UPPER BOUNDS ON DIAGONAL RAMSEY NUMBERS
[after Campos, Griffiths, Morris, and Sahasrabudhe]

by **Yuval Wigderson**

1. Introduction

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 1.1 (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. Even this simple statement has some remarkable consequences. For example, Schur (1917) used Theorem 1.1[2] to prove that for all sufficiently large primes p, there exist non-trivial solutions to the equation $x^n + y^n \equiv z^n \pmod{p}$, that is, that one cannot prove Fermat's last theorem via a local-global argument.

Connections and applications to other fields of mathematics have been an important feature of Ramsey theory from the very beginning. Ramsey himself had an application in mathematical logic in mind when he proved Theorem 1.1 (indeed, his paper is titled "On a problem of formal logic"). The influential paper of Erdős and Szekeres (1935), which helped establish Ramsey theory as a central branch of combinatorics, is titled "A combinatorial problem in geometry"; in it, they reproved Theorem 1.1 in order to deduce a result on convex polygons among sets of points in Euclidean space.

---

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

[2]Alert readers may note that Schur's result precedes Ramsey's by more than a decade. In fact, Schur proved a closely related lemma, which one can now recognize as a consequence of Theorem 1.1, and derived his theorem from that lemma.

Today, Ramsey-theoretic theorems and techniques are of fundamental importance in many different fields, including additive number theory, Banach space theory, discrete geometry, ergodic theory, group theory, and theoretical computer science. These are deep and rich connections, and are difficult to adequately summarize, so we refer to the book of Graham, Rothschild, and Spencer (1990), to the survey of Conlon, Fox, and Sudakov (2015), and to the lecture notes of Wigderson (2024) for more in-depth introductions to the field.

For many applications, such as those of Schur (1917) in number theory, Ramsey (1929) in logic, and Erdős and Szekeres (1935) in geometry mentioned above, qualitative statements such as Theorem 1.1 suffice. However, much of the modern research in Ramsey theory is concerned with *quantitative* statements: how large is the integer N in Theorem 1.1 as a function of k? Formally, we make the following definition.

Definition 1.2. — 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.

Before continuing with the discussion of what is known about the function $r(k)$, let us pause and ask why we should study such quantitative questions, when qualitative statements like Theorem 1.1 are elegant and already suffice for many applications. There are several answers to this question. One answer is that for certain applications, especially in fields such as theoretical computer science (e.g. the lower bound of Razborov (1985) on monotone circuit complexity), qualitative statements are not sufficient, as the application itself is quantitative. A second answer is that a better quantitative understanding of Ramsey-theoretic results can yield new insights and new proofs of existing theorems. For example, recent breakthroughs on the quantitative aspects of the Ramsey-theoretic theorem of Roth (1953), due to Bloom and Sisask (2020) and Kelley and Meka (2023) (see also the exposé of Peluse (2022)), imply that the primes contain infinitely many three-term arithmetic progressions. This result was first proved by van der Corput (1939), and is a special case of the landmark result of Green and Tao (2008). However, in contrast to these earlier proofs, we now know that the primes contain infinitely many three-term arithmetic progressions simply because *there are many prime numbers*. That is, the quantitative improvements yielded a new proof of this theorem, using essentially no properties of the primes other than their density. Finally, and no less importantly, a third reason for studying such quantitative questions is that doing so can reveal a world of deep and beautiful mathematics.

With that said, let us turn to the quantitative aspects of Theorem 1.1, that is, to the determination of the function $r(k)$ from Definition 1.2. The exact value of $r(k)$ is only known for $k \leqslant 4$, and it currently seems completely hopeless[3] to obtain an

---

[3]The following famous anecdote was reported by Spencer (1994): "Erdős asks us to imagine an alien force, vastly more powerful than us, landing on Earth and demanding the value of $r(5)$ or they will destroy our planet. In that case, he claims, we should marshal all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they ask for $r(6)$. In that case, he believes, we should attempt to destroy the aliens." Indeed, results of Exoo (1989) and

exact formula for $r(k)$, so let us content ourselves with asymptotic bounds as $k \to \infty$. Essentially every proof of Theorem 1.1 yields (at least implicitly) an upper bound on $r(k)$, by proving the existence of *some* integer N. The original proof of Ramsey (1929) 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 (1935) proved the following stronger bound.

THEOREM 1.3 (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 (1947) dispelled this belief by proving[4] an exponential lower bound.

THEOREM 1.4 (Erdős, 1947). — $r(k) \geqslant \sqrt{2}^k$ *for every* $k \geqslant 2$.

After this breakthrough, progress stalled for 75 years. There were a number of improvements to these bounds over the years, including important results of Spencer (1975), Graham and Rödl (1987), Thomason (1988), Conlon (2009), and Sah (2023), 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 (2023).

THEOREM 1.5 (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.

The exact constant 3.993 is not particularly important, and a more careful analysis of the same proof yields a slightly better bound[5]. The important thing about this result is that it is the first result, after almost 90 years of intense study, to break the barrier of 4^k.

The new tool introduced by Campos, Griffiths, Morris, and Sahasrabudhe (2023) is the so-called *book algorithm*, an elementary but ingenious technique for finding monochromatic *book graphs* in colorings of K_N. As we will shortly discuss, a book graph is a basic graph-theoretic object, whose study turns out to be closely connected to the study of Ramsey numbers. Every known proof of Theorem 1.1 uses, implicitly or explicitly, monochromatic book graphs.

Angeltveit and McKay (2024) show that $r(5)$ lies takes on one of the four values $\{43, 44, 45, 46\}$, but we remain very far from knowing the value of $r(6)$.

[4] Lower bounds on Ramsey numbers are somewhat beyond the scope of this exposé, so we will not discuss the proof of Theorem 1.4 in detail. However, it would be remiss not to mention that this beautiful proof is extraordinarily influential, and is the origin of the *probabilistic method*, an extremely powerful technique in modern combinatorics.

[5] More recently, Gupta, Ndiaye, Norin, and Wei (2024) recast the proof of Theorem 1.5 in a different language, which allowed them to optimize the technique and obtain a much stronger bound of $r(k) \leqslant 3.8^k$ for sufficiently large k.

As we will see, the proof of Campos, Griffiths, Morris, and Sahasrabudhe (2023) is fairly ad hoc, and relies on the verification of certain complicated numerical inequalities. More recently, however, a new, more conceptual proof of Theorem 1.5[6] was found by Balister et al. (2024). They introduced a modification of the book algorithm, but their crucial new input is a purely geometric lemma, concerning the correlations of probability distributions in high-dimensional Euclidean space. While the proof of Campos, Griffiths, Morris, and Sahasrabudhe (2023) and Balister et al. (2024) have many features in common, they differ in key ways, and we will sketch both proofs.

The rest of this exposé is dedicated to discussing these two proofs of Theorem 1.5, and is organized as follows. We begin in Section 2 with a proof of Theorem 1.3, in the course of which we introduce book graphs as well as several of the key ideas that go into the proof of Theorem 1.5. In Section 3, we introduce and analyze the book algorithm of Campos, Griffiths, Morris, and Sahasrabudhe (2023), and will then *fail* to prove Theorem 1.5. Luckily, we will rescue the argument and complete the proof in Section 4 by introducing two additional ingredients. In Section 5 we introduce and analyze the symmetric book algorithm of Balister et al. (2024), and use it to give another proof of Theorem 1.5. The key new lemma introduced by Balister et al. (2024), and its geometric proof, are discussed in Section 6. We end in Section 7 with an epilogue, discussing the use of book graphs in the original proof of Ramsey (1929) of Theorem 1.1, as well as how our understanding of book graphs and Ramsey theory has developed over the subsequent 95 years.

2. The Erdős–Szekeres theorem and algorithm

In this section, we prove Theorem 1.3 (and thus Theorem 1.1). This proof is elegant and interesting in its own right, and additionally it contains within it several of the important ideas used in the proof of Theorem 1.5. We will actually see three different proofs (or, more precisely, three different ways of viewing the same proof) of Theorem 1.3, in each of the next three subsections. Each proof will help introduce some of the key ideas that go into the proof of Theorem 1.5.

---

[6]Moreover, Balister et al. (2024) were able to prove a more general theorem, which gives a new upper bound on Ramsey numbers in any number of colors. For simplicity, however, we remain with the two-color version of the problem throughout this exposé.

2.1. Off-diagonal Ramsey numbers

We begin with the original proof of Erdős and Szekeres (1935). Before proceeding with the proof, we generalize the notion of Ramsey numbers from Definition 1.2. 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 2.1. — 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*.

With this terminology, we can prove Theorem 1.3. In fact, we will prove the following more precise result.

THEOREM 2.2 (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:

(2.1) $$r(k, \ell) \leqslant r(k - 1, \ell) + r(k, \ell - 1).$$

To prove (2.1), 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 (2.1).

We can now complete the induction. By (2.1) 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. $\qquad\square$

2.2. Enter the book

Definition 2.3. — 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 2.4. — *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. $\qquad\square$

This proof should look familiar—we have already encountered the same idea in the proof of Theorem 2.2, where we implicitly used the $t = 1$ case of Lemma 2.4. 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 2.4 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 2.4 basically goes back to the work of Erdős and Szekeres (1935), it was first formulated in essentially this language by Thomason (1982), who used Lemma 2.4 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 2.4 and conclude that $r(k) \leqslant N$. Again, this is essentially the approach we used in the proof of Theorem 2.2, 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 $r(k)$.

Thomason's idea has been quite successful. The three prior asymptotic improvements to Theorem 2.2, due to Thomason (1988), Conlon (2009), and Sah (2023), all used this idea, roughly showing that if some two-coloring of $E(K_N)$ does *not* contain a monochromatic $B_{t,m}$ (for some fixed t, m), then its structure must be such that the proof of Theorem 2.2 can be made more efficient. A more precise structural result along these lines is given by Conlon, Fox, and Wigderson (2022). However, for fundamental technical reasons, none of these techniques seems capable of finding books with spine larger than $t = O(\log k)$, whereas in order to prove a result like Theorem 1.5 in this way, one would want to take (say) $t = \frac{k}{1000}$. This was where the matter stood for a few years, until the work of Campos, Griffiths, Morris, and Sahasrabudhe (2023).

2.3. The Erdős–Szekeres algorithm

One of the many new ingredients introduced by Campos, Griffiths, Morris, and Sahasrabudhe (2023) 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 2.4. Although this idea is almost a triviality, this change of perspective is crucial for the proof of Theorem 1.5.

Before we define this exploration algorithm—which they termed the *book algorithm*—let us first rephrase the proof of Theorem 1.3 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 1.3.

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.

ALGORITHM 2.5 (Erdős–Szekeres algorithm)

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 Algorithm 2.5, we can give an alternative proof of Theorem 1.3.

"Algorithmic" proof of Theorem 1.3. — 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 Algorithm 2.5 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

$$(2.2) \qquad\qquad |X| \geqslant 2^{-|A|-|B|}N.$$

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$^{(\dagger)}$ $\frac{1}{2}|X|$, thus preserving the validity of (2.2). Similarly, in a blue step, we increase $|B|$ by 1 and decrease $|X|$ to at least $\frac{1}{2}|X|$, again preserving (2.2). By induction, we conclude that (2.2) also holds at the end of the process.

$^{(\dagger)}$Strictly speaking, we should write here $\frac{1}{2}(|X| - 1)$, although the claimed bound (2.2) can also be proved inductively by judicious use of ceiling signs. However, from now on, we will start ignoring such additive ± 1 terms. Of course they need to be carefully dealt with to obtain a correct proof, but they will always contribute a negligible error, which we will ignore. We will add the symbol $(\dagger)$ to mark the places where we omit such additive errors.

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. $\qquad\square$

It is worth noting that, as presented, this argument only proves Theorem 1.3—that is, the bound $r(k) \leqslant 4^k$—rather than the sharper estimate given in Theorem 2.2. It is an interesting and instructive exercise to figure out how to modify Algorithm 2.5 to obtain the stronger bound proved in Theorem 2.2 via a similar "algorithmic" proof.

For future reference (and as a hint to solving the exercise above), it is good to observe that the off-diagonal Erdős–Szekeres bound $r(k, \ell) \leqslant \binom{k+\ell}{\ell}$ can also be obtained in this way. To do so, set $\gamma = \frac{\ell}{k+\ell}$. Then we can modify the Erdős–Szekeres algorithm as follows.

ALGORITHM 2.6 (Off-diagonal Erdős–Szekeres algorithm)
1. If $|X| \leqslant 1$, $|A| \geqslant k$, or $|B| \geqslant \ell$, stop the process.
2. Pick a vertex $v \in X$, and check whether v has at least $(1-\gamma)|X|$ red neighbors in X.
3. If yes, move v to A and shrink X to the red neighborhood of v. That is, update $A \rightarrow A \cup \{v\}$ and $X \rightarrow X \cap N_R(v)$, and keep B the same. Call this a *red step*.
4. If not, then v has at least[†] $\gamma|X|$ blue neighbors in X. We now move v to B, and shrink X to the blue neighborhood of v. That is, we update $B \rightarrow B \cup \{v\}$ and $X \rightarrow X \cap N_B(v)$, and keep A the same. Call this a *blue step*.
5. Return to step 1.

The point now is that we obtain the red K_k or blue K_ℓ if $|A| \geqslant k$ or $|B| \geqslant \ell$, and thus we may assume that we do fewer than k red steps and fewer than ℓ blue steps. X shrinks by a factor of $1 - \gamma$ at every red step, and by a factor of γ at every blue step, so throughout the process we have

$$|X| \geqslant (1-\gamma)^{|A|}\gamma^{|B|}N.$$

On the other hand, the process only terminates if $|X| \leqslant 1$, so this implies $N < (1-\gamma)^{-k}\gamma^{-\ell}$. One can check, by Stirling's approximation, that

$$\binom{k+\ell}{\ell} = 2^{o(k)}\gamma^{-\ell}(1-\gamma)^{-k}$$

for all $\ell \leqslant k$, and hence this gives a contradiction if we choose N of the form $2^{o(k)}\binom{k+\ell}{\ell}$. This recovers Theorem 2.2 up to the subexponential error term.

3. The book algorithm

We are now ready to describe the book algorithm of Campos, Griffiths, Morris, and Sahasrabudhe (2023). 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[7]. 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 $t := |A|$ and $m := |Y|$ satisfy $m \geqslant r(k - t, k)$, so that we can apply Lemma 2.4 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.
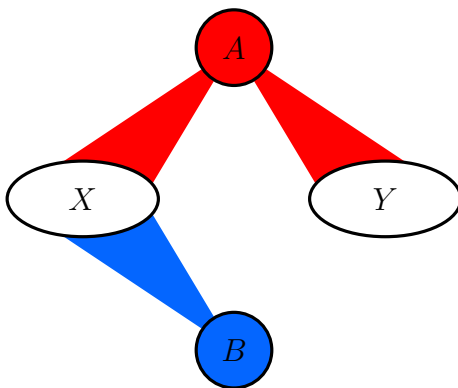
As in the Erdős–Szekeres algorithm, we will iteratively build this picture by moving vertices from X to A or B, and then shrinking X and Y. A move from X to A will be called a red step, and a move from X to B will be called a blue step.

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 (2.2), 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[8] a bound of $|Y| \gtrsim 2^{-|A|}N$.

---

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

[8]If we could really obtain such strong control on $|Y|$, we would show that $r(k) \lesssim 2^k$, a dramatic improvement over Theorem 1.3. 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 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. We denote this density by

$$p := d_R(X, Y) = \frac{e_R(X, Y)}{|X||Y|},$$

where $e_R(X, Y)$ denotes the number of red edges with one endpoint in X and the other in Y. Recall that, by assumption, we have $p \geqslant \frac{1}{2}$ at the beginning of the process. Note that every time we add a vertex to A or to B (and thus have to shrink X and potentially Y), this red density p might change. For our simplified exposition of the proof of Theorem 1.5, we will make the following (completely unjustified) assumption.

ASSUMPTION 3.1. — *At every step of the process, every vertex in X has exactly $p|Y|$ red neighbors in Y, and every vertex in Y has exactly $p|X|$ red neighbors in X. In other words, the bipartite graph of red edges between X and Y is bi-regular.*

We stress again that X, Y, and p change throughout the process, but Assumption 3.1 asserts that whenever such a change happens, we magically end up back with the same bi-regularity.

While Assumption 3.1 is clearly a bogus assumption, it is actually possible to (essentially) make it rigorous. Indeed, the definition of p implies that the vertices in X have, on average, $p|Y|$ red neighbors in Y. A basic but important observation, used frequently in extremal combinatorics, is that one can often convert such average degree conditions to minimum or maximum degree conditions, by deleting a few "outlier" vertices. In the rigorous proof of Theorem 1.5, one must repeatedly "clean" X by removing such outliers, and thus one can indeed maintain an approximate version of Assumption 3.1, at least ensuring that all vertices in X have roughly the same red degree[9]. However, for our exposition, we ignore these important technicalities, and stick with Assumption 3.1.

---

[9]It is *much* harder to ensure degree-regularity in both X and Y simultaneously. Luckily, it turns out that degree-regularity in Y is substantially less important in the argument, and in the formal proof one doesn't even ensure an approximate version of it. In its place, one uses a judicious choice of the vertex v.

3.1. The steps of the book algorithm

The two basic steps in the book algorithm will again be red steps and blue steps, as in the Erdős–Szekeres algorithm. Note that when we perform a blue step (moving $v \in X$ to B and updating $X \to X \cap N_B(v)$), we do not need to update Y at all, since these changes do not affect the fact that (A, Y) is a red book. In particular, thanks to Assumption 3.1, the red density between X and Y remains unchanged during a blue step, since all the remaining vertices in X still have exactly $p|Y|$ red neighbors in Y. However, as discussed above, red steps *can* affect p, since in a red step we update $X \to X \cap N_R(v)$ and $Y \to Y \cap N_R(v)$, and thus our value of p is updated to

$$p' := d_R(X \cap N_R(v), Y \cap N_R(v)).$$

Let us call a vertex *prosperous* if $p' \geqslant p - \alpha$, for some parameter α we will shortly choose. We will then perform a red step only if there is a vertex $v \in X$ which is prosperous, and which has at least $\frac{1}{2}|X|$ red neighbors in X. In such a step, we increase $|A|$ by 1, decrease $|X|$ by a factor of 2, decrease Y by a factor of p (since v has $p|Y|$ red neighbors in Y, by Assumption 3.1), and update p to at least $p - \alpha$.

In Algorithm 2.5, we were always able to do either a red or a blue step, since every vertex in X has at least $\frac{1}{2}|X|$ neighbors in X in one of the colors[†]. However, if we require that our red vertex v be prosperous, then we may be in a position where neither a red nor a blue step is possible. Namely, we get stuck if all vertices in X have at least $\frac{1}{2}|X|$ red neighbors in X, but none of them is prosperous.

In this case, we implement a *density-boost step*, which is one of the other main innovations of Campos, Griffiths, Morris, and Sahasrabudhe (2023). Pick a vertex $v \in X$, and consider the following picture.
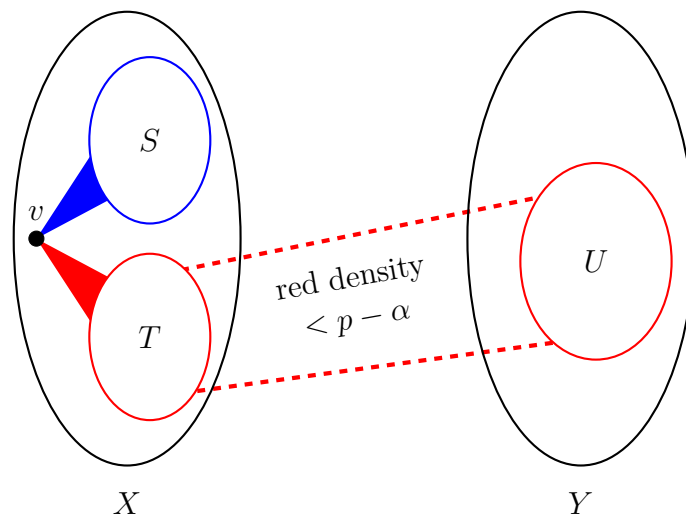
Since v is not prosperous, the red edge density between $T := N_R(v) \cap X$ and $U := N_R(v) \cap Y$ must be less than $p - \alpha$. However, by Assumption 3.1, every vertex in U has $p|X|$ red neighbors in X. Therefore, setting $S := N_B(v) \cap X$, we find that[†]

$$p|X||U| = e_R(X, U) = e_R(T, U) + e_R(S, U) < (p - \alpha)|T||U| + e_R(S, U).$$

Rearranging, we conclude that

$$e_R(S, U) > |U| \left(p|X| - (p - \alpha)|T| \right).$$

Let $\beta := \frac{|S|}{|X|}$, so that β records what fraction of the edges from v to the rest of X are blue. Then $|S| = \beta|X|$ and[†] $|T| = (1 - \beta)|X|$, and the above can be rewritten as

$$e_R(S, U) > |U||S| \left(\frac{p}{\beta} - \frac{(p - \alpha)(1 - \beta)}{\beta} \right) = |S||U| \left(p + \alpha \frac{1 - \beta}{\beta} \right),$$

which implies

$$(3.1) \qquad\qquad d_R(S, U) > p + \alpha \frac{1 - \beta}{\beta}.$$

Note too that since we cannot do a blue step, we must have $\beta \leqslant \frac{1}{2}$, implying that $d_R(S, U) > p + \alpha$. In other words, in the bad situation where we cannot perform a red or a blue step, we *can* perform a *density-boost step*, where we replace X by $S = N_B(v) \cap X$, replace Y by $U = N_R(v) \cap Y$, and thus boost the density from p to at least $p + \alpha \frac{1 - \beta}{\beta} \geqslant p + \alpha$.

Note that density-boost steps are expensive, in that they shrink X and Y, but don't actually make progress by increasing $|A|$ or $|B|$. In particular, we don't *a priori* have any control on how many density-boost steps we perform. Luckily, there is a simple fix to this problem: since we are in any case updating $X \to X \cap N_B(v)$ in a density-boost step, we may add v to B for free, while maintaining the property that (B, X) is a blue book. That is, a density-boost step can also be made a type of blue step, and thus we necessarily perform at most k density-boost steps without creating a blue K_k.

The final piece we need before formally defining the book algorithm is to choose α, which determines the threshold above which a vertex is considered prosperous. Note that every red step may decrease p by α, so if we end up doing up to k red steps, we may decrease p from its initial value of $\frac{1}{2}$ to $\frac{1}{2} - \alpha k$. Moreover, whenever we do a red step, we also shrink Y by a factor of (the current value of) p. In particular, if p ever drops below (say) $\frac{1}{4}$, we are in big trouble: then Y shrinks by a factor of 4 at every step, and we have no real hope of proving a bound stronger than $r(k) \leqslant 4^k$. As such, we want to pick $\alpha \leqslant \frac{\varepsilon}{k}$, so that even after doing k red steps, we have not meaningfully decreased p below its initial value. Here, one can think of ε as a tiny absolute constant, although in the final analysis we will actually pick ε to tend to 0 slowly with k.

Unfortunately, there is a trade-off. Recall that we have very little control over the effect of the density-boost steps, because these are the steps we do as a last resort. In fact, essentially our only way of bounding their total effect is the observation that $p \leqslant 1$ throughout the entire process, which should imply that we cannot do too many density-boost steps, as that would drive the red density up too high. The problem is that a density-boost step only increases p by roughly α, so if we pick $\alpha \leqslant \frac{\varepsilon}{k}$, then even if we do k density-boost steps (the maximum possible number), we will only increase the density by ε. In particular, we have no hope of reaching the threshold of $p = 1$, where we finally gain some control over the density-boost steps.

The way to resolve this apparent contradiction is to pick α adaptively. Indeed, suppose that at some point in the process, we have reached a red density of, say, $p = 0.51$. At this point, it doesn't make sense to have the cutoff be $\alpha = \frac{\varepsilon}{k}$; we wouldn't even mind losing an absolute constant of $1/100$ in the density, since that will only bring us back to our original value of p! So we will instead pick α to be dependent on our current value of p; namely, we set

$$(3.2) \qquad \alpha(p) := \begin{cases} \frac{\varepsilon}{k} & \text{if } p \leqslant \frac{1}{2} + \frac{1}{k}, \\ \varepsilon(p - \frac{1}{2}) & \text{otherwise.} \end{cases}$$

Again, the point of this is that, if we are at some step of the process where $p > \frac{1}{2}$, then we can afford to lose more in the density without every dropping p into the "danger zone" of being substantially smaller than $\frac{1}{2}$. The advantage of this is that the amount we *win* in a density-boost step is itself proportional to $\alpha = \alpha(p)$. So if we have already done some number of density-boost steps, such that $p > \frac{1}{2}$, each subsequent density-boost boosts the density even further, at an exponential rate, thus rapidly bringing us closer to the threshold $p = 1$.

3.2. Formal definition of the book algorithm

With all of these preliminaries, we are finally able to define the book algorithm.

ALGORITHM 3.2 (Book algorithm)

1. If $|X| \leqslant 1$, $|A| \geqslant k$, or $|B| \geqslant k$, stop the process.
2. Let $p = d_R(X, Y)$ be the current red density between X and Y. Define $\alpha = \alpha(p)$ as in (3.2), where ε is some fixed parameter throughout the process.
3. Check whether some vertex $v \in X$ has at least $\frac{1}{2}|X|$ blue neighbors in X. If yes, perform a *blue step*, by updating

$$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X \cap N_B(v), \qquad Y \to Y,$$

   and return to step 1.
4. Check whether some vertex $v \in X$ is prosperous, meaning that $d_R(N_R(v) \cap X, N_R(v) \cap Y) \geqslant p - \alpha$. If yes, perform a *red step*, by updating

$$A \to A \cup \{v\}, \qquad B \to B, \qquad X \to X \cap N_R(v), \qquad Y \to Y \cap N_R(v),$$

   and return to step 1.
5. In the remaining case, pick some vertex $v \in X$. It is not prosperous, and has $\beta|X|$ blue neighbors in X, for some $\beta \leqslant \frac{1}{2}$. We now perform a *density-boost step*, by updating

$$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X \cap N_B(v), \qquad Y \to Y \cap N_R(v),$$

   and return to step 1.

For future reference, the following table records how the key parameters change during the execution of the book algorithm, following the discussion above.

3.3. Analysis of the book algorithm

Suppose that, when the book algorithm ends, we have done t red steps, s density-boost steps, and b blue steps. We may assume that $t < k$ and that $s + b < k$, since otherwise we have found a monochromatic K_k. We now collect a number of estimates on the various parameters associated with the process.

LEMMA 3.3. — *We have $p \geqslant \frac{1}{2} - \varepsilon$ throughout the entire process.*

Proof. — As discussed above, every blue step keeps p constant (by Assumption 3.1), every density-boost step can only increase p, and every red step decreases p by at most $\alpha(p)$. Additionally, the choice of $\alpha(p)$ shows that $p - \alpha(p) \geqslant \frac{1}{2}$ whenever $p > \frac{1}{2} + \frac{1}{k}$, whereas $\alpha(p) = \frac{\varepsilon}{k}$ whenever $p \leqslant \frac{1}{2} + \frac{1}{k}$. Since we do $t \leqslant k$ red steps, p can never drop below $\frac{1}{2} - t \cdot \frac{\varepsilon}{k} \geqslant \frac{1}{2} - \varepsilon$. $\qquad\square$

It will now be convenient to pick $\varepsilon = k^{-1/4}$, although we note that this choice is not particularly important; many functions of k which tend to 0 neither too slowly or too quickly would work.

LEMMA 3.4. — *At the end of the process, we have $|Y| \geqslant 2^{-t-s-o(k)} N$.*

Proof. — Y is unchanged by every blue step. On the other hand, during each red or density-boost step, we decrease Y by a factor of p, by Assumption 3.1. By Lemma 3.3, we have that $p \geqslant \frac{1}{2} - \varepsilon$ at every such step, hence

$$|Y| \geqslant \left(\frac{1}{2} - \varepsilon\right)^{t+s} \cdot \frac{N}{2} = 2^{-t-s-o(k)} N,$$

where we plug in our choice of ε and recall that we start the process with[(†)] $|Y| = \frac{N}{2}$. $\quad\square$

We next turn to bounding $|X|$ at the end of the process. Just as in the Erdős–Szekeres algorithm, the main point of this is to estimate how many steps we do, since we recall that the process terminates when $|X| \leqslant 1$.

Recall that at each density-boost step, we shrink X by a factor of β, where β is defined as the fraction $|N_B(v) \cap X|/|X|$ of blue neighbors of the currently chosen vertex v. Let

	$|A|$	$|B|$	$|X|$	$|Y|$	p
blue step	$-$	$+1$	$\times\frac{1}{2}$	$-$	$-$
red step	$+1$	$-$	$\times\frac{1}{2}$	$\times p$	$-\alpha$
density-boost step	$-$	$+1$	$\times\beta$	$\times p$	$+\alpha\frac{1-\beta}{\beta}$

TABLE 3.1. How the various parameters evolve during Algorithm 3.2. Dashes denote quantities that are unchanged. In general, the entries in the table are *lower bounds*, e.g. a density-boost step may increase p by more than $\alpha\frac{1-\beta}{\beta}$, and a red step may shrink X to more than half of its previous size.

$\beta_1, \dots, \beta_s$ be the sequence of values of β for each of the s blue steps. Let β be the harmonic mean of $\beta_1, \dots, \beta_s$, that is, define β by

$$\frac{1}{\beta} = \frac{1}{s} \sum_{i=1}^{s} \frac{1}{\beta_i}.$$

LEMMA 3.5. — *At the end of the process, we have*

$$|X| \geqslant 2^{-t-b-o(k)} \beta^s N.$$

Proof. — Every red or blue step shrinks X by at most a factor[†] of 2, hence the factor of 2^{-t-b}. On the other hand, the ith density-boost step decreases $|X|$ by a factor of β_i. The inequality of arithmetic and geometric means implies that

$$\frac{1}{\beta} = \frac{1}{s} \sum_{i=1}^{s} \frac{1}{\beta_i} \geqslant \left(\prod_{i=1}^{s} \frac{1}{\beta_i} \right)^{1/s},$$

hence the contribution of the density-boost steps is

$$\prod_{i=1}^{s} \beta_i \geqslant \beta^s.$$

Together with the fact that we begin the process with[†] $|X| = \frac{N}{2}$, this yields the claimed bound. □

The final, and perhaps most important, result we need is an estimate on the number of density-boost steps. As discussed above, we can get a good estimate on this quantity because of the "dynamic" choice of α; this is the content of the next lemma, which is called the *zig-zag lemma* by Campos, Griffiths, Morris, and Sahasrabudhe (2023).

LEMMA 3.6 (Zig-zag lemma). — *We have*

$$\sum_{i=1}^{s} \frac{1 - \beta_i}{\beta_i} \leqslant t + o(k).$$

We won't give a full proof of Lemma 3.6, but the following sketch captures the main ideas.

Proof sketch for Lemma 3.6. — For the moment, let us assume that we stay in the regime $p \geqslant \frac{1}{2} + \frac{1}{k}$. It will be more convenient to reparametrize p, by defining $q := p - \frac{1}{2}$. By our choice of α in (3.2), we have that $\alpha(p) = \varepsilon q$.

Suppose we do one step of the book algorithm, and thus update p to some new value p' (and update q to $q' = p' - \frac{1}{2}$). If the step we do is a blue step, then by Assumption 3.1, the density p does not change, hence $p' = p$ and $q' = q$. If, instead, we do a red step, then v is prosperous, and hence $p' \geqslant p - \alpha(p)$. This implies that $q' \geqslant q - \alpha(p) = q - \varepsilon q = (1 - \varepsilon)q$. Finally, if this step is the ith density-boost step, then by (3.1) we have that

$$p' \geqslant p + \alpha(p) \frac{1 - \beta_i}{\beta_i}$$

and thus

$$q' \geqslant q + \alpha(p)\frac{1-\beta_i}{\beta_i} = q\left(1 + \varepsilon\frac{1-\beta_i}{\beta_i}\right).$$

Putting this all together, we conclude that at each step of the algorithm, we have

$$(3.3) \qquad \frac{q'}{q} \geqslant \begin{cases} 1 & \text{when we do a blue step,} \\ 1 - \varepsilon & \text{when we do a red step,} \\ 1 + \varepsilon\frac{1-\beta_i}{\beta_i} & \text{when we do the } i\text{th density-boost step.} \end{cases}$$

Let q_{final} denote the value of q at the end of the algorithm, and let q_{initial} be the value of q at the beginning of the algorithm. Multiplying (3.3) over all steps of the algorithm, we find that

$$\frac{q_{\text{final}}}{q_{\text{initial}}} \geqslant (1-\varepsilon)^t \prod_{i=1}^{s}\left(1 + \varepsilon\frac{1-\beta_i}{\beta_i}\right),$$

since we get a contribution of $1 - \varepsilon$ from each of the t red steps and a contribution of $1 + \varepsilon\frac{1-\beta_i}{\beta_i}$ from the ith density-boost step. Combining this inequality with the approximation $1 + x \approx e^x$, an approximation that is valid for sufficiently small[10] $|x|$, we find that

$$(3.4) \qquad \frac{q_{\text{final}}}{q_{\text{initial}}} \gtrsim e^{-\varepsilon t}\exp\left(\varepsilon\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i}\right) = \exp\left(\varepsilon\left(-t + \sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i}\right)\right).$$

We have that $q_{\text{final}} \leqslant \frac{1}{2}$, since $p \leqslant 1$ throughout the whole process. On the other hand, since we are assuming that $p \geqslant \frac{1}{2} + \frac{1}{k}$ throughout, we have that $q_{\text{initial}} \geqslant \frac{1}{k}$. Therefore, $\frac{q_{\text{final}}}{q_{\text{initial}}} \leqslant \frac{k}{2} \leqslant k$. Plugging this into (3.4) and taking logarithms, we find that

$$\log k \geqslant \log\left(\frac{q_{\text{final}}}{q_{\text{initial}}}\right) \gtrsim \varepsilon\left(-t + \sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i}\right),$$

implying that

$$\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i} \lesssim t + \frac{\log k}{\varepsilon} = t + o(k),$$

where we plug in our choice of $\varepsilon = k^{-1/4}$ in the final equality.

This proof worked under the assumption that we remain throughout in the range $p \geqslant \frac{1}{2} + \frac{1}{k}$. Let us now work in the complementary regime, where $p < \frac{1}{2} + \frac{1}{k}$ throughout the whole process. In this case, recalling the definition of $\alpha(p)$ from (3.2), we have

$$(3.5) \qquad p' - p \geqslant \begin{cases} 0 & \text{when we do a blue step,} \\ -\frac{\varepsilon}{k} & \text{when we do a red step,} \\ +\frac{\varepsilon}{k}\cdot\frac{1-\beta_i}{\beta_i} & \text{when we do the } i\text{th density-boost step.} \end{cases}$$

---

[10]This approximation can be made rigorous, but we're still cheating in this derivation of (3.4). We have no guarantee that $\varepsilon\frac{1-\beta_i}{\beta_i}$ is small, since we have no control over β_i, and thus no guarantee that the approximation is valid. A correct proof of this lemma would need to separate out the contribution from the steps where β_i is very small, and thus where such an approximation is not accurate.

Adding up (3.5) over all steps of the process, we conclude that

$$(3.6) \qquad p_{\text{final}} - p_{\text{initial}} \geqslant -\frac{\varepsilon}{k}t + \frac{\varepsilon}{k}\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i} = \frac{\varepsilon}{k}\left(-t + \sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i}\right).$$

Recall that we had $p_{\text{initial}} \geqslant \frac{1}{2}$, and we are now assuming that we remain in the regime $p < \frac{1}{2} + \frac{1}{k}$ throughout, hence in particular $p_{\text{final}} < \frac{1}{2} + \frac{1}{k}$. Therefore $p_{\text{final}} - p_{\text{initial}} \leqslant \frac{1}{k}$. Plugging this in to (3.6) and rearranging, we conclude that

$$\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i} \leqslant t + \frac{1}{\varepsilon} = t + o(k),$$

again by our choice of ε.

We have thus proved the desired inequality in the two extreme cases, namely when $p \geqslant \frac{1}{2} + \frac{1}{k}$ throughout the process, and when $p < \frac{1}{2} + \frac{1}{k}$ throughout the process. Of course, in reality, we may move between these two regimes multiple times during the execution of Algorithm 3.2. However, by breaking the execution of the algorithm into intervals in which we remain in one regime or the other, it is not too difficult to combine the arguments above and conclude that the claimed inequality always holds. $\qquad \square$

As an immediate consequence of Lemma 3.6, we obtain an upper bound on the number s of density-boost steps.

LEMMA 3.7. — *We have*

$$s \leqslant \left(\frac{\beta}{1-\beta}\right)t + o(k).$$

Equivalently,

$$(3.7) \qquad \beta \geqslant (1 + o(1))\frac{s}{s+t}.$$

Proof. — By the definition of β, we have that

$$\frac{1}{s}\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i} = \frac{1-\beta}{\beta}.$$

Plugging this into Lemma 3.6 shows that

$$s = \left(\frac{\beta}{1-\beta}\right)\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i} \leqslant \left(\frac{\beta}{1-\beta}\right)(t+o(k)).$$

Moreover, since each β_i is at most $\frac{1}{2}$, we find that $\frac{\beta}{1-\beta} \leqslant 1$, yielding the first claimed bound. The second bound follows by solving for[11] β. $\qquad \square$

---

[11]Again, there is some cheating going on here—one can only obtain the claimed estimate if s is not too small as a function of k, in order to absorb the error terms. In the formal proof, one has to separate into cases: the bound (3.7) is valid if s is not too small, whereas if s is very small one can complete the proof of Theorem 1.5 via a simpler analysis.

3.4. Proof attempt for Theorem 1.5

We are now ready to put everything together. Let C be a constant, which we will optimize later, and let $N = 2^{(1+C)k}$. Our plan is to show that if C is chosen appropriately, then $r(k) \leqslant N = 2^{(1+C)k}$. Since our goal is to prove Theorem 1.5, we thus hope to be able to prove this for some fixed $C < 1$, but for the moment, let us leave C as an unspecified constant.

We proceed by contradiction, so let us assume that there is a two-coloring of $E(K_N)$ with no monochromatic K_k. We now run Algorithm 3.2 on this coloring. We certainly obtain the desired contradiction if the algorithm finds a red or blue K_k, so let us assume that this does not happen. Therefore, the process only ends when $|X| \leqslant 1$, which by Lemma 3.5 implies that

$$N \leqslant \beta^{-s}2^{t+b+o(k)} \leqslant \beta^{-s}2^{t+(k-s)+o(k)},$$

where we plug in the bound $b + s \leqslant k$, arising from the fact that B never becomes a blue K_k. We now plug in the lower bound on β from Lemma 3.7 to find that

$$(3.8) \qquad\qquad N \leqslant \left(\frac{t+s}{s}\right)^s 2^{k+t-s+o(k)}.$$

At this point everything is in terms of the parameters s and t, which we expect to scale linearly in k, so it is more convenient to reparametrize everything in terms of $x := \frac{t}{k}, y := \frac{s}{k}$. In terms of these parameters, and recalling that $N = 2^{(1+C)k}$, we can rewrite (3.8) as

$$C - o(1) \leqslant (x - y) + y \log_2\left(\frac{x+y}{y}\right) =: G(x, y).$$

Recall that our goal is to obtain a contradiction, and the only thing we have not yet specified is the value of C we choose. In particular, if we pick C to be larger than the maximum value of $G(x, y)$ over the square $[0, 1]^2$, then we certainly obtain a contradiction. As our goal is to eventually obtain a contradiction with some fixed $C < 1$, we would be happy if this maximum value were less than 1. However, this is not true, as shown on the following contour plot; the maximum value of G is roughly 1.33.
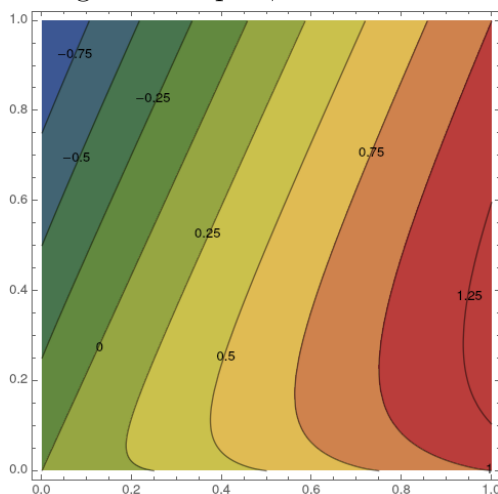
Of course, if we recall our original strategy, it is way too much to hope for that the maximum of G is less than 1. Indeed, the whole point of the book algorithm was to output the book (A, Y), and to ensure that its parameters are good enough to apply Lemma 2.4.

What are the parameters of this book? Well, we have that $|A| = t$ by definition, and

$$m := |Y| \geqslant 2^{-t-s-o(k)} N$$

by Lemma 3.4. By Lemma 2.4, we know that if $m \geqslant r(k-t, k)$, we find a monochromatic K_k, yielding our desired contradiction. Thus, we may assume that $m < r(k-t, k)$, implying that

$$(3.9) \qquad\qquad N \leqslant 2^{t+s+o(k)} m < 2^{t+s+o(k)} r(k-t, k).$$

By Theorem 2.2, we know that

$$r(k-t, k) \leqslant \binom{2k-t}{k-t}.$$

A useful upper bound on binomial coefficients is that $\binom{a}{b} \leqslant 2^{aH(b/a)}$, where $H(z) := -z \log_2 z - (1-z) \log_2(1-z)$ is the *binary entropy function*. Plugging this in, we find that

$$\log_2 r(k-t, k) \leqslant \log_2 \binom{2k-t}{k-t} \leqslant (2k-t)H\left(\frac{k-t}{2k-t}\right) = k\left[(2-x)H\left(\frac{1-x}{2-x}\right)\right].$$

Taking logarithms of (3.9) and dividing by k shows that

$$C - o(1) \leqslant -1 + (x+y) + (2-x)H\left(\frac{1-x}{2-x}\right) =: F(x, y).$$

Putting all of this together, we have shown that either we derive the claimed contradiction, or $C - o(1) \leqslant \min\{F(x, y), G(x, y)\}$. Again, we have the freedom to choose C, so we can obtain the desired contradiction if we set C to be larger than the maximum of $\min\{F(x, y), G(x, y)\}$ on the square $[0, 1]^2$. In particular, as our goal is to pick $C < 1$, we are done if $\min\{F(x, y), G(x, y)\} < 1$ for all $x, y \in [0, 1]$. Here is a contour plot of F:
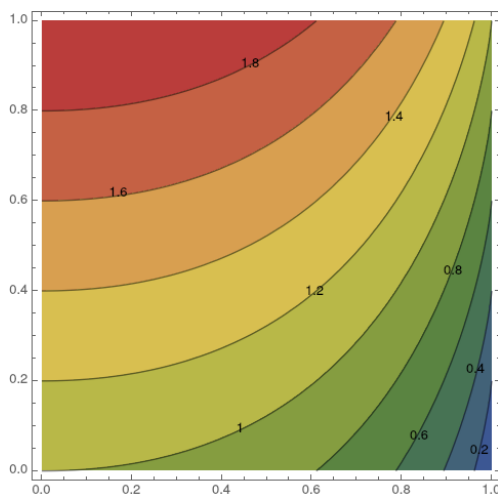
This looks great! The areas where F is large seem to be different from the areas where G is large, so there should be no problem to show that their maximum is always strictly less than 1. In fact, here are the regions where $F > 1$ and $G > 1$.
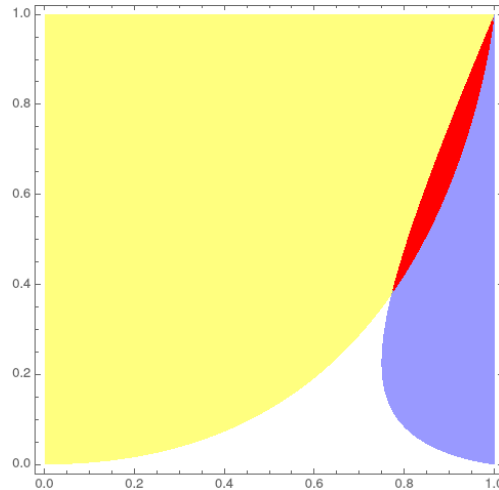
Bad news! There's a big red region where both functions are greater than 1, and our whole proof strategy fails. In fact, one can check that $\min\{F(x, y), G(x, y)\}$ attains a maximum value of roughly 1.054. That is, in order to obtain a contradiction, the smallest value of C we could pick is 1.054, and thus this whole complex proof is only able to show that $r(k) \leqslant 2^{2.054k} \approx 4.15^k$, which is worse than the bound of Theorem 1.3.

4. Rescuing the argument

The fact that $\min\{F(x, y), G(x, y)\} > 1$ for some $(x, y) \in [0, 1]^2$ is a fundamental obstruction to this approach. In order to overcome it, we will use two tricks, both of which involve tweaking the book algorithm.

4.1. Changing the cutoff

The first new idea is to examine our criterion for deciding whether to do red or blue steps. Recall that, as in Algorithm 2.5, we do a blue step if some vertex in X has at least $\frac{1}{2}|X|$ blue neighbors in X, and otherwise we do a red or density-boost step. In the Erdős–Szekeres setting, this is the optimal choice: since the argument is symmetric in the two colors, it would be strictly worse to use any other cutoff.

However, the book algorithm is highly asymmetric, so we should re-examine this assumption. Recall that at the end of the process, we output the red book (A, Y), where $|A| = t$ and $|Y| \geqslant 2^{-t-s-o(k)}N$ by Lemma 3.4. The fact that $|Y|$ decays like $2^{-t}N$ is unavoidable, but the fact that $|Y|$ decays exponentially in s shows that density-boost steps are very expensive, in terms of making this trade-off very bad. As such, we should try to minimize the number s of density-boost steps we do, in terms of t. Since Lemma 3.7 tells us that $s \leqslant \frac{\beta}{1-\beta}t + o(k)$, the natural way to decrease s is to decrease β.

To achieve this, we do the following. We pick a number $\mu \in [0, 1]$, which will be fixed throughout the argument. In step 3 of Algorithm 3.2, we now perform a blue step if some vertex in X has at least $\mu|X|$ blue neighbors in X; otherwise, we proceed to the subsequent steps of the algorithm unchanged. An important effect of this choice is that now, when we perform the ith density-boost step, the parameter β_i is constrained to be at most μ, and thus also $\beta \leqslant \mu$ at the end of the process. In particular, if we pick $\mu < \frac{1}{2}$, we will have accomplished our goal of decreasing s relative to t. This suggests we should pick μ very small, but of course there is a trade-off—if μ is very small then every blue step decreases $|X|$ by a lot, and thus the process will terminate quickly. To balance these two effects, we want to pick μ to be neither too large nor too small. For completeness, here is the description of our modified book algorithm.

ALGORITHM 4.1 (Book algorithm with cutoff μ)

1. If $|X| \leqslant 1$, $|A| \geqslant k$, or $|B| \geqslant k$, stop the process.
2. Let $p = d_R(X, Y)$ be the current red density between X and Y. Define $\alpha = \alpha(p)$ as in (3.2), where ε is some fixed parameter throughout the process.
3. Check whether some vertex $v \in X$ has at least $\mu|X|$ blue neighbors in X. If yes, perform a *blue step*, by updating

$$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X \cap N_B(v), \qquad Y \to Y,$$

   and return to step 1.
4. Check whether some vertex $v \in X$ is prosperous, meaning that $d_R(N_R(v) \cap X, N_R(v) \cap Y) \geqslant p - \alpha$. If yes, perform a *red step*, by updating

$$A \to A \cup \{v\}, \qquad B \to B, \qquad X \to X \cap N_R(v), \qquad Y \to Y \cap N_R(v),$$

   and return to step 1.
5. In the remaining case, pick some vertex $v \in X$. It is not prosperous, and has $\beta|X|$ blue neighbors in X, for some $\beta \leqslant \mu$. We now perform a *density-boost step*, by updating

$$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X \cap N_B(v), \qquad Y \to Y \cap N_R(v),$$

   and return to step 1.

	$|A|$	$|B|$	$|X|$	$|Y|$	p
blue step	$-$	$+1$	$\times\mu$	$-$	$-$
red step	$+1$	$-$	$\times(1-\mu)$	$\times p$	$-\alpha$
density-boost step	$-$	$+1$	$\times\beta$	$\times p$	$+\alpha\frac{1-\beta}{\beta}$

TABLE 4.1. How the various parameters evolve during Algorithm 4.1. The only difference from Table 3.1 is that blue and red steps shrink X by factors of μ and $1 - \mu$, respectively.

In this modified book algorithm, Lemmas 3.3, 3.4, 3.6 and 3.7 remain true; the only change is that Lemma 3.5 needs to be modified to the following statement, reflecting the fact that each blue (resp. red) step shrinks X by a factor$^{(\dagger)}$ of μ (resp. $1 - \mu$) in the worst case. The proof is otherwise identical to that of Lemma 3.5.

LEMMA 4.2 (Modified Lemma 3.5). — *At the end of the process, we have*

$$|X| \geqslant 2^{-o(k)}(1 - \mu)^t \mu^b \beta^s N.$$

In particular, since $b + s \leqslant k$, we have

$$|X| \geqslant 2^{-o(k)}(1 - \mu)^t \mu^{k-s} \beta^s N.$$

Since the process terminates when $|X| \leqslant 1$, we conclude from Lemma 4.2 that

$$(4.1) \qquad N \leqslant 2^{o(k)}(1 - \mu)^{-t}\mu^{-(k-s)}\beta^{-s} \leqslant 2^{o(k)}(1 - \mu)^{-t}\mu^{-(k-s)}\left(\frac{s + t}{s}\right)^s,$$

where the final inequality follows from the lower bound on β in Lemma 3.7. Taking logarithms and dividing by k, we conclude that

$$C - o(1) \leqslant -1 + x\log_2\left(\frac{1}{1 - \mu}\right) + (1 - y)\log_2\frac{1}{\mu} + y\log_2\left(\frac{x + y}{y}\right) =: G_\mu(x, y).$$

Note that in the case $\mu = \frac{1}{2}$, we precisely recover the previous function G, which of course makes sense as we are then recovering Algorithm 3.2. Here are contour plots of G_μ for $\mu \in \left\{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\right\}$.
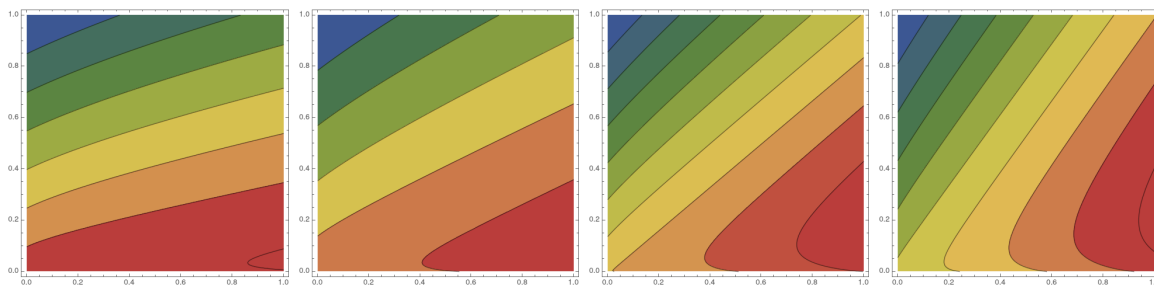
And here are pictures of the regions where $F > 1$ and $G_\mu > 1$, for $\mu \in \left\{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\right\}$.
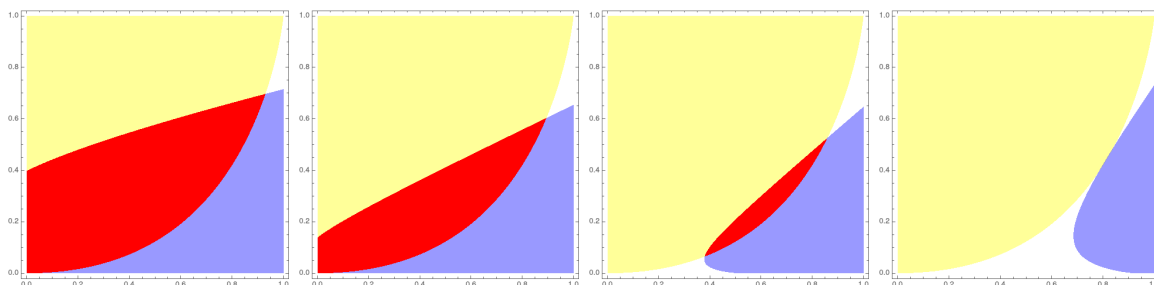
It looks like we're already done at $\mu = \frac{4}{10} = \frac{2}{5}$, but unfortunately we're not: one can check that $\min\{F(x, y), G_{\frac{2}{5}}(x, y)\}$ attains a maximum value of 1.0017, hence we only obtain a bound of $r(k) \leqslant 4.006^k$. Here is a closer view of what happens at $\mu = \frac{2}{5}$:
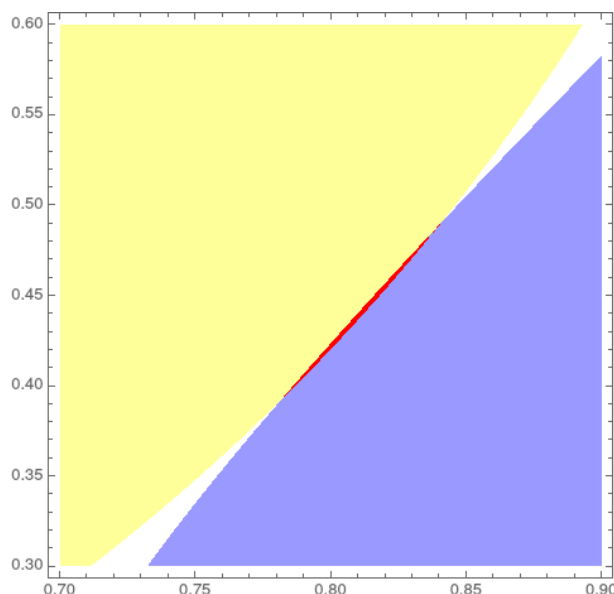
But we're definitely making progress! The bad red region is extremely small now, and our maximum value of $\min\{F, G_\mu\}$ is extraordinarily close to 1. Unfortunately, one can check that no choice of μ will actually decrease this value below 1—which would complete the proof—so another idea is needed.

4.2. Off-diagonal Ramsey numbers

So far, we have played with the parameter μ in order to vary the region where $G_\mu > 1$, and have almost succeeded in making it disjoint from the region where $F > 1$. We will now try to tweak F, in order to move this latter region. Recall that the way we defined F was in terms of an upper bound on $r(k-t, k)$. If we can obtain a better upper bound on $r(k-t, k)$, then F will decrease, and we may be in business. In fact, we don't need to improve the upper bound on $r(k-t, k)$ in all cases; it suffices to improve this upper bound for pairs $(k-t, k)$ near the problematic region where both F and $G_{\frac{2}{5}}$ are greater than 1. Since this problematic region is near $x \approx 0.75$, we could hope to improve the upper bound on $r(k-t, k)$ where $k-t \approx 0.75k$, or equivalently on $r(k, \ell)$ where $\ell \approx \frac{k}{4}$.

There is actually a good reason to expect this to work. Recall Algorithm 2.6, which yields an upper bound on such off-diagonal Ramsey numbers. In that algorithm, we choose whether to do red or blue steps based on the cutoff $\gamma = \frac{\ell}{k+\ell}$. If we just blindly import the same idea into the book algorithm, it makes sense to set $\mu \approx \frac{\ell}{k+\ell}$ in order to upper-bound $r(k, \ell)$. In case $\ell \approx \frac{k}{4}$, we have $\mu \approx \frac{1}{5}$. In our argument above, we saw that it is good to take μ small, except for the trade-off that now X shrinks by a factor of μ for every blue step. However, in this regime, we will do at most ℓ blue steps, and $\mu^\ell \approx (\frac{1}{5})^{\frac{k}{4}} \approx 0.67^k$; in contrast, in the argument above, the blue steps shrink X by $(\frac{2}{5})^k = 0.4^k$, which is a much more significant decrease. Hence we may expect the trade-offs to work well for us.

For completeness, here is our final book algorithm, suited for upper-bounding $r(k, \ell)$. We set $\mu = \frac{\ell}{k+\ell}$ and $\varepsilon = k^{-1/4}$. We initiate $A = B = \varnothing$, and $X \sqcup Y$ an arbitrary partition of $V(K_N)$ into two equally-sized parts. Let $p_{\text{initial}} = d_R(X, Y)$ be the density of red edges between X and Y at the beginning of the process, and define

$$(4.2) \qquad \alpha(p) := \begin{cases} \frac{\varepsilon}{k} & \text{if } p \leqslant p_{\text{initial}} + \frac{1}{k}, \\ \varepsilon(p - p_{\text{initial}}) & \text{otherwise.} \end{cases}$$

The algorithm is then as follows.

ALGORITHM 4.3 (Off-diagonal book algorithm)

1. If $|X| \leqslant 1$, $|A| \geqslant k$, or $|B| \geqslant \ell$, stop the process.
2. Let $p = d_R(X, Y)$ be the current red density between X and Y. Define $\alpha = \alpha(p)$ as in (4.2).
3. Check whether some vertex $v \in X$ has at least $\mu|X|$ blue neighbors in X. If yes, perform a *blue step*, by updating

$$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X \cap N_B(v), \qquad Y \to Y,$$

   and return to step 1.
4. Check whether some vertex $v \in X$ is prosperous, meaning that $d_R(N_R(v) \cap X, N_R(v) \cap Y) \geqslant p - \alpha$. If yes, perform a *red step*, by updating

$$A \to A \cup \{v\}, \qquad B \to B, \qquad X \to X \cap N_R(v), \qquad Y \to Y \cap N_R(v),$$

   and return to step 1.
5. In the remaining case, pick some vertex $v \in X$. It is not prosperous, and has $\beta|X|$ blue neighbors in X, for some $\beta \leqslant \mu$. We now perform a *density-boost step*, by updating

$$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X \cap N_B(v), \qquad Y \to Y \cap N_R(v),$$

   and return to step 1.

Apart from the choice of $\mu = \frac{\ell}{k+\ell}$, this algorithm is identical to Algorithm 4.1, except that we now stop when $|B| \geqslant \ell$, rather than when $|B| \geqslant k$ as before. In particular, Table 4.1 still gives the relevant changes in the parameters. Unfortunately, there is an additional complication introduced by moving to the off-diagonal setting. Before, when we sought to upper-bound $r(k)$, we could assume that the initial red density p_{initial} was at least $\frac{1}{2}$, by simply swapping the roles of the two colors if necessary. However, once we are in the off-diagonal setting, this is no longer allowed, and we may have no control on p_{initial}. Let us make another completely unjustified assumption.

ASSUMPTION 4.4. — *At the beginning of the process, we have* $p_{\text{initial}} \geqslant \frac{k}{k+\ell} = 1 - \mu$.

Note that this is a natural assumption, since Algorithm 2.6 "predicts" that $\frac{k}{k+\ell}$ is roughly the correct red density to expect, in the sense that in the analysis of Algorithm 2.6, this red density is the worst-case occurrence. That is, if Assumption 4.4

is false "robustly", then Algorithm 2.6 should already prove a stronger bound than $r(k, \ell) \leqslant \binom{k+\ell}{\ell}$. In fact, one can essentially force Assumption 4.4 to hold because of such an argument; if we start with $p_{\text{initial}} < \frac{k}{k+\ell}$, we can run a number of steps of the Erdős–Szekeres algorithm, until we end up with $p \geqslant \frac{k}{k+\ell}$. If this never happens, then Algorithm 2.6 itself will prove that $r(k, \ell) \ll \binom{k+\ell}{\ell}$.

Given Assumption 4.4, we can conclude that Lemmas 3.6, 3.7 and 4.2 remain true for Algorithm 4.3. Moreover, we can prove the following modified version of Lemmas 3.3 and 3.4 (combined into a single statement), whose proof is essentially unchanged.

LEMMA 4.5 (Modified Lemmas 3.3 and 3.4). — *We have $p \geqslant p_{\text{initial}} - \varepsilon \geqslant (1 - \mu) - \varepsilon$ throughout the entire process. Therefore, at the end of the process, we have*

$$|Y| \geqslant (1 - \mu)^{t+s+o(k)} N.$$

With all of this setup, we are finally able to prove[12] an exponentially-improved upper bound on $r(k, \ell)$.

THEOREM 4.6. — *We have $r(k, \ell) \leqslant 2^{-\frac{2}{9}\ell + o(k)} \binom{k+\ell}{\ell}$ for all $\ell \leqslant \frac{k}{4}$.*

Note that in this theorem, the gain over Theorem 2.2 is exponential in ℓ, and not in k. This is natural, and the best we could hope for. Indeed, if $\ell = o(k)$, then the bound in Theorem 2.2 is already subexponential in k, so it is impossible to improve it by a factor of $2^{-\delta k + o(k)}$ for any fixed $\delta > 0$.

Proof of Theorem 4.6. — Let C be a constant that we will optimize later, and let $N = 2^{(1+C)k}$. We fix a two-coloring of $E(K_N)$, and assume for contradiction that there is no red K_k or blue K_ℓ in this coloring. We apply Algorithm 4.3, with $\mu = \frac{\ell}{k+\ell} \leqslant \frac{1}{5}$. Note that this choice of μ implies that $\frac{\ell}{k} = \frac{\mu}{1-\mu}$. If we output that A is a red K_k or B is a blue K_ℓ, then we have obtained a contradiction, hence we can assume this does not happen. Therefore, the process only terminates when $|X| \leqslant 1$, and we also have that $b + s \leqslant \ell$. Plugging this into Lemma 4.2, we find that

$$N \leqslant 2^{o(k)}(1-\mu)^{-t}\mu^{-(\ell-s)}\beta^{-s} \leqslant 2^{o(k)}(1-\mu)^{-t}\mu^{-(\ell-s)}\left(\frac{s+t}{s}\right)^s.$$

Note that we obtain a better exponent on μ than we had in (4.1), because the assumption $b + s \leqslant \ell$ is stronger than what we had before; this is precisely the extra strength gained by moving to the off-diagonal setting. Taking logarithms and dividing by k shows that

$$C - o(1) \leqslant -1 + x\log_2\left(\frac{1}{1-\mu}\right) + \left(\frac{\mu}{1-\mu} - y\right)\log_2\frac{1}{\mu} + y\log_2\left(\frac{x+y}{y}\right) =: \widetilde{G}_\mu(x, y),$$

---

[12]One should really write "prove", since everything here is dependent on the unjustified Assumptions 3.1 and 4.4, as well as on the key Lemma 3.6, which we did not rigorously prove. Additionally, the bound in Theorem 4.6 is stronger than any result proved by Campos, Griffiths, Morris, and Sahasrabudhe (2023), and this too is a consequence of the fact that we are being unrigorous, especially with the verification of certain numerical inequalities. However, Theorem 4.6 is true; a rigorous proof of a stronger statement is given by Gupta, Ndiaye, Norin, and Wei (2024, Corollary 6).

where the only difference between G_μ and $\widetilde{G}_\mu$ is the the term $\frac{\mu}{1-\mu}$ in the latter, which is simply 1 in the former. It comes from the ℓ in the exponent; upon dividing by k we obtain $\frac{\ell}{k} = \frac{\mu}{1-\mu}$.

Additionally, by Lemma 4.5, we have

$$|Y| \geqslant (1-\mu)^{t+s+o(k)} N.$$

If $|Y| \geqslant r(k-t, \ell)$, then we obtain a contradiction by Lemma 2.4, so we may assume that $|Y| < r(k-t, \ell)$. Taking logarithms and dividing by k again shows that

$$(4.3) \qquad C - o(1) \leqslant -1 + (x+y)\log_2\left(\frac{1}{1-\mu}\right) + \frac{1}{k}\log_2 r(k-t, \ell).$$

By Theorem 1.3, we have

$$\log_2 r(k-t, \ell) \leqslant \log_2\binom{k-t+\ell}{\ell}$$

$$\leqslant (k-t+\ell)H\left(\frac{\ell}{k-t+\ell}\right)$$

$$= k\cdot\left(1 - x + \frac{\mu}{1-\mu}\right)H\left(\frac{\mu/(1-\mu)}{1 - x + \mu/(1-\mu)}\right).$$

Plugging this into (4.3) shows that

$$C - o(1) \leqslant -1 + (x+y)\log_2\left(\frac{1}{1-\mu}\right) + \left(1 - x + \frac{\mu}{1-\mu}\right)H\left(\frac{\mu/(1-\mu)}{1 - x + \mu/(1-\mu)}\right)$$

$$=: \widetilde{F}_\mu(x, y).$$

We are no longer trying to beat the bound $r(k) \leqslant 4^k$, so our goal is no longer to obtain a contradiction for some $C < 1$. Instead, we are comparing to $\frac{1}{k}\log_2\binom{k+\ell}{\ell}$, which equals $(1 + \frac{\mu}{1-\mu})H(\mu) + o(1)$, and thus our goal is to obtain a contradiction for some fixed $C < (1 + \frac{\mu}{1-\mu})H(\mu) - 1$. That is, we would like to show is that for all $\mu \leqslant \frac{1}{5}$, we have $\min\{\widetilde{F}_\mu(x, y), \widetilde{G}_\mu(x, y)\} < (1 + \frac{\mu}{1-\mu})H(\mu) - 1$ for all $x, y \in [0, 1]$. In fact, we hope to prove this inequality with some slack, so that we gain an improvement in the exponent.

Recall that our goal is to prove a gain over Theorem 2.2 that is exponential in ℓ. As such, the slack we get in this inequality should scale like $\frac{\ell}{k} = \frac{\mu}{1-\mu}$. That is, we would like to prove an inequality of the form

$$\min\{\widetilde{F}_\mu(x, y), \widetilde{G}_\mu(x, y)\} < \left(1 + \frac{\mu}{1-\mu}\right)H(\mu) - 1 - \delta\frac{\mu}{1-\mu},$$

where $\delta > 0$ is some absolute constant; such a bound would prove that $r(k, \ell) \leqslant 2^{-\delta\ell + o(k)}\binom{k+\ell}{\ell}$.

Such an inequality holds! In fact, one can check that for $\mu \leqslant \frac{1}{5}$, we may take δ as large as $\frac{2}{9}$. Indeed, here is a plot of the regions where $\widetilde{F}_\mu > (1 + \frac{\mu}{1-\mu})H(\mu) - 1 - \frac{2}{9}\frac{\mu}{1-\mu}$ and $\widetilde{G}_\mu > (1 + \frac{\mu}{1-\mu})H(\mu) - 1 - \frac{2}{9}\frac{\mu}{1-\mu}$, respectively, for $\mu = \frac{1}{5}$. One can verify that the regions only move further apart as μ decreases, so $\mu = \frac{1}{5}$ is the worst case.
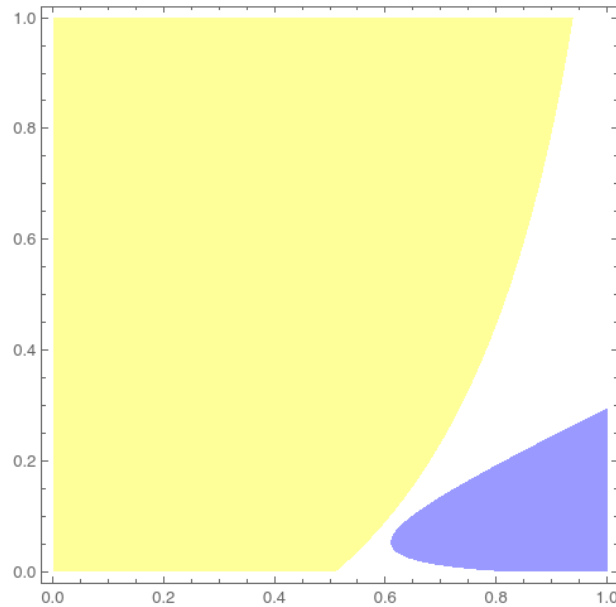
This shows that we do indeed get a contradiction if we set $C = (1 + \frac{\mu}{1-\mu})H(\mu) - \frac{2}{9}\frac{\mu}{1-\mu}$, proving the bound

$$r(k, \ell) \leqslant 2^{\left(1 + \frac{\mu}{1-\mu}H(\mu) - \frac{2}{9}\frac{\mu}{1-\mu} + o(1)\right)k} = 2^{-\frac{2}{9}\ell + o(k)}\binom{k + \ell}{\ell}$$

for all $\ell \leqslant \frac{k}{4}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4.3. Back to the diagonal

Now that we have an upper bound on $r(k, \ell)$ for $\ell \leqslant \frac{k}{4}$, we can finally complete the proof of Theorem 1.5. We will actually prove the following bound, which is a little bit stronger. The only reason we obtain a stronger bound than Campos, Griffiths, Morris, and Sahasrabudhe (2023) is the numerical computations: the authors rigorously justify every numerical inequality, which is often substantially simpler to do if one proves a slightly weaker bound, whereas we content ourselves with "proving" the numerical bounds through pictures. As mentioned in Section 1, much stronger bounds were recently proved by Gupta, Ndiaye, Norin, and Wei (2024) via an alternative analysis. It follows from their results that Theorem 4.7 is true, even though we do not provide a rigorous proof.

THEOREM 4.7. — *We have $r(k) \leqslant 2^{(2 - \frac{3}{200} + o(1))k} \approx 3.96^k$.*

Proof. — Let C be a constant that we will optimize later. Let $N = 2^{(1+C)k}$, and fix a two-coloring of $E(K_N)$, which we assume for contradiction has no monochromatic K_k. We run Algorithm 4.1 with $\mu = \frac{2}{5}$. Thanks to Theorem 4.6 (plus Theorem 2.2), we

know that

$$r(k-t,k) \leqslant \begin{cases} \dbinom{2k-t}{k-t} & \text{if } t < \tfrac{3}{4}k, \\ 2^{-\frac{2}{9}(k-t)+o(k)} \dbinom{2k-t}{k-t} & \text{if } t \geqslant \tfrac{3}{4}k. \end{cases}$$

Recall that we obtain a contradiction if $|Y| \geqslant r(k-t,k)$ at the end of the process, hence we may assume that $|Y| < r(k-t,k)$. Combining this with Lemma 3.4[13], we see that we get a contradiction if

$$C - o(1) \leqslant -1 + (x+y) + \frac{1}{k}\log_2 r(k-t,k)$$

$$\leqslant \begin{cases} -1 + (x+y) + (2-x)H(\frac{1-x}{2-x}) & \text{if } x < \tfrac{3}{4}, \\ -1 + (x+y) - \frac{2}{9}(1-x) + (2-x)H(\frac{1-x}{2-x}) & \text{if } x \geqslant \tfrac{3}{4} \end{cases}$$

$$= F(x,y) - \frac{2}{9}(1-x)\mathbf{1}_{x \geqslant \frac{3}{4}}$$

$$=: \widehat{F}(x,y),$$

where $\mathbf{1}_{x \geqslant \frac{3}{4}}$ denotes the indicator function for the event $x \geqslant \frac{3}{4}$. In particular, it suffices for us to prove that $\min\{\widehat{F}(x,y), G_{\frac{2}{5}}(x,y)\} \leqslant 1 - \delta$ for all $x, y \in [0,1]$, where $\delta > 0$ is a constant that will end up in the exponent in N.

This indeed works! Here are the plots of where $\widehat{F}$ and $G_{\frac{2}{5}}$ are greater than 1; the second plot is just zoomed in to show the "dangerous area", where the two regions no longer intersect.
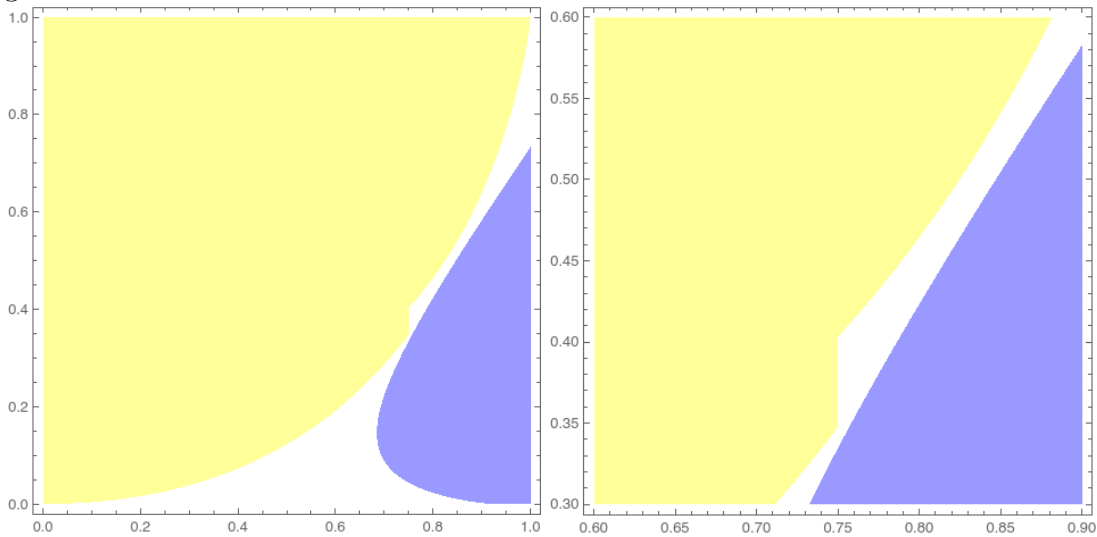
In fact, one can check that $\max_{x,y \in [0,1]} \min\{\widehat{F}(x,y), G_{\frac{2}{5}}(x,y)\} < 0.985$. Therefore, we obtain a contradiction if we set $C = 0.985 = 1 - \frac{3}{200}$, proving that $r(k) \leqslant 2^{(2-\frac{3}{200}+o(1))k}$, as claimed. $\qquad\square$

---

[13]We are back to the diagonal setting, so we may assume that $p_{\text{initial}} \geqslant \frac{1}{2}$. Therefore Lemmas 3.3 and 3.4 are again valid.

5. The symmetric book algorithm

In this section, we describe the new proof that $r(k) \leqslant (4 - \delta)^k$, due to Balister et al. (2024). The main new ingredient in their proof is a lemma about the self-correlation of probability distributions in high-dimensional Euclidean space; this lemma can be used to introduce a new kind of density-boost step, which is much simpler to analyze. As such, their proof is also more conceptual, and does not rely on the verification of complicated numerical inequalities. In this section, we discuss their strategy and prove their bound on $r(k)$, while deferring the geometry, and hence the proof of the key lemma, to Section 6.

Balister et al. (2024) introduce a modified book algorithm, which removes the asymmetry that was inherent to Algorithm 4.1. Roughly speaking, the fact that the colors play symmetric roles is their new approach is the key reason why this argument generalizes to more than two colors, whereas the earlier one of Campos, Griffiths, Morris, and Sahasrabudhe, 2023 did not. However, for simplicity, we will continue to only work with two colors, as before.

5.1. The steps of the symmetric book algorithm

In the symmetric book algorithm of Balister et al. (2024), we maintain five disjoint sets A, B, X, Y, Z, such that $(A, X \cup Y)$ is a red book and $(B, X \cup Z)$ is a blue book. Thus, the only difference from the earlier setup is the new set Z, which is a blue analogue of Y.
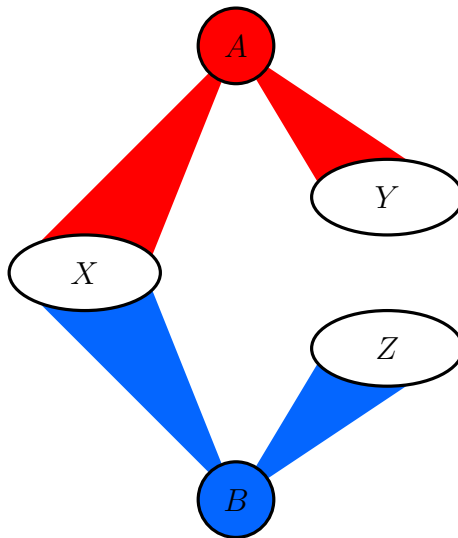
We initialize the process with $A = B = \varnothing$, and $X \sqcup Y \sqcup Z$ an arbitrary equitable partition of $V(K_N)$. As before, the two main types of steps will be red and blue steps, wherein we move a vertex v from X to A (resp. B), and shrink X and Y (resp. X and Z) to their intersection with the red (resp. blue) neighborhood of v.

In our prior analysis of the book algorithm, we had to track four parameters: the sizes of A, B, and X, and the red density between X and Y. In this new setup, we

will naturally also track $|Z|$ and the blue density between X and Z, which controls how much Z shrinks every time we do a blue step. Let us denote the red and blue densities, respectively, as

$$p_R := d_R(X, Y) = \frac{e_R(X, Y)}{|X||Y|} \qquad \text{and} \qquad p_B := d_B(X, Z) = \frac{e_B(X, Z)}{|X||Z|}.$$

In analogy with Assumptions 3.1 and 4.4, we make the following assumptions about the red and blue densities. As before, one can remove these assumptions by working harder, but by imposing them we can simplify the exposition and focus on the key new ideas.

ASSUMPTION 5.1. — *At every step of the process, every vertex in X has exactly $p_R|Y|$ red neighbors in Y and exactly $p_B|Z|$ blue neighbors in Z.*

ASSUMPTION 5.2. — *At the beginning of the process, we have $p_R \geqslant \frac{1}{2}$ and $p_B \geqslant \frac{1}{2}$.*

Note that Assumption 5.1 is a weaker assumption than Assumption 3.1, since we do not assume anything about the degrees of vertices in Y or Z. In our earlier analysis of the book algorithm, such an assumption was necessary in order to prove (3.1), the basic inequality governing the contribution of density-boost steps. In what follows, we will not need this inequality, or indeed any of the earlier analysis of density-boost steps. In its place, we will use the following key lemma, whose proof we defer to Section 6.

LEMMA 5.3 (Refinement lemma). — *Let X, Y, Z be disjoint sets of vertices in a coloring of $E(K_N)$, and let $\alpha_R, \alpha_B \in (0, 1)$ be parameters. There exists a vertex $v \in X$, a subset $X' \subseteq X$, and a real number $\kappa \geqslant 0$ with the following properties.*

*(a) X' **is not much smaller than** X: We have $|X'| \geqslant c2^{-C\kappa}|X|$, where $C = 6$ and $c = 1/8$ are constants.*

*(b) **No density drop in either color:** Letting $Y' = Y \cap N_R(v)$ and $Z' = Z \cap N_B(v)$, we have that*

$$d_R(X', Y') \geqslant d_R(X, Y) - \alpha_R \qquad \text{and} \qquad d_B(X', Z') \geqslant d_B(X, Z) - \alpha_B.$$

*(c) **Density boost in one color:** We have that*

$$d_R(X', Y') \geqslant d_R(X, Y) + (\kappa^2 - 1)\alpha_R \qquad \text{or} \qquad d_B(X', Z') \geqslant d_B(X, Z) + (\kappa^2 - 1)\alpha_B.$$

In the symmetric book algorithm, we apply Lemma 5.3 at every iteration, before deciding whether to take a red, blue, or density-boost step. As before, the quantities α_R, α_B determine how much we are willing to lose on the red and blue density, respectively. As such, we see that Lemma 5.3(b) is extremely helpful, and appears to completely eliminate the need for density-boost steps: every time we take a red or a blue step, Lemma 5.3(b) guarantees that the relevant density does not drop by more than our specified amount of α_R or α_B.

However, we cannot entirely dispense with density-boost steps. Indeed, every time we apply Lemma 5.3, we shrink X to X', and we have little control over how much smaller X' is than X: note that, crucially, κ is an *output* of Lemma 5.3, and not an input, so it may be that X' is much smaller than X. This is where Lemma 5.3(c) comes in: in case κ is very large, so that X' is much smaller than X, we gain a great deal on the density in one of the two colors. We can thus perform a new kind of density-boost step, replacing X by X' and one of Y, Z by Y', Z'. As we will see, the trade-offs in this density-boost step work out in our favor; it is crucial here that X shrinks by a factor of $\exp(-O(\kappa))$, but that the density increases by an amount proportional to κ^2. Because we only do such density-boost steps when κ is large, the amount we gain on the density is substantially larger than the amount we lose on $|X|$.

One final difference between the symmetric book algorithm and the book algorithm studied in Sections 3 and 4 is the stopping condition. We will fix some integers t, m satisfying the assumptions of Lemma 2.4, and our goal is to produce a monochromatic $B_{t,m}$. As such, we will stop the process if $|A| \geqslant t$ or $|B| \geqslant t$, and we aim to prove that at the end of the process, we still have $|Y| \geqslant m$ and $|Z| \geqslant m$. When the process terminates, we have thus found our monochromatic $B_{t,m}$. Note that this stopping condition guarantees that we never do more than t red steps or more than t blue steps.

The two things that remain are to define α_R, α_B, and to specify the cutoff for κ that determines whether we do a density-boost step or a red or blue step. For the former, we use the same definition as before, only adjusted to the fact that we now do at most t red or blue steps, namely we set $\varepsilon = t^{-1/4}$ and define

$$(5.1) \qquad \alpha(p) := \begin{cases} \frac{\varepsilon}{t} & \text{if } p \leqslant \frac{1}{2} + \frac{1}{t}, \\ \varepsilon(p - \frac{1}{2}) & \text{otherwise}, \end{cases}$$

and then set $\alpha_R := \alpha(p_R), \alpha_B := \alpha(p_B)$. Finally, the cutoff κ_{cutoff} for κ will be a fixed constant.

5.2. Formal definition of the symmetric book algorithm

We are now ready to define the symmetric book algorithm of Balister et al. (2024). We begin by fixing various parameters. First, as in Lemma 5.3, we set $C = 6$ and $c = 1/8$. We then define

$$(5.2) \qquad \eta := \frac{1}{8000} \qquad \text{and} \qquad t := \eta k,$$

and set $\varepsilon = t^{-1/4}$, analogously to our choice of ε in Section 3. Finally, we set

$$(5.3) \qquad \kappa_{\text{cutoff}} := 400 = \sqrt{\frac{20}{\eta}}.$$

We remark that in the proof, the only properties we will use of these constants are that

$$\kappa_{\text{cutoff}} \geqslant 8C, \quad \kappa_{\text{cutoff}}^2 \geqslant 8\log_2 \frac{1}{c}, \quad 2\left(\log_2 \frac{1}{c} + C\kappa_{\text{cutoff}}\right) + 5 < \frac{1}{\eta}, \quad \text{and} \quad \kappa_{\text{cutoff}}^2 \geqslant \frac{20}{\eta}.$$

Clearly any sufficiently large κ_{cutoff} will satisfy the first two constraints, and additionally for any sufficiently large κ_{cutoff} we can pick an appropriate η to satisfy the latter two. The final upper bound on $r(k)$ that we will prove (see Theorem 5.11) is $r(k) \leqslant 2^{(2-\eta^2/10)k}$, hence we are interested in choosing η as large as possible while satisfying the constraints above. However, we remark that no real effort was made to optimize any of the constants.

The symmetric book algorithm is then defined as follows, initialized with $A = B = \varnothing$ and $X \sqcup Y \sqcup Z$ an arbitrary partition of $V(K_N)$ into three sets of size[(†)] $N/3$.

ALGORITHM 5.4 (Symmetric book algorithm)

1. If $X \leqslant 1, |A| \geqslant t$, or $|B| \geqslant t$, stop the process.
2. Let $p_R = d_R(X, Y)$ and $p_B = d_B(X, Z)$ be the current red and blue densities. Define $\alpha_R = \alpha(p_R)$ and $\alpha_B = \alpha(p_B)$ according to (5.1).
3. Perform a *refinement step*: apply Lemma 5.3 to obtain a vertex $v \in X$, a set $X' \subseteq X$, and a number $\kappa \geqslant 0$. Let $Y' = Y \cap N_R(v)$ and $Z' = Z \cap N_B(v)$.
4. If $\kappa \geqslant \kappa_{\text{cutoff}}$, perform a *density-boost step* in one of the two colors.
   (a) If $d_R(X', Y') \geqslant (\kappa^2 - 1)p_R$, perform a *red density-boost step*, by updating
   $$A \to A, \qquad B \to B, \qquad X \to X', \qquad Y \to Y', \qquad Z \to Z,$$
   and return to step 1.
   (b) If $d_B(X', Z') \geqslant (\kappa^2 - 1)p_B$, perform a *blue density-boost step*, by updating
   $$A \to A, \qquad B \to B, \qquad X \to X', \qquad Y \to Y, \qquad Z \to Z',$$
   and return to step 1.
5. If $\kappa < \kappa_{\text{cutoff}}$, check whether v has at least $\frac{1}{2}|X'|$ red neighbors in X'.
   (a) If yes, perform a *red step*, by updating
   $$A \to A \cup \{v\}, \qquad B \to B, \qquad X \to X' \cap N_R(v), \qquad Y \to Y', \qquad Z \to Z,$$
   and return to step 1.
   (b) If not, then v has at least[(†)] $\frac{1}{2}|X'|$ blue neighbors in X'. In this case, perform a *blue step*, by updating
   $$A \to A, \qquad B \to B \cup \{v\}, \qquad X \to X' \cap N_B(v), \qquad Y \to Y, \qquad Z \to Z',$$
   and return to step 1.

Note that by Assumption 5.1, we have that $|Y'| = p_R|Y|$ and $|Z'| = p_B|Z|$. Additionally, whenever we perform a red step or a red density-boost step, we shrink X to a subset and do not change Z, hence we do not affect p_B; similarly, we do not affect p_R when doing a red step or red density-boost step. Using these facts, we can fill out Table 5.1, which records how the key parameters change during the execution of Algorithm 5.4.

5.3. Analysis of the symmetric book algorithm

We now collect various lemmas about how the development of the parameters throughout the execution of Algorithm 5.4. Let s_R and s_B denote the total number of red and

	$|A|$	$|B|$	$|X|$	$|Y|$	$|Z|$	p_R	p_B
red density-boost	–	–	$\times c2^{-C\kappa}$	$\times p_R$	–	$+(\kappa^2-1)\alpha_R$	–
blue density-boost	–	–	$\times c2^{-C\kappa}$	–	$\times p_B$	–	$+(\kappa^2-1)\alpha_B$
red step	$+1$	–	$\times \frac{1}{2}c2^{-C\kappa}$	$\times p_R$	–	$-\alpha_R$	–
blue step	–	$+1$	$\times \frac{1}{2}c2^{-C\kappa}$	–	$\times p_B$	–	$-\alpha_B$

TABLE 5.1. How the various parameters evolve during Algorithm 5.4.

blue density-boost steps done during the process. Recall that by the stopping condition in step 1, we do at most t red and at most t blue steps.

The first two lemmas are proved identically to Lemmas 3.3 and 3.4, respectively.

LEMMA 5.5. — *We have $p_R \geqslant \frac{1}{2} - \varepsilon$ and $p_B \geqslant \frac{1}{2} - \varepsilon$ throughout the entire process.*

LEMMA 5.6. — *At the end of the process, we have $|Y| \geqslant 2^{-t-s_R-o(k)}N$ and $|Z| \geqslant 2^{-t-s_B-o(k)}N$.*

Now, let $\kappa_{R,i}, \kappa_{B,i}$ denote the value of κ used during the ith red (resp. blue) density-boost step. Additionally, let κ_R, κ_B denote the average value of these numbers, i.e.

$$\kappa_R := \frac{1}{s_R}\sum_{i=1}^{s_R}\kappa_{R,i} \qquad \text{and} \qquad \kappa_B := \frac{1}{s_B}\sum_{i=1}^{s_B}\kappa_{B,i}.$$

The next lemma is analogous to Lemma 3.5.

LEMMA 5.7. — *At the end of the process, we have*

$$|X| \geqslant \left(\frac{1}{2}c2^{-C\kappa_{\text{cutoff}}}\right)^{2t+o(k)} c^{s_R+s_B}2^{-C(s_R\kappa_R+s_B\kappa_B)}N.$$

Proof. — Every red or blue step shrinks X by at most a factor[†] of $\frac{1}{2}c2^{-C\kappa_{\text{cutoff}}}$, since we only do a red or blue step if the current value of κ is at most κ_{cutoff}. As we do at most t red and at most t blue steps, we obtain the factor of $(\frac{1}{2}c2^{-C\kappa_{\text{cutoff}}})^{2t}$. Similarly, the ith red density-boost step shrinks X by at most a factor of $c2^{-C\kappa_{R,i}}$, hence the total contribution of the red density-boost steps is

$$\prod_{i=1}^{s_R}c2^{-C\kappa_{R,i}} = c^{s_R}2^{-C\sum_{i=1}^{s_R}\kappa_{R,i}} = c^{s_R}2^{-Cs_R\kappa_R},$$

by the definition of κ_R. Adding in the analogous contribution of the blue density-boost steps yields the claimed result. $\square$

Finally, the following lemma is analogous to Lemma 3.6, and is proved identically. Indeed, the only difference from Lemma 3.6 is that, rather than saying that the ith density-boost step boosts the red density by $\frac{1-\beta_i}{\beta_i}\alpha$, we say that the ith red density-boost step increases p_R by at least $(\kappa_{R,i}^2-1)\alpha_R$, and similarly for the blue density-boost steps.

LEMMA 5.8. — *We have*

$$\sum_{i=1}^{s_R}(\kappa_{R,i}^2 - 1) \leqslant t + o(k) \qquad and \qquad \sum_{i=1}^{s_B}(\kappa_{B,i}^2 - 1) \leqslant t + o(k).$$

We remark that in the analysis of the original book algorithm, it was critical to prove Lemma 3.6 with a main term of t, rather than a weaker bound such as $2t$. There is enough slack in the present argument so that a much weaker bound than Lemma 5.8 would suffice (and indeed Balister et al. (2024, Lemma 4.3) prove such a weaker bound, with a constant-factor loss).

Using Lemma 5.8, we can prove the following bound on the number of density-boost steps and on their total contribution to Lemma 5.7, which will play the same role in the argument as Lemma 3.7.

LEMMA 5.9. — *For sufficiently large k, we have*

$$s_R \kappa_R \leqslant \frac{4t}{\kappa_{\text{cutoff}}} \qquad and \qquad s_B \kappa_B \leqslant \frac{4t}{\kappa_{\text{cutoff}}}.$$

Additionally,

$$s_R \leqslant \frac{4t}{\kappa_{\text{cutoff}}^2} \qquad and \qquad s_B \leqslant \frac{4t}{\kappa_{\text{cutoff}}^2}.$$

Proof. — Recall from (5.2) that $t = \eta k$ for a fixed constant η. Hence, for sufficiently large k, the $t + o(k)$ term in Lemma 5.8 is at most $2t$. We henceforth assume that k is sufficiently large for this to hold. Next, we recall by step 4 of Algorithm 5.4 that $\kappa_{R,i} \geqslant \kappa_{\text{cutoff}}$ for all i, and that $\kappa_{\text{cutoff}} \geqslant \sqrt{2}$ by our choice in (5.3). Therefore,

$$2t \geqslant \sum_{i=1}^{s_R}(\kappa_{R,i}^2 - 1) \geqslant \frac{1}{2}\sum_{i=1}^{s_R}\kappa_{R,i}^2 \geqslant \frac{\kappa_{\text{cutoff}}}{2}\sum_{i=1}^{s_R}\kappa_{R,i} = \frac{\kappa_{\text{cutoff}}}{2} \cdot (s_R\kappa_R),$$

where we use Lemma 5.8 in the first inequality, that $x^2 - 1 \geqslant \frac{1}{2}x^2$ for $x \geqslant \sqrt{2}$ in the second, that $\kappa_{R,i} \geqslant \kappa_{\text{cutoff}}$ in the third, and the definition of κ_R in the fourth. This, together with the analogous result for blue, yields the first claimed result.

The second claimed result then follows immediately if we use once more that $\kappa_R \geqslant \kappa_{\text{cutoff}}$, and similarly for blue. $\qquad\square$

Recall from Lemma 5.7 that the total contribution of the density-boost steps to $|X|$ is $c^{s_R+s_B}e^{-C(s_R\kappa_R+s_B\kappa_B)}$. By Lemma 5.9, both of these exponential terms can be bounded as $\exp(O(t/\kappa_{\text{cutoff}}))$. This means that if κ_{cutoff} is sufficiently large in terms of c and C—and our choice in (5.3) is indeed sufficiently large—then the contribution of the density-boost steps is negligible compared to the other exponential terms.

It is here that we see why it is crucial to obtain in Lemma 5.3 a density boost which asymptotically dominates κ. It is thanks to this that the proof of Lemma 5.9 gives us a κ_{cutoff} term in the denominator; thanks to this we can make the density-boost steps negligible by a sufficiently large choice of κ_{cutoff}. That is, if Lemma 5.3 only gave a density-boost of order κ, this argument would not work. Luckily Lemma 5.3 boosts the density by a quadratic quantity in κ, which makes the whole argument go through.

5.4. Proof of the upper bound on $r(k)$

Given our lemmas above on the evolution of the symmetric book algorithm, it is fairly straightforward to prove the following result, which implies that Algorithm 5.4 always outputs a sufficiently large monochromatic book.

LEMMA 5.10. — *Let k be sufficiently large, and let*

$$N \geqslant \max\{2^{t+\eta t/4}m, 2^k\}$$

for some integer m. Then Algorithm 5.4 terminates by finding a monochromatic copy of $B_{t,m}$.

Proof. — It suffices to prove that at the end of the process, we have $|X| > 1$ and $|Y|, |Z| \geqslant m$. Indeed, if this is the case, then Algorithm 5.4 can only terminate when one of A or B has size t, at which point it forms a monochromatic copy of $B_{t,m}$ with Y or Z, respectively.

For the second claim, we have by Lemma 5.6 that $|Y| \geqslant 2^{-t-s_R-o(k)}N$. By Lemma 5.9, we have that

$$s_R \leqslant \frac{4t}{\kappa_{\text{cutoff}}^2} = \frac{\eta t}{5},$$

since $\kappa_{\text{cutoff}}^2 = 20/\eta$ by (5.3). Therefore, for k sufficiently large we have that $s_R + o(k) \leqslant \eta t/4$, hence $|Y| \geqslant 2^{-t-\eta t/4}N \geqslant m$. The proof that $|Z| \geqslant m$ at the end of the process proceeds identically.

It remains to prove the lower bound on $|X|$, for which we use Lemma 5.7. We estimate each of the quantities appearing in it in turn. First, recalling the definitions of c, C, and κ_{cutoff}, we see that

$$\frac{1}{2}c2^{-C\kappa_{\text{cutoff}}} = 2^{-4} \cdot 2^{-6\cdot400} = 2^{-2404}.$$

Therefore, for k sufficiently large, we have that

$$\left(\frac{1}{2}c2^{-C\kappa_{\text{cutoff}}}\right)^{2t+o(k)} \geqslant 2^{-6000t}.$$

Next, by Lemma 5.9, we have that

$$c^{s_R+s_B} = 2^{-3(s_R+s_B)} \geqslant 2^{-24t/\kappa_{\text{cutoff}}^2} \geqslant 2^{-t},$$

since $24/\kappa_{\text{cutoff}}^2 = 24/400^2 \leqslant 1$. Finally, again by Lemma 5.9, we have that

$$C(s_R\kappa_R + s_B\kappa_B) \leqslant \frac{8Ct}{\kappa_{\text{cutoff}}} = \frac{48}{400}t \leqslant t.$$

Plugging this all into Lemma 5.7, we conclude that at the end of the process,

$$|X| \geqslant \left(\frac{1}{2}c2^{-C\kappa_{\text{cutoff}}}\right)^{2t+o(k)} c^{s_R+s_B}2^{-C(s_R\kappa_R+s_B\kappa_B)}N \geqslant 2^{-6000t-t-t}N > 2^{-t/\eta}N = 2^{-k}N,$$

since $6002 < 1/\eta$ and $k = \eta t$. Finally, by our assumption that $N \geqslant 2^k$, we conclude that $|X| > 1$ at the end of the process, as claimed. $\qquad\square$

Finally, the upper bound on diagonal Ramsey numbers is an immediate consequence of Lemmas 2.4 and 5.10.

THEOREM 5.11 (Balister et al., 2024). — *If k is sufficiently large, then*

$$r(k) \leqslant 2^{(2-\eta^2/10)k}.$$

We stress that, as always, the proof we give is not complete, as it relies on Assumptions 5.1 and 5.2, both of which we have not justified. Additionally, we remark that this is not the strongest bound that can be proved by this technique; in particular, the choices of η and κ_{cutoff} are not fully optimized.

Proof of Theorem 5.11. — Let k be sufficiently large, and let $N = 2^{(2-\eta^2/10)k}$. Fix a two-coloring of $E(K_N)$, and run Algorithm 5.4. Recalling that $t = \eta k$, let $m = r(k-t, k)$. By Theorem 2.2, we have that

$$m \leqslant \binom{2k-t}{k} = \binom{(2-\eta)k}{k} \leqslant 2^{(2-\eta)H(\frac{1}{2-\eta})\cdot k}.$$

Now, we have that

$$\left(x + \frac{x^2}{4}\right) + (2-x)H\left(\frac{1}{2-x}\right) \leqslant 2 - \frac{x^2}{10}$$

for all $x \in [0, 1]$; this can be verified by computing the Taylor series of the difference, and noting that all coefficients are non-positive. Applying this fact with $x = \eta$, we conclude that

$$2^{t+\eta t/4}m \leqslant 2^{(\eta+\eta^2/4)\cdot k} \cdot 2^{(2-\eta)H(\frac{1}{2-\eta})\cdot k} \leqslant 2^{(2-\eta^2/10)k} = N.$$

Since we also have $N \geqslant 2^k$, Lemma 5.10 implies that the coloring contains a monochromatic $B_{t,m}$. By Lemma 2.4, we conclude that the coloring also contains a monochromatic K_t, as claimed. $\square$

6. High-dimensional geometry and the proof of Lemma 5.3

6.1. Reduction to a geometric statement

All that remains now is to prove Lemma 5.3, which is really the heart of the proof presented above. As previously mentioned, Lemma 5.3 is a consequence of a purely geometric lemma, about probability distributions in high-dimensional Euclidean space. Before stating this geometric lemma, however, let us build up to it by stating a probabilistic strengthening of Lemma 5.3.

For a vertex $v \in X$, let us denote by $N_Y(v)$ the red neighborhood of v in Y, that is, $N_Y(v) := Y \cap N_R(v)$. Similarly, $N_Z(v) := Z \cap N_B(v)$ will denote the blue neighborhood of v in Z. With this notation, we now state the probabilistic strengthening of Lemma 5.3.

LEMMA 6.1. — *Let X, Y, Z be disjoint sets of vertices in a coloring of $E(K_N)$, and let $\alpha_R, \alpha_B \in (0,1)$. Let $p_R = d_R(X,Y)$ and $p_B = d_B(X,Z)$. Let $\boldsymbol{v}, \boldsymbol{w}$ be independent, uniformly random vertices of X. There exists a real number $\kappa \geqslant 0$ such that*

$$(6.1) \quad \Pr\Big(|N_Y(\boldsymbol{v}) \cap N_Y(\boldsymbol{w})| \geqslant (p_R + (\kappa^2 - 1)\alpha_R)p_R|Y|$$

$$\text{and} \quad |N_Z(\boldsymbol{v}) \cap N_Z(\boldsymbol{w})| \geqslant (p_B - \alpha_B)p_B|Z|\Big) \geqslant c2^{-C\kappa},$$

or else the same holds upon swapping the roles of red and blue and of Y and Z. Here, $c = 1/8$ and $C = 6$ are the same constants as in Lemma 5.3.

Note that Lemma 6.1 is a more symmetric statement than Lemma 5.3: it discusses the common red and blue neighborhoods of a random pair of vertices in X, as opposed to isolating a special vertex $v \in X$ as well as a special subset $X' \subseteq X$. Nonetheless, Lemma 5.3 follows from Lemma 6.1 by a simple averaging argument.

Proof of Lemma 5.3, assuming Lemma 6.1. — Let us suppose that (6.1) holds; the other case, where the roles of the colors are swapped, follows by a symmetric argument. First, by the law of total probability, there exists some fixed $v \in X$ such that

$$(6.2) \quad \Pr\Big(|N_Y(v) \cap N_Y(\boldsymbol{w})| \geqslant (p_R + (\kappa^2 - 1)\alpha_R)p_R|Y|$$

$$\text{and} \quad |N_Z(v) \cap N_Z(\boldsymbol{w})| \geqslant (p_B - \alpha_B)p_B|Z|\Big) \geqslant c2^{-C\kappa},$$

This is the same inequality as (6.1), except that now only $\boldsymbol{w}$ is random. As in the statement of Lemma 5.3, let $Y' = N_Y(v)$ and $Z' = N_Z(v)$. Let $X' \subseteq X$ denote the set of all $w \in X$ such that

$$|N_Y(v) \cap N_Y(w)| \geqslant (p_R + (\kappa^2 - 1)\alpha_R)p_R|Y| \text{ and } |N_Z(v) \cap N_Z(w)| \geqslant (p_B - \alpha_B)p_B|Z|.$$

Then (6.2) is equivalent to the statement that $|X'| \geqslant c2^{-C\kappa}|X|$, proving Lemma 5.3(a). Next, we have that

$$e_R(X', Y') = \sum_{w \in X'} |N_R(w) \cap Y'| = \sum_{w \in X'} |N_Y(v) \cap N_Y(w)| \geqslant |X'| \cdot (p_R + (\kappa^2 - 1)\alpha_R)p_R|Y|,$$

by the definition of X'. Recalling that $|Y'| = p_R|Y|$ by Assumption 5.1, we find that

$$d_R(X', Y') = \frac{e_R(X', Y')}{|X'||Y'|} \geqslant p_R + (\kappa^2 - 1)\alpha_R,$$

which proves Lemma 5.3(c) and also implies the first bound in Lemma 5.3(b). Similarly, we can compute that

$$d_B(X', Z') \geqslant p_B - \alpha_B,$$

which completes the proof of Lemma 5.3(b). $\qquad \square$

It thus only remains to prove Lemma 6.1, which follows from a geometric argument. In order to convert the statement of Lemma 6.1 to a geometric one, we simply observe that the size of the intersection of two sets can be encoded in a linear-algebraic way, as the inner product of their indicator vectors. Namely, let us denote by $\mathbf{1}_Y(v) \in \mathbb{R}^Y$ the

indicator vector of the set $N_Y(v)$; this is the vector whose yth coordinate equals 1 if $y \in N_Y(v)$, and 0 otherwise. Then we have the identity

$$\langle \mathbf{1}_Y(v), \mathbf{1}_Y(w) \rangle = |N_Y(v) \cap N_Y(w)|,$$

since the inner product of two binary vectors is precisely the number of coordinates in which they both have a 1.

In order to precisely arrive at the setup of Lemma 6.1, we center and rescale the vectors $\mathbf{1}_Y(v)$. Namely, letting $\mathbf{1}$ denote the all-ones vector, we define

$$(6.3) \qquad \sigma_Y(v) := \frac{\mathbf{1}_Y(v) - p_R \mathbf{1}}{\sqrt{\alpha_R p_R |Y|}}.$$

Note that by Assumption 5.1, $\mathbf{1}_Y(v)$ has exactly $p_R |Y|$ entries equal to 1, hence by subtracting $p_R \mathbf{1}$ we ensure that the sum of the entries in $\sigma_Y(v)$ is zero. Thus, all the vectors $\{\sigma_Y(v) : v \in X\}$ lie on a sphere of radius $\sqrt{\frac{1-p_R}{\alpha_R}}$ centered at the origin in $\mathbb{R}^Y$. Additionally, we can compute that for all $v, w \in X$, we have

$$\langle \sigma_Y(v), \sigma_Y(w) \rangle = \frac{1}{\alpha_R p_R |Y|} \langle \mathbf{1}_Y(v) - p_R \mathbf{1}, \mathbf{1}_Y(w) - p_R \mathbf{1} \rangle$$

$$= \frac{1}{\alpha_R p_R |Y|} \left(\langle \mathbf{1}_Y(v), \mathbf{1}_Y(w) \rangle - p_R \langle \mathbf{1}_Y(v), \mathbf{1} \rangle - p_R \langle \mathbf{1}, \mathbf{1}_Y(w) \rangle + p_R^2 \langle \mathbf{1}, \mathbf{1} \rangle \right)$$

$$= \frac{1}{\alpha_R p_R |Y|} \left(|N_Y(v) \cap N_Y(w)| - p_R^2 |Y| \right),$$

where we used Assumption 5.1 in the final step, to conclude that $\langle \mathbf{1}_Y(v), \mathbf{1} \rangle = \langle \mathbf{1}, \mathbf{1}_Y(w) \rangle = p_R |Y|$. Note that if $N_Y(v), N_Y(w)$ were random 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 $\langle \sigma_Y(v), \sigma_Y(w) \rangle$ measures (up to scaling) how much their true intersection size deviates from this average value. In particular, for any $\lambda \in \mathbb{R}$, we have that $\langle \sigma_Y(v), \sigma_Y(w) \rangle \geqslant \lambda$ if and only if

$$\frac{1}{\alpha_R p_R |Y|} \left(|N_Y(v) \cap N_Y(w)| - p_R^2 |Y| \right) \geqslant \lambda,$$

which in turn happens if and only if

$$|N_Y(v) \cap N_Y(w)| \geqslant (p_R + \lambda \alpha_R) p_R |Y|.$$

This shows that Lemma 6.1 is a special case of the following geometric result of Balister et al. (2024).

THEOREM 6.2 (Geometric lemma). — *Let X, Y, Z be finite sets, and let $\sigma_Y : X \to \mathbb{R}^Y$ and $\sigma_Z : X \to \mathbb{R}^Z$ be arbitrary functions. Let $\boldsymbol{v}, \boldsymbol{w}$ be independent, uniformly random elements of X. There exists a real number $\kappa \geqslant 0$ such that*

$$(6.4) \qquad \Pr \left(\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle \geqslant \kappa^2 - 1 \quad and \quad \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle \geqslant -1 \right) \geqslant c2^{-C\kappa},$$

or else the same holds upon swapping the roles of Y and Z.

Indeed, the statement of Lemma 6.1 is precisely that of Theorem 6.2 when one defines σ_Y as in (6.3) (and analogously for σ_Z). Hence it only remains to prove Theorem 6.2.

Before turning to the proof, a few remarks about Theorem 6.2 are in order. First, note that the sets Y and Z play no role in the statement: we could equally well take σ_Y and σ_Z to be valued in the same high-dimensional Euclidean space $\mathbb{R}^n$, or alternately in the infinite-dimensional Hilbert space ℓ^2. Second, as we will see, the set X also does not really matter: the same statement holds for arbitrary coupled probability distributions on $\mathbb{R}^Y \times \mathbb{R}^Z$, and not only those distributions that are uniform on a finite support. Finally, and perhaps most importantly, we stress that the statement of Theorem 6.2 is *not* scale-invariant, unlike many other results about inner products. That is, because the outcome of Theorem 6.2 involves constants like -1, the statement of the lemma changes if we replace, say, σ_Y by $100\sigma_Y$: we cannot simply take the same value of κ upon such a rescaling. In fact, this lack of scale-invariance is crucial in the deduction of Lemma 6.1 and hence Lemma 5.3: recall from (6.3) that we chose a careful scaling of σ_Y, in order to incorporate the parameter α_R.

6.2. The one-color version of Theorem 6.2

Theorem 6.2 is a "two-color" statement, in that it involves two functions σ_Y, σ_Z, as is necessary for the application to two-color Ramsey numbers. Balister et al. (2024, Lemma 3.1) prove a more general result, which gives essentially the same outcome but for an arbitrary number of functions $\sigma_1, \ldots, \sigma_q$; this is, naturally, necessary for their application to upper bounds on multicolor Ramsey numbers. While we will continue to work in the two-color setting, the proof of Theorem 6.2 is easier to understand if one begins by proving a simpler statement, which is the *one-color* version of this result.

PROPOSITION 6.3 (One-color geometric lemma). — *Let X, Y be finite sets, and let $\sigma : X \to \mathbb{R}^Y$ be an arbitrary function. Let $\boldsymbol{v}, \boldsymbol{w}$ be independent, uniformly random elements of X. There exists a real number $\kappa \geqslant 0$ such that*

$$\Pr\left(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle \geqslant \kappa^2 - 1\right) \geqslant \frac{1}{(\kappa^2 + 2)^2}.$$

Note that the quantitative dependencies in the one-color statement are much better: we obtain a *polynomial* relationship between the lower bound on the inner products and the success probability, rather than the exponential relationship appearing in Theorem 6.2. It remains open whether there is such a polynomial strengthening of Theorem 6.2; if there is, one would obtain a strengthened version of Lemma 5.3, which would in all likelihood lead to better upper bounds on diagonal Ramsey numbers.

In the proof of Proposition 6.3, we will use two simple and well-known observations. The first states that the expected inner product of independent, identically distributed random vectors is non-negative.

LEMMA 6.4. — *Let $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$ be independent, identically distributed random vectors in $\mathbb{R}^n$, sampled from an arbitrary probability distribution. Then*

$$\mathbb{E}[\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle] \geqslant 0.$$

Proof. — By linearity of expectation and the independence of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$, we have

$$\mathbb{E}[\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle] = \langle \mathbb{E}[\boldsymbol{\sigma}], \mathbb{E}[\boldsymbol{\sigma}'] \rangle.$$

$\mathbb{E}[\boldsymbol{\sigma}]$ is some fixed vector in $\mathbb{R}^n$, and as $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are identically distributed, it equals $\mathbb{E}[\boldsymbol{\sigma}']$. Hence $\mathbb{E}[\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle]$ is simply the squared norm of some vector in $\mathbb{R}^n$, which is non-negative. $\square$

The second well-known observation we need is an immediate consequence of Fubini's theorem.

LEMMA 6.5. — *For any random variable $\boldsymbol{V}$, we have*

$$\mathbb{E}[\boldsymbol{V}] \leqslant \int_0^\infty \Pr(\boldsymbol{V} \geqslant \lambda) \, \mathrm{d}\lambda.$$

Proof. — Let the underlying probability space of $\boldsymbol{V}$ be (Ω, μ). Then $\Pr(\boldsymbol{V} \geqslant \lambda) = \int_\Omega \mathbf{1}_{\boldsymbol{V}(\omega) \geqslant \lambda} \, \mathrm{d}\mu(\omega)$. Additionally, for any fixed real number V, we have $\int_0^\infty \mathbf{1}_{V \geqslant \lambda} \, \mathrm{d}\lambda$ equals V if $V \geqslant 0$, and equals 0 otherwise. In particular, for all V, we have $\int_0^\infty \mathbf{1}_{V \geqslant \lambda} \, \mathrm{d}\lambda \geqslant V$. Therefore, by Fubini's theorem,

$$\begin{aligned}
\int_0^\infty \Pr(\boldsymbol{V} \geqslant \lambda) \, \mathrm{d}\lambda &= \int_0^\infty \int_\Omega \mathbf{1}_{\boldsymbol{V}(\omega) \geqslant \lambda} \, \mathrm{d}\mu(\omega) \, \mathrm{d}\lambda \\
&= \int_\Omega \int_0^\infty \mathbf{1}_{\boldsymbol{V}(\omega) \geqslant \lambda} \, \mathrm{d}\lambda \, \mathrm{d}\mu(\omega) \\
&\geqslant \int_\Omega \boldsymbol{V}(\omega) \, \mathrm{d}\mu(\omega) \\
&= \mathbb{E}[\boldsymbol{V}]. \qquad \square
\end{aligned}$$

Proposition 6.3 is now a simple consequence of these two observations.

Proof of Proposition 6.3. — Let $\boldsymbol{V} = \langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle + 1$, so that $\mathbb{E}[\boldsymbol{V}] \geqslant 1$ by Lemma 6.4. By Lemma 6.5, we find that

$$1 \leqslant \mathbb{E}[\boldsymbol{V}] \leqslant \int_0^\infty \Pr(\boldsymbol{V} \geqslant \lambda) \, \mathrm{d}\lambda = \int_{-1}^\infty \Pr(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle \geqslant \lambda) \, \mathrm{d}\lambda.$$

If we had that $\Pr(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle \geqslant \lambda) \leqslant 1/(\lambda + 3)^2$ for all $\lambda \geqslant -1$, then we would conclude that

$$1 \leqslant \int_{-1}^\infty \Pr(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle \geqslant \lambda) \, \mathrm{d}\lambda \leqslant \int_{-1}^\infty \frac{1}{(\lambda + 3)^2} \, \mathrm{d}\lambda = \frac{1}{2},$$

a contradiction. Therefore, there exists some $\lambda \geqslant -1$ for which $\Pr(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle \geqslant \lambda) \geqslant 1/(\lambda + 3)^2$. Substituting $\kappa = \sqrt{\lambda + 1}$, we conclude that

$$\Pr(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle \geqslant \kappa^2 - 1) \geqslant \frac{1}{(\kappa^2 + 2)^2},$$

as claimed. $\square$

6.3. Proof of Theorem 6.2

Although the proof of Proposition 6.3 was very simple, it is helpful to reconsider it from a high-level view, as a very similar strategy will be used to prove Theorem 6.2. In the proof of Proposition 6.3, we considered the function $f(x) = x + 1$, and then defined a random variable

$$\boldsymbol{V} \coloneqq f(\langle \sigma(\boldsymbol{v}), \sigma(\boldsymbol{w}) \rangle).$$

The proof of Proposition 6.3 relies on three important properties of this function f, although they are all so simple that they were not explicitly noted in the proof. First, f is non-positive on the region $\{x : x \leqslant -1\}$; this was implicitly used to restrict our attention to $\lambda \geqslant -1$, which ends up meaning that we obtain $\kappa \geqslant 0$. Second, f does not grow too fast, which was used to obtain a good trade-off between λ and the probability $\Pr(\boldsymbol{V} \geqslant \lambda)$. Finally, and most importantly, we had the inequality $\mathbb{E}[\boldsymbol{V}] \geqslant 1$, which followed from Lemma 6.4.

In the proof of Theorem 6.2, we will pick a more complicated function $f : \mathbb{R}^2 \to \mathbb{R}$, and use it to define a random variable $\boldsymbol{V}$ by

$$(6.5) \qquad \boldsymbol{V} \coloneqq f(\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle, \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle).$$

In addition to requiring f to satisfy analogues of the three properties discussed above, we also want the choice of f to allow us to relate $\Pr(\boldsymbol{V} \geqslant \lambda)$ to the probability we wish to estimate in (6.4). The main difficulty, of course, is that we now want to ensure that two inner products are simultaneously large.

To that end, we define

$$(6.6) \qquad \boldsymbol{M} \coloneqq \max\{\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle, \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle\}$$

and let $\mathcal{E}$ be the event

$$(6.7) \qquad \mathcal{E} \coloneqq \{\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle \geqslant -1 \qquad \text{and} \qquad \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle \geqslant -1\}.$$

Then our first goal is to obtain a lower bound on $\Pr(\{\boldsymbol{M} \geqslant \kappa^2 - 1\} \cap \mathcal{E})$, for some κ; such a lower bound can be directly converted to a lower bound on the probability in (6.4) (or on the analogous quantity upon interchanging the roles of Y and Z). As such, it is natural to try defining the random variable $\boldsymbol{V}$ as $\boldsymbol{M}\boldsymbol{1}_{\mathcal{E}}$, since then the probability we are interested in is precisely of the form $\Pr(\boldsymbol{V} \geqslant \lambda)$. The problem with this choice is that it is not clear how to lower-bound $\mathbb{E}[\boldsymbol{V}]$, or indeed why one should expect any non-trivial lower bound to hold. Indeed, while $\mathbb{E}[\boldsymbol{M}] \geqslant 0$ by Lemma 6.4, it is possible that much of the positive contribution to $\mathbb{E}[\boldsymbol{M}]$ is attained on the complement of $\mathcal{E}$.

Thus, in order for $\boldsymbol{V}$ to "know about" the event $\mathcal{E}$, we will ensure that $f(y, z) \leqslant 0$ if $y \leqslant -1$ or $z \leqslant -1$. This ensures that $\boldsymbol{V}$ is non-positive on the complement of $\mathcal{E}$, so that bounds on $\Pr(\boldsymbol{V} \geqslant \lambda)$ can be converted to bounds on the quantity we are interested in, $\Pr(\{\boldsymbol{M} \geqslant \kappa^2 - 1\} \cap \mathcal{E})$. Next, as before, we want a global upper bound on $f(y, z)$ in terms of y and z, and we would like this upper bound to be "reasonable": we will use this bound to convert information of the form $\boldsymbol{V} \geqslant \lambda$ to information of form $\boldsymbol{M} \geqslant \kappa^2 - 1$, and we want good control on the dependencies between κ and λ. Finally,

and most importantly, we want that $\mathbb{E}[\boldsymbol{V}] \geqslant 1$, so that we can apply the same argument as in the proof of Proposition 6.3.

This last condition seems like the most difficult one to satisfy, since we have very little information about the random variables $\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle$ and $\langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle$: if f is some complicated function of these variables, why should we expect $\mathbb{E}[\boldsymbol{V}] \geqslant 1$?

The way to ensure this is via a very simple but remarkably powerful trick, as follows. First, we will ensure that f is analytic with $f(0,0) = 1$. We then consider the Taylor series of f, that is, we write $f(y,z) = \sum_{a,b \geqslant 0} r_{a,b} y^a z^b$, and apply analyticity and linearity of expectation to conclude that

$$\mathbb{E}[\boldsymbol{V}] = \sum_{a,b=0}^{\infty} r_{a,b} \mathbb{E}[\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle^a \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle^b].$$

The utility of this expansion comes from the following simple lemma, a generalization of Lemma 6.4, which states that all the expectations appearing in the expression above are non-negative. Thus, a lower bound on $\mathbb{E}[\boldsymbol{V}]$ follows immediately, as long as we ensure that f has only non-negative Taylor coefficients.

LEMMA 6.6. — *Fix any probability distribution on $\mathbb{R}^Y \times \mathbb{R}^Z$, and let $(\boldsymbol{\sigma}_Y, \boldsymbol{\sigma}_Z)$ and $(\boldsymbol{\sigma}'_Y, \boldsymbol{\sigma}'_Z)$ be independent samples from this distribution. We have*

$$\mathbb{E}[\langle \boldsymbol{\sigma}_Y, \boldsymbol{\sigma}'_Y \rangle^a \langle \boldsymbol{\sigma}_Z, \boldsymbol{\sigma}'_Z \rangle^b] \geqslant 0$$

for all non-negative integers a, b.

Proof. — Let $\boldsymbol{\sigma}$ be a random vector in $(\mathbb{R}^Y)^{\otimes a} \otimes (\mathbb{R}^Z)^{\otimes b}$ defined by

$$\boldsymbol{\sigma} = (\underbrace{\boldsymbol{\sigma}_Y \otimes \cdots \otimes \boldsymbol{\sigma}_Y}_{a \text{ times}}) \otimes (\underbrace{\boldsymbol{\sigma}_Z \otimes \cdots \otimes \boldsymbol{\sigma}_Z}_{b \text{ times}}),$$

and define $\boldsymbol{\sigma}'$ analogously. Note that $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are independent and identically distributed, and that

$$\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle = \langle \boldsymbol{\sigma}_Y, \boldsymbol{\sigma}'_Y \rangle^a \langle \boldsymbol{\sigma}_Z, \boldsymbol{\sigma}'_Z \rangle^b,$$

hence the result follows from Lemma 6.4. $\qquad\square$

So all that remains is to pick a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is non-positive on the set $\{(y,z) : y \leqslant -1 \text{ or } z \leqslant -1\}$, which does not grow too quickly, and which is analytic with non-negative Taylor coefficients. The following lemma, whose proof is a simple calculus exercise that we omit, provides such a function.

LEMMA 6.7. — *The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by*

$$(6.8) \qquad f(y,z) := 1 + y(2 + \cosh\sqrt{2z}) + z(2 + \cosh\sqrt{2y})$$

is non-positive on the set $\{(y,z) : y \leqslant -1 \text{ or } z \leqslant -1\}$, is analytic with non-negative Taylor coefficients, has $f(0,0) = 1$, and satisfies

$$(6.9) \qquad f(y,z) \leqslant e^{3\sqrt{\max\{y,z\}+1}}$$

for all $y, z \geqslant -1$.

With this, we have all of the ingredients in place to prove Theorem 6.2, following the strategy outlined above.

Proof of Theorem 6.2. — We define

$$\boldsymbol{V} := f(\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w})\rangle, \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w})\rangle)$$

as in (6.5), where $f : \mathbb{R}^2 \to \mathbb{R}$ is given by (6.8). By Lemma 6.6, the fact that f is analytic with non-negative Taylor coefficients, and the fact that $f(0,0) = 1$, we conclude that $\mathbb{E}[\boldsymbol{V}] \geqslant 1$.

Let $\mathcal{E}$ be the event defined in (6.7), and let $\overline{\mathcal{E}}$ be its complement. The fact that f is non-positive on the set $\{(y, z) : y \leqslant -1 \text{ or } z \leqslant -1\}$ implies that on the event $\overline{\mathcal{E}}$, the random variable $\boldsymbol{V}$ is non-positive. Therefore $\mathbb{E}[\boldsymbol{V}\mathbf{1}_{\overline{\mathcal{E}}}] \leqslant 0$, and hence

$$1 \leqslant \mathbb{E}[\boldsymbol{V}] = \mathbb{E}[\boldsymbol{V}\mathbf{1}_{\mathcal{E}}] + \mathbb{E}[\boldsymbol{V}\mathbf{1}_{\overline{\mathcal{E}}}] \leqslant \mathbb{E}[\boldsymbol{V}\mathbf{1}_{\mathcal{E}}].$$

By Lemma 6.5, we now have

$$1 \leqslant \mathbb{E}[\boldsymbol{V}\mathbf{1}_{\mathcal{E}}] \leqslant \int_0^\infty \Pr(\boldsymbol{V}\mathbf{1}_{\mathcal{E}} \geqslant \lambda)\,\mathrm{d}\lambda = \int_0^\infty \Pr(\{\boldsymbol{V} \geqslant \lambda\} \cap \mathcal{E})\,\mathrm{d}\lambda.$$

On the interval $[0, 1]$, the integrand is at most $\Pr(\mathcal{E})$, so

$$1 \leqslant \Pr(\mathcal{E}) + \int_1^\infty \Pr(\{\boldsymbol{V} \geqslant \lambda\} \cap \mathcal{E})\,\mathrm{d}\lambda.$$

We now recall (6.9), which implies that on the event $\mathcal{E}$, we have $\boldsymbol{V} \leqslant e^{3\sqrt{M+1}}$, hence

$$1 - \Pr(\mathcal{E}) \leqslant \int_1^\infty \Pr(\{\boldsymbol{V} \geqslant \lambda\} \cap \mathcal{E})\,\mathrm{d}\lambda \leqslant \int_1^\infty \Pr\left(\left\{e^{3\sqrt{M+1}} \geqslant \lambda\right\} \cap \mathcal{E}\right)\,\mathrm{d}\lambda.$$

We now change variables to $\lambda = e^{3\kappa}$, so that $e^{3\sqrt{M+1}} \geqslant \lambda$ if and only if $\boldsymbol{M} \geqslant \kappa^2 - 1$, and conclude that

$$(6.10) \qquad 1 - \Pr(\mathcal{E}) \leqslant \int_0^\infty \Pr(\{\boldsymbol{M} \geqslant \kappa^2 - 1\} \cap \mathcal{E}) \cdot 3e^{3\kappa}\,\mathrm{d}\kappa.$$

We now claim that for some $\kappa \geqslant 0$, we have

$$(6.11) \qquad \Pr(\{\boldsymbol{M} \geqslant \kappa^2 - 1\} \cap \mathcal{E}) \geqslant \frac{e^{-4\kappa}}{4}.$$

Indeed, if this does not hold, then continuing (6.10), we conclude that

$$1 - \Pr(\mathcal{E}) \leqslant \int_0^\infty \frac{e^{-4\kappa}}{4} \cdot 3e^{3\kappa}\,\mathrm{d}\kappa = \frac{3}{4}\int_0^\infty e^{-\kappa}\,\mathrm{d}\kappa = \frac{3}{4},$$

hence $\Pr(\mathcal{E}) \geqslant \frac{1}{4}$. But on the event $\mathcal{E}$ we have $\boldsymbol{M} \geqslant -1$, hence (6.11) holds with $\kappa = 0$, a contradiction. We conclude that, as claimed, (6.11) holds for some $\kappa \geqslant 0$.

Fixing such a κ, recalling the definitions of $\boldsymbol{M}$ and $\mathcal{E}$ from (6.6) and (6.7), and applying the union bound, we find that

$$\frac{e^{-4\kappa}}{4} \leqslant \Pr(\{\boldsymbol{M} \geqslant \kappa^2 - 1\} \cap \mathcal{E})$$

$$= \Pr\Big(\max\{\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle, \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle\} \geqslant \kappa^2 - 1 \text{ and}$$

$$\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle \geqslant -1 \text{ and } \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle \geqslant -1 \Big)$$

$$\leqslant \Pr\Big(\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle \geqslant \kappa^2 - 1 \text{ and } \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle \geqslant -1 \Big)$$

$$+ \Pr\Big(\langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle \geqslant \kappa^2 - 1 \text{ and } \langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle \geqslant -1 \Big).$$

Therefore, either

$$\Pr\Big(\langle \sigma_Y(\boldsymbol{v}), \sigma_Y(\boldsymbol{w}) \rangle \geqslant \kappa^2 - 1 \text{ and } \langle \sigma_Z(\boldsymbol{v}), \sigma_Z(\boldsymbol{w}) \rangle \geqslant -1 \Big) \geqslant \frac{e^{-4\kappa}}{8}$$

or the same holds upon reversing the roles of Y and Z. This concludes the proof of Theorem 6.2 upon recalling that $c = 1/8$ and $2^{-C} = 1/64 \leqslant e^{-4}$. $\qquad\square$

7. Epilogue: Ramsey's original proof of Theorem 1.1

As mentioned in Section 1, there is no known proof of Theorem 1.1 that does not use book graphs in some way. As a hopefully fitting end to this exposé, let us see the original proof of Ramsey (1929), which uses book graphs in a rather different way from Lemma 2.4, yet whose proof shares certain ideas with the ones we have already seen.

Let us denote by $r(B_{t,m})$ the least integer N such that every two-coloring of $E(K_N)$ contains a monochromatic copy of $B_{t,m}$. Ramsey (1929) proved the following upper bound on $r(B_{t,m})$.

THEOREM 7.1 (Ramsey, 1929). — $r(B_{t,m}) \leqslant (t+1)! \cdot m$ for all $t, m \geqslant 1$.

Note that since $K_k = B_{k-1,1}$, this immediately implies the bound $r(k) \leqslant k!$, and hence yields a proof of Theorem 1.1.

Proof of Theorem 7.1. — We proceed by induction on t. For the base case $t = 1$, we wish to prove that $r(B_{1,m}) \leqslant 2m$, which is immediate: in any two-coloring of $E(K_{2m})$, any vertex is incident to $2m - 1$ edges, at least m of which must have the same color by the pigeonhole principle. This yields a monochromatic copy of $B_{1,m}$.

For the inductive step, suppose the result has been proved for $t-1$, and fix a coloring of $E(K_N)$ where $N = (t+1)! \cdot m = t! \cdot ((t+1)m)$. By the inductive hypothesis, this coloring contains a monochromatic copy of $B_{t-1,(t+1)m}$, which we may assume to be red without loss of generality. That is, there exist disjoint sets $A, X \subseteq V(K_N)$ with $|A| = t - 1$ and $|X| = (t+1)m$, such that all edges inside A and between A and X are red. If there is a vertex $v \in X$ with at least m red neighbors in X, we may perform a "red step" by updating $A \to A \cup \{v\}$ and $X \to X \cap N_R(v)$, yielding a red $B_{t,m}$. So we

may assume that every vertex in X has at most $m - 1$ red neighbors in X. Let v_1 be an arbitrary vertex of X, and let $X_1 = N_B(v_1) \cap X$, so that $|X_1| \geqslant |X| - m$. Let v_2 be an arbitrary vertex of X_1, and let $X_2 = N_B(v_2) \cap X_1$, so that $|X_2| \geqslant |X_1| - m \geqslant |X| - 2m$. Continuing in this way, we may build a sequence of vertices $v_1, \ldots, v_t$ as well as a set X_t with $|X_t| \geqslant |X| - tm = m$, such that each v_i is adjacent in blue to all v_j with $j > i$, as well as to all vertices in X_t. But this precisely means that we have constructed a blue $B_{t,m}$, completing the inductive step. $\qquad\square$

Given Theorem 7.1, it is natural to wonder what the true value of $r(B_{t,m})$ is. This question was first explicitly raised by Erdős, Faudree, Rousseau, and Schelp (1978) and Thomason (1982), who independently proved the bounds $(2^t - o(1))m \leqslant r(B_{t,m}) \leqslant 4^t m$. Thomason in particular was motivated by Lemma 2.4, as discussed in Section 2.2, and made the following bold conjecture.

CONJECTURE 7.2 (Thomason, 1982). — $r(B_{t,m}) \leqslant 2^t(m + t - 2) + 2$ for all $m, t \geqslant 1$.

This conjecture is known to be true (and optimal) for $t \in \{1, 2\}$, but it is wide open for $t \geqslant 3$ (and may well be false). Moreover, Conjecture 7.2 is likely to be very difficult: even the $m = 1$ case would yield $r(k) \leqslant 2^{k+o(k)}$, a bound far stronger than anything currently known. However, a beautiful result of Conlon (2019) confirms this conjecture asymptotically for any fixed t.

THEOREM 7.3 (Conlon, 2019). — $r(B_{t,m}) = (2^t + o(1))m$ as $m \to \infty$, for any fixed t.

Conlon's result addresses a question that arguably goes back 90 years to the original work of Ramsey (1929), yet uses highly sophisticated tools developed in the interim, such as the regularity lemma of Szemerédi (1978), and this question is also closely related to a famous conjecture of Burr and Erdős (1975), which was recently resolved by Lee (2017). In fact, Ramsey theory has seen a number of recent breakthroughs on old, seemingly intractable problems by the introduction of remarkable new techniques: three other examples from the past two years are the works of Li (2023) on explicit constructions, of Mattheus and Verstraete (2024) on off-diagonal Ramsey numbers, and of Reiher and Rödl (2023) on restricted Ramsey graphs. There is every reason to hope and expect this trend to continue.

References

Vigleik Angeltveit and Brendan D. McKay (2024). $R(5,5) \leqslant 46$. arXiv: 2409.15709 [math.CO].

Paul Balister, Béla Bollobás, Marcelo Campos, Simon Griffiths, Eoin Hurley, Robert Morris, Julian Sahasrabudhe, and Marius Tiba (2024). *Upper bounds for multicolour Ramsey numbers.* arXiv: 2410.17197 [math.CO].

Thomas F. Bloom and Olof Sisask (2020). *Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions.* arXiv: 2007.03528 [math.NT].

Stefan A. Burr and Paul Erdős (1975). "On the magnitude of generalized Ramsey numbers for graphs", in: *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vols. I, II, III*. Vol. 10. Colloq. Math. Soc. János Bolyai. North-Holland, Amsterdam-London, pp. 215–240.

Marcelo Campos, Simon Griffiths, Robert Morris, and Julian Sahasrabudhe (2023). *An exponential improvement for diagonal Ramsey.* arXiv: 2303.09521 [math.CO].

David Conlon (2009). "A new upper bound for diagonal Ramsey numbers", *Ann. of Math. (2)* **170** (2), pp. 941–960.

——— (2019). "The Ramsey number of books", *Adv. Comb.*, Paper No. 3, 12pp.

David Conlon, Jacob Fox, and Benny Sudakov (2015). "Recent developments in graph Ramsey theory", in: *Surveys in combinatorics 2015*. Vol. 424. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 49–118.

David Conlon, Jacob Fox, and Yuval Wigderson (2022). "Ramsey numbers of books and quasirandomness", *Combinatorica* **42** (3), pp. 309–363.

Paul Erdős (1947). "Some remarks on the theory of graphs", *Bull. Amer. Math. Soc.* **53**, pp. 292–294.

Paul Erdős, Ralph J. Faudree, Cecil C. Rousseau, and Richard H. Schelp (1978). "The size Ramsey number", *Period. Math. Hungar.* **9** (1-2), pp. 145–161.

Paul Erdős and George Szekeres (1935). "A combinatorial problem in geometry", *Compositio Math.* **2**, pp. 463–470.

Geoffrey Exoo (1989). "A lower bound for $R(5,5)$", *J. Graph Theory* **13** (1), pp. 97–98.

Ronald L. Graham and Vojtěch Rödl (1987). "Numbers in Ramsey theory", in: *Surveys in combinatorics 1987 (New Cross, 1987)*. Vol. 123. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, pp. 111–153.

Ronald L. Graham, Bruce L. Rothschild, and Joel H. Spencer (1990). *Ramsey theory.* 2nd ed. Wiley-Interscience Series in Discrete Mathematics and Optimization. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, pp. xii+196.

Ben Green and Terence Tao (2008). "The primes contain arbitrarily long arithmetic progressions", *Ann. of Math. (2)* **167** (2), pp. 481–547.

Parth Gupta, Ndiame Ndiaye, Sergey Norin, and Louis Wei (2024). *Optimizing the CGMS upper bound on Ramsey numbers.* arXiv: 2407.19026 [math.CO].

Zander Kelley and Raghu Meka (2023). "Strong bounds for 3-progressions", in: *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Soc., Los Alamitos, CA, pp. 933–973.

Choongbum Lee (2017). "Ramsey numbers of degenerate graphs", *Ann. of Math. (2)* **185** (3), pp. 791–829.

Xin Li (2023). "Two source extractors for asymptotically optimal entropy, and (many) more", in: *2023 IEEE 64th Annual Symposium on Foundations of Computer Science— FOCS 2023*. IEEE Computer Soc., Los Alamitos, CA, pp. 1271–1281.

Sam Mattheus and Jacques Verstraete (2024). "The asymptotics of $r(4,t)$", *Ann. of Math. (2)* **199** (2), pp. 919–941.

Sarah Peluse (2022). "Recent progress on bounds for sets with no three terms in arithmetic progression", in: Séminaire Bourbaki, Volume 2021/2022, Astérisque, no. 438 (2022), Exposé no. 1196, 547–581.

Frank P. Ramsey (1929). "On a problem of formal logic", *Proc. London Math. Soc. (2)* **30** (4), pp. 264–286.

Alexander A. Razborov (1985). "Lower bounds on the monotone complexity of some Boolean functions", *Dokl. Akad. Nauk SSSR* **281** (4), pp. 798–801.

Christian Reiher and Vojtěch Rödl (2023). *The girth Ramsey theorem.* arXiv: 2308. 15589 [math.CO].

Klaus F. Roth (1953). "On certain sets of integers", *J. London Math. Soc.* **28**, pp. 104–109.

Ashwin Sah (2023). "Diagonal Ramsey via effective quasirandomness", *Duke Math. J.* **172** (3), pp. 545–567.

Issai Schur (1917). "Über die Kongruenz $x^m + y^m \equiv z^m \pmod p$)", *Jahresber. Dtsch. Math.-Ver.* **25**, pp. 114–117.

Joel Spencer (1975). "Ramsey's theorem—a new lower bound", *J. Combin. Theory Ser. A* **18**, pp. 108–115.

——— (1994). *Ten lectures on the probabilistic method.* 2nd ed. Vol. 64. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, pp. vi+88.

Endre Szemerédi (1978). "Regular partitions of graphs", in: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976).* Vol. 260. Colloq. Internat. CNRS. CNRS, Paris, pp. 399–401.

Andrew Thomason (1982). "On finite Ramsey numbers", *European J. Combin.* **3** (3), pp. 263–273.

——— (1988). "An upper bound for some Ramsey numbers", *J. Graph Theory* **12** (4), pp. 509–517.

Johannes G. van der Corput (1939). "Über Summen von Primzahlen und Primzahlquadraten", *Math. Ann.* **116** (1), pp. 1–50.

Yuval Wigderson (2024). "Ramsey theory—lecture notes". URL: https://n.ethz.ch/~ywigderson/math/static/RamseyTheory2024LectureNotes.pdf.

Yuval Wigderson

Institute for Theoretical Studies,
ETH Zürich
E-mail: yuval.wigderson@eth-its.ethz.ch