moving it to either A or B (a red or blue step, respectively). When we do a red or blue step, we of course need to shrink X, as well as the corresponding set Y or Z, in order to maintain the book picture above. This naturally means we now need to track *two* densities, namely p_R, the red edge density between X and Y, and p_B, the blue edge density between X and Z.

As before, the densities p_R and p_B control how much Y and Z, respectively, shrink during red and blue steps, respectively. As such, we want to maintain that p_R and p_B stay large throughout the process. This means that, as before, our failure mode is when we cannot take a red step or a blue step, as this would mean that the red or blue density drops too much. Suppose, say, that the red density is the problem. This means that for every potential choice of $v \in X$, the red density between $N_R(v) \cap X$ and $N_R(v) \cap Y$ is substantially smaller than p_R.

Let us actually for the moment forget that we need to restrict X to $N_R(v) \cap X$, and simply study the red density between X and $N_R(v) \cap Y$. If we denote $U_R(v) \coloneqq N_R(v) \cap Y$, the fact that this red density is low means that there is "negative correlation" between the sets $U_R(v)$ and $U_R(w)$ over different choices of $v, w \in X$. Indeed, the red density between X and $U_R(v)$ is precisely

$$\frac{e_R(X, U_R(v))}{|X||U_R(v)|} = \frac{1}{|X|} \sum_{w \in X} \frac{|U_R(w) \cap U_R(v)|}{|U_R(v)|}.$$

Thus, the drop in red density that we need to worry about precisely corresponds to the pairwise intersections $|U_R(w) \cap U_R(v)|$ being unusually small for all choices of w, v.

Let us suppose (with essentially no loss of generality) that every vertex in X has precisely $p_R|Y|$ red neighbors in Y, i.e. that $|U_R(v)| = p_R|Y|$ for all $v \in X$. In this case, the "negative correlation" discussed above can be naturally encoded as a geometric property in $|Y|$-dimensional Euclidean space. Indeed, let us associate to every $v \in X$ a vector $\tau_R(v) \in \mathbb{R}^Y$, which is simply the indicator vector of its neighborhood $U_R(v)$: the yth coordinate of $\tau_R(v)$ is 1 if v is adjacent to y in red, and zero otherwise. By assumption $\tau_R(v)$ has exactly $p_R|Y|$ entries equal to 1, and all other entries equal to 0. Moreover, we have that

$$|U_R(v) \cap U_R(w)| = \langle \tau_R(v), \tau_R(w) \rangle,$$

since the inner product of $\tau_R(v)$ and $\tau_R(w)$ precisely equals the number of 1 coordinates they have in common, which is exactly the number of $y \in Y$ that are adjacent in red to both u and w.

It will be more convenient to "center" these vectors so that the average value of their entries is 0, so let us define $\sigma_R(v) \coloneqq \tau_R(v) - p_R\mathbf{1}$, where $\mathbf{1}$ is the all-ones vector. By our assumption that $\tau_R(v)$ has exactly $p_R|Y|$ entries equal to 1, we see that the average value of the entries of $\sigma_R(v)$ is indeed zero. Moreover,

$$\begin{aligned}
\langle \sigma_R(v), \sigma_R(w) \rangle &= \langle \tau_R(v) - p_R\mathbf{1}, \tau_R(w) - p_R\mathbf{1} \rangle \\
&= \langle \tau_R(v), \tau_R(w) \rangle - \langle \tau_R(v), p_R\mathbf{1} \rangle - \langle p_R\mathbf{1}, \tau_R(w) \rangle + \langle p_R\mathbf{1}, p_R\mathbf{1} \rangle \\
&= |U_R(v) \cap U_R(w)| - p_R|U_R(v)| - p_R|U_R(w)| + p_R^2|Y| \\
&= |U_R(v) \cap U_R(w)| - p_R^2|Y|.
\end{aligned}$$

Note that if $U_R(v)$ and $U_R(w)$ were randomly chosen subsets of Y, each of size $p_R|Y|$, then their expected intersection size would be exactly $p_R^2|Y|$. Hence, the inner product $\langle \sigma_R(v), \sigma_R(w) \rangle$ records how much the true intersection size differs from this expected amount; the negative correlation behavior we were worried about earlier is precisely the statement that all (or most) of these inner products are "quite negative".

An obvious way to get many such pairs of negatively correlated vectors is to split the set X into two halves, and then have the vectors associated to one half lying close to the north pole of the sphere in $\mathbb{R}^Y$, and the other half lying close to the south pole. However, if the situation looks like this, we win for another reason: we can throw away half the set, and obtain a lot of *positive* correlation. The striking geometric lemma proved by Balister et al. essentially states that such an operation is always possible in one of the two colors, without completely messing up the other color. Their actual result is somewhat more general in a number of respects; I am simply stating a consequence that suffices for the framework discussed above.

Lemma 10 (Balister, Bollobás, Campos, Griffiths, Hurley, Morris, Sahasrabudhe, and Tiba, 2024)**.** *There exist constants $C, c > 0$ such that the following holds. Let X, Y be finite sets, and let $\sigma_R, \sigma_B : X \to \mathbb{R}^Y$ be arbitrary functions. There exist $v \in X$ and $X' \subseteq X$, as well as a number $\mu \geqslant 0$, such that $|X'| \geqslant ce^{-C\sqrt{\mu}}|X|$ and for all $w \in X'$, we have*

$$\langle \sigma_R(v), \sigma_R(w) \rangle \geqslant -1 + \mu \qquad \text{and} \qquad \langle \sigma_B(v), \sigma_B(w) \rangle \geqslant -1$$

(or the same statement holds after reversing the roles of red and blue).

There are a few remarks to make about this statement. First, note that it essentially agrees with the intuition we were discussing above: we gain some positive correlation in red, without affecting too much the correlation in blue (because $\mu \geqslant 0$). Second, we don't get to control how much positive correlation we obtain (since μ is an output of the statement, not an input), but we do have some information on it: the amount that we shrink X also depends on μ, so that if we obtain fairly weak positive correlation, we also don't shrink X too much. It is crucial—although the reason why is a bit too technical to explain here—that the loss in the size of X is subexponential in μ; the proof could not go through if we could only guarantee $|X'| \geqslant ce^{-C\mu}|X|$, say. Finally, we remark that this lemma is *not* scale-invariant: if we multiply each vector $\sigma_R(v)$ by 10, we cannot simply output the same v, X', and μ. In fact, this lack of scale-invariance is crucial in the way this lemma is used; it is necessary to rescale σ_R and σ_B by appropriate factors depending on the current values of p_R and p_B.

The proof of Lemma 10 is actually quite simple, although it feels like a magic trick. The first observation is that, if we let v, w be independent, uniformly random elements of X, then all the moments $\mathbb{E}[\langle \sigma_R(v), \sigma_R(w)\rangle^a \langle \sigma_B(v), \sigma_B(w)\rangle^b]$ are non-negative, for all non-negative integers a, b. This is a generalization of (and in fact follows from) the well-known fact that the expected inner product of two iid random vectors is non-negative.

The second step, which is more involved but still quite simple, is to come up with a function $f : \mathbb{R}^2 \to \mathbb{R}$ all of whose Taylor coefficients are non-negative, which is positive on the positive quadrant but bounded from above away from it, and which does not grow too fast at infinity. It turns out that a good choice of such a function is

$$f(x, y) \coloneqq x(2 + \cosh \sqrt{y}) + y(2 + \cosh \sqrt{x}).$$

The point is that, since all the Taylor coefficients of f are non-negative, and by the non-negativity of moments discussed above, we have that $\mathbb{E}[f(\langle \sigma_R(v), \sigma_R(w)\rangle, \langle \sigma_B(v), \sigma_B(w)\rangle)] \geqslant 0$. On the other hand, since f is bounded from above away from the positive quadrant, we can effectively bound the (negative) contributions to this expectation arising from very negative inner products. We then conclude that there must be a fair amount of probability mass which contributes positively to this expectation, which, after some rearrangements and after using the bounds on the growth rate of f, proves the lemma.