Ramsey theory—lecture notes

Yuval Wigderson

Spring semester 2024
Last updated: June 18, 2024

Contents

Chapter 1

Introduction

Ramsey theory is the study of structure and of disorder. The main message of Ramsey theory, which underlies all results we'll study in this course, is that *complete disorder is impossible*—any sufficiently large system, no matter how disordered, must contain within it some highly structured component. This general, highly unintuitive, philosophy manifests itself in topics as diverse as computer science, number theory, geometry, functional analysis, and, of course, graph theory, which is the topic we will mostly be focused on.

However, as Ramsey theory has connections to so many other areas of mathematics and beyond, we will also frequently pause to see how the results we have proved connect to these other fields. This is, in fact, how we begin the course, with perhaps the first-ever Ramsey-theoretic result, published by Issai Schur [123] while Frank Ramsey was only fourteen years old.

1.1 Ramsey theory before Ramsey

Like many other people, Schur was interested in Fermat's last theorem, the statement that the equation $x^q + y^q = z^q$ has no non-trivial integer solutions x, y, z for any fixed $q \geqslant 3$, where a solution is *trivial* if $0 \in \{x, y, z\}$ and *non-trivial* otherwise.

Proving Fermat's last theorem is (very) hard, so let's start with something simpler. There are, of course, non-trivial integer solutions to the Pythagoras equation $x^2 + y^2 = z^2$. What if we change the equation slightly, to, say, $x^2 + y^2 = 3z^2$? After playing around with it for a bit, you might be tempted to conjecture that now, there are no non-trivial integer solutions.

This conjecture is indeed true, and there is a standard technique in number theory for proving such results. Namely, if there *were* some non-trivial solution $x, y, z \in \mathbb{Z}$ to the equation $x^2 + y^2 = 3z^2$, then there would also be a non-trivial[1] solution to the same equation modulo 3, namely the equation $x^2 + y^2 \equiv 0 \pmod 3$. However, we know that that $1^2 \equiv 2^2 \equiv 1 \pmod 3$, and we can conclude that there *do not* exist non-trivial solutions modulo 3.

[1] One has to be a bit careful here, as a non-trivial solution over $\mathbb{Z}$ may become trivial in $\mathbb{Z}/3$. However, it is not hard to get around this issue, as one can argue that a *minimal* non-trivial solution over $\mathbb{Z}$ cannot have all three of x, y, z divisible by 3.

A similar argument can be used to prove that many other polynomial equations have no non-trivial integer solutions, and a general phenomenon called the *Hasse principle* very roughly says that in many instances, such a technique is guaranteed to work. So it is natural to wonder whether Fermat's last theorem can also be proved in this way. This is the question that motivated Schur[2], who proved that this technique *cannot* work for Fermat's last theorem.

Theorem 1.1.1 (Schur [123]). *For any integer $q \geqslant 3$, there exists an integer $N = N(q)$ such that the following holds for any prime $p > N$. There exist non-zero $x, y, z \in \mathbb{Z}/p$ with*

$$x^q + y^q \equiv z^q \pmod p.$$

As Schur himself realized, despite proving an important and impressive result in number theory, his proof used almost no number theory! He wrote "daß [Theorem 1.1.1] sich fast unmittelbar aus einem sehr einfachen Hilfssatz ergibt, der mehr der Kombinatorik als der Zahlentheorie angehört."[3] This Hilfssatz is the following.

Theorem 1.1.2 (Schur [123]). *For any positive integer q, there exists an integer $N = N(q)$ such that the following holds. If $[\![N]\!]$ is colored in q colors, then there exist $x, y, z \in [\![N]\!]$, all receiving the same color, such that $x + y = z$.*

In this theorem, and throughout the course, we use the notation $[\![N]\!] := \{1, \ldots, N\}$, and the terminology of *coloring*. By a coloring of $[\![N]\!]$ with q colors, we just mean a partition of $[\![N]\!]$ into q sets $A_1, \ldots, A_q$, where we think of the elements of A_1 as receiving a first color, the elements of A_2 as receiving some second, distinct, color, and so on. We will also frequently use the shorthand *monochromatic* for "receiving the same color", so the conclusion of Theorem 1.1.2 could also be stated as the existence of a monochromatic solution to $x + y = z$.

As Schur wrote, the derivation of Theorem 1.1.1 from Theorem 1.1.2 is almost immediate, but as it requires a few ideas from number theory and group theory, we will defer it for the moment. Let us first see how to prove Theorem 1.1.2. Schur proved Theorem 1.1.2 directly, but the modern, Ramsey-theoretic, perspective is to reduce Theorem 1.1.2 to an even more combinatorial lemma, which we now state.

Lemma 1.1.3. *For any positive integer q, there exists an integer $N = N(q)$ such that the following holds. If the edges of the complete graph K_N are q-colored, then there exists a monochromatic triangle.*

Proof. We will actually prove something stronger, namely an explicit upper bound on $N(q)$; we will show that $N(q) = 3q!$ satisfies the desired condition. We proceed by induction on q.

[2] In fact, the same question had motivated Dickson [32] a few years earlier, and he was the first to prove Theorem 1.1.1. However, his technique used very messy casework and does not at all connect to Ramsey theory, so we won't discuss it any further.

[3] "that [Theorem 1.1.1] follows almost immediately from a very simple lemma, which belongs more to combinatorics than to number theory."

The base case $q = 1$ is immediate. We are claiming that any 1-coloring of the edges of K_N, where $N = 3 \cdot 1! = 3$, contains a monochromatic triangle. But as there is only one color, and the complete graph we are "coloring" is itself a triangle, this is certainly true.

For the inductive step, suppose the result is true for $q-1$, i.e. that any $(q-1)$-coloring of $E(K_{3(q-1)!})$ contains a monochromatic triangle. Fix a q-coloring of $E(K_N)$, where $N = 3q!$, and let v be any vertex of K_N. v is incident to $N-1$ edges, each of which receives one of q colors. Therefore, by the pigeonhole principle, there is some color, say red, which appears on at least

$$\left\lceil \frac{N-1}{q} \right\rceil = \left\lceil \frac{3q!-1}{q} \right\rceil = \left\lceil 3(q-1)! - \frac{1}{q} \right\rceil = 3(q-1)!$$

edges incident to v. Let R denote the set of endpoints of these red edges, and consider the coloring restricted to R. If there is any red edge appearing in R, then it forms a red triangle together with v, and we are done. If not, then R is a set of at least $3(q-1)!$ vertices that are colored by at most $q-1$ colors, and we can find a monochromatic triangle in R by the inductive hypothesis. In either case we are done. $\square$

With Lemma 1.1.3 in hand, the proof of Theorem 1.1.2 is almost immediate. All we need to do is to translate the number-theoretic coloring into a graph-theoretic coloring.

Proof of Theorem 1.1.2. Let $N(q) = 3q!$ be chosen so that Lemma 1.1.3 holds. We are given a q-coloring χ of $[\![N]\!]$, which we convert to a q-coloring $\hat{\chi}$ of $E(K_N)$ as follows. Identify the vertices of K_N with $[\![N]\!]$, and then color an edge ab, where $1 \leqslant a < b \leqslant N$, according to the color of $b - a \in [\![N]\!]$ in χ.

As $\hat{\chi}$ is a q-coloring of $E(K_N)$, by Lemma 1.1.3, there is a monochromatic triangle in $\hat{\chi}$. Let the vertices of this triangle be a, b, c, where $a < b < c$. Let $x = b - a, y = c - b$, and $z = c - a$, and note that these satisfy $x + y = z$. Finally, note that they all receive the same color under χ, since $\chi(x) = \hat{\chi}(ab), \chi(y) = \hat{\chi}(bc)$, and $\chi(z) = \hat{\chi}(ac)$, and we assumed that a, b, c is a monochromatic triangle under $\hat{\chi}$. $\square$

This completes the combinatorial part of Schur's work. For completeness, let's see how to derive Theorem 1.1.1 from Theorem 1.1.2. As this topic is somewhat outside the main narrative of the class, it will not be covered in lecture; throughout the notes we use a gray box, as follows, to indicate material that was skipped.

Deduction of Theorem 1.1.1 from Theorem 1.1.2

Proof of Theorem 1.1.1. Let $N = N(q)$ be as in Theorem 1.1.2, and fix a prime $p > N$. We recall the well-known fact that the set $\Gamma \coloneqq \{x^q : 0 \neq x \in \mathbb{Z}/p\}$ forms a subgroup of the multiplicative group $(\mathbb{Z}/p)^\times$, and the index of this subgroup is at most[†] q. Therefore, there are at most q cosets of Γ which partition the non-zero elements of $\mathbb{Z}/p$. By identifying the non-zero elements of $\mathbb{Z}/p$ with $[\![p-1]\!] \supseteq [\![N]\!]$, we obtain a q-coloring of $[\![N]\!]$ according to these cosets.

Now, by Theorem 1.1.2, there must exist monochromatic $a, b, c \in [\![N]\!]$ such that $a + b = c$. As these three numbers receive the same color, they must lie in some single coset $\alpha\Gamma$ of Γ, for some $\alpha \in (\mathbb{Z}/p)^\times$. By the definition of Γ, this means that we can write

$$a \equiv \alpha x^q \pmod p, \qquad b \equiv \alpha y^q \pmod p, \qquad c \equiv \alpha z^q \pmod p,$$

for some non-zero $x, y, z \in \mathbb{Z}/p$. The equation $a + b = c$ remains true when we reduce it mod p, so we conclude that

$$\alpha x^q + \alpha y^q \equiv \alpha z^q \pmod p.$$

As α is invertible in $\mathbb{Z}/p$, and as $x, y, z \neq 0$, we obtained the desired non-trivial solution $x^q + y^q \equiv z^q \pmod p$. □

[†]More precisely, the index is exactly $\gcd(q, p-1)$.

Chapter 2

Classical Ramsey numbers

2.1 Ramsey's theorem and upper bounds on Ramsey numbers

While Schur's theorem can be seen as an early example of Ramsey theory, the theory did not really get going until Frank Ramsey's pioneering work [109] in 1929. Ramsey's theorem, as it is now called, is a generalization of Lemma 1.1.3 from triangles to arbitrary cliques.

Theorem 2.1.1 (Ramsey [109])**.** *For all positive integers k, q, there exists an integer $N = N(k, q)$ such that the following holds. If the edges of the complete graph K_N are q-colored, then there exists a monochromatic K_k, that is, k vertices such that all the $\binom{k}{2}$ edges between them receive the same color.*

Given this theorem, which we will shortly prove, we can make a definition that will be central for much of the rest of the course.

Definition 2.1.2. Given positive integers k, q, the *q-color Ramsey number of K_k*, denoted $r(k; q)$, is the least N such that the conclusion of Theorem 2.1.1 is true. That is, $r(k; q)$ is the minimum integer N such that every q-coloring of $E(K_N)$ contains a monochromatic K_k.

In case $q = 2$, we usually abbreviate $r(k; 2)$ as simply $r(k)$, and usually refer to the 2-color Ramsey number as simply the *Ramsey number*.

In this language, Theorem 2.1.1 can equivalently be stated as saying that $r(k; q) < \infty$ for all k, q. In fact, for much of this course, we will be interested not just in the fact that such Ramsey numbers are finite, but in quantitative estimates on how large they are.

For now, let's focus on the case $q = 2$. Ramsey's original proof of Theorem 2.1.1 showed that $r(k) \leqslant k!$ for all k. But a few years later, a different proof was found by Erdős and Szekeres [52], in another foundational paper of the field. In order to present their proof, we need to define a slightly more general notion of Ramsey number.

Definition 2.1.3. Given positive integers k, ℓ, we denote by $r(k, \ell)$ the *off-diagonal Ramsey number*, defined to be the least N such that every 2-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 $r(k) = r(k, k)$.

Theorem 2.1.4 (Erdős–Szekeres). *For all positive integers k, ℓ, we have*

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

In particular, we have

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

Proof. We proceed by induction on $k + \ell$, with the base case[1] $k = \ell = 1$ being trivial. For the inductive step, the key claim is that the following inequality holds:

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

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

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

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

A similar argument works when the number of colors is more than 2. If we denote by $r(k_1, \ldots, k_q)$ the *off-diagonal multicolor Ramsey number* (defined in the natural way), we obtain the following generalization of Theorem 2.1.4, which you will prove on the homework.

Theorem 2.1.5. *For all positive integers q and $k_1, \ldots, k_q$, we have*

$$r(k_1, \ldots, k_q) \leqslant \binom{k_1 + \cdots + k_q - q}{k_1 - 1, \ldots, k_q - 1},$$

where the right-hand side denotes the multinomial coefficient. In particular,

$$r(k; q) < q^{qk}.$$

[1] If you don't like starting the induction with $k = \ell = 1$—what does a monochromatic K_1 mean, exactly?—you should convince yourself that the base case $k = \ell = 2$ also works.

2.2 Lower bounds on Ramsey numbers

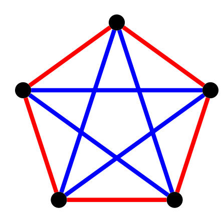
The Erdős–Szekeres bound, Theorem 2.1.4, gives us the upper bound $r(k) < 4^k$, which improves on Ramsey's earlier bound of $r(k) \leqslant k!$. To understand how good this bound is, we would like to obtain some *lower bounds* on $r(k)$.

Thinking about the definition of Ramsey numbers, we see that proving a lower bound of $r(k) > N$ boils down to exhibiting a 2-coloring of $E(K_N)$ with no monochromatic K_k. Perhaps the simplest such coloring is the *Turán coloring*, which proves the following result (and which we will meet again later in the course).

Proposition 2.2.1. *For any positive integer k, we have $r(k) > (k-1)^2$.*

Proof. Let $N = (k-1)^2$. We split the vertex set of K_N into $k-1$ parts, each of size $k-1$. We color all edges within a part red, and all edges between parts blue. The red graph is a disjoint union of $k-1$ copies of K_{k-1}, so there is certainly no red K_k. On the other hand, as there are only $k-1$ parts, the pigeonhole principle implies that any set of k vertices must include two vertices in one part; these two vertices span a red edge, and thus there is no blue K_k either. □

Is Proposition 2.2.1 tight? It's not too hard to see that the answer is no. Indeed, already for $k = 3$, Proposition 2.2.1 implies that $r(3) > 4$, and it is not hard to show that in fact $r(3) > 5$, as witnessed by the following coloring.



Nonetheless, it is not clear how to do much better than Proposition 2.2.1 in general. Indeed, I've heard that in the 1940s, Turán believed that the Erdős–Szekeres bound is way off, and that the truth is $r(k) = \Theta(k^2)$ (i.e. that Proposition 2.2.1 is best possible up to a constant factor). As it turns out, this belief was *way* off.

Theorem 2.2.2 (Erdős [51])**.** *For any $k \geqslant 2$, we have $r(k) \geqslant 2^{k/2}$.*

Together with Theorem 2.1.4, this proves that $r(k)$ really does grow as an exponential function of k, although these theorems do not tell us the precise growth rate. Theorem 2.2.2 was a major breakthrough not only—or even primarily—because of the result itself. In proving Theorem 2.2.2, Erdős introduced the so-called *probabilistic method* to combinatorics. This method would quickly become one of the most important tools in combinatorics, and will recur frequently throughout this course.

Proof of Theorem 2.2.2. Fix k, and let[2] $N = 2^{k/2}$. The claimed bound is trivial for $k = 2$, so let's assume $k \geqslant 3$. Consider a *random* 2-coloring of $E(K_N)$. Namely, for each edge of K_N, we assign it color red or blue with probability $\frac{1}{2}$, making these choices independently over all edges. We begin by estimating the probability that this coloring contains a monochromatic K_k.

For any fixed set of k vertices, the probability that it forms a monochromatic K_k is precisely $2^{1-\binom{k}{2}}$. This is because we have $\binom{k}{2}$ coin tosses, which we need to all agree, and we have two options for the shared outcome (hence the extra $+1$ in the exponent). Moreover, there are exactly $\binom{N}{k}$ possible k-sets we need to consider. Therefore,

$$\Pr(\text{there is a monochromatic } K_k) \leqslant \binom{N}{k} 2^{1-\binom{k}{2}},$$

where we have applied the *union bound* $\binom{N}{k}$ times; this is the bound that says that the probability that A *or* B happens is at most the sum of the probability that A happens and the probability that B happens.

Note that $\binom{N}{k} < N^k/k!$ and that $k! > 2^{1+k/2}$ for all $k \geqslant 3$. Therefore, we have

$$\binom{N}{k} 2^{1-\binom{k}{2}} < \frac{N^k}{k!} \cdot 2^{1-\frac{k^2-k}{2}} < \frac{N^k}{2^{1+\frac{k}{2}}} \cdot 2^{1+\frac{k}{2}-\frac{k^2}{2}} = \left(N \cdot 2^{-\frac{k}{2}}\right)^k = 1, \tag{2.2}$$

where the final equality is our choice of N.

Putting this all together, we find that in this random coloring, the probability that there is a monochromatic K_k is *strictly less than one.* Therefore, there must exist *some* coloring of $E(K_N)$ with no monochromatic K_k, as if such a coloring did not exist, the probability above would be exactly one. This completes the proof. □

It's worth stressing the miraculous magic trick that takes place in the proof of Theorem 2.2.2. Unlike in Proposition 2.2.1, Erdős does not give any sort of explicit description of a coloring on $2^{k/2}$ vertices with no monochromatic K_k. Instead, he argues that such a coloring must exist for probabilistic reasons, but this argument gives absolutely no indication of what such a coloring looks like. In fact, the following remains a major open problem.

Open problem 2.2.3 (Erdős). *For some $\varepsilon > 0$ and all sufficiently large k, explicitly construct a 2-coloring on $(1+\varepsilon)^k$ vertices with no monochromatic K_k.*

There was a great deal of partial progress over the years, much of it exploiting a deep and surprising connection to the topic of *randomness extraction* in theoretical computer science. Just last year, there was a major breakthrough on this problem.

[2]The astute reader will notice that $2^{k/2}$ is not an integer unless k is even. Thus, we should really write here $N = \lceil 2^{k/2} \rceil$. However, once the computations we do become more complicated, keeping track of such floor and ceiling signs becomes not just annoying, but actively confusing. Therefore, for the rest of the course, we'll omit floor and ceiling signs unless they are actually crucial, and it will be understood that any quantity that should be an integer but doesn't look like one should be rounded up or down to an integer.

Theorem 2.2.4 (Li [90]). *For some absolute constant $\varepsilon > 0$ and all sufficiently large k, there is an explicit 2-coloring on 2^{k^ε} vertices with no monochromatic K_k.*

By using a random q-coloring, one can adapt the proof of Theorem 2.2.2 and prove that for any $k, q \geqslant 3$, we have

$$r(k; q) > q^{k/2}.$$

Together with Theorem 2.1.5, this shows that for any fixed $r(k; q)$ grows exponentially as a function of k for any fixed q. However, for fixed k, the upper and lower bounds are rather far apart—the lower bound is merely polynomial in q, whereas the upper bound is super-exponential in q. For several decades this was the state of the art, until Abbott[3] [1] noticed a simple trick that does much better.

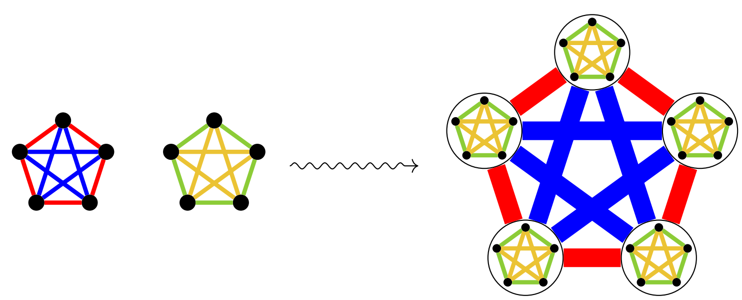
Proposition 2.2.5 (Abbott [1]). *For all positive integers k, q_1, q_2, we have*

$$r(k; q_1 + q_2) - 1 \geqslant (r(k; q_1) - 1)(r(k; q_2) - 1). \tag{2.3}$$

As a consequence, we have

$$r(k; q) > 2^{\frac{k}{2}\lfloor \frac{q}{2} \rfloor}.$$

Proof. Let $N_1 = r(k; q_1) - 1$ and $N_2 = r(k; q_2) - 1$. By assumption, we have colorings $\chi_i : V(K_{N_i}) \to [\![q_i]\!]$, for $i = 1, 2$, both of which avoid monochromatic K_k. Let $N = N_1 N_2$, and identify the vertex set of K_N with $V(K_{N_1}) \times V(K_{N_2})$. We can now define a coloring $\chi : E(K_N) \to [\![q_1 + q_2]\!]$ as follows. It is easiest to understand with the following picture, which shows how to convert two 2-colorings of $E(K_5)$ into a 4-coloring of $E(K_{25})$, maintaining the property of having no monochromatic triangle.



Formally, given a pair of vertices $(a_1, b_1), (a_2, b_2) \in V(K_{N_1}) \times V(K_{N_2}) \cong V(K_N)$, we define

$$\chi((a_1, b_1), (a_2, b_2)) = \begin{cases} \chi_1(a_1, a_2) & \text{if } a_1 \neq a_2, \\ q_1 + \chi_2(b_1, b_2) & \text{otherwise.} \end{cases}$$

This is a $(q_1 + q_2)$-coloring of $E(K_N)$, and one can readily verify that there is no monochromatic K_k, as such a monochromatic clique could be used to obtain a monochromatic K_k in either χ_1 or χ_2. Thus proves the claimed inequality (2.3).

To use it, we recall that we proved in Theorem 2.2.2 that $r(k; 2) \geqslant 2^{k/2} + 1$. Applying (2.3) $\lfloor q/2 \rfloor$ times, we conclude that $r(k; q) > (2^{k/2})^{\lfloor q/2 \rfloor}$, as claimed. $\square$

[3] An earlier version of these notes, as well as many papers on the topic, attribute this result to Lefmann [87], and note that the trick goes back at least to work of Chung [16]. But the paper of Abbott is even earlier.

2.3 The past and the future

Let us now zoom out a bit and discuss both some history, and a preview of what is to come in (part of) the rest of the course. Until five years ago, the results stated above were essentially the state of the art. In the case of two colors, we knew

$$2^{\frac{k}{2}} < r(k) < 4^k,$$

and more generally for q colors we had (say for simplicity that q is even)

$$2^{\frac{qk}{4}} < r(k;q) < q^{qk}.$$

There were a number of important papers that obtained slight improvements on some of these bounds [19, 119, 129], but no one knew how to improve any of the exponential constants appearing above. But in recent years there have been a number of important breakthroughs on the problems discussed above.

The first concerns the lower bound on $r(k;q)$ when $q \geqslant 3$ is fixed. Here, there was a major breakthrough of Conlon and Ferber in 2020 [24], followed shortly thereafter by improvements of myself [142] and Sawin [122]. The current state of the art, due to Sawin, is the following result.

Theorem 2.3.1 (Sawin [122])**.** *For fixed $q \geqslant 3$, we have*

$$r(k;q) > \left(2^{0.383796q-0.267592}\right)^{k-o(k)},$$

where the $o(k)$ term grows asymptotically slower than k as $k \to \infty$.

This is better than what is given by Proposition 2.2.5, because $0.384 > \frac{1}{4}$. The proof is ingeneous but quite simple, and we will see it later in the course. We remark that while these recent breakthroughs have improved the lower bound given in Proposition 2.2.5, they have so far been unable to answer the main question about multicolor Ramsey numbers, which Erdős offered \$100 for.

Open problem 2.3.2 (Erdős, \$100)**.** *For fixed $k \geqslant 3$, does $r(k;q)$ grow exponentially or super-exponentially as a function of q?*

The next breakthrough, chronologically, came in March 2023, when Campos, Griffiths, Morris, and Sahasrabudhe [13] obtained the first improvement to the exponential constant in Theorem 2.1.4.

Theorem 2.3.3 (Campos–Griffiths–Morris–Sahasrabudhe [13])**.** *$r(k) < 3.9999^k$ for all sufficiently large k.*

This might seem like a small improvement, but this was a really major breakthrough, since this problem had been intractably stuck for almost 90 years. The proof of Theorem 2.3.3

is completely elementary, but rather involved; we will hopefully see a sketch of the argument later in the course, time permitting.

The final breakthrough that I want to talk about came out just three months later, in June 2023, and was a result of Mattheus and Verstraëte [93] about off-diagonal Ramsey numbers. Before stating their result, let's back up and learn a bit about off-diagonal Ramsey numbers, which we have not yet seriously discussed.

Generally speaking, when we talk about off-diagonal Ramsey numbers, we are interested in the function $r(s, k)$ (as in Definition 2.1.3), where we think of s as fixed and $k \to \infty$. If we specialize Theorem 2.1.4 to this setting, we find that for any fixed $s \geqslant 2$, we have

$$r(s, k) \leqslant \binom{k + (s-2)}{s-1} = O_s(k^{s-1}).$$

Here, and throughout the course, we use the big-O notation $f = O(g)$ to mean that $f(x) \leqslant C \cdot g(x)$ for an absolute constant $C > 0$. In case we use a subscript, as the subscript s above, this means that the constant C may depend on the parameter s, i.e. that this bound should be thought of for fixed s. It is easy to see that $r(2, k) = k$ for all k, hence this bound is tight for $s = 2$. For all larger s, a polylogarithmic improvement to the upper bound was obtained by Ajtai, Komlós, and Szemerédi [2], who proved that for fixed $s \geqslant 3$, we have

$$r(s, k) = O_s\left(\frac{k^{s-1}}{(\log k)^{s-2}}\right).$$

We will see a proof of this result later in the course. In particular, in the case $s = 3$, their result says that

$$r(3, k) = O\left(\frac{k^2}{\log k}\right).$$

Even before the Ajtai–Komlós–Szemerédi theorem was proved, Erdős [37] used a very sophisticated and intricate probabilistic argument to obtain a nearly matching lower bound,

$$r(3, k) = \Omega\left(\frac{k^2}{(\log k)^2}\right),$$

where the big-Ω notation $f = \Omega(g)$ is equivalent to $g = O(f)$. Erdős's result was re-proved by Spencer [130] using a different (and simpler) probabilistic technique, but the logarithmic gap remained for a long time until Kim [77] finally managed to prove that the upper bound is correct, that is

$$r(3, k) = \Theta\left(\frac{k^2}{\log k}\right),$$

where the big-Θ notation $f = \Theta(g)$ means that $f = O(g)$ and $f = \Omega(g)$. More recent improvements to the lower and upper bounds [10, 53, 125] have been able to almost completely determine the asymptotics of $r(3, k)$; we now know that

$$\left(\frac{1}{4} - o(1)\right)\frac{k^2}{\ln k} \leqslant r(3, k) \leqslant (1 + o(1))\frac{k^2}{\ln k},$$

where ln denotes the natural logarithm, and where the little-o notation $f = o(g)$ means that $\lim_{x\to\infty} f(x)/g(x) = 0$.

Despite this string of successes, very little remains known about the asymptotics of $r(s,k)$ for fixed $s \geqslant 4$. The best known lower bound, again due to Spencer [130] (with polylogarithmic improvements due to Bohman–Keevash [9]) is of the form $r(s,k) \geqslant k^{\frac{1}{2}(s+1)+o(1)}$, compared to the upper bound of $r(s,k) \leqslant k^{s-1-o(1)}$. In particular, for $s = 4$, there is a gap of $1/2$ in the exponent. Or at least, there *was*, until the Mattheus–Verstraëte breakthrough [93].

Theorem 2.3.4 (Mattheus–Verstraëte [93])**.** *We have*

$$r(4,k) = \Omega\left(\frac{k^3}{(\log k)^4}\right).$$

This matches the Ajtai–Komlós–Szemerédi upper bound up to a factor of $\Theta((\log k)^2)$. Their proof builds on a long line of recent work [3, 20, 95], and happens to be closely related to the techniques used to prove Theorem 2.3.1 (the improved lower bound for $r(k;q)$). As such, we will see the proof of Theorem 2.3.4 later in the course.

Chapter 3

Lower bounds on multicolor Ramsey numbers

Recall that Abbott [1] proved that $r(k;q) > 2^{\frac{k}{2}\lfloor \frac{q}{2} \rfloor}$; for even q we can write this as $r(k;q) > (2^{q/4})^k$. We will now see how to improve this bound for all $q \geqslant 3$. In so doing, we will also lay the groundwork for proving lower bounds on the off-diagonal Ramsey numbers $r(3,k)$ and $r(4,k)$. The ideas in this section go back at least to work of Alon–Rödl [3], and were crystallized in a series of works [24, 73, 95, 122, 142].

3.1 Random sampling and random homomorphisms

Let's suppose we wish to prove a lower bound on the two-color Ramsey number $r(s,k)$. If we can find a graph G that has no clique of order s and no independent set of order k, then we've found such a lower bound: $r(s,k)$ is greater than the number of vertices of G, since we can color the edges of G red and the non-edges blue. But since finding such graphs is hard, it would be nice to be able to lower-bound $r(s,k)$ by finding a graph G with some weaker property.

It turns out that this is possible. Suppose we now have a graph G with no K_s, but let's not assume that it has no independent sets of order k. Instead, let's suppose that G has "few" independent sets of order k. Concretely, assume that G has at most M^k independent sets of order k, for some parameter M (note that it is natural to parametrize things in this way, since there are exponentially many k-sets of vertices in G). It turns out that as long as M is not too big, we can use this G to get a good lower bound on $r(s,k)$, by *random sampling.*

Lemma 3.1.1 (Random sampling)**.** *Let G be a K_s-free graph on N vertices, and suppose that G has at most M^k independent sets of order k. Then*

$$r(s,k) \geqslant \frac{N}{4M}.$$

Proof. We will randomly sample a subgraph H of G, by keeping each vertex of G independently with probability p, to be chosen later. Since G is K_s-free, its subgraph H is

K_s-free as well. Additionally, each independent set of order k in G will survive in H with probability p^k. So the expected number of independent sets of order k in H is at most $p^k M^k = (pM)^k$. By choosing $p = 1/(2M)$, this number is less than $1/2$, so the probability that H has no independent set of order k is at least $1/2$. Additionally, with high probability, H has at least $pN/2$ vertices, by standard probabilistic tail bounds[1]. So we find that with positive probability, H is a graph on at least $N/(4M)$ vertices with no K_s or $\overline{K_k}$, proving that $r(s,k) \geqslant N/(4M)$, as claimed. $\square$

In order to extend these ideas further, it will be convenient to take a different perspective on Lemma 3.1.1. Specifically, rather than keeping each vertex of G with probability p, we will pick a random function from a set of pN vertices to $V(G)$, and "pull back" the graph structure. Of course, if $p \ll 1$, then this random function will have no collisions with high probability, and so we will exactly get the random induced subgraph we got before, except that we'll have exactly pN vertices (rather than a binomial distribution on the number of vertices), but this difference is immaterial. The reason for taking this change of perspective is that it is much more amenable to using more than two colors: we can just pick more random functions and overlay them, as we'll soon see.

Concretely, suppose that G is a K_s-free graph on N vertices with at most M^k independent sets of order at most[2] k. Let $n = pN$ for some parameter p, and pick a uniformly random function $f : [\![n]\!] \to V(G)$. Define a graph H on vertex set $[\![n]\!]$ by setting $\{u, v\} \in E(H)$ if $\{f(u), f(v)\} \in E(G)$; note that in particular we only connect u and v if $f(u) \neq f(v)$, which implies that H is also K_s-free. Then for any given set $K \subset [\![n]\!]$ of order $|K| = k$, and any fixed $U \subseteq V(G)$ of order $|U| \leqslant k$, the probability that $f(K) \subseteq U$ is at most $(k/N)^k$. Thus, the probability that K is independent in H is at most $(kM/N)^k$, as there are at most M^k choices for such a U that is independent in G. As there are $\binom{n}{k}$ choices for this K, we see by the union bound that

$$\Pr(H \text{ has an independent set of order } k) \leqslant \binom{n}{k}\left(\frac{kM}{N}\right)^k \leqslant \left(\frac{epN}{k}\frac{kM}{N}\right)^k = (epM)^k,$$

and we can recover the result of Lemma 3.1.1—up to the constant factor—by setting $p = 1/(2eM)$.

However, as indicated above, the power of this perspective is that it easily extends to more colors. Indeed, suppose that we instead pick independent uniformly random functions $f_1, \ldots, f_r : [\![n]\!] \to V(G)$. We color the edges of K_n in $r+1$ colors, as follows. If there is some $i \in [r]$ such that $\{f_i(u), f_i(v)\} \in E(G)$, then we color $\{u, v\}$ by the minimum such i. If

[1]As I am not assuming any probabilistic background in this course, I won't get into exactly what this means, but if you're curious you should look up Chebyshev's inequality or the Chernoff bound, and verify that either of them will suffice to prove this statement. Strictly speaking, to make this argument work, we'd have to assume that $N > 10M$ (or some other similar bound). But given that the lemma statement is uninteresting if M and N have the same order, let's not worry about this technicality.

[2]Note that we've slightly strengthened this assumption, bounding the number of independent sets of order *at most* k. As it turns out, this is usually OK: many techniques that bound the number of independent sets of order exactly k will also work here.

not, we color $\{u, v\}$ by color $r+1$. Then each of the first r colors is K_s-free, by the above. Additionally, the probability that some fixed k-set K is monochromatic in the last color is at most $(kM/N)^{rk}$, since we have a probability $(kM/N)^k$ for each function f_i, and these probabilities are independent. Therefore, by the union bound, we find that the probability that the last color has a clique of order k is at most

$$\binom{n}{k}\left(\frac{kM}{N}\right)^{rk} \leqslant \left(\frac{epN}{k}\right)^k \left(\left(\frac{kM}{N}\right)^r\right)^k = \left(\frac{epk^{r-1}M^r}{N^{r-1}}\right)^k. \tag{3.1}$$

We conclude the following generalization of Lemma 3.1.1.

Lemma 3.1.2 (Random homomorphisms)**.** *Let G be a K_s-free graph on N vertices, and suppose that G has at most M^k independent sets of order at most k. Then*

$$r(\underbrace{s,\dots,s}_{r \text{ times}}, k) \geqslant \frac{N^r}{2ek^{r-1}M^r}.$$

Proof. We set $p = N^{r-1}/(2ek^{r-1}M^r)$, so that the quantity in (3.1) is less than 1. Then we see that the coloring described above has no K_s in the first r colors, and no K_k in the final color, and has $n = pN$ vertices. □

Of course, even this isn't the most general form of this lemma that we could prove, since there's no real reason to have $f_1, \dots, f_r$ all have the same codomain. Indeed, in [73], this idea was used to obtain lower bounds on many off-diagonal multicolor Ramsey numbers.

The crucial thing to observe about Lemma 3.1.2 is that p is not a probability, and in particular, it does not need to be less than 1! If $p > 1$, then $n = pN$ will be larger than N, and the functions $f_1, \dots, f_r$ will no longer be making random subgraphs of G. Instead, they will be forming random *blowups* of G, and thus the coloring we use in Lemma 3.1.2 is obtained by randomly overlaying r random blowups of G, and then coloring all uncolored edges with the final color. This idea of overlaying random blowups to obtain lower bounds on multicolor Ramsey numbers goes back to Alon and Rödl [3], though they didn't use the perspective of random homomorphisms. The observation that the Alon–Rödl approach and the Mubayi–Verstraëte approach are both instances of the same general technique is due to Xiaoyu He, and our paper [73] uses this observation to combine the Alon–Rödl and Mubayi–Verstraëte approaches and obtain unified bounds on multicolor Ramsey numbers. In my opinion, the fact that random induced subgraphs and random blowups are "the same thing" is a very powerful observation, and it's the main message I'd like to get across in this section.

3.2 The Conlon–Ferber argument

In Lemma 3.1.2, we gave all remaining edges the same color, and then used a simple union bound to estimate the probability of a monochromatic K_k. The Conlon–Ferber idea is to actually use two colors for these remaining edges, choosing randomly for each edge. Since we know that random colorings generally have small monochromatic cliques, it stands to reason that doing this will improve the lower bound on the Ramsey number. Of course, doing this is costly, in the sense that we have to add a new color, so we are obtaining a strengthened bound on a different Ramsey number. The precise statement, implicit in [24, 142] and explicit in [122], is as follows.

Lemma 3.2.1. *Let G be a K_s-free graph on N vertices, and suppose that G has at most M^k independent sets of order at most k. Then*

$$r(\underbrace{s,\dots,s}_{r\ \textit{times}},k,k) \geqslant \frac{2^{k/2}N^r}{4k^rM^r}.$$

Proof. As indicated above, we pick a parameter p, set $n = pN$, and choose r random functions $f_1,\dots,f_r : [\![n]\!] \to V(G)$. We color $E(K_n)$ by assigning the first r colors as before, with $\{u,v\}$ getting color i only if $\{f_i(u), f_i(v)\} \in E(G)$. For the uncolored edges, we assign one of the colors $r+1, r+2$ uniformly at random, independently for each uncolored edge. Then as above, we know that the first r colors are K_s-free. For the final two colors, let's estimate the probability that a k-set $K \subset [\![n]\!]$ is monochromatic. For K to be monochromatic, it must first not contain any edges of the first r colors, which we know happens with probability at most $(kM/N)^{rk}$. Then, there is a probability $2^{1-\binom{k}{2}}$ that all the pairs of K get assigned the same color among $\{r+1, r+2\}$. Putting this all together with the union bound, we see that the probability that K_n has a monochromatic K_k in one of the last two colors is at most

$$\binom{n}{k}2^{1-\binom{k}{2}}\left(\frac{kM}{N}\right)^{rk} \leqslant \left(pN \cdot 2^{1-\frac{k}{2}} \cdot \frac{k^rM^r}{N^r}\right)^k = \left(p\frac{2k^rM^r}{2^{k/2}N^{r-1}}\right)^k.$$

So if we take $p = 2^{k/2}N^{r-1}/(4k^rM^r)$, this probability will be less than 1, and we'll obtain a coloring with no K_s in the first r colors and no K_k in the final two colors. This gives that

$$r(\underbrace{s,\dots,s}_{r\text{ times}},k,k) \geqslant n = pN = \frac{2^{k/2}N^r}{4k^rM^r}. \qquad \square$$

3.3 Actually getting a lower bound on $r(k;q)$

So far, all of the results proved above are of the form "if a graph G with certain properties exists, then we obtain a lower bound on some Ramsey number". But we haven't yet proved that any such graph G exists!

In many settings, such as the lower bounds on $r(3,k)$ and $r(4,k)$ that we will discuss shortly, finding such graphs is quite difficult, and is basically the crux of the argument (see also [3, 95]). In their work improving the lower bound on $r(k;q)$, Conlon and Ferber [24] used an ingenious linear-algebraic construction of a graph with such properties, and you will have the opportunity to explore this graph in the homework. However, as observed by Sawin [122], it is more efficient to use a *random* graph.

Lemma 3.3.1. *For every $k \geqslant 10$, there exists a K_k-free graph G on $N = 2^{k/2}$ vertices with at most M^k independent sets of order at most k, where $M = 2 \cdot 2^{k/8}$.*

Proof. We consider a uniformly random graph G on N vertices, i.e. where each pair is included as an edge of G with probability $\frac{1}{2}$, independently over all choices. By the same computation as in Theorem 2.2.2, we see that G is K_k-free with probability at least $\frac{2}{3}$. Any set of order m is an independent set in G with probability $2^{-\binom{m}{2}}$, hence the expected number of independent sets of order at most k in G is

$$\sum_{m=1}^{k} \binom{N}{m} 2^{-\binom{m}{2}} \leqslant N + \binom{N}{2} + \sum_{m=3}^{k} \binom{N}{m} 2^{-\binom{m}{2}} \leqslant N^2 + \sum_{m=3}^{k} \left(N \cdot 2^{-\frac{m}{2}}\right)^m,$$

where the final step is the same computation as in equation (2.2). Recalling that we chose $N = 2^{k/2}$, we can write

$$\left(N \cdot 2^{-\frac{m}{2}}\right)^m = \left(2^{\frac{k-m}{2}}\right)^m = \left(2^{\frac{k-m}{2} \cdot \frac{m}{k}}\right)^k.$$

The function $(k-m)m/(2k)$ is a quadratic function of m, and it is easy to see that it is maximized at $m = k/2$, where it takes on the value $k/8$. Therefore, the expected number of independent set of order at most k in G is upper-bounded by

$$N^2 + (k-2) \cdot \left(2^{\frac{k}{8}}\right)^k \leqslant \frac{1}{3}\left(2 \cdot 2^{\frac{k}{8}}\right)^k = \frac{1}{3} M^k,$$

where the inequality holds by our assumption that $k \geqslant 10$.

Now, an application of Markov's inequality shows that with probability at least $\frac{2}{3}$, G has at most M^k independent sets of order at most k. As we also said that G is K_k-free with probability at least $\frac{2}{3}$, we conclude that there exists a graph G with the claimed properties. □

Plugging this result into Lemma 3.2.1 (with $s = k$ and $r = q-2$), we obtain

$$r(k;q) \geqslant \frac{2^{k/2} N^{q-2}}{4k^{q-2} M^{q-2}} = \frac{(2^{(q-1)/2})^k}{4(2k)^{q-2}(2^{(q-2)/8})^k} = \left(2^{\frac{q-1}{2} - \frac{q-2}{8}}\right)^{k-o(k)} = \left(2^{\frac{3}{8}q - \frac{1}{4}}\right)^{k-o(k)}, \tag{3.2}$$

where in the second equality we use the fact that $4(2k)^{q-2}$ is a polynomial in k, and thus is of the form $2^{o(k)}$ as $k \to \infty$ (with q fixed). Note that this bound is already better than that given by Proposition 2.2.5 for any $q \geqslant 3$, whereas (3.2) matches Proposition 2.2.5 (and Theorem 2.2.2) for $q = 2$. This should not be surprising, since for $q = 2$ we use $r = q - 2 = 0$ random homomorphisms, and thus this construction is simply the same as in Theorem 2.2.2!

The bound (3.2) was proved in [142], by using the linear-algebraic graph of Conlon–Ferber rather than the random graph G constructed in Lemma 3.3.1 (both constructions end up having the same value of N/M, and thus yield the same bound). One of Sawin's main observations in [122] is that by running the same argument with a random graph of edge density p slightly less than $1/2$, one can get the better exponent in Theorem 2.3.1.

Lemma 3.3.2 (Sawin [122])**.** *For any $k \geqslant 4$ and $p \in (0,1)$, there exists a K_k-free graph G on $N = p^{-k/2}$ vertices with at most M^k independent sets of order at most k, where*

$$M = N \cdot 2^{k \cdot \frac{(4\log(1-p)-\log(p))\log(p)}{8\log(1-p)} - o(k)},$$

and the logarithms are to base 2*.*

The proof of this lemma is the same as that of Lemma 3.3.1, except that now when defining the random graph G, we include every pair as an edge with probability p. As our goal is to pick M as small as possible, to obtain as strong a lower bound from Lemma 3.2.1 as possible, we wish to minimize the exponent as a function of p; one can check that this minimum is attained at $p \approx 0.455$, and plugging this into Lemma 3.2.1 yields the bound in Theorem 2.3.1.

Chapter 4

Off-diagonal Ramsey numbers

Let us now turn our attention to the off-diagonal Ramsey number $r(3,k)$. Already in Lemma 3.1.1, we saw the basic tool that we will use to obtain lower bounds on this function (the idea of applying Lemma 3.1.1 to this problem is due to Mubayi–Verstraëte [95]). However, before we get there, let us discuss upper bounds.

4.1 Upper bounds on off-diagonal Ramsey numbers

Recall that as a consequence of the general Erdős–Szekeres bound, Theorem 2.1.4, we have

$$r(3,k) \leqslant \binom{k+1}{2} \leqslant k^2. \tag{4.1}$$

In this section, we will prove a better upper bound, of the form $r(3,k) = O(k^2/\log k)$, originally due to Ajtai–Komlós–Szemerédi [2] (improving on earlier work of Graver–Yackel [69]), although we will follow a somewhat more streamlined proof due to Shearer [125]. Before we do that, let's spend a moment thinking about the bound $r(3,k) \leqslant k^2$. Setting $n = k^2$, this bound can equivalently phrased as follows: any n-vertex triangle-free graph G contains an independent set of order $\sqrt{n}$. In this language, this is rather easy to prove, as follows. If G has a vertex v of degree at least $\sqrt{n}$, then the triangle-free condition implies that the neighborhood of v is an independent set of order at least $\sqrt{n}$. On the other hand, if all vertices of G have degree strictly less than $\sqrt{n}$, we can greedily build up an independent A set as follows. We pick a vertex v_1, place v_1 into A, and then delete v_1 and all its neighbors from G. We then pick another vertex v_2, place it into A, and delete it and all its neighbors from G. We continue this process as long as we can. Note that no matter what, we definitely create an independent set at the end of this process, since the step where we delete all neighbors of v_i guarantees that no pair of vertices in A are adjacent. Moreover, as every vertex in G has degree less than $\sqrt{n}$, we delete at most $\sqrt{n}$ vertices at each step of the process, and hence we can continue the process for at least $n/\sqrt{n} = \sqrt{n}$ steps. Thus, we end by producing an independent set A with $|A| \geqslant \sqrt{n}$.

We can thus split the proof of (4.1) into two lemmas. We denote by $\alpha(G)$ the *independence number* of G, that is, the size of the largest independent set in G.

Lemma 4.1.1. *If a triangle-free graph G has average degree d, then $\alpha(G) \geqslant d$.*

Lemma 4.1.2. *If an n-vertex graph G has average degree d, then $\alpha(G) \geqslant n/(1+d)$.*

Note that Lemma 4.1.1 follows directly from the argument above, since if the average degree is d, then certainly there is *some* vertex with degree at least d. In contrast, Lemma 4.1.2 actually *doesn't* follow from the above; the argument presented above only really works if G has *maximum* degree d. Nonetheless, Lemma 4.1.2 is true, and is one of the many equivalent formulations of Turán's theorem; you'll prove it on the homework. Note that in this lemma, and in the argument above, we didn't actually use the assumption that G is triangle-free—that assumption only came into Lemma 4.1.1, whereas the inductive procedure for building A works in any graph with bounded maximum degree.

The basic idea underlying the Ajtai–Komlós–Szemerédi theorem is that Lemma 4.1.2, while tight in general, is far from tight for triangle-free graphs. Basically, one can use the triangle-freeness to pick *intelligent* choices for v_i, which ensure that the process can continue for somewhat longer than the naïve analysis above suggests. A very slick formulation of this idea, due to Shearer [125], is the content of the following lemma.

Lemma 4.1.3 (Shearer [125])**.** *Define*

$$f(d) := \frac{d \ln d - d + 1}{(d-1)^2},$$

extended continuously to $f(0) := 1, f(1) := \frac{1}{2}$. If G is an n-vertex triangle-free graph with average degree d, then

$$\alpha(G) \geqslant n \cdot f(d) = (1 - o(1)) \frac{n \ln d}{d},$$

where the $o(1)$ tends to 0 as $d \to \infty$.

Proof. Note that $f(d) = (1-o(1))\frac{\ln d}{d}$, so it suffices to prove the first inequality. We prove the statement by induction on n, where for every fixed n we prove it simultaneously for all d. For the base case, note that the result is trivial if $n = 1$, as the only 1-vertex graph has no edges and $d = 0$, and independence number $1 = 1 \cdot f(0)$. We now proceed with the inductive step, and assume that the result has been proved for all smaller values of n.

For a vertex $v \in V(G)$, let us denote by $\deg(v)$ its degree, and by $r(v)$ the average degree of its neighbors, namely

$$r(v) := \frac{1}{\deg(v)} \sum_{u \sim v} \deg(u),$$

where $u \sim v$ denotes that u is adjacent to v. The reason we care about these quantities is that we plan to pick a carefully-chosen vertex v, and then define $\widehat{G}$ to be the induced subgraph obtained by deleting v and all its neighbors. When we do this, we have that

$$v(\widehat{G}) = n - (\deg(v) + 1),$$

since we deleted $\deg(v)+1$ vertices, and that

$$e(\widehat{G}) = e(G) - \sum_{u\sim v} \deg(u) = e(G) - \deg(v)r(v).$$

This is the only step in which we use that G is triangle-free, to ensure that indeed $\deg(v)r(v)$ edges are deleted—if there were triangles in the graph, then neighbors of v might be adjacent, and then this might be an overcount of the number of deleted edges. Recalling that the average degree of G is d, so that $e(G) = nd/2$, we compute that the average degree of $\widehat{G}$ is

$$\widehat{d} = 2\frac{e(\widehat{G})}{v(\widehat{G})} = 2\frac{\frac{1}{2}nd - \deg(v)r(v)}{n-\deg(v)-1} = \frac{nd - 2\deg(v)r(v)}{n-\deg(v)-1}.$$

Let us note for future reference that

$$\begin{aligned}(n-\deg(v)-1)(\widehat{d}-d) &= (nd - 2\deg(v)r(v)) - (nd - d\deg(v) - d) \\ &= d\deg(v) + d - 2\deg(v)r(v). \end{aligned} \tag{4.2}$$

Note that we may add v to any independent set in $\widehat{G}$ to obtain an independent set in G, and therefore the inductive hypothesis implies that

$$\alpha(G) \geqslant 1 + \alpha(\widehat{G}) \geqslant 1 + (n-\deg(v)-1)f(\widehat{d}).$$

One can check that $f''(x) \geqslant 0$ for all $x > 0$, which implies that

$$f(\widehat{d}) \geqslant f(d) + (\widehat{d}-d)f'(d).$$

Therefore, continuing the computation above, we have that

$$\begin{aligned}\alpha(G) &\geqslant 1 + (n-\deg(v)-1)f(\widehat{d}) \\ &\geqslant 1 + (n-\deg(v)-1)f(d) + (n-\deg(v)-1)(\widehat{d}-d)f'(d) \\ &= 1 + (n-\deg(v)-1)f(d) + (d\deg(v) + d - 2\deg(v)r(v))f'(d), \end{aligned} \tag{4.3}$$

where the final equality uses (4.2).

Recall that we have yet to pick v. From the computation above, it is clear that we should pick v so that $A(v) \coloneqq (d\deg(v) + d - 2\deg(v)r(v))f'(d)$ is large relative to $B(v) \coloneqq (\deg(v)+1)f(d)$. In order to do this, let us compute the average values of both of these quantities, averaged over all $v \in V(G)$. The quantity B is easy, as

$$\frac{1}{n}\sum_{v\in V(G)} (\deg(v)+1)f(d) = (d+1)f(d). \tag{4.4}$$

For the first quantity, we first compute that

$$\sum_{v\in V(G)} \deg(v)r(v) = \sum_{v\in V(G)}\sum_{u\sim v}\deg(u) = \sum_{u\in V(G)}\sum_{v\sim u}\deg(u) = \sum_{u\in V(G)} \deg(u)^2 \geqslant nd^2, \tag{4.5}$$

where the final inequality uses the Cauchy–Schwarz inequality and the assumption that the average degree in G is d. Therefore, the average value of $A(v)$ is

$$\frac{1}{n}\sum_{v\in V(G)}(d\deg(v)+d-2\deg(v)r(v))f'(d)\geqslant(d^2+d-2d^2)f'(d)=(d-d^2)f'(d),\qquad(4.6)$$

where we *reversed* the direction of the inequality from (4.5) because $f'(x)\leqslant 0$ for all $x>0$. We now observe that the definition of f implies that it solves the differential equation

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

Thus, (4.4) and (4.6) imply that one plus the average value of $A(v)$ is at least the average value of $B(v)$. This implies that we can pick some vertex $v\in V(G)$ such that $1+A(v)\geqslant B(v)$. Plugging this into (4.3) shows that

$$\begin{aligned}\alpha(G)&\geqslant 1+(n-\deg(v)-1)f(d)+(d\deg(v)+d-2\deg(v)r(v))f'(d)\\&=1+nf(d)-B(v)+A(v)\\&\geqslant nf(d),\end{aligned}$$

completing the proof. □

Given this lemma, the proof of the improved upper bound on $r(3,k)$ is straightforward.

Theorem 4.1.4 (Ajtai–Komlós–Szemerédi [2], Shearer [125])**.** *With the function f as defined in Lemma 4.1.3, we have*

$$r(3,k)\leqslant\frac{k}{f(k)}=(1+o(1))\frac{k^2}{\ln k},$$

where the $o(1)$ term tends to 0 as $k\to\infty$.

Proof. Note that $f(k)=(1+o(1))\frac{\ln k}{k}$ as $k\to\infty$, so it suffices to prove the first inequality. Let $n=k/f(k)$. The statement that $r(3,k)\leqslant n$ is equivalent to saying that every n-vertex graph G contains a triangle or an independent set of order at least k. So fix an n-vertex graph, and let us assume that G is triangle-free (for otherwise we are done). If the average degree d of G is at least k, then we have $\alpha(G)\geqslant d\geqslant k$ by Lemma 4.1.1, so we may assume that $d<k$. Therefore, by Lemma 4.1.3 and the monotonicity of the function f, we have

$$\alpha(G)\geqslant nf(d)\geqslant nf(k)=k.$$ □

We remark that a simple induction argument, together with (2.1), can be used to deduce from Theorem 4.1.4 that for any fixed s, we have

$$r(s,k)=O_s\left(\frac{k^{s-1}}{\log k}\right).$$

Ajtai, Komlós, and Szemerédi used the same idea, but with a more involved induction, to prove that in fact

$$r(s,k) = O_s\left(\frac{k^{s-1}}{(\log k)^{s-2}}\right).$$

This remains the best known upper bound on off-diagonal Ramsey numbers, and it may well be asymptotically best possible.

Another direction one can consider is what happens if we replace the triangle-free assumption in Lemma 4.1.3 by the assumption that G avoids a copy of some other graph. For example, if G is C_4-free, then Li and Rousseau [91] proved that $\alpha(G) \geqslant (1-o(1))(n \ln d)/d$, the same conclusion as in Lemma 4.1.3; proving this is a homework problem. However, if G is K_4-free (or K_s-free for any $s \geqslant 4$), then the optimal bound is not known; the strongest bound known is the following result of Shearer [126].

Theorem 4.1.5 (Shearer [126]). *For every $s \geqslant 4$, there exists a constant $c_s > 0$ such that the following holds. If G is an n-vertex K_s-free graph with average degree $d > 2$, then*

$$\alpha(G) \geqslant c_s \cdot \frac{n \log d}{d \log \log d}.$$

It is widely believed that the $\log \log d$ term in this theorem can be removed, but this remains open even for $s = 4$.

4.2 Interlude: an application to sphere packing

Before we continue the discussion of off-diagonal Ramsey numbers by seeing lower bounds on $r(3,k)$ and $r(4,k)$, let's discuss a recent and striking application of Lemma 4.1.3 (or rather, a strengthening of it) to a geometric problem.

A *sphere packing* in d dimensions is a collection of unit balls in $\mathbb{R}^d$ whose interiors are disjoint. The *density* of a sphere packing is, informally, the fraction of $\mathbb{R}^d$ that is contained in one of the spheres; more formally, if we let S be the union of the balls, then the density is

$$\theta(S) := \limsup_{N \to \infty} \frac{\mathrm{vol}(S \cap [-N,N]^d)}{(2N)^d}.$$

The *sphere packing constant* in dimension d, denoted $\theta(d)$, is the supremum of $\theta(S)$ over all d-dimensional sphere packings; it captures the most efficient way of filling $\mathbb{R}^d$ with disjoint unit balls.

The exact value of $\theta(d)$ is only known in dimensions $d \in \{1,2,3,8,24\}$. Dimension 1 is trivial, and dimension 2 was resolved by Thue in the 19th century; the triangular lattice gives the densest packing in $\mathbb{R}^2$, which is what you would expect from playing around with circle packings. The correct answer in dimension 3 was conjectured by Kepler in 1611, but remained open for hundreds of years until finally being proved by Hales [72], via an extremely long and heavily computer-assisted proof; more recently, the proof was fully formalized in a proof assistant [71]. Even more recently, Viazovska [141] determined $\theta(8)$, and Cohn–Kumar–Miller–Radchenko–Viazovska determined $\theta(24)$. The densest packings in these dimensions are determined by very special lattices called the E_8 and Leech lattice, respectively.

For general dimensions, much less is known. There is a simple general lower bound of $\theta(d) \geqslant 2^{-d}$, which was improved by Rogers [115] to $\theta(d) = \Omega(d2^{-d})$. There have been a number of constant-factor improvements to this bound over the years, but no one was able to prove that $\theta(d) = \omega(d2^{-d})$ as $d \to \infty$. This changed very recently with a breakthrough of Campos, Jenssen, Michelen, and Sahasrabudhe [14], who improved Rogers' bound by a factor of $\Omega(\log d)$.

Theorem 4.2.1 (Campos–Jenssen–Michelen–Sahasrabudhe [14])**.**

$$\theta(d) \geqslant (1 - o(1))\frac{d \ln d}{2^{d+1}}.$$

Their proof is too complicated (and too off-topic) to do in any sort of detail, but let's see a very rough sketch. They begin by randomly selecting a set of points $X \subset \mathbb{R}^d$, which will be potential centers of spheres in the packing. This random choice is done in a very carefully-defined manner, which we will not describe, but which ensures that X satisfies certain desirable properties. Having defined X, one can define a graph G_X whose vertex set is X, and where two vertices are adjacent if their Euclidean distance is less than 2. Because of this choice, an independent set in G_X is precisely a collection of centers of disjoint unit balls. Hence, the task boils down to proving a lower bound on $\alpha(G_X)$, which is where the connection to Lemma 4.1.3 comes in. Unfortunately, G_X is not triangle-free in general, so Campos–Jenssen–Michelen–Sahasrabudhe proved a strengthening of Lemma 4.1.3 to the setting of graphs with "few" triangles (or, more precisely, to the setting when all pairs of vertices have few common neighbors).

Theorem 4.2.2 (Campos–Jenssen–Michelen–Sahasrabudhe [14])**.** *Let G be a graph with n vertices and maximum degree Δ. Suppose that every pair of distinct vertices in G has at most $\Delta/(2\ln\Delta)^7$ common neighbors. Then*

$$\alpha(G) \geqslant (1 - o(1))\frac{n \ln \Delta}{\Delta},$$

where the $o(1)$ tends to 0 as $\Delta \to \infty$.

The proof of Theorem 4.2.2 can be viewed as a generalization of the proof of Lemma 4.1.3. Basically, rather than deleting a single carefully-chosen v (as well as its neighbors) at every step, they instead pick a random set of $\varepsilon n/\Delta$ vertices at every step, and delete them and all their neighbors from G, where ε is a small constant. By carefully adding edges back to G after such a step, in order to ensure that the edge density stays constant, they can continue this process for $(\frac{1}{\varepsilon} - o(1))\ln\Delta$ steps, and thus find the desired independent set.

4.3 Lower bounds on off-diagonal Ramsey numbers

Let us recall the statement of Lemma 3.1.1. It said that if G is a K_s-free graph on N vertices, such that the number of independent sets of order k is at most M^k, then

$$r(s,k) \geqslant \frac{N}{4M}.$$

Thus, in order to prove a lower bound on $r(3,k)$ (for example), we need to find a triangle-free graph G where we have good control over the number of independent sets of order k in G.

The tool we'll use to estimate the number of independent sets of order k is the following result, which says that if a graph is "locally dense"—any reasonably large set contains many edges—then it has few independent sets of a given order. This specific lemma is due to Kohayakawa, Lee, Rödl, and Samotij [79], although the proof technique goes back to work of Kleitman and Winston [78], and the same idea was first applied in this setting by Alon and Rödl [3]. An excellent survey on this topic, including a detailed proof of Lemma 4.3.1, was written by Samotij [120].

Lemma 4.3.1. *Fix positive integers n, r, R and a parameter $\beta \in [0,1]$, which satisfy $Re^{\beta r} \geqslant n$. Suppose that G is an n-vertex graph with the property that for every $X \subseteq V(G)$ with $|X| \geqslant R$, we have*

$$e(X) \geqslant \beta \frac{|X|^2}{2}.$$

Then for any $k \geqslant r$, the number of independent sets in G of order k is at most

$$rn^r \binom{R}{k-r}.$$

If $r \ll k$, then the term rn^r will be subexponential in k, whereas the binomial coefficient is at most $(eR/k)^k$. Thus, we are roughly in the setting of Lemma 3.1.1 with $M \approx R/k$.

Proof of Lemma 4.3.1. We run the following algorithm (called the Kleitman–Winston algorithm) to enumerate the independent sets of order k in G. At a given step of the algorithm, we have chosen some vertices $v_1, \dots, v_i$ which are in our independent set, and we have a remaining set C_{i+1} of candidate vertices. We begin with $C_1 = V(G)$, and we stop the iteration if ever $|C_{i+1}| < R$.

At every step of the algorithm, we look at a maximal-degree vertex v in $G[C_{i+1}]$, the subgraph of G induced by C_{i+1}. As we have not yet stopped, we know that $|C_{i+1}| \geqslant R$, and therefore $e(C_{i+1}) \geqslant \beta \frac{|C_{i+1}|^2}{2}$ by assumption. Equivalently, this condition says that the average degree in $G[C_{i+1}]$ is at least $\beta |C_{i+1}|$. As v was chosen to have maximal degree in C_{i+1}, we conclude that

$$|N(v) \cap C_{i+1}| \geqslant \beta |C_{i+1}|.$$

We now decide whether to include v in our independent set. If yes, we set $v_{i+1} = v$ and $C_{i+2} = C_{i+1} \setminus N(v)$, to ensure that C_{i+2} is still a valid set of candidates for forming an independent set. If no, we discard v from C_{i+1} and repeat the process above with v replaced by a new maximum-degree vertex.

As stated above, we continue this process until $|C_{i+1}| < R$. At that point, we arbitrarily select $w_{i+1}, \ldots, w_k \in C_{i+1}$ such that $\{v_1, \ldots, v_i, w_{i+1}, \ldots, w_k\}$ forms an independent set.

We claim that we can run this process only up to step r, that is, once we select $v_1, \ldots, v_r$, our candidate set C_{r+1} has necessarily shrunk to $|C_{r+1}| < R$. Indeed, every time we select v_i, we have that

$$\frac{|C_{i+1}|}{|C_i|} = \frac{|C_i \setminus N(v_i)|}{|C_i|} \leqslant \frac{(1-\beta)|C_i|}{|C_i|} = 1 - \beta.$$

Therefore,

$$|C_{r+1}| = \frac{|C_{r+1}|}{|C_r|} \cdot \frac{|C_r|}{|C_{r-1}|} \cdots \frac{|C_2|}{|C_1|} \cdot |C_1| \leqslant (1-\beta)^r n < e^{-\beta r} n \leqslant R,$$

where the final inequality is our assumption that $Re^{\beta r} \geqslant n$.

Note that the procedure above necessarily generates every independent set of order k in G. Therefore, we can bound the number of such independent sets by estimating how many choices we have. The process may stop at any index $0 \leqslant i \leqslant r$, and we have at most n^i choices for $v_1, \ldots, v_i$. At that point, as the candidate set has shrunk to size at most R, we have at most $\binom{R}{k-i}$ choices for $w_{i+1}, \ldots, w_k$. Therefore, the total number of independent sets of order k in G is at most

$$\sum_{i=0}^{r} n^i \binom{R}{k-i}.$$

It is easy to see that the summand is maximized at $i = r$, and hence the total number is at most $r \cdot n^r \binom{R}{k-r}$, as claimed. $\square$

4.3.1 Lower bounds on $r(3, k)$

Given Lemma 4.3.1, our task is now to find a triangle-free graph that is locally dense, in the sense of satisfying Lemma 4.3.1 with appropriate parameters. The construction we present is inspired by work of Conlon [20, 22], but is not quite the same as his, and the analysis also uses ideas from [30, 93]. Several alternative constructions are presented in [7].

Let q be a prime power, and consider the finite field $\mathbb{F}_q$, as well as the three-dimensional vector space $\mathbb{F}_q^3$ over it. We begin by defining a bipartite graph Γ_q as follows. The vertex set of Γ_q has two parts P, L, whose names stand for *points* and *lines*. We identify P with $\mathbb{F}_q^3$, and think of the vertices in P as points in this vector space. L, in turn, consists of all lines in $\mathbb{F}_q^3$ whose *direction* is of the form $(1, z, z^2)$, namely all lines of the form

$$\{x + y \cdot (1, z, z^2) : y \in \mathbb{F}_q\} \subseteq \mathbb{F}_q^3,$$

where $x \in \mathbb{F}_q^3$ and $z \in \mathbb{F}_q$. Note that there are exactly q^3 such lines, since we have q options for the direction (from the q options for z), and each such direction gives exactly q^2 parallel lines. Thus $|P| = |L| = q^3$. Finally, we define edges in Γ_q by incidence: we set a vertex $p \in P$ adjacent to a vertex $\ell \in L$ if and only if the point p lies on the line ℓ.

The first key fact we need about Γ_q is the following lemma.

Lemma 4.3.2. *Γ_q is C_4-free and C_6-free.*

Proof. First suppose that there is a C_4 in Γ_q. As Γ_q is bipartite, this means that there are distinct $p_1, p_2 \in P, \ell_1, \ell_2 \in L$ such that p_1, p_2 are both incident to both ℓ_1, ℓ_2. But this is impossible, as any two lines in $\mathbb{F}_q^3$ intersect in at most one point. (This is what we expect from our geometric intuition in $\mathbb{R}^3$, and it's not hard to prove that the same holds in $\mathbb{F}_q^3$.)

Similarly, if there is a C_6 in Γ_q, then there exist distinct $p_1, p_2, p_3 \in P$ and $\ell_1, \ell_2, \ell_3 \in L$ such that p_i and p_{i+1} are both incident to ℓ_i for all i, where the indices are taken modulo 3. Let $z_1, z_2, z_3 \in \mathbb{F}_q$ be such that ℓ_i has direction $(1, z_i, z_i^2)$ for $i \in [\![3]\!]$. Then as both p_i and p_{i+1} are on line ℓ_i, we see that $p_i - p_{i+1}$ is a non-zero multiple of $(1, z_i, z_i^2)$, say $p_i - p_{i+1} = y_i \cdot (1, z_i, z_i^2)$ for some non-zero y_1, y_2, y_3. Therefore,

$$0 = (p_1 - p_2) + (p_2 - p_3) + (p_3 - p_1) = \sum_{i=1}^{3} y_i \cdot (1, z_i, z_i^2).$$

In other words, we've found that the vectors $\{(1, z_i, z_i^2)\}_{i=1}^3$ are linearly dependent. However, the well-known Vandermonde determinant formula implies that this is impossible unless $z_i = z_j$ for some $i \neq j$. But, for example, if $z_1 = z_2$, then this means that ℓ_1 and ℓ_2 are parallel. But as they both pass through p_2, they must be the same line, a contradiction. The same argument applies if $z_2 = z_3$ or $z_1 = z_3$, and we conclude that Γ_q is C_6-free. $\square$

We now (randomly) define a graph G_q as follows. The vertex set of G_q is L, the second vertex part of Γ_q. The edges of G_q are defined as follows. For each $p \in P$, let $N(p)$ denote the neighborhood of p in Γ_q, i.e. the set of lines in L incident to p. For each $p \in P$, we pick a uniformly random bipartition of $N(p)$ into $A(p) \sqcup B(p)$. Then, for every $\ell_1 \in A(p), \ell_2 \in B(p)$, we add an edge between ℓ_1 and ℓ_2 in G_q. Doing this for all $p \in P$, we obtain the random graph G_q. In other words, G_q is the edge-union of complete bipartite graphs, where each $p \in P$ contributes a complete bipartite graph between $A(p)$ and $B(p)$.

Recall that Γ_q is C_4-free by Lemma 4.3.2. This means that for every $\ell_1, \ell_2 \in L$, there is at most one choice of p such that $\ell_1, \ell_2 \in N(p)$. Hence, to every edge $(\ell_1, \ell_2) \in E(G_q)$, we can associate a *label* p, which is the unique $p \in P$ such that $\ell_1, \ell_2 \in N(p)$.

Lemma 4.3.3. *G_q is triangle-free with probability* 1 *(i.e. regardless of the random choices).*

Proof. Suppose for contradiction that there exist distinct $\ell_1, \ell_2, \ell_3 \in L = V(G_q)$ that form a triangle in G_q. Let $p_1, p_2, p_3 \in P$ be the labels of $(\ell_1, \ell_2), (\ell_2, \ell_3)$, and (ℓ_3, ℓ_1), respectively.

We split into two cases. First, suppose that two of the p_i are equal, say $p_1 = p_2$. This implies that ℓ_1, ℓ_2, ℓ_3 all lie in $N(p_1)$. This then implies that $p_3 = p_1$ as well. But recall that the only edges we add with label p_1 are a complete bipartite graph between $A(p_1)$ and $B(p_1)$, and these edges can contain no triangle as this graph is bipartite. This concludes this case.

So we may now assume that p_1, p_2, p_3 are distinct. But then the fact that $\ell_i, \ell_{i+1} \in N(p_i)$ for all $i \in [\![3]\!]$ implies that $\ell_1, p_1, \ell_2, p_2, \ell_3, p_3$ forms a copy of C_6 in Γ_q. By Lemma 4.3.2 no such copy can exist, a contradiction. $\square$

The final result we need about G_q is that it satisfies the local density condition we need to apply Lemma 4.3.1. It is here where the randomness in the definition of G_q is crucial.

We first prove that for any large set of vertices X of G_q, there are many "potential edges" of G_q, namely many pairs $\ell_1, \ell_2 \in X$ such that $\ell_1, \ell_2 \in N(p)$ for some $p \in P$. Once we have this, the randomness will imply that a good fraction of these potential edges will become true edges of G_q.

Lemma 4.3.4. *For any $X \subseteq L$, the number of pairs $(\ell_1, \ell_2) \in X^2$ such that $\ell_1, \ell_2 \in N(p)$ for some $p \in P$ is at least $|X|^2/q$.*

Proof. The quantity we are interested in is precisely

$$\sum_{p \in P} |N(p) \cap X|^2.$$

By the Cauchy–Schwarz inequality, we have

$$\sum_{p \in P} |N(p) \cap X|^2 \geqslant \frac{1}{|P|} \left(\sum_{p \in P} |N(p) \cap X| \right)^2.$$

Note that the quantity in parentheses is precisely the number of edges in Γ_q incident to $X \subseteq L$. Since every vertex in L is incident to precisely q edges (as every line in $\mathbb{F}_q^3$ contains exactly q points), we have that

$$\frac{1}{|P|} \left(\sum_{p \in P} |N(p) \cap X| \right)^2 = \frac{1}{|P|} \left(\sum_{\ell \in X} q \right)^2 = \frac{1}{q^3} (q|X|)^2 = \frac{|X|^2}{q},$$

where we also plug in that $|P| = q^3$. □

Since Lemma 4.3.4 counts unordered pairs (ℓ_1, ℓ_2), we find that X contains at least $|X|^2/(2q)$ "potential edges". On average, a set $X \subseteq L$ will keep roughly half of its "potential edges" when we sample the random graph G_q. The reason is that each potential edge corresponds to a pair $\ell_1, \ell_2 \in N(p)$ for some p, and there is a probability $1/2$ that these two vertices will be placed on opposite sides of the bipartition $A(p) \cup B(p)$, thus yielding a true edge in G_q. Of course, not every set X will receive *exactly* half of its potential edges, and we expect some random fluctuations. Nonetheless, it is intuitively reasonable that all *large* sets X will receive roughly half of the potential edges, and thus we expect to be in the setting of Lemma 4.3.1 with the parameter $\beta \approx 1/(2q)$.

Before making this formal, let's think about how small of an X we can expect this to hold for. Note that in G_q, a typical vertex ℓ has $\Theta(q^2)$ neighbors. The reason is that ℓ lies in $N(p)$ for exactly q choices of p, and each such p will yield, on average, $|N(p)|/2 = \Theta(q)$ edges of G_q incident to ℓ. As G_q is triangle-free, clearly the neighborhood of any ℓ must actually contain *zero* edges. Hence, we cannot expect $e(X) \geqslant \beta |X|^2/2$ to hold for all sets X of order $\Theta(q^2)$. Thus, again using the terminology of Lemma 4.3.1, we should expect to pick R of order at least q^2.

In fact, one can really obtain such a result with $R = \Theta(q^2)$, as noted in [93, Section 3]. However, doing this requires a somewhat involved argument based on a certain dyadic partitioning. We will prove the following weaker statement, which establishes that we are in the setting of Lemma 4.3.1 with $R = \Theta(q^2 \log q)$ and $\beta = \Theta(1/q)$.

Lemma 4.3.5. *With positive probability, G_q has the following property. For every $X \subseteq L$ with $|X| \geqslant R := 200q^2 \ln q$, we have that*

$$e(X) \geqslant \beta \frac{|X|^2}{2},$$

where $\beta := 1/(10q)$.

Proof of Lemma 4.3.5

For the proof, we will need the following probabilistic concentration inequality, which is a convenient form of the Azuma–Hoeffding inequality. A proof can be found in [74, Corollary 2.27 and Remark 2.28] or [4, Section 7.2]. Let us say that a function $f : \{0,1\}^m \to \mathbb{R}$ is $\{L_i\}$-*Lipschitz* if its value changes by at most L_i whenever the input is changed on only the ith coordinate, that is, for all $i \in [\![m]\!]$ and all $z_1, \dots, z_m \in \{0,1\}$, we have

$$|f(z_1, \dots, z_i, \dots, z_m) - f(z_1, \dots, 1 - z_i, \dots, z_m)| \leqslant L_i.$$

Lemma 4.3.6. *Let $Z_1, \dots, Z_m$ be independent random variables taking values in $\{0,1\}$. Let $f : \{0,1\}^m \to \mathbb{R}$ be $\{L_i\}$-Lipschitz, and let $Z = f(Z_1, \dots, Z_m)$. Then*

$$\Pr\left(Z \leqslant \frac{1}{2}\mathbb{E}[Z]\right) \leqslant \exp\left(-\frac{\mathbb{E}[Z]^2}{2\sum_{i=1}^m L_i^2}\right).$$

With this in hand, we are ready to prove Lemma 4.3.5.

Proof of Lemma 4.3.5. First, let us fix some set $X \subseteq L$ with $|X| \geqslant R$. For every $\ell \in X$ and every $p \in P$ such that $\ell \in N(p)$, let us make a random variable $Z_{\ell,p}$ with value 1 if $\ell \in A(p)$, and value 0 if $\ell \in B(p)$. Let $Z = e(X)$, which is a random variable depending on the random choices of the bipartition. In fact, we see that Z is a function of the random variables $Z_{\ell,p}$. Note that flipping $Z_{\ell,p}$ corresponds to changing whether $\ell \in A(p)$ or $\ell \in B(p)$, and this can affect the number of edges in X by at most $|N(p) \cap X|$. Hence, this function is Lipschitz with parameters

$$L_{\ell,p} := |N(p) \cap X|.$$

From the proof of Lemma 4.3.4, we see that $S := \sum_{\ell,p} L_{\ell,p}^2$ is precisely equal to the number of pairs $(\ell_1, \ell_2) \in X^2$ with $\ell_1, \ell_2 \in N(p)$ for some $p \in P$.

We now claim that $\mathbb{E}[Z] \geqslant \frac{1}{5}S \geqslant |X|^2/(5q)$, where the final inequality is simply the statement of Lemma 4.3.4. The reason is that, as discussed above, every unordered pair of distinct ℓ_1, ℓ_2 counted by S becomes an edge of G_q with probability $\frac{1}{2}$. S counts ordered pairs, so we need to divide by 2, and need to subract off the contribution of $|X|$ pairs (ℓ, ℓ). But since $|X| \geqslant R > 10q$, the number of such pairs is at most $S/10$.

Therefore, by Lemma 4.3.6 and the definition of β, we find that

$$\begin{aligned}\Pr\left(e(X) < \beta\frac{|X|^2}{2}\right) &\leqslant \Pr\left(Z \leqslant \frac{1}{2}\mathbb{E}[Z]\right)\\ &\leqslant \exp\left(-\frac{\mathbb{E}[Z]^2}{2S}\right)\\ &\leqslant \exp\left(-\frac{\mathbb{E}[Z]}{10}\right)\\ &\leqslant \exp\left(-\frac{|X|^2}{50q}\right).\end{aligned}$$

We may now take a union bound over the $\binom{q^3}{|X|}$ choices for such an X, and sum this up over all choices of $|X|$, to find that the probability that the claimed property does not hold is at most

$$\sum_{|X|=R}^{q^3}\binom{q^3}{|X|}e^{-|X|^2/(50q)} \leqslant \sum_{|X|=R}^{q^3} q^{3|X|}e^{-|X|^2/(50q)} = \sum_{|X|=R}^{q^3}\left(e^{3\ln q-|X|/(50q)}\right)^{|X|}.$$

Note that our choice of $R = 200q^2\ln q$ implies that

$$e^{3\ln q-|X|/(50q)} \leqslant e^{3\ln q-R/(50q)} \leqslant \frac{1}{q}.$$

Hence, the sum above is at most $2q^{-R}$, which is less than 1. Thus, G_q has the claimed property with positive probability. □

By plugging the result of Lemma 4.3.5 into Lemma 4.3.1, we can estimate the number of independent sets of a given size in G_q. For completeness, we actually estimate the number of independent sets of size at most k.

Lemma 4.3.7. *Let q be a prime power and let $30q(\ln q)^2 \leqslant k \leqslant q^2$. There exists a triangle-free graph G_q with $N := q^3$ vertices such that the number of independent sets of order at most k in G_q is at most M^k, where $M = 200q/\ln q$.*

Proof. By Lemmas 4.3.3 and 4.3.5, there exists an N-vertex graph G_q that is triangle-free and satisfies the conditions of Lemma 4.3.1 with $R = 200q^2\ln q$ and $\beta = 1/(10q)$. Let $r := 10q\ln q$, and note that

$$Re^{\beta r} = (200q^2\ln q)e^{(10q\ln q)/(10q)} = 200q^3\ln q \geqslant N.$$

We also have that $k \geqslant 3r\ln q$, which implies that $k \geqslant r$ and that $N^r = q^{3r} \leqslant e^k$. We are in a position to apply Lemma 4.3.1. We conclude that the number of independent sets in G_q of order *exactly* k is at most

$$rN^r\binom{R}{k-r} \leqslant \left(\frac{e^2R}{k}\right)^k \leqslant \left(\frac{200e^2q^2\ln q}{30q(\ln q)^2}\right)^k \leqslant \left(\frac{50q}{\ln q}\right)^k.$$

To obtain an estimate on the number of independent sets of order *at most* k, we argue as follows. For every $r \leqslant t \leqslant k$, the number of independent sets of order exactly t is at most

$$rN^r \binom{R}{t-r} \leqslant rN^r \binom{R}{k-r} \leqslant \left(\frac{50q}{\ln q}\right)^k.$$

On the other hand, for every $t < r$, the number of independent sets of order t in G_q is certainly at most $\binom{N}{t}$. Adding this all up, we conclude that the total number of independent sets in G_q of order at most k is at most

$$\sum_{t=0}^{r-1} \binom{N}{t} + k\left(\frac{50q}{\ln q}\right)^k \leqslant N^r + k\left(\frac{50q}{\ln q}\right)^k \leqslant e^k + k\left(\frac{50q}{\ln q}\right)^k \leqslant \left(\frac{200q}{\ln q}\right)^k = M^k. \qquad \square$$

We are finally ready to deduce a lower bound on $r(3, k)$.

Theorem 4.3.8. *We have*

$$r(3, k) > \frac{k^2}{C(\log k)^3}$$

for an absolute constant $C > 0$.

By being more careful (specifically, by proving a version of Lemma 4.3.5 without the logarithmic loss in the value of R), one can improve this result to $r(3, k) = \Omega(k^2/(\log k)^2)$. It is not known whether one can use such a technique to obtain the optimal result, of $r(3, k) = \Omega(k^2/\log k)$.

Proof. By Bertrand's postulate, we can find a prime power q satisfying $k/(60(\ln k)^2) \leqslant q \leqslant k/(30(\ln k)^2)$, which implies that $k \geqslant 30q(\ln q)^2$. By Lemma 4.3.7, there exists a triangle-free graph G_q on $N = q^3$ vertices with at most M^k independent sets of order at most k, where $M = 200q/\ln q$. We may therefore apply Lemma 3.1.1 to conclude that

$$r(3, k) \geqslant \frac{N}{4M} = \frac{q^3}{800q/\ln q} = \frac{q^2 \ln q}{800} \geqslant 10^{-7} \frac{k^2}{(\ln k)^3}. \qquad \square$$

4.3.2 Lower bounds on $r(4, k)$

Given everything we have done so far, it becomes very simple to explain the new ingredient introduced by Mattheus and Verstraëte to obtain a good lower bound on $r(4, k)$. Of course, this is really doing them a disservice, since the presentation above is heavily inspired by their work, and a major contribution of theirs is realizing how to implement such an approach.

The key new ingredient we need is a construction of a graph Λ_q, which is also a point-line incidence graph in a certain finite geometry, which does not contain the so-called *O'Nan configuration*. In graph-theoretic terms, this is simply a *subdivision* of K_4, and can be explicitly described as a set of four distinct lines $\ell_1, \ldots, \ell_4$ and four distinct points $p_{12}, p_{13}, p_{14}, p_{23}, p_{24}, p_{34}$ such that each p_{ij} is incident to both ℓ_i and ℓ_j.

We fix a prime power q, and work over the finite field $\mathbb{F}_{q^2}$ and in the two-dimensional vector space $\mathbb{F}_{q^2}^2$ over it. Note that this is also, of course, a four-dimensional vector space over $\mathbb{F}_q$, but we won't think of it like this; our base field will always be $\mathbb{F}_{q^2}$, and then when we discuss e.g. lines, we will always mean one-dimensional $\mathbb{F}_{q^2}$-affine-linear subspaces. We define

$$P \coloneqq \{(x_1, x_2) \in \mathbb{F}_{q^2}^2 : x_1^{q+1} + x_2^{q+1} + 1 = 0\} \subseteq \mathbb{F}_{q^2}^2.$$

One can show that $|P| = (1 + o(1))q^3$; there are q^2 choices for x_1, and having fixed x_1, there are $(1 + o(1))q$ choices for a $(q + 1)$th root of $-1 - x_1^{q+1}$, that is, $(1 + o(1))q$ choices for x_2.

We also define L to consist of all lines in $\mathbb{F}_{q^2}^2$ which intersect P in at least two points. There are $(1 + o(1))q^4$ lines in $\mathbb{F}_{q^2}^2$, and one can show that at most q^3 of them intersect P in fewer than two points, so $|L| = (1+o(1))q^4$. We define Λ_q to be the incidence graph between P and L, i.e. the bipartite graph with parts $P \cup L$, in which a pair (p, ℓ) is an edge if and only if p lies in the line ℓ.

The following lemma, which is analogous to Lemma 4.3.2, shows that this is a good graph to use for lower-bounding $r(4, k)$ using the technique discussed above. This result was first proved by O'Nan [101], which is why O'Nan configurations are so named.

Lemma 4.3.9. *The graph Λ_q is C_4-free and contains no O'Nan configuration.*

The fact that Λ_q is C_4-free follows from the exact same reason as in Lemma 4.3.2. Namely, a C_4 in Λ_q would correspond to two lines intersecting in two distinct points, and that cannot happen. The proof that Λ_q has no O'Nan configuration is also based on elementary linear algebra—just as the proof in Lemma 4.3.2 that Γ_q is C_6-free—but we will skip it because it is somewhat more involved and not particularly interesting. An elementary proof can be found in [93, Proposition 1].

We now form a random graph H_q on vertex set L by picking, for each $p \in P$, a random bipartition $N(p) = A(p) \cup B(p)$ of its neighborhood in Λ_q, and adding to H_q all edges between $A(p)$ and $B(p)$. From Lemma 4.3.9, it is not hard to prove the following statement, analogous to Lemma 4.3.3.

Lemma 4.3.10. *H_q is K_4-free with probability* 1.

The final ingredient we need, analogously to Lemma 4.3.5, is the following statement.

Lemma 4.3.11. *With positive probability, H_q has the following property. For every $X \subseteq L$ with $|X| \geqslant R \coloneqq 10^7 q^2$, we have that*

$$e(X) \geqslant \beta \frac{|X|^2}{2},$$

where $\beta \coloneqq 1/(300q)$.

Unfortunately, for technical reasons arising from the fact that $|L| \approx q^4$ is much larger than $|P| \approx q^3$, it seems impossible to prove Lemma 4.3.11 (or even a weaker version with some logarithmic losses) by blindly following the proof of Lemma 4.3.5. Instead, one has to partition P into three parts, depending on how large $|N(p) \cap X|$ is, and then apply the argument of Lemma 4.3.5 to each part in turn. As such, we will skip the proof; it can be found in [93, Theorem 3]. A somewhat more general (and somewhat simpler) result, with a logarithmic loss in the value of R, is proved in [30, Lemma 2].

However, once we have these preliminaries, we can follow the proof technique used above for $r(3,k)$. Namely, we can plug Lemma 4.3.11 into Lemma 4.3.1 to bound the number of independent sets of order k there are in H_q, as follows.

Lemma 4.3.12. *Let q be a prime power and let $2400q(\ln q)^2 \leqslant k \leqslant q^2$. There exists a K_4-free graph H_q with N vertices, where $q^4/2 \leqslant N \leqslant q^4$, such that the number of independent sets of order at most k in H_q is at most M^k, where $M = 10^7 q/(\ln q)^2$.*

Proof. By Lemmas 4.3.10 and 4.3.11, there exists an N-vertex[1] graph H_q that is K_4-free and satisfies the conditions of Lemma 4.3.1 with $R = 10^7 q^2$ and $\beta = 1/(300q)$. Let $r \coloneqq 600q \ln q$, and note that

$$Re^{\beta r} = (10^7 q^2) e^{(600q \ln q)/(300q)} = 10^7 q^4 \geqslant N.$$

We also have that $k \geqslant 4r \ln q$, which implies that $k \geqslant r$ and that $N^r \leqslant q^{4r} \leqslant e^k$. We are in a position to apply Lemma 4.3.1. We conclude that for any $r \leqslant t \leqslant k$, the number of independent sets in H_q of order t is at most

$$rN^r \binom{R}{t-r} \leqslant rN^r \binom{R}{k-r} \leqslant \left(\frac{e^2 R}{k}\right)^k \leqslant \left(\frac{10^8 q^2}{2400q(\ln q)^2}\right)^k \leqslant \left(\frac{10^6 q}{(\ln q)^2}\right)^k.$$

On the other hand, for every $t < r$, the number of independent sets of order t in H_q is certainly at most $\binom{N}{t}$. Adding this all up, we conclude that the total number of independent sets in H_q of order at most k is at most

$$\sum_{t=0}^{r-1} \binom{N}{t} + k\left(\frac{10^6 q}{(\ln q)^2}\right)^k \leqslant N^r + k\left(\frac{10^6 q}{(\ln q)^2}\right)^k \leqslant e^k + k\left(\frac{10^6 q}{(\ln q)^2}\right)^k \leqslant \left(\frac{10^7 q}{(\ln q)^2}\right)^k = M^k. \qquad \square$$

We can now plug this result into Lemma 3.1.1 to obtain the theorem of Mattheus and Verstraëte [93].

Theorem 4.3.13 (Mattheus–Verstraëte [93])**.** *We have*

$$r(4,k) > \frac{k^3}{C(\log k)^4}$$

for an absolute constant $C > 0$.

Proof. Pick a prime power q with $k/(4800(\ln k)^2) \leqslant q \leqslant k/(2400(\ln k))^2$. By Lemmas 3.1.1 and 4.3.12, we have

$$r(4,k) \geqslant \frac{N}{4M} \geqslant \frac{q^4/2}{4 \cdot 10^7 q/(\ln q)^2} \geqslant 10^{-8} q^3 (\ln q)^2 \geqslant 10^{-20} \frac{k^3}{(\ln k)^4}. \qquad \square$$

[1]One can verify that $|V(H_q)| = |L| = q^4 - q^3 + q^2$, which is between $q^4/2$ and q^4 for all prime powers q.

Chapter 5

Graph Ramsey numbers

5.1 Introduction

We will now move away to a more general topic than we have considered so far, that of *graph Ramsey numbers*.

Definition 5.1.1. Given graphs $H_1, \ldots, H_q$, their *Ramsey number* $r(H_1, \ldots, H_q)$ is defined as the minimum N such that any q-coloring of $E(K_N)$ contains a monochromatic copy of H_i in color i, for some $i \in [\![q]\!]$. Here, by a monochromatic copy, we mean a subgraph of K_N isomorphic to H_i, all of whose edges receive color i.

In case $H_1 = \cdots = H_q = H$, we denote this Ramsey number by $r(H; q)$. In case $q = 2$, we use the shorthand $r(H) \coloneqq r(H; 2)$.

Of course, everything we have studied so far is a special case of these more general graph Ramsey numbers, as $r(k)$ is simply $r(K_k)$, and $r(k, \ell) = r(K_k, K_\ell)$, etc. However, it turns out that there is an extremely rich theory of Ramsey numbers of graphs H which are not necessarily complete graphs; moreover, most of the interesting results actually arise when H is extremely far from being a complete graph.

We begin with a simple observation, which is that if H_i is a subgraph of H_i', then $r(H_1, \ldots, H_q) \leqslant r(H_1', \ldots, H_q')$, since any monochromatic copy of H_i' also yields a monochromatic copy of H_i. Thus, $r(H) \leqslant r(H')$ whenever $H \subseteq H'$. Since every n-vertex graph is a subgraph of K_n, we conlude that

$$r(H) \leqslant r(K_n) < 4^n \qquad \text{for every } n\text{-vertex graph } H.$$

Thus, in the worst case, an n-vertex graph may have Ramsey number that is exponential in n.

On the other hand, the most general lower bound we can get is that $r(H) \geqslant n$ if H is an n-vertex graph. Indeed, we need at least n vertices to be able to "fit" a copy of H. Moreover, this trivial lower bound is best possible in general, for if H has no edges (or even one edge), then $r(H) = n$.

Thus, for a general n-vertex graph H, we know $n \leqslant r(H) \leqslant 4^n$, and both behaviors—linear in n and exponential in n—are possible, for the empty graph and the complete graph, respectively. Based on our experience for cliques, we might expect that the exponential bound should be closer to the truth for most graphs. However, the striking result that we will see is that for many "natural" classes of graphs—and, in fact, for all *sparse* graphs—the lower bound is much closer to the truth.

5.2 Ramsey numbers of trees

Let us begin with the following simple result, which was probably first observed by Erdős and Graham [42]; it says that the lower bound is close to tight for trees.

Theorem 5.2.1. *If T is an n-vertex tree, then $r(T) \leqslant 4n - 3$.*

To prove this, we will use two simple lemmas from elementary graph theory.

Lemma 5.2.2. *If a graph G has average degree d, then it has a subgraph G' with minimum degree at least $d/2$.*

Proof. Let G have m vertices, so that it has $md/2$ edges. Repeatedly delete from G a vertex of degree less than $d/2$, as long as such a vertex exists. Since we delete fewer than $d/2$ edges at each step, and continue for at most m steps, we delete fewer than $md/2$ edges in total. As G has exactly $md/2$ edges, when we terminate this process, there must be at least one edge—and thus at least one vertex—remaining. However, the process only terminates once we've produced a subgraph of minimum degree at least $d/2$, completing the proof. □

Lemma 5.2.3. *Let T be an n-vertex tree. If G is a graph with minimum degree at least $n - 1$, then $T \subseteq G$.*

Proof. We proceed by induction on n, with the base case $n = 1$ being trivial since the only 1-vertex tree is a subgraph of every non-empty graph. Inductively, suppose this is true for all $(n-1)$-vertex trees. Let T' be obtained from T by deleting a leaf v, and let u be the unique neighbor of v in T. By the inductive hypothesis, $T' \subseteq G$, so let us pick a copy of T' in G, and let w be the vertex of G filling the role of u. As G has minimum degree at least $n-1$, w has at least $n - 1$ neighbors, and at most $n - 2$ of these neighbors were used in embedding the other $n-2$ vertices of T'. Thus, there is at least one unused neighbor of w, which means that we can extend the T'-copy to a T-copy by adding in this unused neighbor. □

With this lemmas, it is straightforward to prove Theorem 5.2.1.

Proof of Theorem 5.2.1. Let $N = 4n - 3$, and fix a 2-coloring of $E(K_N)$. Without loss of generality, we may assume that at least half the edges are red. Let $G \subseteq K_N$ be the graph comprising the red edges. Since G has at least half the edges of K_N, it has average degree at least $2n-2$. By Lemma 5.2.2, there is a subgraph $G' \subseteq G$ of minimum degree at least $n-1$. By Lemma 5.2.3, we have $T \subseteq G'$, and this yields a monochromatic red copy of T. □

5.3 Ramsey numbers of complete bipartite graphs

Recall that $K_{s,t}$ denotes the *complete bipartite graph* with parts of sizes s, t. We will always assume, without loss of generality, that $s \leqslant t$. Let us begin by proving the following upper bound on $r(K_{s,t})$.

Theorem 5.3.1. *For any $s \leqslant t$, we have*

$$r(K_{s,t}) \leqslant 2^{s+1}t.$$

Note that, if we plug in $s = t = n$, then we obtain that $r(K_{n,n}) = O(n2^n)$. Since $K_{n,n}$ has $2n$ vertices, this is a much better, although still exponential, bound than the naïve one of

$$r(K_{n,n}) \leqslant r(2n) < 4^{2n} = 16^n.$$

We remark that $r(K_{n,n})$ really does grow exponentially in n, and that the lower bound

$$r(K_{n,n}) > 2^{\frac{n-1}{2}}$$

will follow from a more general result, Proposition 5.4.1, which we will prove shortly. On the other hand, if we think of s as a constant, we obtain that $r(K_{s,t}) = O_s(t)$ as $t \to \infty$. Since $K_{s,t}$ has $s + t \leqslant 2t$ vertices, this shows that for fixed s, $K_{s,t}$ has a Ramsey number which is linear in its number of vertices—the same behavior as we saw for trees.

Proof of Theorem 5.3.1. The case $s = 1$ follows from a homework problem; it also follows, up to an additive constant of 1, from Theorem 5.2.1, since $K_{1,t}$ is a tree. We henceforth assume that $t \geqslant s \geqslant 2$.

Let $N = 2^{s+1}t$, and fix a red/blue coloring of $E(K_N)$. For every vertex $v \in V(K_N)$, let $\deg_R(v), \deg_B(v)$ denote the red and blue degrees, respectively, of v. Let S denote the number of monochromatic copies of $K_{1,s}$ in the coloring. We can count S by summing over all N choices for the central vertex, and then picking s distinct neighbors; this shows that

$$S = \sum_{v \in V(K_N)} \left(\binom{\deg_R(v)}{s} + \binom{\deg_B(v)}{s} \right).$$

Note that $\deg_R(v) + \deg_B(v) = N - 1$ for every v, and that the sum $\binom{x}{s} + \binom{N-1-x}{s}$ is minimized[1] when $x = N - 1 - x$, i.e. $x = \frac{N-1}{2}$. Therefore, we find that

$$S \geqslant N \cdot 2\binom{\frac{N-1}{2}}{s}.$$

On the other hand, another way of counting S is by counting over all options for the s leaves of the star. Let us assume for contradiction that there is no monochromatic $K_{s,t}$. Then

[1] This is a special case of a much more general fact, that the function $x \mapsto \binom{x}{s}$ is convex for any fixed $s \geqslant 1$. This special case can also be proved directly without appealing to convexity.

every s-set of vertices forms the set of leaves of fewer than t red stars $K_{1,s}$, and of fewer than t blue stars $K_{1,s}$. Thus,

$$S < 2t\binom{N}{s}.$$

Comparing our lower and upper bounds on S, we find that

$$2t\binom{N}{s} > 2N\binom{\frac{N-1}{2}}{s}$$

or equivalently

$$t \cdot N(N-1)\cdots(N-s+1) > N \cdot \frac{N-1}{2}\left(\frac{N-1}{2}-1\right)\cdots\left(\frac{N-1}{2}-s+1\right).$$

Rearranging, this is equivalent to

$$\frac{2^s t}{N} > \left(\frac{N-1}{N}\right)\left(\frac{N-3}{N-1}\right)\left(\frac{N-5}{N-2}\right)\cdots\left(\frac{N-2s+1}{N-s+1}\right) = \prod_{i=0}^{s-1}\frac{N-2i-1}{N-i}.$$

However, we have that

$$\prod_{i=0}^{s-1}\frac{N-2i-1}{N-i} = \prod_{i=0}^{s-1}\left(1-\frac{i+1}{N-i}\right) \geqslant 1-\sum_{i=0}^{s-1}\frac{i+1}{N-i} \geqslant 1-\frac{2\binom{s+1}{2}}{N} \geqslant \frac{1}{2},$$

where the second inequality uses that $N \geqslant 2s$, hence $N-i \geqslant N/2$ for all $i \leqslant s-1$, and the third inequality uses that $2\binom{s+1}{2} = (s+1)s \leqslant (s+1)t \leqslant 2^s t = N/2$, since $2^s \geqslant s+1$ for all $s \geqslant 2$. Putting this all together, we conclude that

$$\frac{2^s t}{N} > \frac{1}{2},$$

which contradicts our choice of N. This contradiction completes the proof. □

In fact, if one carefully examines the proof of Theorem 5.3.1, it becomes evident that we are not *really* using the fact that there are two colors. Instead, all we are really doing is noticing that any graph with sufficiently many edges must contain a copy of $K_{s,t}$. Such a result was first proved by Kővári, Sós, and Turán [82], and it has become one of the most fundamental results in extremal graph theory; they also introduced the technique which we used in the proof of Theorem 5.3.1.

Theorem 5.3.2 (Kővári–Sós–Turán [82])**.** *Let $s \leqslant t$ be integers, and let G be an N-vertex graph with at least $t^{\frac{1}{s}}N^{2-\frac{1}{s}} + sN$ edges. Then $K_{s,t}$ is a subgraph of G.*

Proof. Let d be the average degree of G, and note that

$$d = \frac{2e(G)}{N} \geqslant 2t^{\frac{1}{s}} N^{1-\frac{1}{s}}. \tag{5.1}$$

Also, since $e(G) \geqslant sN$, we have that $d \geqslant 2s$. For a real number x, we define

$$f(x) = \begin{cases} \frac{x(x-1)\ldots(x-s+1)}{s!} & \text{if } x \geqslant s-1 \\ 0 & \text{otherwise.} \end{cases}$$

The function f is convex (this can be verified by computing its second derivative), and agrees with the binomial coefficient $\binom{x}{s}$ whenever x is a non-negative integer. Note too that since $d \geqslant 2s$, we have

$$f(d) = \frac{1}{s!} \cdot d(d-1)\cdots(d-s+1) \geqslant \frac{1}{s!} \cdot \left(\frac{d}{2}\right)^s = \frac{d^s}{2^s s!}. \tag{5.2}$$

Let S denote the number of copies of $K_{1,s}$ in G. We can count S by summing over all N choices for the central vertex, and then picking s distinct neighbors; this shows that

$$S = \sum_{v \in V(G)} \binom{\deg(v)}{s} = \sum_{v \in V(G)} f(\deg(v)).$$

Now, since f is convex, Jensen's inequality and (5.2) imply that

$$S \geqslant N \cdot f(d) \geqslant \frac{Nd^s}{2^s s!}.$$

On the other hand, another way of counting S is by counting over all options for the s leaves of the star. Let us assume for contradiction that $K_{s,t} \not\subseteq G$. Then every s-set of vertices forms the set of leaves of fewer than t stars $K_{1,s}$. Hence,

$$S < \binom{N}{s} t \leqslant \frac{N^s t}{s!}.$$

Comparing the lower and upper bounds on S, we find that

$$\frac{Nd^s}{2^s} < N^s t,$$

or equivalently

$$d < 2t^{\frac{1}{s}} N^{1-\frac{1}{s}}.$$

This contradicts (5.1), completing the proof. □

5.4 The Burr–Erdős conjecture

So far, we have seen several examples of graph Ramsey numbers, and observed different growth rates. First, as we know from Chapter 2, $r(K_n)$ grows exponentially in n. Similarly, $r(K_{n,n})$ grows exponentially in n (and thus in $2n$, which is its number of vertices). On the other hand, all trees, as well as complete bipartite graphs in which one side has constant size, have Ramsey numbers *linear* in the number of vertices. Can we figure out a general rule explaining these extremely different growth rates?

Looking at the examples above, it is natural to guess that the major difference has to do with *density*. Both K_n and $K_{n,n}$ are very dense graphs, namely graphs with a quadratic number of edges. On the other hand, trees and complete bipartite graphs with one side of constant size are very sparse, in that their number of edges is only linear in their number of vertices. Equivalently, the average degree of the former graphs is large—linear in the number of vertices—whereas it is *constant* for the latter graphs. Perhaps this explains the difference in the Ramsey numbers?

As it turns out, this is close to the correct explanation. One direction really is true; if a graph has high average degree, then its Ramsey number is large, as shown in the following simple proposition.

Proposition 5.4.1. *If H has average degree d, then $r(H) > 2^{\frac{d-1}{2}}$.*

Proof. The proof is very similar to that of Theorem 2.2.2. Let H have $k \geqslant 2$ vertices, and thus $kd/2$ edges. Let $N = 2^{\frac{d-1}{2}}$, and consider a uniformly random 2-coloring of $E(K_N)$. Every tuple of k vertices in K_N forms a monochromatic copy of H with probability $2^{1-kd/2}$, and there are $k!\binom{N}{k}$ such tuples[2]. Therefore, the probability that the coloring has a monochromatic copy of H is at most

$$k!\binom{N}{k} \cdot 2^{1-\frac{kd}{2}} < N^k \cdot 2^{1-\frac{kd}{2}} = 2^{k\frac{d-1}{2}+1-\frac{kd}{2}} = 2^{1-\frac{k}{2}} \leqslant 1,$$

and thus there must exist a coloring with no monochromatic copies of H. $\square$

Thus, we find that if H has average degree which is linear in its number of vertices $v(H)$, then $r(H)$ is exponential in $v(H)$. Is it possible that the same holds at the opposite extreme, namely that if H has constant average degree, then $r(H)$ is linear in $v(H)$, as happened for trees and complete bipartite graphs? It is not hard to see that the answer is no.

Proposition 5.4.2. *There exists an n-vertex graph H with average degree at most 1 and with $r(H) > 2^{\sqrt{n}/2}$.*

Proof. Let H be obtained from a complete graph $K_{\sqrt{n}}$ by adding $n - \sqrt{n}$ isolated vertices. Then H has $\binom{\sqrt{n}}{2}$ edges, and thus average degree $\frac{2}{n}\binom{\sqrt{n}}{2} \leqslant 1$. However,

$$r(H) \geqslant r(K_{\sqrt{n}}) > 2^{\sqrt{n}/2},$$

[2] Note that we include an extra factor of $k!$, which was not present in the proof of Theorem 2.2.2. The reason is that K_k is highly symmetric; for a general H, we need to consider not only the k vertices that can define it, but also the potentially $k!$ different ways of identifying $V(H)$ with these k vertices.

by Theorem 2.2.2. □

Given this example, it's clear why the naïve conjecture "constant average degree implies linear Ramsey number" cannot be true. Namely, the graph H above has constant average degree, but it contains a subgraph (namely $K_{\sqrt{n}}$) with much higher average degree, and it is this subgraph that really determines $r(H)$. This shows that rather than considering the global average degree, we need to consider a more refined parameter that takes into account subgraphs that are denser than H itself. There are several different ways of formalizing such a parameter, and they end up giving essentially identical results; we will use the following.

Definition 5.4.3. The *degeneracy* of a graph H is defined as the maximum, over all subgraphs $H' \subseteq H$, of the minimum degree of H'. H is said to be *d-degenerate* if its degeneracy is at most d.

Equivalently, H is d-degenerate if its vertices can be ordered as $v_1, \ldots, v_n$ with the property that, for all $i \in [\![n]\!]$, v_i has at most d neighbors which precede it in the ordering, that is, at most d neighbors v_j with $j < i$.

From Lemma 5.2.2, we see that a d-degenerate graph has average degree at most $2d$. On the other hand, the H in Proposition 5.4.2 is an example of a graph with constant average degree and degeneracy $\sqrt{n} - 1$. Thus, we see that having bounded degeneracy is a strictly stronger condition than having bounded average degree. In particular, Proposition 5.4.1 implies that graphs with high degeneracy have large Ramsey numbers, as shown in the following result.

Theorem 5.4.4. *Let H be a graph of degeneracy d. Then $r(H) > 2^{\frac{d-1}{2}}$.*

Proof. By the definition of degeneracy, there is a subgraph $H' \subseteq H$ with minimum degree at least d, and thus also average degree at least d. Then Proposition 5.4.1 implies that

$$r(H) \geqslant r(H') > 2^{\frac{d-1}{2}}.$$

□

Given this, we can now amend our naïve conjecture to the following fundamental conjecture of Burr and Erdős [12].

Conjecture 5.4.5 (Burr–Erdős [12])**.** *Graphs of bounded degeneracy have linear Ramsey numbers.*

More precisely, for every $d \geqslant 1$ there exists $C \geqslant 1$ such that the following holds. If an n-vertex graph H is d-degenerate, then $r(H) \leqslant Cn$.

The Burr–Erdős conjecture has a long history, with many important partial results. The first major breakthrough in this direction was a theorem of Chvatál, Rödl, Szemerédi, and Trotter [17], which established the Burr–Erdős conjecture under the stronger assumption that H has bounded maximum degree.

Theorem 5.4.6 (Chvatál–Rödl–Szemerédi–Trotter [17])**.** *Graphs of bounded maximum degree have linear Ramsey numbers.*

More precisely, for every $\Delta \geqslant 1$, there exists $C \geqslant 1$ such that the following holds. If an n-vertex graph H has maximum degree at most Δ, then $r(H) \leqslant Cn$.

This result was extremely important, and so was the proof technique they introduced; this theorem is the first result in Ramsey theory to be proved via the so-called *regularity method*, whose basis is the fundamental *regularity lemma* of Szemerédi [137]. This method has become one of the most important techniques in Ramsey theory and in extremal graph theory more broadly, and we will discuss it in more depth shortly. For now, let us only remark that this proof technique gives truly enormous bounds on how large C has to be as a function of Δ; namely their proof showed that Theorem 5.4.6 is true for

$$C = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}} \Big\} 2^{100\Delta}.$$

This enormous bound is one of several reasons why many researchers attempted to find alternative proofs of Theorem 5.4.6.

5.4.1 Greedy embedding

There are now (at least) two other techniques known for proving Theorem 5.4.6, both of which are very important in their own right. The first is the *greedy embedding technique*, which was developed in this context by Graham, Rödl, and Ruciński [66, 67], although it goes back in some form at least to work of Erdős and Hajnal [45]. We will unfortunately not have time to discuss this technique in detail in this course, but let us see a high-level overview of how it works.

Proof sketch of Theorem 5.4.6 using greedy embedding. Let H be an n-vertex graph of maximum degree at most Δ, and let $N = Cn$ for a large constant $C = C(\Delta)$ that we choose appropriately. Fix a red/blue coloring of $E(K_N)$. Our goal is to attempt to find a red copy of H in a greedy manner; we'll then show that, if we fail, we will be able to find a blue copy of H.

Let us label the vertices of H as $v_1, \ldots, v_n$. Define $V_1 = V_2 = \cdots = V_n = V(K_N)$. We think of V_i as the set of candidate vertices for v_i, and will attempt to embed the vertices of H one at a time, at each step updating the set of candidate vertices. We fix some small parameter $\varepsilon > 0$.

Note that if we pick where to embed v_i into V_i, we need to update our candidate sets. Indeed, since our goal is to build a red copy of H, if we choose where to place v_i, we need to shrink each V_j, for all j such that $v_iv_j \in E(H)$, to only include the red neighbors of the chosen embedding of v_i. Let us call a vertex $w \in V_i$ *prolific* if it has the following property: if we choose to embed v_i as w, then each candidate set shrinks by at most a factor of ε. In other words, w is prolific if its red neighborhood in V_j has size at least $\varepsilon|V_j|$, for every j such that $v_iv_j \in E(H)$.

Our embedding rule is now as follows. If there is a prolific vertex in V_1, we embed v_1 there and update all the candidate sets appropriately. If there is now a prolific vertex in V_2, we embed v_2 there and update all the candidate sets. We continue in this way as long as we can.

If this process gets to the end, that is, if we embed v_n into V_n, then we have found a red copy of H. So we may assume that the process gets stuck at some step i. Note that every candidate set shrinks at most Δ times, since H has maximum degree at most Δ, and moreover every time it shrinks it does so by at most a factor of ε. Thus, when we get stuck, we still have that $|V_j| \geqslant \varepsilon^\Delta N$ for all j. In particular, $|V_i| \geqslant \varepsilon^\Delta N$. Moreover, since we got stuck, there is no prolific vertex in V_i. That is, for every vertex $w \in V_i$, there is some j such that the red neighborhood of w in V_j has size less than $\varepsilon|V_j|$. There are at most Δ options for this choice of j, so by the pigeonhole principle, there is some fixed $j \in [\![n]\!]$ and some set $W_i \subseteq V_i$ with $|W_i| \geqslant \frac{1}{\Delta}|V_i|$ such that every $w \in W_i$ has a red neighborhood in V_j of size less than $\varepsilon|V_j|$.

We have thus proved the following lemma. If this greedy embedding procedure ever gets stuck, we find two sets W_i, V_j with $|W_i| \geqslant \frac{1}{\Delta}\varepsilon^\Delta N$ and $|V_j| \geqslant \varepsilon^\Delta N$, and with the property that the density of red edges between W_i and V_j is less than ε. In other words, we have found two sets A_1, A_2 with $|A_1|, |A_2| \geqslant \frac{1}{\Delta}\varepsilon^\Delta N$, and such that the density of *blue* edges between A_1 and A_2 is at least $1 - \varepsilon$.

We now iterate this lemma, as follows. Inside A_1, we run the same procedure to attempt to greedily embed H in red. If we succeed, we are done. If we fail, we find two sets $A_{11}, A_{12} \subseteq A_1$ with blue density between them at least $1 - \varepsilon$, where $|A_{11}|, |A_{12}| \geqslant (\frac{1}{\Delta}\varepsilon^\Delta)^2 N$. We also run the same procedure inside A_2 to find two such sets A_{21}, A_{22}. Moreover, since the blue density between A_1 and A_2 was at least $1 - \varepsilon$, we can ensure[3] that the blue density between A_{1i} and A_{2j} is at least $1 - \varepsilon$, for all $i, j \in [\![2]\!]$.

In other words, we've now found *four* sets, each of size at least $(\frac{1}{\Delta}\varepsilon^\Delta)^2 N$, such that the blue density between every pair is at least $1 - \varepsilon$. Continuing in this manner k times, we can find 2^k such sets, each with size at least $(\frac{1}{\Delta}\varepsilon^\Delta)^k N$, and with all pairwise blue densities at least $1 - \varepsilon$. We now do this until $2^k \geqslant \Delta + 1$ (i.e. pick $k = \lceil \log(\Delta + 1) \rceil$), and we thus obtain at least $\Delta + 1$ sets, which we rename $B_1, \ldots, B_{\Delta+1}$.

Since H has maximum degree at most Δ, it is $(\Delta + 1)$-colorable, i.e. it can be partitioned into $\Delta + 1$ independent sets $C_1, \ldots, C_{\Delta+1}$. Note that

$$|B_i| \geqslant \left(\frac{1}{\Delta}\varepsilon^\Delta\right)^k N \geqslant n,$$

where we can ensure the final inequality by picking C sufficiently large as a function of Δ and ε (and thus k, which is itself a function of Δ). Thus, each set B_i is large enough to accommodate embedding C_i. Moreover, one can check that if ε is sufficiently small (e.g. $\varepsilon = \Delta^{-2}$ suffices), then the greedy embedding strategy we tried for red is now *guaranteed* to work in blue. Namely, we greedily embed H in blue, ensuring that all vertices of C_i get embedded into B_i, and updating all candidate sets at every step. The strong density conditions we know about blue imply that we will never get stuck. □

Examining the proof sketch above, we see that it gives a bound of the form $C \leqslant 2^{O(\Delta(\log \Delta)^2)}$. Moreover, in case H is bipartite, the iteration step is unnecessary, and we

[3]There is some subtlety in doing this step correctly; since A_{1i} and A_{2j} are rather small subsets of A_1, A_2, one needs an extra argument to ensure that the blue density remains high when we restrict to them.

can simply take $k = 1$ in the proof above, and thus obtain a bound of $C \leqslant 2^{O(\Delta \log \Delta)}$. In other words, the greedy embedding technique allowed Graham, Rödl, and Ruciński to prove the following more refined version of Theorem 5.4.6.

Theorem 5.4.7 (Graham–Rödl–Ruciński [66, 67]). *There exists an absolute constant $M > 0$ such that the following holds. If H is an n-vertex graph with maximum degree at most Δ, then*

$$r(H) \leqslant 2^{M\Delta(\log \Delta)^2} n.$$

Moreover, if H is bipartite, we have the stronger bound

$$r(H) \leqslant 2^{M\Delta \log \Delta} n.$$

Remarkably, Graham, Rödl, and Ruciński also proved that their upper bound is nearly tight, even for bipartite graphs.

Theorem 5.4.8 (Graham–Rödl–Ruciński [67]). *There exists an absolute constant $c > 0$ such that the following holds. For every $n > \Delta > 1$, there is an n-vertex bipartite graph H with maximum degree Δ which satisfies*

$$r(H) \geqslant 2^{c\Delta} n.$$

We will not prove Theorem 5.4.8 in this course, but let us briefly remark on the technique. The bipartite graph H in the theorem is defined randomly; for example, one can pick it uniformly at random among all bipartite n-vertex Δ-regular graphs. One then wants to show that, with high probability over the random choice of H, it satisfies $r(H) \geqslant 2^{c\Delta} n$. To do this, one needs to exhibit a coloring on $N = 2^{c\Delta} n$ vertices with no monochromatic copy of H. This is done as follows. First, for an appropriate constant $a > c > 0$, one picks a uniformly random red/blue coloring χ of $E(K_A)$, where $A = 2^{a\Delta}$. One then "blows up" χ to a coloring of $E(K_N)$ as follows. We partition $V(K_N)$ into A parts $V_1, \ldots, V_A$, each of size $N/A = 2^{(c-a)\Delta} n$. We then color all edges between parts V_i and V_j according to the color $\chi(v_i, v_j)$, where $v_i, v_j \in V(K_A)$. Finally, all edges within a part V_i are colored red. Note that since $a > c$, each part V_i has size $N/A = 2^{(c-a)\Delta} n$, which is much smaller than n (assuming a and c are chosen appropriately). Thus, any monochromatic copy of H must use vertices from many different parts V_i. However, since the coloring χ is random, one can show that with high probability, any large collection of parts—or equivalently, any large collection of vertices of K_A—includes many edges of both colors. Moreover, since H is random, one can show that its edges are extremely well-distributed. Because of this, one can show that any potential embedding of H into K_N cannot entirely avoid one of the two colors.

Looking back at the greedy embedding proof sketch above, one might be struck by the fact that the colors play such asymmetrical roles; we keep trying, insistently, to embed H in red, and only when we have failed many times do we relent and succeed in embedding it in blue. This asymmetry is in fact a weakness of the proof technique, and Conlon, Fox, and Sudakov [26] were able to improve the bound of Theorem 5.4.7 to $r(H) \leqslant 2^{O(\Delta \log \Delta)} n$ for every n-vertex graph H with maximum degree Δ, by modifying the greedy embedding technique so that both colors play roughly the same role. Unfortunately, it is still not known if this technique can be used to remove the final logarithmic factor, and thus match the lower bound of Theorem 5.4.8.

Moreover, this discussion hints at another, more fundamental, weakness of the greedy embedding technique, which is that it is tailor-made for the two-color case. Indeed, the entire upshot of the technique is that *failing* to find H in red tells us something about the blue edges. In case there are three or more colors, it is not at all clear how to obtain useful information from the failure of the first embedding. As far as I am aware, no one has been able to use the greedy embedding technique to prove any results on $r(H;q)$ for any H and any $q \geqslant 3$.

5.4.2 Dependent random choice

An extremely powerful technique for proving results like Theorem 5.4.6 was introduced about two decades ago, and is called *dependent random choice*. This technique is very flexible, but at a high level, it allows one to find, in any reasonably dense graph, a large "rich" set of vertices. Here, "rich" means that most, or all, of the r-tuples of vertices in the set have many common neighbors. A survey on dependent random choice and its many applications can be found in [57], and a much less in-depth, but hopefully gentler, introduction can be found in [143]. The development of the dependent random choice technique led to a number of results building towards the Burr–Erdős conjecture. Eventually, this culminated in a breakthrough result of Lee [86], who completely resolved Conjecture 5.4.5.

Theorem 5.4.9 (Lee [86])**.** *For every $d \geqslant 1$, there exists a constant $C \leqslant 2^{2^{1000d}}$ such that the following holds. If H is an n-vertex, d-degenerate graph, then $r(H) \leqslant Cn$.*

Theorem 5.4.9 is too complex to cover in this course. However, we will see a relatively simple application of the dependent random choice technique, which will indicate how it works and why it is useful for proving such theorems. The result we will prove is essentially implicit in two of the earliest applications of dependent random choice to Ramsey-theoretic problems, due to Kostochka–Rödl [81] and Kostochka–Sudakov [80].

Theorem 5.4.10. *For every $\Delta \geqslant 1$, the following holds for all sufficiently large n. Let H be a bipartite graph with bipartition $A \cup B$, where $|A| = |B| = n$. If every vertex of B has degree at most Δ, then $r(H) \leqslant n2^{5\Delta\sqrt{\log n}}$.*

Note that this theorem is weaker than Theorem 5.4.6, in that it only applies for bipartite graphs, and rather than giving a linear bound, proves only that $r(H) \leqslant n2^{5\Delta\sqrt{\log n}} = n^{1+o(1)}$. On the other hand, it is stronger than Theorem 5.4.6 because it only assumes that vertices in B have bounded degree, whereas the vertices in A may have arbitrarily large degree.

Proof. Let $N = n2^{5\Delta\sqrt{\log n}}$, and consider an arbitrary red/blue coloring of $E(K_N)$. We may assume without loss of generality that at least half the edges are red. Let $t = \sqrt{\log n}$, where the logarithm is to base 2. Let $u_1, \ldots, u_t$ be t uniformly random vertices of K_N, chosen with repetition (that is, each u_i is a uniformly random vertices, and the choices are indpendent over all i). Let S be the common red neighborhood of $u_1, \dots, u_t$.

For a given vertex $v \in V(K_N)$, we have that $v \in S$ if and only if $u_1, \dots, u_t$ are all red neighbors of v. The probability that this happens is therefore exactly $(\deg_R(v)/N)^t$, where $\deg_R(v)$ denotes the red degree of v. Therefore, by linearity of expectation,

$$\mathbb{E}[|S|] = \sum_{v \in V(K_N)} \Pr(v \in S) = \sum_{v \in V(K_N)} \left(\frac{\deg_R(v)}{N} \right)^t.$$

Let $d \coloneqq \frac{1}{N} \sum_{v \in V(K_N)} \deg_R(v)$ denote the average red degree, and note that $d \geqslant (N-1)/2$ since we assumed that at least half the edges are red. By Jensen's inequality, applied to the convex function $x \mapsto x^t$, we have that

$$\mathbb{E}[|S|] = \sum_{v \in V(K_N)} \left(\frac{\deg_R(v)}{N} \right)^t \geqslant N \left(\frac{d}{N} \right)^t \geqslant N \left(\frac{N-1}{2N} \right)^t \geqslant \frac{N}{3^t},$$

where the final inequality uses that $N - 1 \geqslant \frac{2}{3}N$ for all $N \geqslant 3$, and we assumed that n (and thus N) is sufficiently large.

Let us call a set $\{v_1, \dots, v_\Delta\}$ of vertices of K_N *unfriendly* if they have fewer than $2n$ common red neighbors. The total number of unfriendly sets in K_N is at most $\binom{N}{\Delta}$. Moreover, for a given unfriendly set $\{v_1, \dots, v_\Delta\}$, the probability that it is contained in S is at most $(2n/N)^t$. Indeed, for this unfriendly set to be contained in S, we must have that all the random vertices $u_1, \dots, u_t$ lie in the common red neighborhood of $v_1, \dots, v_\Delta$, and there are at most $2n$ such common red neighbors by the definition of unfriendliness. Therefore, if we let Z denote the number of unfriendly sets in S, we find by linearity of expectation that

$$\mathbb{E}[Z] \leqslant \binom{N}{\Delta} \left(\frac{2n}{N} \right)^t \leqslant (2n)^\Delta \left(\frac{2n}{N} \right)^{t-\Delta}.$$

Recall that $t = \sqrt{\log n}$. If n is sufficiently large in terms of Δ, then $t \geqslant 2\Delta$, and thus $t - \Delta \geqslant \frac{t}{2}$. Thus,

$$\mathbb{E}[Z] \leqslant (2n)^\Delta \left(\frac{2n}{N} \right)^{t-\Delta} \leqslant n^{2\Delta} \left(\frac{2n}{N} \right)^{t/2}.$$

Now note that by our choice of N, we have $2n/N = 2 \cdot 2^{-5\Delta\sqrt{\log n}} \leqslant 2^{-4\Delta\sqrt{\log n}}$, again for sufficiently large n. Thus,

$$\mathbb{E}[Z] \leqslant n^{2\Delta} \left(\frac{2n}{N} \right)^{t/2} \leqslant n^{2\Delta} \left(2^{-4\Delta\sqrt{\log n}} \right)^{\frac{1}{2}\sqrt{\log n}} = n^{2\Delta} \cdot 2^{-2\Delta \log n} = n^{2\Delta} \cdot n^{-2\Delta} = 1.$$

Again by linearity of expectation, we conclude that

$$\mathbb{E}[|S| - Z] = \mathbb{E}[|S|] - \mathbb{E}[Z] \geqslant \frac{N}{3^t} - 1 \geqslant \frac{N}{4^t}.$$

Therefore, in the random experiment, there must exist some outcome of $u_1, \ldots, u_t$ such that the corresponding quantities $|S|$ and Z satisfy $|S| - Z \geqslant N/4^t$. Fix such an outcome, and let S be their common red neighborhood. Define T by deleting one vertex from every unfriendly set in S. Then $|T| = |S| - Z \geqslant N/4^t$. Moreover, since we deleted one vertex from every unfriendly set, we see that T has no unfriendly sets.

In other words, we've found a set T of size $|T| \geqslant N/4^t$, in which every set of Δ vertices is friendly. All that remains in the proof is to use this "rich set" to find a monochromatic copy of H. Recall that H is a bipartite graph on parts $A \cup B$, each of size n. Note that

$$|T| \geqslant \frac{N}{4^t} = \frac{n 2^{5\Delta\sqrt{\log n}}}{2^{2\sqrt{\log n}}} \geqslant n = |A|.$$

Let us arbitrarily embed A into T; that is, if we write $A = \{a_1, \ldots, a_n\}$, then we pick arbitrary $x_1, \ldots, x_n \in T$, and we will find a monochromatic copy of H in which x_i plays the role of a_i.

To do this, all that remains is embed the vertices of B, which we call $b_1, \ldots, b_n$. We embed them one by one, in this order. For a given vertex b_i, we know that it has at most Δ neighbors $a_{i_1}, \ldots, a_{i_\Delta}$ in A. By construction, the set $\{x_{i_1}, \ldots, x_{i_\Delta}\} \subseteq T$ is friendly, and thus these vertices have at least $2n$ common red neighbors. Fewer than $2n$ of these have been used in embedding $a_1, \ldots, a_n, b_1, \ldots, b_{i-1}$, so at least one remains available for embedding b_i. We pick one of these arbitrarily. At the end of the process, we have constructed a monochromatic red copy of H. $\square$

One should note that in this proof, we never used any property of the coloring besides that one of the color classes is dense, that is, has at least half of the edges. Thus, this proof naturally extends to an arbitrary number of colors, and moreover can be used to prove results like Theorem 5.3.2, showing that certain bipartite graphs appear in any graph with sufficiently many edges. We also record here a general form of the lemma we used implicitly in the proof of Theorem 5.4.10, which can be used to deduce such more general results.

Lemma 5.4.11 (Dependent random choice lemma)**.** *Let G be an N-vertex graph with average degree d. Let t, Δ, r, s be integers satisfying the inequality*

$$\frac{d^t}{N^{t-1}} - \binom{N}{\Delta}\left(\frac{r}{N}\right)^t \geqslant s.$$

Then there is a set $T \subseteq V(G)$ with $|T| \geqslant s$ such that every set of Δ vertices in T has at least r common neighbors.

In the proof above, we implicitly applied this lemma with $t = \sqrt{\log n}, \Delta = \Delta, r = 2n$, and $s = N/4^t$. The general case is proved in exactly the same way, by selecting t random vertices $u_1, \ldots, u_t$ and considering their common neighborhood; the details are left for the homework.

Using the same ideas, but with one extra trick, Kostochka and Sudakov [80] were able to prove that if H is a d-degenerate bipartite graph on n vertices, then $r(H) \leqslant n 2^{C(\log n)^{2/3}}$, where $C = C(d)$ is a constant depending only on d. By combining this with yet one more idea, they were also able to remove the bipartiteness assumption, and thus prove that $r(H) \leqslant n^{1+o(1)}$ for any d-degenerate n-vertex graph H, where the $o(1)$ term tends to 0 as $n \to \infty$ (while d stays fixed). Thus, this result "almost" confirms the Burr–Erdős conjecture, Conjecture 5.4.5.

After these early results of Kostochka–Rödl [81] and Kostochka–Sudakov [80], there were a series of improvements on the dependent random choice technique, culminating in Lee's theorem [86], Theorem 5.4.9, proving the Burr–Erdős conjecture. We will not state or prove most of these intermediate results, but let us mention just one, due to Conlon, Fox, and Sudakov [28].

Theorem 5.4.12 (Conlon–Fox–Sudakov [28])**.** *If H is an n-vertex bipartite graph with maximum degree Δ, then*

$$r(H) \leqslant 2^{\Delta+6} n.$$

Note that this bound is very close to best possible, thanks to Theorem 5.4.8 discussed above.

Before ending this section, let us briefly discuss two important open problems related to Theorems 5.4.7–5.4.9 and 5.4.12. Recall that for an n-vertex graph H with degeneracy d, we have that $r(H) \geqslant 2^{\frac{d-1}{2}}$ by Theorem 5.4.4, and that $r(H) \geqslant n$ since this is true for all n-vertex graphs. Therefore,

$$r(H) \geqslant \max\left\{n, 2^{\frac{d-1}{2}}\right\} \geqslant \sqrt{n \cdot 2^{\frac{d-1}{2}}},$$

and thus

$$\log r(H) = \Omega(d + \log n).$$

On the other hand, Theorem 5.4.12 implies that if H is bipartite with maximum degree (and thus degeneracy) at most d, then $r(H) \leqslant 2^{d+6} n$, implying that in this case

$$\log r(H) = O(d + \log n).$$

Conlon, Fox, and Sudakov [27] conjectured that in fact, such an upper bound holds in general.

Conjecture 5.4.13 (Conlon–Fox–Sudakov [27])**.** *If H is an n-vertex graph of degeneracy d, then*

$$\log r(H) = \Theta(d + \log n).$$

In words, this conjecture says that in order to understand the approximate order of $r(H)$, the structure of H is almost completely irrelevant: all that matters is its degeneracy and its number of vertices. Conjecture 5.4.13 remains open in general, but is known to be true up to logarithmic factors [56, Theorem 3.1].

Finally, let us briefly discuss the *hypercube* graph, Q_d. This is the graph with vertex set $\{0, 1\}^d$, in which two vertices are adjacent if and only if they differ in a single coordinate. Note that Q_d is a d-regular bipartite graph, and thus has degeneracy d; moreover, it has

$n = 2^d$ vertices, so its degeneracy is logarithmic in its vertex count. This is precisely the interesting regime where the two terms in Conjecture 5.4.13 are of the same order. By Theorem 5.4.12, we know that

$$r(Q_d) \leqslant 2^{d+6}n = 64n^2,$$

and thus $r(Q_d)$ is at most quadratic in its number of vertices. However, Burr and Erdős [12] conjectured that the answer is in fact linear, that is, that $r(Q_d) = O(n)$. This important conjecture remains open despite a great deal of effort. However, very recently, Tikhomirov [139] improved the bound coming from Theorem 5.4.12, and showed that

$$r(Q_d) \leqslant n^{1.97}$$

for all sufficiently large d. His proof also uses dependent random choice, as well as an intricate embedding scheme that is tailor-made for the hypercube (and is thus able to improve upon the general-purpose result in Theorem 5.4.12).

Chapter 6

The regularity method

In this section, we will develop the regularity method, which is one of the most powerful techniques in modern graph theory. This technique has its origins in work of Szemerédi [135, 136], who used it as a crucial ingredient in his proof of his eponymous theorem in arithmetic progressions in dense sets of integers. In Ramsey theory, it was first applied by Chvatál, Rödl, Szemerédi, and Trotter [17] in their proof of Theorem 5.4.6; indeed, a full proof of that theorem will be our first main application of the regularity method.

Warning! The "regularity method", as I will present it here, differs in crucial ways from the standard way it is presented. In particular, we will not state or prove Szemerédi's regularity lemma, but instead a weaker result that is still sufficient for many applications. For an excellent introduction to the "real" regularity method, see e.g. [145, Chapter 2].

6.1 Definitions and key lemmas

Definition 6.1.1. If G is an N-vertex graph, we define its *edge density* to be

$$d(G) \coloneqq \frac{e(G)}{\binom{N}{2}}.$$

In words, the edge density is the fraction of all possible edges that are present in G.

If $S \subseteq V(G)$, we denote by $d(S)$ the edge density of the induced subgraph on S, or equivalently define

$$d(S) \coloneqq \frac{e(S)}{\binom{|S|}{2}}.$$

Definition 6.1.2. Let $\varepsilon > 0$. An N-vertex graph G is called ε-*quasirandom* if, for every $S \subseteq V(G)$ with $|S| \geqslant \varepsilon N$, we have that

$$|d(S) - d(G)| \leqslant \varepsilon.$$

In words, G is ε-quasirandom if every large vertex subset has roughly the same edge density as the whole graph.

The reason for the name *quasirandom* is that it is not hard to show that random graphs are ε-quasirandom. Indeed, if N is sufficiently large with respect to ε, and we form a random N-vertex graph by making each pair an edge independently with probability $p \in [0,1]$, then one can show that with high probability, every subset of size at least εN has edge density $p \pm \varepsilon$.

One important property of large random graphs is that they contain all "small" graphs as subgraphs. This property extends to quasirandom graphs, which is the content of the next lemma. This property is very useful for Ramsey-theoretic applications, as we can often use it to guarantee the existence of monochromatic copies of some graph H in certain colorings.

The next lemma is often called the *embedding lemma*, and also traces back to the work of Chvatál–Rödl–Szemerédi–Trotter [17], although in some form it goes back at least to earlier work of Ruzsa–Szemerédi [118].

Lemma 6.1.3 (Embedding lemma)**.** *Let H be an n-vertex graph with maximum degree $\Delta \geqslant 1$. Let $0 < \varepsilon < \frac{1}{2\Delta}$ be a real number, and let $N \geqslant \frac{2n}{\varepsilon}$ be an integer.*

Let G be an N-vertex graph with edge density $d(G) \geqslant (2\Delta\varepsilon)^{1/\Delta}$, and suppose that G is ε-quasirandom. Then H is a subgraph of G.

The proof of Lemma 6.1.3 follows the exact same strategy as the greedy embedding argument sketched in Section 5.4.1. Namely, one attempts to greedily embed H into G, one vertex at a time. The ε-quasirandomness (as well as the lower bound on $d(G)$) then guarantees that in each set of candidate vertices most vertices are prolific, guaranteeing that each step of the embedding process can proceed. Eventually the process terminates, and produces a copy of H. A detailed proof can be found[1] in [145, Theorem 2.6.4].

Thus, we see that ε-quasirandom graphs are very special, and they seem to be very useful for finding copies of subgraphs H. For example, if we are given a coloring of $E(K_N)$ and are promised that the red graph (say) is ε-quasirandom, then we might hope to find a monochromatic red copy of some H. But of course, in Ramsey-theoretic settings, we are dealing with *arbitrary* colorings of K_N, so there is no reason to expect one of the two colors to form an ε-quasirandom graph.

Remarkably, it turns out that we essentially *can* reduce to this case. As we will shortly see, every graph contains a large, ε-quasirandom induced subgraph. The powerful *regularity lemma* of Szemerédi [137], which we won't state, says something even stronger, namely that any graph can be partitioned into quasirandom pieces. We will make do with a weaker statement, which was first proved by Conlon and Fox [25].

Lemma 6.1.4 (Quasirandom subset lemma; Conlon–Fox [25])**.** *For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. If G is an n-vertex graph, then there exists $Q \subseteq V(G)$ with $|Q| \geqslant \delta n$ such that the induced subgraph $G[Q]$ is ε-quasirandom.*

Moreover, if $\varepsilon < \frac{1}{6}$, we may take $\delta = 2^{-2^{\varepsilon^{-7}}}$.

[1] Actually, [145, Theorem 2.6.4] proves a substantially more general result, but it is not too hard to check that it implies Lemma 6.1.3. A key observation is that ε-quasirandomness, as defined in Definition 6.1.2, implies (2ε)-regularity, as defined in [145, Definition 2.1.2].

Note that Lemma 6.1.4 can itself be viewed as a Ramsey-theoretic statement: it says that, although the graph G may be very complicated globally, it has a large induced subgraph that is extremely well-behaved, in the sense of being ε-quasirandom.

An important word of warning about applications of Lemma 6.1.4. In the embedding lemma, and in all other applications of the regularity method, one needs not only that the graph is ε-quasirandom, but also that its density is not too small. Indeed, if G and all its subgraphs have edge density smaller than ε, then G is ε-quasirandom, but we learn very little about G. In particular, if G is very sparse, then Lemma 6.1.4 is vacuously true—we may take $Q = V(G)$—but it is not useful. As we will see, however, in Ramsey-theoretic contexts one can usually handle this issue by picking one of the two colors intelligently. To do this, we will often apply the following simple lemma, which implies that in a 2-coloring of $E(K_N)$, sets that are quasirandom in one color are quasirandom in the other color as well.

Lemma 6.1.5. *Suppose we are given a red/blue coloring of $E(K_N)$, and denote by G_R, G_B the graphs of red and blue edges, respectively. If $Q \subseteq V(K_N)$ is such that $G_R[Q]$ is ε-quasirandom, then $G_B[Q]$ is ε-quasirandom as well.*

Proof. Denoting by d_R, d_B the edge densities in G_R, G_B, respectively, we note that for any set S, we have $d_R(S) = 1 - d_B(S)$, since every edge in S is either red or blue. The fact that $G_R[Q]$ is ε-quasirandom means that $|d_R(S) - d_R(Q)| \leqslant \varepsilon$ for every $S \subseteq Q$ with $|S| \geqslant \varepsilon|Q|$. But by the above, we have that $|d_R(S) - d_R(Q)| = |d_B(S) - d_B(Q)|$, hence $G_B[Q]$ is also ε-quasirandom. □

Proof of Lemma 6.1.4

The original proof of Conlon–Fox [25] relied on an auxiliary result, the *cylinder regularity lemma* of Duke, Lefmann, and Rödl [35], as well as an application of Ramsey's theorem. We will see an alternative proof, due to Fox (private communication), which is more elementary (although not necessarily simpler or shorter).

We will use the following structural lemma about graphs that are not ε-quasirandom. If A, B are two disjoint sets of vertices in a graph G, then we denote by $e(A, B)$ the number of edges between them, and by $d(A, B) := e(A, B)/(|A||B|)$ their edge density.

Lemma 6.1.6. *Fix $\varepsilon \in (0, \frac{1}{6})$. Let G be a graph with $d(G) = \alpha$ which is not ε-quasirandom. Either there exists a subset $A \subseteq V(G)$ with $|A| \geqslant \varepsilon|V(G)|$ and $d(A) \geqslant \alpha + \varepsilon^4$, or else there exist disjoint $A, B \subseteq V(G)$ with $|A|, |B| \geqslant \varepsilon|V(G)|$ and $d(A, B) \geqslant \alpha + \varepsilon^3$.*

Proof. It is a simple exercise to show that if $d(G) \geqslant 1 - \varepsilon^4$, then G is automatically ε-quasirandom, so we may assume that $\alpha + \varepsilon^4 \leqslant 1$. We thus see that if $|V(G)| \leqslant \frac{2}{\varepsilon}$ then we may take A to be the endpoints of any edge and obtain the claimed result; thus we assume henceforth that $|V(G)| \geqslant \frac{2}{\varepsilon}$.

By the definition of ε-quasirandomness, there exists $S_0 \subseteq V(G)$ with $|S_0| \geqslant \varepsilon|V(G)|$ and $d(S_0) \leqslant \alpha - \varepsilon$ or $d(S_0) \geqslant \alpha + \varepsilon$. In the latter case we may set $A = S_0$ and obtain the claimed result, so let us assume that $d(S_0) \leqslant \alpha - \varepsilon$.

Let S be a random subset of S_0, chosen uniformly at random among all subsets of size exactly $\varepsilon|V(G)|$. All edges of S_0 are included in S with equal probability, so the expected edge density of S equals $d(S_0)$. Thus, there exists some (deterministic) choice of $S \subseteq S_0$ with $|S| = \varepsilon|V(G)|$ and $d(S) \leqslant d(S_0) \leqslant \alpha - \varepsilon$. Let $T = V(G) \setminus S$, and note that $|T| = (1-\varepsilon)|V(G)| \geqslant \varepsilon|V(G)|$ since we assumed $\varepsilon < \frac{1}{2}$.

If $d(T) \geqslant \alpha + \varepsilon^4$, we may set $A = T$ and obtain the claimed result, so we may assume that $d(T) \leqslant \alpha + \varepsilon^4$. We now observe that

$$\begin{aligned}\alpha\binom{V(G)}{2} = e(G) = e(S) + e(T) + e(S,T) \\ &\leqslant (\alpha - \varepsilon)\binom{|S|}{2} + (\alpha + \varepsilon^4)\binom{|T|}{2} + d(S,T)|S||T| \\ &= \alpha\left[\binom{|S|}{2} + \binom{|T|}{2} + |S||T|\right] - \varepsilon\binom{|S|}{2} + \varepsilon^4\binom{|T|}{2} + (d(S,T) - \alpha)|S||T| \\ &= \alpha\binom{V(G)}{2} - \varepsilon\binom{|S|}{2} + \varepsilon^4\binom{|T|}{2} + (d(S,T) - \alpha)|S||T|,\end{aligned}$$

where the final step uses that $S \sqcup T$ partitions $V(G)$, so $\binom{|V(G)|}{2} = \binom{|S|}{2} + \binom{|T|}{2} + |S||T|$. Subtracting $\alpha\binom{|V(G)|}{2}$ from both sides, we learn that

$$(d(S,T) - \alpha)|S||T| \geqslant \varepsilon\binom{|S|}{2} - \varepsilon^4\binom{|T|}{2}$$

which implies

$$d(S,T) - \alpha \geqslant \varepsilon\frac{|S|-1}{2|T|} - \varepsilon^4\frac{|T|-1}{2|S|}.$$

We have that $|S| \geqslant 2$ since we assumed $|V(G)| \geqslant \frac{2}{\varepsilon}$. Thus $|S| - 1 \geqslant \frac{1}{2}|S|$, and therefore

$$\frac{|S|-1}{2|T|} \geqslant \frac{|S|}{4|T|} = \frac{\varepsilon}{4(1-\varepsilon)} \geqslant \frac{\varepsilon}{4}.$$

Additionally,

$$\frac{|T|-1}{2|S|} \leqslant \frac{|T|}{2|S|} = \frac{1-\varepsilon}{2\varepsilon} \leqslant \frac{1}{2\varepsilon}.$$

Putting this together, we find that

$$d(S,T) - \alpha \geqslant \frac{\varepsilon^2}{4} - \frac{\varepsilon^3}{2} = \varepsilon^2\left(\frac{1}{4} - \frac{\varepsilon}{2}\right) \geqslant \varepsilon^2\left(\frac{1}{4} - \frac{1}{12}\right) = \frac{\varepsilon^2}{6} \geqslant \varepsilon^3,$$

where the final two inequalities use our assumption that $\varepsilon \leqslant \frac{1}{6}$. Setting $A = S, B = T$ concludes the proof. $\square$

With this lemma in hand, we are ready for the proof of Lemma 6.1.4. We will actually prove a stronger statement, which lends itself naturally to a *density increment* argument.

Lemma 6.1.7. *For every $\alpha \in [0,1], \varepsilon \in (0, \frac{1}{6})$ there exists some $\gamma(\alpha, \varepsilon) > 0$ such that the following holds. For every graph G, there exists a subset $S \subseteq V(G)$ with $|S| \geqslant \gamma(\alpha,\varepsilon)|V(G)|$ such that $G[S]$ is ε-quasirandom, or $d(S) \geqslant \alpha$.*

Moreover, we may take

$$\gamma(\alpha, \varepsilon) = 2^{-2^{\alpha/\varepsilon^7}}.$$

Before proving this, let's see how it immediately implies Lemma 6.1.4.

Proof of Lemma 6.1.4. Note that a set that is ε-quasirandom is also ε'-quasirandom for any $\varepsilon' > \varepsilon$, so it suffices to prove this statement for $\varepsilon < \frac{1}{6}$. Apply Lemma 6.1.7 with $\alpha = 1$, and let $\delta = \gamma(1, \varepsilon)$. Note that the claimed bound on δ follows from the bound on γ in Lemma 6.1.7. By Lemma 6.1.7, we know that for any graph G, there exists $S \subseteq V(G)$ with $|S| \geqslant \delta|V(G)|$, such that $G[S]$ is ε-quasirandom or $d(S) \geqslant 1$. In the former case, we may set $Q = S$ and are done. In the latter case, as $d(S) = 1$, we see that S defines a complete subgraph of G. Every subgraph of $G[S]$ is thus also complete, and hence S is ε-quasirandom, so we may again set $Q = S$ and conclude the proof. □

All that remains now is the proof of Lemma 6.1.7.

Proof of Lemma 6.1.7. We fix $\varepsilon \in (0, \frac{1}{6})$. Our proof will be by "induction" on α, except that of course induction doesn't make sense since α is a real parameter. Nonetheless, it is not hard to make this make sense. Note that the statement we are trying to prove is monotone in α, in the sense that if we prove the existence of $\gamma(\alpha, \varepsilon)$, we also prove the existence of $\gamma(\alpha', \varepsilon)$ for any $\alpha' < \alpha$. We will show that the statement for the pair (α, ε) implies the statement for the pair $(\alpha + \varepsilon^6, \varepsilon)$, which then also yields the result for all $\alpha' \in [\alpha, \alpha + \varepsilon^6]$ by the monotonicity discussed above.

To begin the induction, note that we may take $\gamma(0, \varepsilon) = \varepsilon^6$. Indeed, letting S be an arbitrary subset of $V(G)$ of size $\varepsilon^6|V(G)|$ shows the existence of the desired subset of edge density at least 0, since any set has edge density at least 0. Note that we could have set $\gamma(0, \varepsilon)$ to be any number in $(0, 1]$, but this choice will be useful for simplifying some later computations.

Now suppose that we have proved the existence of $\gamma(\alpha, \varepsilon)$. Let G be a graph; we wish to prove the existence of a large set $S \subseteq V(G)$ which is either ε-quasirandom or else has edge density at least $\alpha + \varepsilon^6$. By the definition of $\gamma(\alpha, \varepsilon)$, we may find a set $S_0 \subseteq V(G)$ with $|S_0| \geqslant \gamma(\alpha, \varepsilon)|V(G)|$ such that $G[S_0]$ is ε-quasirandom or $d(S_0) \geqslant \alpha$. If the former case happens we are done, as long as we ensure that $\gamma(\alpha + \varepsilon^6, \varepsilon) \leqslant \gamma(\alpha, \varepsilon)$. So let us assume that $d(S_0) \geqslant \alpha$, and that $G[S_0]$ is *not* ε-quasirandom.

We now apply Lemma 6.1.6 to $G[S_0]$. If we find a set $A \subseteq S_0$ with $|A| \geqslant \varepsilon|S_0| \geqslant \varepsilon \cdot \gamma(\alpha, \varepsilon)|V(G)|$ and $d(A) \geqslant d(S_0) + \varepsilon^4 \geqslant \alpha + \varepsilon^6$, we are done, so long as we ensure that

$$\gamma(\alpha + \varepsilon^6, \varepsilon) \leqslant \varepsilon \cdot \gamma(\alpha, \varepsilon).$$

So we may assume that we are in the second case of Lemma 6.1.6, that is, that there exist $A, B \subseteq S_0$ with $|A|, |B| \geqslant \varepsilon|S_0|$ and $d(A, B) \geqslant \alpha + \varepsilon^3$.

Let $A_1 \subseteq A$ comprise all vertices in A with at least $(\alpha + \varepsilon^4)|B|$ neighbors in B. Note that

$$\begin{aligned}(\alpha + \varepsilon^3)|A||B| &\leqslant d(A, B)|A||B| \\ &= e(A, B) \\ &= e(A_1, B) + e(A \setminus A_1, B) \\ &\leqslant |A_1||B| + (\alpha + \varepsilon^4)|A \setminus A_1||B| \\ &\leqslant |A_1||B| + (\alpha + \varepsilon^4)|A||B|.\end{aligned}$$

Rearranging, we find that

$$|A_1| \geqslant (\varepsilon^3 - \varepsilon^4)|A| \geqslant \varepsilon^4|A| \geqslant \varepsilon^5|S_0| \geqslant \varepsilon^5 \cdot \gamma(\alpha, \varepsilon)|V(G)|. \tag{6.1}$$

We now again apply the definition of $\gamma(\alpha, \varepsilon)$, now to the induced subgraph $G[A_1]$. We find a subset $X \subseteq A_1$, with $|X| \geqslant \gamma(\alpha, \varepsilon)|A_1| \geqslant \varepsilon^5 \cdot \gamma(\alpha, \varepsilon)^2|V(G)|$, such that $G[X]$ is ε-quasirandom or $d(X) \geqslant \alpha$. We are done in the former case, as long as we ensure that

$$\gamma(\alpha + \varepsilon^6, \varepsilon) \geqslant \varepsilon^5 \cdot \gamma(\alpha, \varepsilon)^2.$$

So let us assume instead that $d(X) \geqslant \alpha$. Recall that since $X \subseteq A_1$, every vertex in X has at least $(\alpha+\varepsilon^4)|B|$ neighbors in B, and thus $d(X, B) \geqslant \alpha+\varepsilon^4$. As in the proof of Lemma 6.1.6, we may pass to a random subset of X to ensure that $d(X) \geqslant \alpha$ and that $|X| = \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)^2|V(G)|$. We now let B_1 comprise all vertices in B with at least $(\alpha + \varepsilon^5)|X|$ neighbors in X. Essentially the same argument that proved (6.1) shows that

$$|B_1| \geqslant (\varepsilon^4 - \varepsilon^5)|B| \geqslant \varepsilon^5|B| \geqslant \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)|V(G)|.$$

We now apply the definition of $\gamma(\alpha, \varepsilon)$ to $G[B_1]$ to obtain a set Y with $|Y| \geqslant \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)^2|V(G)|$, such that $G[Y]$ is ε-quasirandom or $d(Y) \geqslant \alpha$. In the former case we are done if we ensure that

$$\gamma(\alpha + \varepsilon^6, \varepsilon) \geqslant \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)^2,$$

so let us assume that the latter case holds. We may again assume by passing to a random subset that $|Y| = \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)^2|V(G)|$. Since $Y \subseteq B_1$, we know that every vertex in Y has at least $(\alpha + \varepsilon^5)|X|$ neighbors in X, hence $d(X, Y) \geqslant \alpha + \varepsilon^5$. Moreover, $d(X), d(Y) \geqslant \alpha$, and $|X| = |Y|$ by construction. Thus,

$$e(X\cup Y) = e(X)+e(Y)+e(X, Y) \geqslant \alpha\binom{|X|}{2}+\alpha\binom{|Y|}{2}+(\alpha+\varepsilon^5)|X||Y| \geqslant \left(\alpha + \frac{\varepsilon}{2}\right)\binom{|X \cup Y|}{2},$$

which implies that

$$d(X \cup Y) \geqslant \alpha + \frac{\varepsilon^5}{2} \geqslant \alpha + \varepsilon^6.$$

Moreover, $|X \cup Y| \geqslant |X| = \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)^2|V(G)|$. Comparing all the restrictions we placed on $\gamma(\alpha + \varepsilon^6, \alpha)$, we see that we can define

$$\gamma(\alpha + \varepsilon^6, \varepsilon) = \varepsilon^6 \cdot \gamma(\alpha, \varepsilon)^2$$

and have the desired property continue inductively. Note that this definition implies that γ is monotonically decreasing in α (for fixed ε), hence our choice of $\gamma(0,\varepsilon) = \varepsilon^6$ implies that $\gamma(\alpha,\varepsilon) \leqslant \varepsilon^6$ for all α. Thus, we have that

$$\gamma(\alpha + \varepsilon^6, \varepsilon) = \varepsilon^6 \gamma(\alpha,\varepsilon)^2 \geqslant \gamma(\alpha,\varepsilon)^3.$$

Applying this bound iteratively, subtracting ε^6 at every step, we find that for every α,

$$\gamma(\alpha,\varepsilon) \geqslant \gamma(0,\varepsilon)^{3^{\alpha/\varepsilon^6}} = (\varepsilon^6)^{3^{\alpha/\varepsilon^6}} \geqslant 2^{-2^{\alpha/\varepsilon^7}},$$

as claimed. □

6.2 Application I: Proof of Theorem 5.4.6

Let us now see how these powerful tools—the embedding lemma and the quasirandom subset lemma—can be used to give a short proof of Theorem 5.4.6.

Proof of Theorem 5.4.6. Note that any graph of maximum degree at most 1 is a forest, so the $\Delta = 1$ case follows from Theorem 5.2.1. Hence we may assume that $\Delta \geqslant 2$.

First we pick some parameters depending on Δ. Let $\varepsilon = 2^{-\Delta}/(2\Delta)$, which is chosen so that $\frac{1}{2} = (2\Delta\varepsilon)^{1/\Delta}$; note that $\varepsilon < \frac{1}{6}$ since $\Delta \geqslant 2$. Let $\delta = 2^{-2^{\varepsilon^{-7}}}$ be the constant from Lemma 6.1.4. Finally, let $C = 2/(\varepsilon\delta)$, and note that C depends only on Δ.

Fix an n-vertex graph H with maximum degree at most Δ, and let $N = Cn$. Consider a red/blue coloring of $E(K_N)$, and let G_R, G_B be the red and blue graphs, respectively. Applying Lemma 6.1.4 to G_R (and recalling Lemma 6.1.5), we find a subset $Q \subseteq V(K_N)$ with $|Q| \geqslant \delta N$ such that $G_R[Q]$ and $G_B[Q]$ are both ε-quasirandom. Assume without loss of generality that at least half the edges in Q are red, so that $d(G_R[Q]) \geqslant \frac{1}{2} = (2\Delta\varepsilon)^{1/\Delta}$. Note that

$$|Q| \geqslant \delta N = \delta C n = \frac{2n}{\varepsilon}.$$

Thus, we are in the setting of Lemma 6.1.3, which immediately tells us that H is a subgraph of $G_R[Q]$. Thus, we have found a monochromatic red copy of H, implying that $r(H) \leqslant N$. □

As discussed on page 43, the original proof of Chvatál–Rödl–Szemerédi–Trotter of Theorem 5.4.6 also used the regularity method, and gave tower-type bounds on the constant C. In contrast, the proof above shows that we can take $C \leqslant 2^{2^{2^{10\Delta}}}$, which is still huge, but substantially smaller. The reason is that Chvatál–Rödl–Szemerédi–Trotter used the full regularity lemma of Szemerédi, whereas we used the weaker result Lemma 6.1.4, which is sufficient for this application. The fact that the full regularity lemma is not needed in this approach, and thus that tower-type bounds can be avoided, was first observed by Eaton [36].

Note that this proof, as written, only works when the number of colors is 2, since we crucially used Lemma 6.1.5 to deduce that $G_B[Q]$ is ε-quasirandom, even though we applied Lemma 6.1.4 to G_R. To deal with more colors, we can use the following generalization of Lemma 6.1.4, which produces a subset that is quasirandom in any graph in a fixed collection.

Lemma 6.2.1. *For every $\varepsilon > 0$ and $q \geqslant 1$, there exists $\delta > 0$ such that the following holds. Let $G_1, \dots, G_q$ be graphs on the same vertex set V. There exists $Q \subseteq V$ with $|Q| \geqslant \delta|V|$ such that $G_1[Q], \dots, G_q[Q]$ are ε-quasirandom.*

Note that we do not assume that $G_1, \dots, G_q$ are edge-disjoint, although we will usually have that property in applications, as we will let $G_1, \dots, G_q$ be the color classes in a q-coloring of $E(K_N)$.

Lemma 6.2.1 can be proved in the same way as Lemma 6.1.4, but it can also be deduced as a direct consequence of Lemma 6.1.4. For this deduction, we will need a simple observation, called the *hereditary property* of quasirandomness. It says that any large induced subgraph of a quasirandom graph is still quasirandom, albeit with a worse value of ε.

Lemma 6.2.2. *Let G be an ε-quasirandom graph. If $S \subseteq V(G)$ satisfies $|S| \geqslant \eta|V(G)|$ for some $\eta > 0$, then $G[S]$ is (ε')-quasirandom, where $\varepsilon' = \max\{2\varepsilon, \varepsilon/\eta\}$.*

Proof. Note that if $\eta \geqslant \varepsilon$, then there is nothing to prove as every graph is 1-quasirandom. Hence we may assume that $\eta < \varepsilon$. Then $|S| \geqslant \varepsilon|V(G)|$, so the quasirandomness of G implies that $|d(G) - d(S)| \leqslant \varepsilon$.

To prove that $G[S]$ is ε'-quasirandom, we need to show that $|d(S) - d(T)| \leqslant \varepsilon'$ for every $T \subseteq S$ with $|T| \geqslant \varepsilon'|S|$. Note that for any such T, we have $|T| \geqslant \varepsilon'|S| \geqslant (\varepsilon/\eta)|S| \geqslant \varepsilon|V(G)|$, hence the ε-quasirandomness of G implies that $|d(G) - d(T)| \leqslant \varepsilon$. Therefore,

$$|d(S) - d(T)| \leqslant |d(S) - d(G)| + |d(G) - d(T)| \leqslant \varepsilon + \varepsilon \leqslant \varepsilon'. \qquad \square$$

Using this simple observation, we can prove Lemma 6.2.1 by induction on q.

Proof of Lemma 6.2.1. We prove the existence of $\delta = \delta(\varepsilon, q)$ by induction on q. The base case $q = 1$ is precisely the statement of Lemma 6.1.4. Inductively, having defined $\delta(\varepsilon, q-1)$ for all ε, we define $\gamma \coloneqq \frac{1}{2}\varepsilon \cdot \delta(\varepsilon, 1)$ and

$$\delta(\varepsilon, q) \coloneqq \delta(\varepsilon, 1)\delta(\gamma, q-1).$$

Now suppose we are given graphs $G_1, \dots, G_q$ on the same vertex set V. By the definition of $\delta(\gamma, q-1)$, we can find a set $Q_1 \subseteq V$ with $|Q_1| \geqslant \delta(\gamma, q-1)|V|$ such that $G_1[Q_1], \dots, G_{q-1}[Q_1]$ are all γ-quasirandom. We now apply the base case, that is Lemma 6.1.4, to the graph $G_q[Q_1]$, to conclude that there exists $Q \subseteq Q_1$ with $|Q| \geqslant \delta(\varepsilon, 1)|Q_1|$ such that $G_q[Q]$ is ε-quasirandom. By Lemma 6.2.2, $G_1[Q], \dots, G_{q-1}[Q]$ are all γ'-quasirandom, where

$$\gamma' = \max\{2\gamma, \gamma/\delta(\varepsilon, 1)\} \leqslant \frac{2\gamma}{\delta(\varepsilon, 1)} = \varepsilon.$$

Thus, $G_1[Q], \dots, G_q[Q]$ are all ε-quasirandom. To conclude the proof, we note that

$$|Q| \geqslant \delta(\varepsilon, 1)|Q_1| \geqslant \delta(\varepsilon, 1)\delta(\gamma, q-1)|V| = \delta(\varepsilon, q)|V|. \qquad \square$$

Using Lemma 6.2.1, it is now straightforward to prove the following multicolor version of Theorem 5.4.6; the proof is left for the homework.

Theorem 6.2.3. *For all integers $\Delta \geqslant 1, q \geqslant 2$, there exists a constant C such that if H is an n-vertex graph with maximum degree at most Δ, then $r(H; q) \leqslant Cn$.*

6.3 Application II: Rödl's theorem

A graph G is called *induced-H-free* if it does not have an induced subgraph isomorphic to H. One of the most important questions in structural graph theory is to understand the structure of induced-H-free graphs.

For certain choices of H, one can get very precise results fairly easily. For example, if $H = K_2$, then an induced-K_2-free graph is the same as a graph with no edges. A slightly less trivial example is when $H = K_{1,2}$; in that case, one can show that a graph G is induced-$K_{1,2}$-free if and only if G is a disjoint union of complete graphs. In other cases, such as $H = K_3$, it is essentially impossible to get such a strong characterization, but we still know a lot; for example, Theorem 4.1.4 implies that every N-vertex induced-K_3-free graph contains an independent set of order $(1 - o(1))\sqrt{N \ln N}$.

For general H, much less is known. One of the most important conjectures in this field is due to Erdős and Hajnal [44], and states that induced-H-free graphs contain very large cliques or independent sets.

Conjecture 6.3.1 (Erdős–Hajnal [44])**.** *For every graph H, there exists a constant $c > 0$ such that the following holds. If G is an N-vertex induced-H-free graph, then G contains a clique or independent set of order at least N^c.*

Note that, by Theorem 2.1.4, every N-vertex graph contains a clique or an independent set of order at least $\frac{1}{2} \log N$. The Erdős–Hajnal conjecture states that if we impose the condition that G is induced-H-free, this result can be substantially improved.

In recent years, there have been a number of important breakthroughs related to the Erdős–Hajnal conjecture, most of which we will not discuss. Let us only mention a recent result of Bucić, Nguyen, Scott, and Seymour [11], which gives the best known result for general H. Their result improves on that of Erdős and Hajnal [45], who proved the same bound without the $\log \log N$ term.

Theorem 6.3.2 (Bucić–Nguyen–Scott–Seymour [11])**.** *For every graph H, there exists a constant $c > 0$ such that the following holds. If G is an N-vertex induced-H-free graph, then G contains a clique or independent set of order at least $e^{c\sqrt{\log N \log \log N}}$.*

The full Erdős–Hajnal conjecture remains wide open, and even improving on the bound in Theorem 6.3.2 seems like it would require substantial new ideas.

However, there are other things one can say about induced-H-free graphs. In particular, we will shortly prove the following surprising and useful theorem of Rödl [112], which can be viewed as an approximate form of the Erdős–Hajnal conjecture. It states that every induced H-free graph contains a *linearly-sized* subset whose edge density is very close to 0 or 1.

Theorem 6.3.3 (Rödl [112])**.** *For every graph H and every $\sigma > 0$, there exists $\delta > 0$ such that the following holds. If G is an induced-H-free graph, then there is a subset $S \subseteq V(G)$ with $|S| \geqslant \delta |V(G)|$ such that $d(S) < \sigma$ or $d(S) > 1 - \sigma$.*

In order to prove Theorem 6.3.3, we will need a form of the embedding lemma suited to embedding induced copies of H. Later we will need an even more general form of such a result, so we state the most general form now; we will shortly see how to use it for induced copies of H.

Lemma 6.3.4 (Multicolor embedding lemma)**.** *Let H be an n-vertex graph with maximum degree $\Delta \geqslant 1$. Let $0 < \varepsilon < \frac{1}{2\Delta}$ be a real number, and let $N \geqslant \frac{2n}{\varepsilon}$ be an integer. Additionally, let q be an integer and fix a q-coloring $\chi : E(H) \to [\![q]\!]$ of the edges of H.*

Let $G_1, \ldots, G_q$ be graphs on a common vertex set V with $|V| = N$. Suppose that each G_i is ε-quasirandom and has edge density $d(G_i) \geqslant (2\Delta\varepsilon)^{1/\Delta}$. Then there is a copy of H in $G_1 \cup \cdots \cup G_q$ such that if an edge of H has color $i \in [\![q]\!]$, then it appears in G_i.

We will omit the proof of Lemma 6.3.4. It can be proved in the same way as Lemma 6.1.3, but it can also be deduced directly from (a slightly more general form of) Lemma 6.1.3. See e.g. [145, Remark 2.6.3] for details. With this tool, we can quickly prove Theorem 6.3.3.

Proof of Theorem 6.3.3. Let H have n vertices. Let us define a 2-coloring $\chi_H : E(K_n) \to [\![2]\!]$ by coloring the edges of H with color 1 and the non-edges of H with color 2. Let $\varepsilon = \frac{\sigma^n}{2n} > 0$, and let $\delta_0 = \delta_0(\varepsilon)$ be the parameter from Lemma 6.1.4. Let $\delta = \frac{\varepsilon \delta_0}{2n}$.

We now fixed an induced-H-free graph G, and we wish to prove that G contains a subset S with $|S| \geqslant \delta |V(G)|$ such that $d(S) < \sigma$ or $d(S) > 1 - \sigma$. If $|V(G)| \leqslant \frac{1}{\delta}$ then this is trivially true as we may set S to comprise a single vertex, so we henceforth assume that $|V(G)| \geqslant \frac{1}{\delta}$. Let $G_1 = G$ and let $G_2 = \overline{G}$ be the complement of G.

We apply Lemma 6.1.4 to G to find a set $Q \subseteq V(G)$ with $|Q| \geqslant \delta_0 |V(G)|$ such that $G_1[Q]$ is ε-quasirandom; Lemma 6.1.5 then implies that $G_2[Q]$ is ε-quasirandom as well. Note that

$$|Q| \geqslant \delta_0 |V(G)| \geqslant \frac{\delta_0}{\delta} = \frac{2n}{\varepsilon}.$$

If $d(G[Q]) = d(G_1[Q]) < \sigma$, then we can set $S = Q$ and be done. Similarly, if $d(G_2[Q]) < \sigma$, then $d(G[Q]) > 1 - \sigma$, and we are again done. So we may assume that $d(G_i[Q]) \geqslant \sigma$ for $i = 1, 2$. By our choice of ε, this implies that $d(G_i[Q]) \geqslant (2n\varepsilon)^{1/n}$ for $i = 1, 2$.

We now apply Lemma 6.3.4, where we are trying to embed the graph K_n with the coloring χ_H defined above. By Lemma 6.3.4, which we may apply since K_n has maximum degree $n - 1 \leqslant n$, we find that there is a copy of K_n in Q such that all edges colored 1 appear in G_1, and all edges colored 2 appear in G_2. But this precisely means that the edges of H appear in G, and that the non-edges of H appear in $\overline{G}$. Hence we have found an induced copy of H in G, contradicting our assumption that G is induced-H-free; this contradiction completes the proof. $\square$

Chapter 7

Restricted Ramsey graphs

7.1 Folkman's theorem and beyond

We started this course with Ramsey's theorem: for every k, there exists an N such that if the edges of K_N are two-colored, then there exists a monochromatic K_k. In Chapter 5, we generalized the conclusion: rather than finding a monochromatic K_k, we found a monochromatic copy of H, for some not-necessarily-complete graph H. We will now generalize the first part of the statement.

Definition 7.1.1. Given two graphs G, H, we say that *G is Ramsey for H in q colors* (or *G is q-color Ramsey for H*) if, whenever the edges of G are q-colored, there is a monochromatic copy of H. In case $q = 2$, we simply say that *G is Ramsey for H*.

Thus, Ramsey's theorem simply states that K_N is q-color Ramsey for K_k whenever N is sufficiently large (as a function of q and k).

To gain some intuition for this definition, let's think of the case when $H = K_3$. If G is Ramsey for K_3, then certainly G must contain at least one triangle. But in fact, the definition of G being Ramsey for K_3 tells us that G contains triangles "very robustly". Indeed, another way of saying Definition 7.1.1 is that, no matter how we try to split G into the union of two subgraphs, we cannot destroy all triangles in G. This idea of robustness is one of the reasons that Definition 7.1.1 is interesting.

That being said, it's not at all obvious that this definition actually gives us any new information. Indeed, we know that $r(3) = 6$, or equivalently that K_6 is Ramsey for K_3 while K_5 is not. In particular, we find that if G is a graph containing K_6 as a subgraph, then G is Ramsey for K_3. Indeed, given any 2-coloring of $E(G)$, ignore all the edges except for those in the K_6 subgraph; among those $\binom{6}{2}$ edges, we are guaranteed to find a monochromatic triangle, regardless of how the other edges are colored.

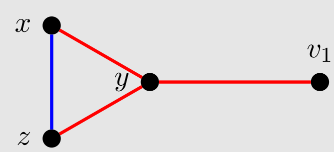
If you spend some time trying to construct graphs that are Ramsey for K_3, you may start to wonder if this is the *only* reason a graph can be Ramsey for K_3. In other words, you might be tempted to conjecture that G is Ramsey for K_3 if and only if $K_6 \subseteq G$. The question of whether this is true was raised by Erdős and Hajnal [43], and was rapidly answered in the negative independently by Cherlin (unpublished), Graham [65], and van Lint (unpublished). The following slick construction is due independently to Galluccio–Simonovits–Simonyi [61] and to Szabó [133], and generalizes Graham's original argument. Given two graphs G_1, G_2, their *join*, denoted $G_1 * G_2$, is the graph obtained from their disjoint union by adding all edges with one endpoint in G_1 and one in G_2.

Proposition 7.1.2 (Galluccio–Simonovits–Simonyi [61], Szabó [133]). *Let $G = K_3 * C_\ell$, where $\ell \geqslant 3$ is an odd integer. Then G is Ramsey for K_3. Moreover, if $\ell \geqslant 5$, then $K_6 \nsubseteq G$.*

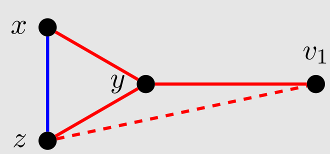
Proof. Let the vertices of G be $x, y, z, v_1, \ldots, v_\ell$, where x, y, z form a triangle, $v_1, \ldots, v_\ell$ form a cycle C_ℓ, and all edges between $\{x, y, z\}$ and $\{v_1, \ldots, v_\ell\}$ are present. Note that if $K_6 \subseteq G$, then at least three of the vertices of this K_6 must come from $v_1, \ldots, v_\ell$ (and they must form a triangle), so the second statement of the proposition is immediate since C_ℓ is triangle-free whenever $\ell \geqslant 5$.

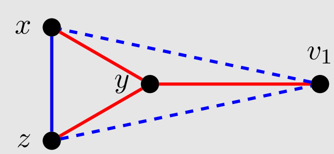
It remains to show that G is Ramsey for K_3, so fix a two-coloring of $E(G)$. If $\{x, y, z\}$ form a monochromatic triangle then we are done, so two of the edges xy, xz, yz receive one color and the third edge receives the other color. Without loss of generality, we may assume that xy, yz are red and xz is blue.

Now consider the edges between $\{x, y, z\}$ and v_1. First suppose yv_1 is red.

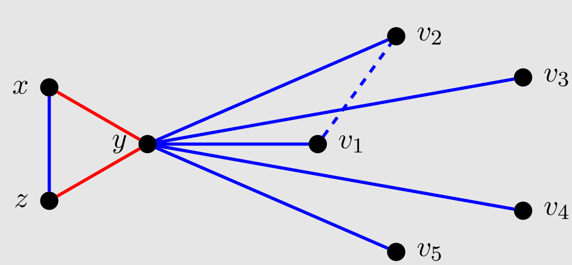


If xv_1 or zv_1 is red, then we close a red triangle xyv_1 or zyv_1, so we may assume that both these edges are blue. But that also creates a blue triangle, xzv_1.

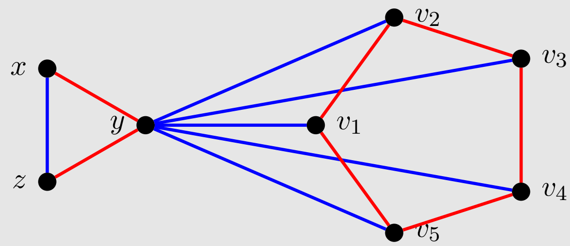




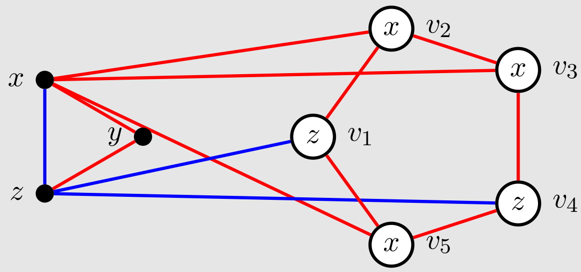
So we may assume that yv_1 is blue. By the same logic, yv_i is blue for all $i \in [\![\ell]\!]$. Note that if any of the edges v_iv_{i+1} in the cycle is blue, then we create a blue triangle yv_iv_{i+1}.



Therefore, we may assume that all the edges in the cycle are red.



Recall that xv_i and zv_i cannot both be blue, as this would create a blue triangle xzv_i. Let us label v_i by either the label x or z, depending on whether xv_i or zv_i is red (picking a label arbitrarily if both are red). By the above, every v_i receives a label.



Since ℓ is odd, the cycle C_ℓ is not bipartite. Hence, two adjacent vertices v_i, v_{i+1} must receive the same label (like v_2 and v_3 in the picture above). But then they create a red triangle together with their label. $\square$

Note that $K_3 * K_3 = K_6$, so this result also gives a new (and more complicated) proof that K_6 is Ramsey for K_3. But it also shows that the set of graphs Ramsey for K_3 is surprisingly rich.

Note that each of the graphs $K_3 * C_\ell$ considered above *does* contain K_5 as a subgraph. So there is a natural weakening of our previous question: does every graph which is Ramsey for K_3 contain K_5 as a subgraph? The answer to this question also turns out to be negative, as proved by Pósa (unpublished, but included in [43]). So we may weaken our question further: does every graph Ramsey for K_3 contain a K_4? The answer to this also turns out to be no, as shown by the following remarkable theorem of Folkman [54].

Theorem 7.1.3 (Folkman [54])**.** *For every $k \geqslant 2$, there exists a graph G such that G is Ramsey for K_k, but $K_{k+1} \nsubseteq G$.*

This is pretty astonishing, even in the case $k = 3$. As discussed above, a graph that is Ramsey for K_3 must contain triangles "very robustly", in the sense that we cannot destroy all the triangles by splitting the graph into two subgraphs. Yet Folkman's theorem shows that such a graph can exist even though, locally, the triangles have almost no overlap.

Folkman's proof only worked for the case of two-colors, but the general case was shortly thereafter established by Nešetřil and Rödl [96], who proved the following generalization. We denote by $\omega(H)$ the *clique number* of H, that is, the maximum k such that $K_k \subseteq H$.

Theorem 7.1.4 (Nešetřil–Rödl [96])**.** *For every graph H and every $q \geqslant 2$, there exists a graph G which is q-color Ramsey for H with $\omega(G) = \omega(H)$.*

In their proof, Nešetřil and Rödl introduced a very powerful technique, called the *partite construction*, which is a very general-purpose way of producing graphs G that are Ramsey for a given graph H, while satisfying certain local sparsity conditions. We will not cover the partite construction in this course, but we refer to the excellent introduction in [85].

The partite construction (as well as the earlier construction of Folkman) is completely explicit, so we can get a complete description of what the graph G in Theorem 7.1.4 looks like. Unfortunately, these constructions are iterative in nature, and each step of the iteration is complicated, so the *size* of the graph G constructed is unbelievably huge.

There is now an alternative approach to constructing such restricted Ramsey graphs, which uses randomness. It has a number of advantages over the partite construction, including giving much better bounds on how large G has to be in results like Theorem 7.1.4. However, as we will discuss shortly, it also seems to be less flexible than the partite construction, and there are results that the random approach seems incapable of proving.

The main result in this direction is the *random Ramsey theorem* of Rödl and Ruciński [114]. To state it, we define the *maximal* 2*-density* of a graph H to be

$$m_2(H) \coloneqq \max_{\substack{J \subseteq H \\ v(J) \geqslant 3}} \frac{e(J)-1}{v(J)-2}.$$

Theorem 7.1.5 (Rödl–Ruciński [114])**.** *Let H be a graph which is not a forest, and let $q \geqslant 2$. There exist constants $C > c > 0$ such that the following holds. Form an N-vertex graph G by including each edge independently with probability p. Then*

$$\lim_{N \to \infty} \Pr(G \text{ is Ramsey for } H \text{ in } q \text{ colors}) = \begin{cases} 1 & \text{if } p \geqslant CN^{-1/m_2(H)}, \\ 0 & \text{if } p \leqslant cN^{-1/m_2(H)}. \end{cases}$$

In other words, $p \asymp N^{-1/m_2(H)}$ is a *threshold* for the property of G being Ramsey for H. If p is substantially smaller than this value, then G is extremely unlikely to be Ramsey for H, whereas if p is substantially larger than this value, then G is extremely likely to be Ramsey for H. The heuristic reason why this value of p controls the threshold is the following. One can check that at this value, a typical edge of G lies in a constant number of copies of H[1]. Thus, if $p \leqslant cN^{-1/m_2(H)}$ for a small constant c, then the majority of edges of G lie in *zero* copies of H, and thus it is not surprising that G does not "robustly" contain H; we should be able to color $E(G)$ and destroy all copies of H. On the other hand, if $p \geqslant CN^{-1/m_2(H)}$ for a large constant C, then most edges of G lie in very many copies of H, and we expect a great deal of interaction between the copies, such that destroying all of them becomes impossible no matter how we color the edges. While this is a good heuristic explanation, turning it into a proof is substantially harder, and we will not do so in this course.

However, Theorem 7.1.5 does allow us to easily prove results along the lines of Theorem 7.1.3. One can actually prove Theorem 7.1.3 as a consequence of (a more precise

[1] I am cheating a bit here; really, I should be counting copies of the subgraph $J \subseteq H$ achieving the maximum in the definition of $m_2(H)$.

version of) Theorem 7.1.5, but we will content ourselves with proving the following weakening of Theorem 7.1.3, which generalizes Proposition 7.1.2 (which corresponds to the case $k = 3, q = 2$).

Proposition 7.1.6. *For every $k \geqslant 3$ and $q \geqslant 2$, there exists a graph G which is q-color Ramsey for K_k, but $K_{k+3} \nsubseteq G$.*

Proof. We begin by observing that

$$\frac{e(K_k) - 1}{v(K_k) - 2} = \frac{\binom{k}{2} - 1}{k - 2} = \frac{k^2 - k - 2}{2(k-2)} = \frac{k+1}{2}.$$

It is not hard to check that $\frac{e(J)-1}{v(J)-2}$ is strictly smaller for any proper subgraph $J \subsetneq K_k$, hence $m_2(K_k) = \frac{k+1}{2}$. By Theorem 7.1.5, there is a constant $C > 0$ such that the following holds. If we pick an N-vertex graph randomly by including each edge independently with probability $p := CN^{-\frac{2}{k+1}}$, then G is q-color Ramsey for H with probability tending to 1 as $N \to \infty$. In particular, if N is sufficiently large, then this probability is at least $\frac{2}{3}$.

On the other hand, by the union bound, the probability that $K_{k+3} \subseteq G$ is at most

$$\binom{N}{k+3} p^{\binom{k+3}{2}} < C^{\binom{k+3}{2}} \cdot N^{k+3} \cdot N^{-\frac{2}{k+1}\binom{k+3}{2}} = C^{\binom{k+3}{2}} \cdot N^{-\left(\frac{2}{k+1}\binom{k+3}{2} - (k+3)\right)}. \tag{7.1}$$

We have that

$$\frac{2}{k+1}\binom{k+3}{2} - (k+3) = \frac{(k+3)(k+2)}{k+1} - (k+3) = (k+3)\left(\frac{k+2}{k+1} - 1\right) > 0.$$

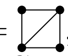
Hence, the exponent on N is negative in (7.1), so the probability that $K_{k+3} \subseteq G$ tends to 0 as $N \to \infty$. In particular, by picking N sufficiently large, we can ensure that $K_{k+3} \nsubseteq G$ with probability at least $\frac{2}{3}$.

Therefore, with positive probability, G satisfies both the desired properties, proving the claimed result. $\square$

Before ending this section, let us briefly discuss one further recent breakthrough on the structure of restricted Ramsey graphs, due to Reiher and Rödl [110].

Definition 7.1.7. Let H be a graph. We say that another graph F is *Ramsey obligatory for H* if the following holds. For every sufficiently large q and every graph G which is q-color Ramsey for H, we have $F \subseteq G$.

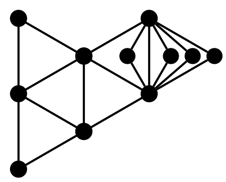
In this language, we can restate Proposition 7.1.6 as saying that K_{k+3} is not Ramsey obligatory for K_k, and Theorem 7.1.3 (or more precisely its multicolor extension, which follows from Theorem 7.1.4) states that K_{k+1} is not Ramsey obligatory for H. On the other hand, we can easily show that certain graphs *are* Ramsey obligatory for H. For example, H itself is Ramsey obligatory for H—if G is Ramsey for H, then certainly G contains H as a subgraph!

To keep things concrete, let's specialize to the case $H = K_3$. Then we know that K_3 is Ramsey obligatory for K_3, but K_4 is not. On the other hand, the graph $F =$ , obtained by gluing two triangles along an edge, is also Ramsey obligatory. Indeed, if G is an F-free graph, then all the triangles in G are edge-disjoint, so certainly we can color $E(G)$ and avoid all monochromatic triangles. More generally, we make the following definition.

Definition 7.1.8. *Triangle trees* are the class of graphs defined recursively as follows.

- K_3 is a triangle tree.
- Given a triangle tree T, we can obtain a new triangle tree T' by picking an edge of T and gluing a new triangle to it.

A typical triangle tree might look something like the following.



It is not hard to show the following fact; the proof is left for the homework.

Proposition 7.1.9. *If F is a subgraph of a triangle tree, then F is Ramsey obligatory for K_3.*

The astonishing theorem of Reiher and Rödl [110] is that this sufficient condition is also necessary.

Theorem 7.1.10 (Reiher–Rödl [110])**.** *A graph F is Ramsey obligatory for K_3 if and only if F is a subgraph of a triangle tree.*

Said differently, given any graph F which is not a subgraph of a triangle tree, Reiher and Rödl are able to construct a graph G which is q-color Ramsey for K_3, yet does not contain F as a subgraph. In particular, since one can check that K_4 is not a subgraph of a triangle tree, this implies the $k = 3$ case of Theorem 7.1.3.

In fact, their theorem is vastly more general than this, and implies many strengthenings of Theorem 7.1.4. Somewhat more surprisingly, it appears that even for proving a result like Theorem 7.1.10, one actually has to prove these much more general results; their proof is based on a very complicated inductive argument, and in order to make the induction work one has to maintain a very general inductive statement.

7.2 The induced Ramsey theorem

So far, we have been studying graphs G which are Ramsey for H and which have a restricted structure (e.g. containing no large cliques). However, we have not at all restricted the way in which a monochromatic H can appear. We now turn to such a restriction.

Definition 7.2.1. A graph G is *q-color induced Ramsey for H* if, in any q-coloring of $E(G)$, there is an induced subgraph of G, isomorphic to H, all of whose edges receive the same color.

In other words, we wish to find a monochromatic copy of H, but such that every non-edge of H is also not present in G. In the case $H = K_k$, note that G is Ramsey for K_k if and only if G is induced Ramsey for K_k. The reason is that, whenever K_k is a subgraph of G, it is also an induced subgraph of G. Thus, for cliques, this new notion is the same as the old notion.

However, if H is not a clique, it is not at all obvious that there *exists* some G which is induced Ramsey for H. Indeed, while Ramsey's theorem guarantees that K_N is Ramsey for H for sufficiently large H, K_N is certainly *not* induced Ramsey for H.

In fact, the existence of induced Ramsey graphs is a highly non-trivial result, which was first proved independently by Deuber [31], Erdős–Hajnal–Pósa [46], and Rödl [111].

Theorem 7.2.2 (Deuber [31], Erdős–Hajnal–Pósa [46], Rödl [111])**.** *For every graph H and every $q \geqslant 2$, there exists a graph G such that G is q-color induced Ramsey for H.*

Shortly thereafter, Nešetřil and Rödl [97] gave a simplified proof using the partite construction. We will give a short proof using the regularity method, the idea of which can be traced back to [98], and which is more closely inspired by [55].

Proof. Let H have n vertices. Define $\varepsilon = \frac{(3q)^{-n}}{2n}$, and let $\delta = \delta(\varepsilon, q)$ be the parameter from Lemma 6.2.1. Let $\rho = \delta\varepsilon$.

Let N be a sufficiently large, and let G be a random N-vertex graph obtained by making each pair an edge independently with probability $\frac{1}{2}$. The only property we will need about G, which we mentioned after Definition 6.1.2, is that G is highly quasirandom. More precisely, assuming N is sufficiently large, then with positive probability G has the property that every subset of at least ρN vertices has edge density in $[\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}]$. This implies, in particular, that for any set $Q \subseteq V(G)$ with $|Q| \geqslant \delta N$, the induced subgraph $G[Q]$ is ε-quasirandom. We now fix G to have this property, which we can do since the probability it does is positive for sufficiently large N. Note that the complement $\overline{G}$ also has this property. We claim that G is induced Ramsey for H.

Indeed, fix a q-coloring of $E(G)$, that is, a partition of $E(G)$ into subgraphs $G_1, \ldots, G_q$. By Lemma 6.2.1, there exists $Q \subseteq V(G)$ such that $G_1[Q], \ldots, G_q[Q]$ are all ε-quasirandom. Moreover, by the property of G above, we know that $\overline{G}[Q]$ is ε-quasirandom, and that $d(\overline{G}[Q]) \leqslant \frac{1+\varepsilon}{2} \leqslant \frac{2}{3}$. Therefore, by the pigeonhole principle, there exists an $i \in [\![q]\!]$ such that $d(G_i[Q]) \geqslant \frac{1}{3q}$. We claim that we can find an induced copy of H in G all of whose edges receive color i, implying the claimed result.

Indeed, we apply Lemma 6.3.4. Note that we may assume $|Q| \geqslant \frac{2n}{\varepsilon}$ since we chose N sufficiently large. Define a coloring $\chi : K_n \to \{1, 2\}$ by coloring all edges of H with color 1 and all non-edges of H with color 2. We have that

$$d(\overline{G}[Q]) \geqslant \frac{1}{3} \geqslant \frac{1}{3q} = (2n\varepsilon)^{1/n} \qquad \text{and} \qquad d(G_i[Q]) \geqslant \frac{1}{3q} = (2n\varepsilon)^{1/n},$$

by our choice of ε. Therefore, Lemma 6.3.4 implies that there is a copy of K_n in $(G_i \cup \overline{G})[Q]$ such that all edges of H are in G_i, and all non-edges of H are in $\overline{G}$. But this is precisely saying that we have found an induced copy of H in G, all of whose edges have color i, as claimed. $\square$

Let us define the *induced Ramsey number* $r_{\mathrm{ind}}(H; q)$ to be the least N such that there exists an N-vertex graph G which is q-color induced Ramsey for H. The proof above shows that $r_{\mathrm{ind}}(H; 2) \leqslant 2^{2^{2^{10n}}}$ for any n-vertex graph H. For more colors, the proof above gives a worse bound, because the value of δ in Lemma 6.2.1 depends quite poorly on q. However, Erdős [38] made the following conjecture.

Conjecture 7.2.3. *For every $q \geqslant 2$, there exists an absolute constant $C > 0$ such that $r_{\mathrm{ind}}(H; q) \leqslant 2^{Cn}$ for every n-vertex graph H.*

Note that this result, if true, is best possible, since $r_{\mathrm{ind}}(K_n; q) = r(K_n; q)$, and we know that $r(K_n; q)$ grows exponentially with n for any fixed q. In the case of two colors, Conjecture 7.2.3 is almost known; Conlon, Fox, and Sudakov [26] proved that $r_{\mathrm{ind}}(H; 2) \leqslant 2^{Cn \log n}$ for every n-vertex graph H, where C is an absolute constant. However, their proof uses the greedy embedding approach discussed in Section 5.4.1, and as such does not extend to more than two colors. The best known upper bound in general is due to Balogh and Samotij [5] (using quite a different analysis, although still applied to the same random graph G), who showed that $r_{\mathrm{ind}}(H; q) \leqslant 2^{C_q n^2}$ for any n-vertex graph H, where $C_q > 0$ is a constant depending only on q.

Chapter 8

C-Ramsey graphs

8.1 The Erdős–Szemerédi theorem

The following question was first raised by Erdős, Hajnal, and Rado [47], and was brought to prominence by an important result of Erdős and Szemerédi [50], which is the main topic of this section.

Definition 8.1.1. Let $k, q \geqslant 2$ be integers. The *omission Ramsey number* $r_o(k; q)$ is the minimum N such that, in any q-coloring of $E(K_N)$, there is a set $S \subseteq V(K_N)$ with $|S| = k$ such that the edges within S are colored with at most $q-1$ colors.

In other words, rather than searching for a *monochromatic* K_k, as in the definition of $r(k; q)$, we are searching for a copy of K_k which *omits* at least one of the q colors. In particular, if $q = 2$, then $r(k; 2)$ and $r_o(k; 2)$ are equal, since omitting one of the two colors is the same as only using one of the two colors. In general, we always have the inequality

$$r_o(k; q) \leqslant r(k; q)$$

since if we can find a monochromatic K_k, then in particular we have found a K_k that omits $q - 1 \geqslant 1$ of the colors. Note too that the question is not interesting if $q > \binom{k}{2}$, since then every k-set is colored with at most $q-1$ colors.

For large q, it is natural to expect that $r_o(k; q)$ is much smaller than $r(k; q)$. Indeed, if we have 1000 colors, then finding a K_k colored with only 999 colors seems much easier than finding a K_k colored with only 1 color. And indeed, while we proved in Chapter 3 that $r(k; q) = 2^{\Omega(kq)}$, the best known lower bound for $r_o(k; q)$ is much smaller—of order $\exp(\Omega(\frac{k}{q}))$, rather than $\exp(\Omega(kq))$.

Proposition 8.1.2. *If $k \geqslant 4$ and $2 \leqslant q \leqslant \binom{k}{2}$, then*

$$r_o(k; q) > e^{\frac{k}{2q}}.$$

Proof. Set $N = e^{\frac{k}{2q}}$, and consider a random q-coloring of $E(K_N)$. The probability that a given k-set receives at most $q - 1$ colors is at most $q \cdot (\frac{q-1}{q})^{\binom{k}{2}}$, since there are q choices for the omitted color, and then we need each of the $\binom{k}{2}$ edges to receive one of the $q-1$ non-omitted colors, which happens with probability $\frac{q-1}{q}$. Hence, by the union bound, the probability that there is a k-set receiving at most $q-1$ colors is at most

$$\binom{N}{k} \cdot q \cdot \left(\frac{q-1}{q}\right)^{\binom{k}{2}} < \frac{q}{k!(1-\frac{1}{q})^{\frac{k}{2}}} \cdot N^k \left(1-\frac{1}{q}\right)^{\frac{k^2}{2}}.$$

We have that $1-\frac{1}{q} \geqslant \frac{1}{2}$ since $q \geqslant 2$, so

$$\frac{q}{k!(1-\frac{1}{q})^{\frac{k}{2}}} \leqslant \frac{q2^{\frac{k}{2}}}{k!} \leqslant \frac{\binom{k}{2}2^{\frac{k}{2}}}{k!} \leqslant 1,$$

where the final inequality holds since $k \geqslant 4$. On the other hand, using the inequality $1-x \leqslant e^{-x}$, we have that

$$N^k \left(1-\frac{1}{q}\right)^{\frac{k^2}{2}} \leqslant \left(Ne^{-\frac{k}{2q}}\right)^k = 1,$$

by our choice of N. Hence, the probability that there is a k-set receiving at most $q-1$ colors is strictly less than 1, proving that there is a coloring with no such set, hence $r_o(k;q) > N$. $\square$

Combining this with Theorem 2.1.5, we find that

$$\exp\left(\Omega\left(\frac{k}{q}\right)\right) \leqslant r_o(k;q) \leqslant r(k;q) \leqslant \exp(O(kq \log q)).$$

This shows that $r(k;q)$ does grow exponentially in k, but there is a huge gap in the dependence on q. In particular, we don't even know if the exponent tends to 0 or ∞ as $q \to \infty$!

This situation was remedied by the following theorem of Erdős and Szemerédi [50].

Theorem 8.1.3 (Erdős–Szemerédi [50])**.** *For all $k \geqslant q \geqslant 2$, we have that*

$$r_o(k;q) \leqslant 2^{A\frac{\log q}{q}k},$$

where $A > 0$ is an absolute constant.

Note that this matches the lower bound in Proposition 8.1.2 up to the factor of $\Theta(\log q)$. It is a major open problem to close this logarithmic gap, and (as far as I'm aware) no one even has a conjecture for where the truth lies.

In order to prove Theorem 8.1.3, Erdős and Szemerédi proved the following result, which has since become one of the fundamental tools in Ramsey theory, going far beyond the original scope of Theorem 8.1.3. Recall that, by Theorem 2.1.4, every N-vertex graph contains a clique or an independent set of order at least $\frac{1}{2}\log N$. The following result shows that a much stronger result is true if we assume that G is sparse. Recall that $d(G)$ denotes the edge density of G.

Theorem 8.1.4 (Erdős–Szemerédi [50]). *There is an absolute constant $\tau > 0$ such that the following holds for every $\varepsilon > 0$ and every integer $N \geqslant \frac{1}{\varepsilon}$. If G is an N-vertex graph with $d(G) \leqslant \varepsilon$, then G contains a clique or an independent set of order at least*

$$\frac{\tau}{\varepsilon \log \frac{1}{\varepsilon}} \log N.$$

Note that, as $\varepsilon \to 0$, the function $(\varepsilon \log \frac{1}{\varepsilon})^{-1}$ tends to infinity. So for small ε, this yields a much larger clique or independent set than what is guaranteed by Theorem 2.1.4 alone.

The assumption that $N \geqslant \frac{1}{\varepsilon}$ is not particularly important. However, note that we must assume some lower bound on N in terms of ε (or equivalently a lower bound on ε in terms of N). The reason is that, if N is fixed, then the quantity $\tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N$ will eventually exceed N if we choose ε sufficiently small, and it is of course not possible for an N-vertex graph to contain a clique or independent set of order more than N. In fact, the assumption $N \geqslant \frac{1}{\varepsilon}$ is essentially best possible, since we already have $\tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N > N$ if ε is smaller than $\frac{1}{N}$ by some appropriate constant factor.

Before proving Theorem 8.1.4, let us see how it allows us to prove Theorem 8.1.3.

Proof of Theorem 8.1.3. Let $A = \max\{\frac{1}{\tau}, 1\}$, let $N = 2^{A \frac{\log q}{q} k}$, and fix a q-coloring of $E(K_N)$. Note that since $k \geqslant q$ and $A \geqslant 1$, we have that $N \geqslant q^A \geqslant q$. Suppose without loss of generality that red is the color containing the fewest number of edges, and let G be the graph of red edges. In particular, since at most $\frac{1}{q}\binom{N}{2}$ of the edges are red, we conclude that $d(G) \leqslant \frac{1}{q}$.

We now apply Theorem 8.1.4 with $\varepsilon = \frac{1}{q}$, which we may do since $N \geqslant q = \frac{1}{\varepsilon}$, and find that G contains a clique or independent set of order

$$\frac{\tau}{\varepsilon \log \frac{1}{\varepsilon}} \log N = \frac{\tau q}{\log q} \log \left(2^{A \frac{\log q}{q} k}\right) = \tau A k \geqslant k.$$

An independent set in G is a collection of vertices such that the color red does not appear among them, so if we've found an independent set of order k then we are done. On the other hand, a clique in G is a set of vertices receiving *only* the color red, so we are also done if we've found a clique of order k. □

Let us now prove Theorem 8.1.4.

Proof of Theorem 8.1.4. We fix a sufficiently small constant τ. We will not explicitly determine its value, but we will use in several places that it is sufficiently small so that certain inequalities hold. The assumption $d(G) \leqslant \varepsilon$ implies that the average degree of G is at most εN. Therefore Lemma 4.1.2 and our assumption that $N \geqslant \frac{1}{\varepsilon}$ imply that

$$\alpha(G) \geqslant \frac{N}{\varepsilon N + 1} = \frac{1}{\varepsilon} - \frac{1/\varepsilon}{\varepsilon N + 1} = \frac{1}{\varepsilon} - \frac{1}{\varepsilon^2 N + \varepsilon} \geqslant \frac{1}{\varepsilon} - \frac{1}{2\varepsilon} = \frac{1}{2\varepsilon}.$$

Thus, the result is automatically true if $\frac{1}{2\varepsilon} \geqslant \tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N$. Thus, we may assume that $\frac{1}{2\varepsilon} < \tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N$, or equivalently that $N > (\frac{1}{\varepsilon})^{1/(2\tau)}$. By choosing $\tau \leqslant \frac{1}{4}$, we may in particular assume that $N \geqslant \varepsilon^{-2}$, which we will do henceforth. Let $k = \tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N$, and note that our assumption that $N \geqslant \varepsilon^{-2}$ implies that $k \leqslant \frac{N}{4}$. Note too that, by picking τ sufficiently small, we may assume that $\varepsilon \leqslant \frac{1}{10}$; for if $\varepsilon > \frac{1}{10}$ and τ is sufficiently small, then we have that $\tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} < \frac{1}{2}$, and then the claimed result follows from Theorem 2.1.4.

Let A denote the set of vertices in G of degree at least $2\varepsilon N$. Note that the total number of edges incident to A is at least $\varepsilon N|A|$, where we lose a factor of two since each edge may be counted twice. Since the total number of edges in G is at most $\varepsilon\binom{N}{2}$, we conclude that $\varepsilon N|A| \leqslant \varepsilon\binom{N}{2}$, implying that $|A| < \frac{N}{2}$. Let G_1 be obtained from G by deleting all vertices in A, so that $|V(G_1)| \geqslant \frac{N}{2}$ and that every vertex in G_1 has degree at most $2\varepsilon N$.

Now, let X be a maximum-sized independent set in G_1. If $|X| \geqslant k$, then we are done, so we may assume that this is not the case. Let $Y = V(G_1) \setminus X$. Recalling that $k \leqslant \frac{N}{4}$ and $|V(G_1)| \geqslant \frac{N}{2}$, we find that $|Y| \geqslant \frac{N}{4}$.

Every vertex in X has degree at most $2\varepsilon N$, so there are at most $2\varepsilon N|X|$ edges between X and Y. Let $B \subseteq Y$ be the set of vertices in Y with at least $s \coloneqq 10\varepsilon|X|$ neighbors in X. As $e(X, Y) \leqslant 2\varepsilon N|X|$, and each vertex in B contributes at least s such edges, we conclude that

$$|B| \leqslant \frac{2\varepsilon N|X|}{s} = \frac{2\varepsilon N|X|}{10\varepsilon|X|} = \frac{N}{5}.$$

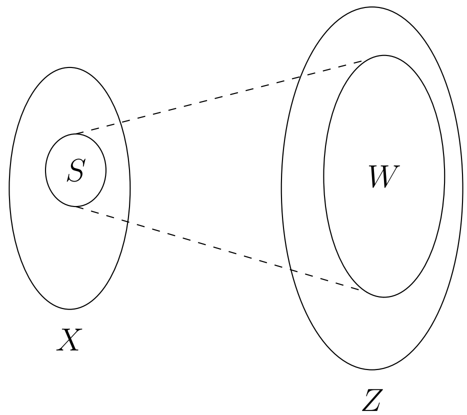
Let $Z = Y \setminus B$, and note that $|Z| \geqslant \frac{N}{4} - \frac{N}{5} = \frac{N}{20}$. Moreover, each vertex in Z has fewer than s neighbors in X. However, the number of subsets of X of size less than s is

$$\sum_{i=0}^{s-1} \binom{|X|}{i} \leqslant s\binom{|X|}{s} \leqslant s\left(\frac{e|X|}{s}\right)^s = s\left(\frac{e}{10\varepsilon}\right)^s < \left(\frac{1}{\varepsilon}\right)^{10\varepsilon k}.$$

We now recall that $10\varepsilon k = \frac{10\tau}{\log\frac{1}{\varepsilon}} \log N$, and therefore

$$\left(\frac{1}{\varepsilon}\right)^{10\varepsilon k} = N^{10\tau} \leqslant N^{\frac{1}{3}}, \tag{8.1}$$

so long as $\tau \leqslant \frac{1}{30}$. In other words, every vertex in Z has fewer than s neighbors in X, and there are at most $N^{\frac{1}{3}}$ choices for the value of this neighborhood. By the pigeonhole principle, this implies that there is a set $W \subseteq Z$ of size at least $|Z|/N^{\frac{1}{3}}$ such that all vertices in W have exactly the same neighborhood in X. Let this neighborhood be $S \subseteq X$.



The key point of this is that every vertex in W is non-adjacent to every vertex in $X \setminus S$. So if we find any independent set $T \subseteq W$, we may "swap" S for T and obtain another independent set; that is, the set $(X \setminus S) \cup T$ is a new independent set. Since we assumed that X was a maximum-sized independent set and since $|S| < s$, that implies that W contains no independent set of order s. The final step is to note that

$$\binom{s+k}{s} \leqslant \binom{(1+10\varepsilon)k}{10\varepsilon k} \leqslant \left(e\frac{1+10\varepsilon}{10\varepsilon}\right)^{10\varepsilon k} \leqslant \left(\frac{1}{\varepsilon}\right)^{10\varepsilon k} \leqslant N^{\frac{1}{3}},$$

by (8.1). Combining this bound with Theorem 2.1.4 and our lower bound on $|W|$, we find that

$$r(s,k) \leqslant \binom{s+k}{s} \leqslant N^{\frac{1}{3}} \leqslant \frac{N^{\frac{2}{3}}}{20} \leqslant \frac{|Z|}{N^{\frac{1}{3}}} \leqslant |W|,$$

assuming that $N \geqslant 20^3$, an assumption we can make by choosing τ sufficiently small. Since W contains no independent set of order s, we must have a clique of order k in W, completing the proof. □

In the setting of Theorem 8.1.4, since the density of G is very small, it is natural to expect that its largest independent set is much larger than its largest clique. Thus, it would be reasonable to suppose that we might be able to strengthen Theorem 8.1.4—finding a clique of order $\tau(\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N$ or an independent set of somewhat larger size. However, the following simple construction, also due to Erdős and Szemerédi [50], shows that this is not possible, and that Theorem 8.1.4 is best possible up to the value of τ.

Proposition 8.1.5. *For every ε and every sufficiently large N, there is an N-vertex graph G with $d(G) \leqslant \varepsilon$ such that the largest clique and largest independent set in G both have size $O((\varepsilon \log \frac{1}{\varepsilon})^{-1} \log N)$.*

Proof. Let m be some integer, and let G_0 be a random graph on m vertices with every edge appearing independently with probability $p = 1 - \sqrt{\varepsilon}$. Let $a = (8 \ln m)/\sqrt{\varepsilon}$ and let $b =$

$(16 \ln m)/\ln \frac{1}{\varepsilon}$. By the union bound, the probability that G_0 has a clique of order a is at most

$$\binom{m}{a} p^{\binom{a}{2}} < m^a p^{\frac{a^2}{4}} = \left(m p^{\frac{a}{4}}\right)^a.$$

Plugging in the definitions of p and a, and using the bound $1 - x \leqslant e^{-x}$, we find that

$$m p^{\frac{a}{4}} \leqslant m e^{-\sqrt{\varepsilon} \cdot \frac{a}{4}} = m e^{-2 \ln m} = \frac{1}{m}.$$

Hence, for sufficiently large m, we find that G_0 has no clique of order a with probability at least $\frac{2}{3}$.

Similarly, the probability that G_0 has an independent set of order b is at most

$$\binom{m}{b} (1-p)^{\binom{b}{2}} < m^b (1-p)^{\frac{b^2}{4}} = \left(m(1-p)^{\frac{b}{4}}\right)^b.$$

Plugging in the definitions of p and b, we see that

$$m(1-p)^{\frac{b}{4}} = m\left(\sqrt{\varepsilon}\right)^{\frac{b}{4}} = m \cdot \varepsilon^{\frac{b}{8}} = m e^{-2 \ln m} = \frac{1}{m}.$$

So the probability that G_0 has no independent set of order b is at least $\frac{2}{3}$ for sufficiently large m. Thus, for sufficiently large m, we can fix a graph G_0 with no clique of order a and no independent set of order b.

Now, fix $\varepsilon > 0$ a sufficiently large integer N. Let $m = \varepsilon N/2$, and let G be the disjoint union of N/m copies of the graph G_0 constructed above. Since G is a disjoint union, any clique in G must be a clique in some copy of G_0, hence the largest clique in G is of order at most

$$a = \frac{8 \ln m}{\sqrt{\varepsilon}} = \frac{8 \ln(\varepsilon N/2)}{\sqrt{\varepsilon}} < \frac{8 \ln N}{\sqrt{\varepsilon}} \leqslant \frac{8 \ln N}{\varepsilon \ln \frac{1}{\varepsilon}},$$

where the final inequality uses the fact that $\sqrt{x} \geqslant x \ln \frac{1}{x}$ for all $x \in (0,1)$.

On the other hand, every independent set in G is a disjoint union of independent sets, one from each copy of G_0. Thus, the order of the largest independent set in G is at most

$$\frac{N}{m} \cdot b = \frac{2b}{\varepsilon} = \frac{32 \ln m}{\varepsilon \ln \frac{1}{\varepsilon}} < \frac{32 \ln N}{\varepsilon \ln \frac{1}{\varepsilon}}.$$

The final thing to prove is that $d(G) \leqslant \varepsilon$. Indeed, the total number of edges in G is

$$e(G) = \frac{N}{m} e(G_0) \leqslant \frac{N}{m} \binom{m}{2} \leqslant \frac{Nm}{2}$$

and therefore

$$d(G) = \frac{e(G)}{\binom{N}{2}} \leqslant \frac{Nm/2}{N^2/4} = \frac{2m}{N} = \varepsilon. \qquad \square$$

8.2 The structure of C-Ramsey graphs

Although Theorem 8.1.4 was originally developed to prove bounds on omission Ramsey numbers, it has since become one of the most important tools in a rather different area of Ramsey theory, which is the study of C-Ramsey graphs.

Recall that the Erdős–Szekeres bound $r(k) \leqslant 4^k$ implies that every N-vertex graph G contains a clique or an independent set of order at least $\frac{1}{2}\log N$, where the logarithm is to base 2. On the other hand, Erdős's bound $r(k) \geqslant 2^{k/2}$ implies that there exist N-vertex graphs whose largest clique and independent set are both of order at most $2\log N$. However, as discussed in Chapter 2, essentially the only technique we have for finding such graphs involves randomness. So a natural (albeit vague) question is whether *all* graphs whose largest clique and independent set are both of order $O(\log N)$ are "random-like". To formalize this question, we make the following definition.

Definition 8.2.1. Let $C > 0$ be a real number. A *C-Ramsey graph* is a graph whose largest clique and independent set are both of order at most $C\log|V(G)|$.

In this definition, we should think of C as a fixed constant, and of G as a very large graph. In general, we are interested in proving that C-Ramsey graphs have certain structure, and in particular, that they have structure that is "similar" to that of a random graph.

We have already discussed one "random-like" notion, that of being ε-quasirandom. Unfortunately, this is a very strong definition, and it is much too strong to be true for all C-Ramsey graphs, as shown in the following proposition.

Proposition 8.2.2. *For every sufficiently large N, there exists an N-vertex 4-Ramsey graph which is not ε-quasirandom for any $\varepsilon \leqslant \frac{1}{5}$.*

Proof sketch. Let G_0 be a random graph, with edge probability $\frac{1}{2}$, on $N/2$ vertices. Let G be the disjoint union of two copies of G_0.

We recall from Theorem 2.2.2 that G_0 is 2-Ramsey with high probability. Since the largest clique in G is a largest clique in G_0, and since every independent set in G is the union of two independent sets, one from each copy of G_0, we conclude that G is 4-Ramsey.

Additionally, we know that with high probability, $d(G_0)$ is very close to $\frac{1}{2}$. Since G has no edges between the two copies of G_0, this implies that $d(G)$ is very close to $\frac{1}{4}$. Hence G has an induced subgraph on half of its vertices whose edge density deviates from that of G by roughly $\frac{1}{4}$. Thus, for sufficiently large N, we have that G is not ε-quasirandom for any ε bounded away from $\frac{1}{4}$; in particular it is not ε-quasirandom for any $\varepsilon \leqslant \frac{1}{5}$. □

ε-quasirandomness is a way of saying that the edges of a graph are very "well-distributed", and this condition is too strong to be true for all C-Ramsey graphs. However, weaker "edge well-distribution" results are true for C-Ramsey graphs; in fact, such results are immediate consequences of Theorem 8.1.4. There are many such results that one can state (and they are all proved in the same way); we stick with the following fairly simple statement, which will suffice for our later applications.

Theorem 8.2.3. *For every $C > 0$, there exists $\sigma > 0$ such that the following holds for all sufficiently large N. If G is an N-vertex C-Ramsey graph, then every $S \subseteq V(G)$ with $|S| \geqslant \sqrt{N}$ satisfies $\sigma \leqslant d(S) \leqslant 1 - \sigma$.*

Note that this "edge well-distribution" result is simultaneously weaker and stronger than ε-quasirandomness. It is substantially weaker, in that rather than saying that $d(S)$ is very close to $d(G)$, we only say that $d(S)$ is not too close to 0 or to 1. On the other hand, we obtain such a conclusion for sets as small as $|S| = \sqrt{N}$, whereas in ε-quasirandomness we are restricted to sets of size $|S| \geqslant \varepsilon N$. It turns out that in certain applications, weak estimates on $d(S)$ are still useful, especially if they hold for fairly small S.

Proof of Theorem 8.2.3. Let τ be the constant from Theorem 8.1.4. Pick $\sigma > 0$ to satisfy

$$\sigma \log \frac{1}{\sigma} < \frac{\tau}{2C},$$

and note that we can pick such a σ since the left-hand side tends to zero as $\sigma \to 0$.

Now let G be an N-vertex C-Ramsey graph, and suppose for contradiction that $S \subseteq V(G)$ satisfies $|S| \geqslant \sqrt{N}$ and $d(S) \notin [\sigma, 1-\sigma]$. Suppose first that $d(S) < \sigma$. Applying Theorem 8.1.4 to the induced subgraph $G[S]$ (which we may do since N, and thus $|S|$, is sufficiently large), we conclude that $G[S]$ contains a clique or an independent set of order at least

$$\frac{\tau}{\sigma \log \frac{1}{\sigma}} \log |S| \geqslant \frac{\tau}{\sigma \log \frac{1}{\sigma}} \log \left(\sqrt{N}\right) = \frac{\tau}{2\sigma \log \frac{1}{\sigma}} \log N > C \log N,$$

where the final inequality holds by our choice of σ. Now, we note that any clique or independent set in $G[S]$ is also a clique or independent set in G, contradicting the assumption that G is a C-Ramsey graph.

The other case is when $d(S) \geqslant 1 - \sigma$. We now apply Theorem 8.1.4 to the graph $\overline{G}[S]$, which satisfies $d(\overline{G}[S]) = 1 - d(S) \leqslant \sigma$. Then the exact same computation as above shows that $\overline{G}[S]$ contains a clique or an independent set of order greater than $C \log N$. But the complement of a clique is an independent set, and vice versa, so this in turn implies that $G[S]$, and thus G, contains a clique or an independent set of order greater than $C \log N$, again a contradiction. □

As a consequence of Theorem 8.2.3, we can prove the following result, which was first proved by Erdős and Hajnal [44]. It shows that C-Ramsey graphs do share another property in common with random graphs, namely that they contain all fixed graphs H as induced subgraphs. Given the tools we have already developed, this powerful and surprising result is almost immediate.

Theorem 8.2.4 (Erdős–Hajnal [44])**.** *For every $C > 0$ and every graph H, the following holds for all sufficiently large N. If G is an N-vertex C-Ramsey graph, then H is an induced subgraph of G.*

Proof. Let $\sigma = \sigma(C) > 0$ be the parameter from Theorem 8.2.3. Let $\delta > 0$ be the parameter from Theorem 6.3.3, applied with this choice of σ and H.

Let N be sufficiently large, and let G be an N-vertex C-Ramsey graph. Suppose for contradiction that G is induced-H-free. By Theorem 6.3.3, there exists some $S \subseteq V(G)$ with $|S| \geqslant \delta N$ such that $d(S) < \sigma$ or $d(S) > 1 - \sigma$. For sufficiently large N, we have that $\delta N \geqslant \sqrt{N}$. But this contradicts Theorem 8.2.3, which asserts that any S with $|S| \geqslant \sqrt{N}$ must satisfy $\sigma \leqslant d(S) \leqslant 1 - \sigma$. $\square$

Another way of phrasing Theorem 8.2.4 is as follows. Let us say that a graph is k-*universal* if it contains as an induced subgraph every graph H with at most k vertices. Then Theorem 8.2.4 can equivalently be stated as the fact that for every $C, k > 0$, every sufficiently large C-Ramsey graph is k-universal.

Over the years, stronger versions of Theorem 8.2.4 were proved, which give better control on how large N must be to ensure that every N-vertex C-Ramsey graph is k-universal. The optimal result was finally proved by Prömel and Rödl [106].

Theorem 8.2.5 (Prömel–Rödl [106])**.** *For every $C > 0$, there exists $c > 0$ such that the following holds. If G is an N-vertex C-Ramsey graph, then G is $(c \log N)$-universal.*

This is a pretty remarkable theorem! Indeed, the assumption is that G is C-Ramsey, meaning that G *does not* contain a clique or an independent set of order $C \log N$. The conclusion is then that G *does* contain as an induced subgraph *all* graphs of order $c \log N$. In particular, this result is best possible up to the value of c, since the assumption and conclusion contradict one another if we try to take $c \geqslant C$.

We will not prove Theorem 8.2.5 in this course, although the proof is actually not very hard. The idea is to apply the greedy embedding technique, rather than the regularity method, to prove a quantitatively stronger version of Theorem 6.3.3. At that point, one can combine it with Theorem 8.2.3 as we did above to conclude that every C-Ramsey graph is $(c \log N)$-universal. For more details, see [55, Section 2].

In recent years, there have been a number of further results on the "random-like" structure of C-Ramsey graphs, most of which we will not discuss. The most recent, however, is a breakthrough of Kwan–Sah–Sauermann–Sawhney [84], which in particular proved an old conjecture of Erdős–McKay [39].

Theorem 8.2.6 (Kwan–Sah–Sauermann–Sawhney [84])**.** *For every $C, \eta > 0$, the following holds for all sufficiently large N. Let G be an N-vertex C-Ramsey graph. Then for any integer $0 \leqslant x \leqslant (1 - \eta)e(G)$, there exists a subset $X \subseteq V(G)$ such that $G[X]$ has exactly x edges.*

The proof of Theorem 8.2.6 is quite complicated, and relies heavily on non-Ramsey-theoretic tools. In fact, somewhat remarkably, the proof uses very little about the structure of C-Ramsey graphs—results like Theorem 8.2.3 are essentially the only properties needed of C-Ramsey graphs in the proof of Theorem 8.2.6.

To end this section, we briefly mention a beautiful conjecture which remains open. Recall Proposition 8.2.2, which states that C-Ramsey graphs need not be quasirandom. However, we might hope that truly extremal Ramsey graphs—that is, those graphs with $r(k) - 1$ vertices and no clique or independent set of size k—actually are quasirandom. This is the content of the next conjecture, which was raised by Sós (see [128]).

Conjecture 8.2.7 (Sós)**.** *For every $\varepsilon > 0$, the following holds for all sufficiently large k. Let $N = r(k) - 1$, and let G be an N-vertex graph with no clique or independent set of order k. Then G is ε-quasirandom.*

This question is quite possibly very hard; in particular, it may be at least as hard as determining $\lim_{k\to\infty} \log r(k)$.

Chapter 9

The Hales–Jewett theorem

9.1 Van der Waerden's theorem

We will now turn our attention away from graph theory, and discuss Ramsey-theoretic results in other areas of mathematics. We have already encountered one such result—Schur's theorem, Theorem 1.1.2—already in Chapter 1. Recall that this theorem states that for any q-coloring of $[\![N]\!]$, where N is sufficiently large in terms of q, we can find a monochromatic solution to the equation $x + y = z$.

Schur's theorem is perhaps the most basic result in *additive Ramsey theory*. We will now discuss a related, and substantially more complicated, result, originally due to van der Waerden [140]. Recall that a *k-term arithmetic progression* (or *k-AP* for short) is a sequence of k integers, of the form

$$a, a + r, a + 2r, \ldots, a + (k - 1)r.$$

Theorem 9.1.1 (van der Waerden [140]). *For every $k, q \geqslant 2$, there exists some N such that the following holds. Any q-coloring of $[\![N]\!]$ contains a monochromatic k-AP.*

An equivalent, but perhaps pithier, statement is that any q-coloring of $\mathbb{N}$ contains arbitrarily long monochromatic arithmetic progressions.

The original proof of van der Waerden used a very clever, and intricate, double induction argument. For any fixed k, the result is proved simultaneously for all q, and the result for a fixed (k, q) is proved by using the validity of the result for $(k-1, q')$, where q' is an enormous number depending on k and q.

Eventually, Hales and Jewett [70] realized that van der Waerden's proof is not "really" about arithmetic progressions in the integers. They were able to adapt his proof and prove a similar result in a more abstract combinatorial setting. However, their result, now called the Hales–Jewett theorem, ends up being far more than a simple restatement of van der Waerden's theorem. Indeed, working in this more abstract setting allows one to immediately prove several other powerful results Ramsey-theoretic results in a variety of different settings. We will now state the Hales–Jewett theorem, and will then see a number of applications of it before proceeding with the proof.

We will work in the d-dimensional grid $[\![k]\!]^d$, of side length k. The most important object of study for us is a *combinatorial line*, which we now define.

Definition 9.1.2. A *combinatorial line* in $[\![k]\!]^d$ is a collection of k points $x^{(1)}, \ldots, x^{(k)} \in [\![k]\!]^d$ with the following property. For each coordinate $i \in [\![d]\!]$, either $x_i^{(1)} = \cdots = x_i^{(k)}$, or

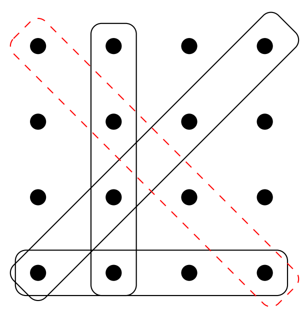
$$x_i^{(1)} = 1, x_i^{(2)} = 2, \ldots, x_i^{(k)} = k. \tag{9.1}$$

Additionally, we require $x^{(1)}, \ldots, x^{(k)}$ to be distinct elements of $[\![k]\!]^d$, which is equivalent to saying that (9.1) holds for at least one coordinate $i \in [\![d]\!]$.

Equivalently, a combinatorial line can be identified with a *root*, which is an element $\rho \in \{1, 2, \ldots, k, *\}^d$ with at least one $*$. Given a root ρ and $j \in [\![k]\!]$, we define $\rho(j) \in [\![k]\!]^d$ to be obtained from ρ by substituting the symbol j for every instance of $*$ in ρ. Then a combinatorial line can be obtained from ρ by setting

$$x^{(1)} = \rho(1), \ldots, x^{(k)} = \rho(k).$$

As an example, let $k = 4, d = 2$. Three examples of combinatorial lines, and one non-example (in red), are in the picture below:



Thus, for example, the bottom horizontal line contains the points $(1,1), (2,1), (3,1), (4,1)$, and corresponds to the root $*1$. The vertical line has points $(2,1), (2,2), (2,3), (2,4)$, and corresponds to the root $2*$. The diagonal line of slope 1 corresponds to the root $**$. Finally, the diagonal line of slope -1 (in red) is *not* a combinatorial line. The reason is that in any combinatorial line, the "moving coordinates" have to move in sync—every instance of $*$ in the root must be replaced with the *same* element of $[\![k]\!]$ to obtain a point.

With this setup, we are ready to state the Hales–Jewett theorem.

Theorem 9.1.3 (Hales–Jewett [70])**.** *For every $k, q \geqslant 1$, there exists some d such that, in any q-coloring $[\![k]\!]^d \to [\![q]\!]$, there is some monochromatic combinatorial line.*

Thus, rather than coloring the edges of a graph, as we were before, we are now coloring the points of the grid $[\![k]\!]^d$, and rather than looking for a monochromatic clique or subgraph, we are looking for a monochromatic combinatorial line.

We will defer the proof of Theorem 9.1.3 to Section 9.6. In the meantime, we will see a number of applications of this powerful theorem. It will be convenient to make the following definition, analogous to that of Ramsey numbers.

Definition 9.1.4. The *Hales–Jewett number* $\mathrm{HJ}(k;q)$ is the minimum d such that every q-coloring of $[\![k]\!]^d$ contains a monochromatic combinatorial line.

Just as Ramsey's theorem guarantees that $r(k)$ exists for all k, the Hales–Jewett theorem implies that $\mathrm{HJ}(k;q)$ is well-defined.

Our first application of Theorem 9.1.3 is a proof of van der Waerden's theorem, Theorem 9.1.1. The basic idea is that, if we write an integer in its base-k representation, then combinatorial lines in $[\![k]\!]^d$ yield arithmetic progressions in $\mathbb{N}$. For example, we saw above the combinatorial lines $\{(1,1),(2,1),(3,1),(4,1)\}$; $\{(2,1),(2,2),(2,3),(2,4)\}$; and $\{(1,1),(2,2),(3,3),(4,4)\}$. If we view each pair (x,y) as a base-10 number by concatenating the entries, we get the arithmetic progressions $\{11,21,31,41\};\{21,22,23,24\}$; and $\{11,22,33,44\}$. This simple idea is almost a complete proof; the details are below.

Proof of Theorem 9.1.1. Let $d = \mathrm{HJ}(k;q)$. We define a function $f : [\![k]\!]^d \to \mathbb{N}$ by viewing an element of $[\![k]\!]^d$ as the base-k representation[1] of an integer; formally, we define

$$f(x_1,\dots,x_d) \coloneqq \sum_{i=1}^d x_i k^{i-1}.$$

Note that f is injective, and that its image is contained in $[\![N]\!]$, where $N = k^{d+1}$. We claim that this choice of N works in Theorem 9.1.1, that is, that every q-coloring of $[\![N]\!]$ contains a monochromatic k-AP.

Indeed, fix a q-coloring $\chi : [\![N]\!] \to [\![q]\!]$. By composing with f, we obtain a q-coloring $\psi : [\![k]\!]^d \to [\![q]\!]$, defined by $\psi(x) = \chi(f(x))$. By the definition of d, there exists a monochromatic combinatorial line in this coloring, say $x^{(1)},\dots,x^{(k)}$.

We claim that $f(x^{(1)}),\dots,f(x^{(k)})$ is a monochromatic k-AP under χ. First, by the way we defined ψ, we certainly have that this sequence is monochromatic, so it suffices to prove that it is an arithmetic progression. But indeed, for any $1 \leqslant j \leqslant k-1$, we have that

$$f(x^{(j+1)}) - f(x^{(j)}) = \sum_{i=1}^d (x_i^{(j+1)} - x_i^{(j)})k^{i-1} = \sum_{i\in M} k^{i-1},$$

where $M \subseteq [\![d]\!]$ denotes the set of "moving coordinates" in the combinatorial line, or equivalently the set of indices where the root has the symbol $*$. Indeed, in every coordinate $i \in M$ we have $x_i^{(j+1)} - x_i^{(j)} = 1$, and in every coordinate $i \notin M$ we have $x_i^{(j+1)} = x_i^{(j)}$. Note that M is non-empty since we insist that our combinatorial lines are non-constant. Hence if we set $r \coloneqq \sum_{i\in M} k^{i-1}$, we conclude that $f(x^{(1)}),\dots,f(x^{(k)})$ is a monochromatic k-AP with common difference $r \neq 0$. □

If we define the *van der Waerden* number $W(k;q)$ to be the least N such that every q-coloring of $[\![N]\!]$ contains a monochromatic k-AP, this proof shows that

$$W(k;q) \leqslant k^{\mathrm{HJ}(k;q)+1}. \tag{9.2}$$

[1] Strictly speaking, we first reverse the vector before viewing it as the base-k representation. Also, rather than using the "digits" $0,\dots,k-1$, we use the digits $1,\dots,k$.

As we will shortly discuss, this is not a very good bound, but it was essentially the best known for many years.

9.2 The Gallai–Witt theorem

We will now discuss a higher-dimensional generalization of van der Waerden's theorem, due independently to Gallai[2] (quoted in [107]) and to Witt [144].

Let $S \subseteq \mathbb{Z}^t$ be a finite subset of the t-dimensional integer lattice. A *homothetic copy* of S is a set of the form

$$a + r \cdot S := \{a + r \cdot s : s \in S\},$$

where $a \in \mathbb{Z}^t$ and $r \in \mathbb{N}$ is a positive integer. In other words, we obtain a homothetic copy by translating and dilating S, but not rotating it. Note that if $t = 1$ and $S = [\![k]\!]$, then a homothetic copy of S is simply a k-AP. Thus, the following result is a natural generalization of Theorem 9.1.1 to arbitrary dimensions.

Theorem 9.2.1 (Gallai–Witt [107, 144])**.** *For every finite $S \subseteq \mathbb{Z}^t$, and $q \geqslant 1$, there exists some N such that any q-coloring of $[\![N]\!]^t$ contains a monochromatic homothetic copy of S.*

To get a sense of how astonishing this theorem is, note that every pixel in a computer screen can display one of 2^{24} colors. Thus, applying Theorem 9.2.1, we conclude that any sufficiently large computer screen, regardless of what it displays, contains a monochromatic scaled copy of the words "Ramsey theory", in this font (but note that *scaling* means that the pixels making up the letters may be very far apart).

Proof of Theorem 9.2.1. Let $S = \{s_1, \ldots, s_k\}$, and let $d = \mathrm{HJ}(k; q)$. By translating S, we may assume that all coordinates of all the points s_i are positive (and any homothetic copy of a translate of S is a homothetic copy of S, so we lose nothing by doing this translation).

We define a function $g : [\![k]\!]^d \to \mathbb{N}^t$ by

$$g(x_1, \ldots, x_d) = \sum_{i=1}^{d} s_{x_i}.$$

Thus, for example, if $x = (3, 1, 2, 2, 1, 2)$, then

$$g(x) = s_3 + s_1 + s_2 + s_2 + s_1 + s_2 = 2s_1 + 3s_2 + s_3.$$

In other words, g just counts how many entries of x are equal to 1, adds up that many copies of s_1, then does the same for $2, 3, \ldots, k$. Note that $s_1, \ldots, s_k$ are elements of $\mathbb{N}^t$, so addition here denotes vector addition. g is not injective, but as we will see, this does not matter[3].

[2] In [107], this result is attributed to Grünwald, which was Gallai's birth name.

[3] The astute reader may have already noticed that we never used the injectivity of f in the proof of Theorem 9.1.1.

Let $N = d \cdot \max\{\|s_1\|_\infty, \ldots, \|s_k\|_\infty\}$, and note that the image of g is contained in $[\![N]\!]^t$. Note too that N depends only on S and on d, which in turn depends only on S and q. We claim that this choice of N suffices, so we fix some coloring $\chi : [\![N]\!]^t \to [\![q]\!]$, and wish to find a monochromatic homothetic copy of S under χ.

We define $\psi : [\![k]\!]^d \to [\![q]\!]$ by $\psi(x) = \chi(g(x))$. By the choice of d, we can find a monochromatic combinatorial line $x^{(1)}, \ldots, x^{(k)}$ under ψ. Let $M \subseteq [\![d]\!]$ denote the set of moving coordinates in this combinatorial line. For every $i \notin M$, there exists some $x_i \in [\![k]\!]$ such that $x_i^{(j)} = x_i$ for all j. Define

$$a = \sum_{i \notin M} s_{x_i}.$$

Now, for any $j \in [\![k]\!]$, we have that

$$g(x^{(j)}) = \sum_{i=1}^{d} s_{x_i^{(j)}} = \sum_{i \notin M} s_{x_i^{(j)}} + \sum_{i \in M} s_{x_i^{(j)}} = a + |M| s_j, \tag{9.3}$$

where the final equality uses the definition of a and the fact that $x_i^{(j)} = j$ for every $i \in M$, by the definition of a combinatorial line—in every moving coordinate, we plug in the value j to obtain $x^{(j)}$. Let $r \coloneqq |M|$, and note that $r > 0$ since M is non-empty. But then (9.3) guarantees that

$$\{g(x^{(1)}), \ldots, g(x^{(k)})\} = \{a + rs_1, \ldots, a + rs_k\} = a + rS,$$

hence these points form a homothetic copy of S in $[\![N]\!]^t$. The final observation is that, by our definition of ψ, all these points receive the same color under χ, completing the proof. □

As discussed above, Theorem 9.2.1 contains Theorem 9.1.1 as a special case. However, if one specializes the proof above to the setting where $S = [\![k]\!] \subseteq \mathbb{Z}$, one actually gets a (slightly) different proof of Theorem 9.1.1! In particular, this proof shows that

$$W(k; q) \leqslant k \cdot \mathrm{HJ}(k; q),$$

which is a different (and better) bound than that given in (9.2).

Just as van der Waerden's theorem is one of the most basic results in the rich field of arithmetic Ramsey theory, the Gallai–Witt theorem is one of the most basic results in the rich field of Euclidean Ramsey theory, a topic which we will not discuss. But we simply state the following theorem, whose proof is identical to that of Theorem 9.2.1.

Theorem 9.2.2. *For every finite $S \subseteq \mathbb{R}^t$ and every $q \geqslant 2$, every q-coloring $\chi : \mathbb{R}^t \to [\![q]\!]$ contains a monochromatic homothetic copy of S.*

9.3 Monochromatic subset sums

Recall Schur's theorem, Theorem 1.1.2; it states that if N is sufficiently large, then any q-coloring of $[\![N]\!]$ contains a monochromatic solution to $x + y = z$. Equivalently, it says that we can find $x, y \in [\![N]\!]$ such that x, y, and $x + y$ all have the same color. A natural generalization of this is the following result, proved independently by Folkman (quoted in [68]), Rado [108], and Sanders [121].

Theorem 9.3.1 (Folkman–Rado–Sanders [68, 108, 121])**.** *For every $m, q \geqslant 1$, there exists N such that in any q-coloring of $[\![N]\!]$, there exist distinct $x_1, \ldots, x_m \in [\![N]\!]$ such that all the subset sums $\sum_{i \in I} x_i$, for $\varnothing \neq I \subseteq [\![m]\!]$, receive the same color.*

Note that Schur's theorem is simply the $m = 2$ case of Theorem 9.3.1. In the first homework, you were asked to prove a weaker statement, which guarantees the same conclusion but only for *subintervals* I.

We will prove Theorem 9.3.1 as a consequence of van der Waerden's theorem. We will need the following lemma, which is an instance of what is sometimes called a *canonization result* (or a *canonical Ramsey* result). Such results do not guarantee monochromatic substructures, but rather substructrues where the coloring is constrained in some way. We will also (implicitly) use such a lemma in the proof of Theorem 9.1.3, and in the proof of the hypergraph Ramsey theorem.

If $x_1, \ldots, x_t$ are integers and $I \subseteq [\![t]\!]$ is a set, we use the notation $x(I)$ to denote $\sum_{i \in I} x_i$.

Lemma 9.3.2. *For every $t, q \geqslant 1$, there exists $M = M(t; q)$ such that the following holds for every q-coloring of $[\![M]\!]$. There exist $x_1 < x_2 < \cdots < x_t$ such that, for all $\varnothing \neq I \subseteq [\![t]\!]$, the color of $x(I)$ depends only on* $\max I$*.*

In other words, this lemma guarantees that the numbers $x_3, x_1 + x_3, x_2 + x_3, x_1 + x_2 + x_3$ all have the same color, but it says nothing about whether these have the same color as, say, $x_1 + x_2$.

Proof. For any fixed q, we proceed by induction on t. The base case $t = 1$ is trivial (we may simply take $M(1; q) = 1$). Inductively, suppose we have proved the existence of $M(t - 1; q)$. Define

$$k \coloneqq tM(t-1; q) \qquad \text{and} \qquad M(t; q) \coloneqq 2W(k + 1; q),$$

where $W(k; q)$ is the van der Waerden number. For brevity, we write $M = M(t; q)$.

Now, fix a q-coloring $\chi : [\![M]\!] \to [\![q]\!]$. Since $\frac{M}{2} = W(k; q)$, we can apply Theorem 9.1.1 to the coloring restricted to† $\{\frac{M}{2} + 1, \ldots, M\}$ to find a monochromatic $(k + 1)$-AP, that is, a sequence

$$a, a + r, a + 2r \ldots, a + (k - 1)r, a + kr$$

such that all these numbers receive the same color under χ.

We now define $\chi' : [\![M(t - 1; q)]\!] \to [\![q]\!]$ by

$$\chi'(x) \coloneqq \chi(xr).$$

By the inductive hypothesis, we can find $y_1, \ldots, y_{t-1} \in [\![M(t - 1; q)]\!]$ such that the color of $y(I)$ under χ' depends only on $\max I$, for all $\varnothing \neq I \subseteq [\![t - 1]\!]$. Defining

$$x_i = ry_i,$$

we find that $x_1 < \cdots < x_{t-1}$, and the color $\chi(x(I))$ depends only on $\max I$, for all $\varnothing \neq I \subseteq [\![t-1]\!]$.

We have that $y_{t-1} \leqslant M(t-1;q) \leqslant k$ and $r \leqslant \frac{M}{2k}$, which imply that $x_{t-1} = ry_{t-1} \leqslant \frac{M}{2}$. We define $x_t \coloneqq a$, and thus have that $x_1 < \cdots < x_{t-1} < x_t$. We claim that this sequence satisfies the desired property. Indeed, we already know that for all $\varnothing \neq I \subseteq [\![t-1]\!]$, we have that $\chi(x(I))$ depends only on $\max I$. So it suffices to consider those I with $\max I = t$. But in that case we have that

$$x(I) = x_t + \sum_{i \in I \setminus \{t\}} x_i = x_t + r \sum_{i \in I \setminus \{t\}} y_i.$$

The final observation is that $\sum_{i \in I \setminus \{t\}} y_i \leqslant tM(t-1;q) = k$. Hence, this final quantity is one of the elements in our monochromatic k-AP. In particular, all such elements have the same color, namely the common color of the arithmetic progression. □

With this lemma, Theorem 9.3.1 is fairly straightforward.

Proof of Theorem 9.3.1. Let $N = M(mq;q)$, where M is the quantity from Lemma 9.3.2, and fix a q-coloring of $[\![N]\!]$. By Lemma 9.3.2, there exist $x_1 < \cdots < x_{mq}$ such that $x(I)$ depends only on $\max I$. Among these mq numbers, at least m of them must receive the same color, say $x_{j_1}, ..., x_{j_m}$. But then for every $\varnothing \neq I \subseteq [\![m]\!]$, we have that the color of $\sum_{i \in I} x_{j_i}$ equals the color of $x_{j_{\max I}}$, and all of these colors are equal. Hence all non-empty subset sums of $x_{i_1}, \ldots, x_{i_m}$ receive the same color. □

†More formally, we can obtain a coloring $\chi' : [\![M]\!] \to [\![q]\!]$ by defining $\chi'(x) \coloneqq \chi(x + \frac{M}{2})$, and apply Theorem 9.1.1 to this new coloring. The point, which we will use again later in this proof, is that if we translate or dilate an arithmetic progression, we end up with another arithmetic progression; hence we are always free to apply Theorem 9.1.1 to colorings of translations and dilations of $[\![W(k+1;q)]\!]$.

9.4 Communication complexity

We now present a beautiful application of the Gallai–Witt theorem in theoretical computer science, more precisely in the subfield of communication complexity.

In the most simple communication complexity question, there are two players, Alice and Bob. Alice receives an input $a \in [\![N]\!]$, Bob receives an input $b \in [\![N]\!]$, and their goal is to evaluate some function $f : [\![N]\!] \times [\![N]\!] \to \{0,1\}$, that is, to compute $f(a,b) \in \{0,1\}$. In contrast to typical complexity-theoretic problems, we are not interested in the *computational* difficulty of this problem; we assume that Alice and Bob are all-powerful, and can perform arbitrary computations instantaneously. Insted, we are interested in the *communication complexity*: Alice and Bob would like to communicate as few bits as possible to one another before determining the value of $f(a,b)$.

More precisely, Alice and Bob can agree on a *protocol* before the game starts. Once they receive their inputs, they start transmitting bits to one another according to the protocol; crucially, their decisions can depend on their inputs, as well as on the already-transmitted bits. At the end of the protocol, one of the two players must announce an answer, and the protocol is successful if the answer equals $f(a,b)$ for all inputs a, b. The *complexity* of the protocol is the number of bits transmitted in the worst case (that is, the maximum over all

a, b of the number of bits transmitted), and the *communication complexity* of f is defined as the minimum complexity of any protocol computing f.

Note that any function f has communication complexity at most $\lceil \log N \rceil$. Indeed, by using $\lceil \log N \rceil$ bits, Alice can simply send Bob her entire input $a \in [\![N]\!]$. Then Bob has all the information, so he can compute $f(a, b)$, and announce the (correct) answer. For some functions, such as the *equality* function which returns 1 if and only if $a = b$, one can show that this bound is actually best possible, that is, that no protocol can compute equality by transmitting fewer than $\lceil \log N \rceil$ bits. However, for other functions, there are (often extremely clever and involved) protocols that do much better than this simple one. For more on the basics of communication complexity, see the book [83].

We will instead be focused on *multiparty* communication complexity, where there are $t \geqslant 3$ players. Our inputs now come from $[\![N]\!]^t$, and our players wish to compute some function $f : [\![N]\!]^t \to \{0, 1\}$. The most natural generalization of the two-party model is to give the ith player the input $a_i \in [\![N]\!]$. However, it turns out that a more useful and interesting model is the so-called *number on the forehead* model, introduced by Chandra, Furst, and Lipton [15]. In this model, we imagine that the ith player holds the input a_i on her forehead; thus, she has access to all the inputs a_j *except* for a_i. Note that in the case of two players, this is really the same as the model we discussed above, since each of the two players knows the value of one of the two inputs. However, once the number of players is at least three, things become substantially more complicated, since now each pair of players has some amount of shared knowledge. In particular, this is a more powerful model of computation, and hence one for which it is substantially more difficult to prove lower bounds. As it turns out, this model also has connections to many other important topics in complexity theory, and lower bounds on this model can be used to give lower bounds for other computational models such as branching programs. In general, proving lower bounds on computational problems is arguably the most important problem in complexity theory (the P vs. NP problem is the most famous example of such a question—it is equivalent to proving a super-polynomial lower bound on the complexity of checking whether a graph is 3-colorable, say). As the most general form of this problem is extremely hard, there is a great deal of interest in proving such results for certain restricted models of computation, such as branching programs.

The following result, due to Chandra, Furst, and Lipton [15] gives a lower bound (albeit a rather modest one) for the multiparty communication complexity of a natural function in the number on the forehead model. Define the *Exactly-N* function $f_N : [\![N]\!]^t \to \{0, 1\}$ to take the value 1 on input $(x_1, \ldots, x_t) \in [\![N]\!]^t$ if and only if $x_1 + \cdots + x_t = N$.

Theorem 9.4.1 (Chandra–Furst–Lipton [15])**.** *There is no constant-communication protocol for f_N in the number on the forehead model.*

More precisely, for every $t \geqslant 3$ and every constant $C \geqslant 1$, the following holds for sufficiently large N. There is no communication protocol using at most C bits of communication that correctly computes f_N on all inputs $(x_1, \ldots, x_t) \in [\![N]\!]^t$.

Proof. Consider a protocol P using at most C bits of communication. This protocol produces a *transcript*, which is simply the ordered list of all communicated bits, as well as which of the t players said them. Note that there are at most $q \coloneqq (2t)^C$ possible transcripts; for each of C positions, we have t choices for which player is speaking and 2 choices for which bit they communicate. Let $T_1, \dots, T_q$ be the list of all possible transcripts.

Thus, given the protocol P, we can define a coloring $\chi_P : [\![N]\!]^t \to [\![q]\!]$ as follows: for each input $(x_1, \dots, x_t) \in [\![N]\!]^t$, a transcript is produced by P, and we color $(x_1, \dots, x_t)$ with color i if the transcript produced is T_i.

We now define a coloring $\psi : [\![N/t]\!]^{t-1} \to [\![q]\!]$ by setting

$$\psi(x_1, \dots, x_{t-1}) \coloneqq \chi_P(x_1, \dots, x_{t-1}, N - x_1 - \dots - x_{t-1}).$$

Let $S \subseteq \mathbb{Z}^{t-1}$ consist of the zero vector, as well as each of the $t-1$ standard basis vectors. By Theorem 9.2.1, as long as N/t is sufficiently large with respect to t and q (or equivalently, N is sufficiently large with respect to t and C), there is a homothetic copy of S in $[\![N/t]\!]^{t-1}$ which is monochromatic under ψ. That is, there exist $a = (a_1, \dots, a_{t-1}) \in [\![N/t]\!]^{t-1}$ and $r \geqslant 1$ such that

$$\psi(a_1, \dots, a_{t-1}) = \psi(a_1 + r, a_2, \dots, a_{t-1}) = \psi(a_1, a_2 + r, \dots, a_{t-1}) = \dots = \psi(a_1, \dots, a_{t-1} + r).$$

If we let $s = a_1 + \dots + a_{t-1}$, this says that the protocol P produces *exactly* the same transcript on the inputs

$$\begin{aligned} I_1 &\coloneqq (a_1 + r, a_2, \dots, a_{t-1}, N - s - r), \\ I_2 &\coloneqq (a_1, a_2 + r, \dots, a_{t-1}, N - s - r), \\ &\vdots \\ I_{t-1} &\coloneqq (a_1, a_2, \dots, a_{t-1} + r, N - s - r), \\ I_t &\coloneqq (a_1, \dots, a_{t-1}, N - s). \end{aligned}$$

The key claim is that P also produces the same transcript on the input $I_0 \coloneqq (a_1, \dots, a_{t-1}, N - s - r)$. Indeed, from the perspective of player i, input I_0 is indistinguishable from input I_i. Thus, whoever speaks first will act the same whether the input is I_0 or one of the inputs $I_i, i \in [\![t]\!]$. Then whoever speaks next still can't distinguish I_0 from one of the I_i (and the first bit communicated doesn't help), so the next player to speak will also act the same. Continuing in this fashion, we see that P will produce exactly the same transcript on I_0 as on each of the $I_i, i \in [\![t]\!]$.

In particular, at the end of the process, whoever announces the answer will announce the *same* answer on input I_0 as they would have on each of the inputs $I_i, i \in [\![t]\!]$. But this shows that the protocol P does not correctly compute f_N! Indeed, $f_N(I_i) = 1$ for all $i \in [\![t]\!]$, whereas $f_N(I_0) = 0$, since

$$a_1 + \dots + a_{t-1} + (N - s - r) = s + (N - s - r) = N - r \neq N.$$

This shows that any protocol P using at most C bits of communication cannot correctly compute f_N if N is sufficiently large. □

We remark that the connection to the Gallai–Witt theorem is not simply an artifact of the proof. On the homework, you will see that the communication complexity of f_N is in fact very closely tied to the minimum number of colors needed to color $[\![N]\!]^t$ while avoiding monochromatic homothetic copies of the set S above.

9.5 The induced Ramsey theorem for bipartite graphs

Recall that a graph G is q-color induced Ramsey for a graph H if every q-coloring of $E(G)$ contains a monochromatic copy of H, which is moreover an induced subgraph of G; the induced Ramsey theorem, Theorem 7.2.2, states that such a G exists for every graph H. As we discussed in Chapter 7, there are basically two techniques for proving "restricted Ramsey" results such as the induced Ramsey theorem: an approach using random graphs, and an approach using extremely large explicit graphs which are defined in some recursive way. In the latter family, the *partite construction* of Nešetřil–Rödl is the most flexible. One of the key building blocks in the partite construction is the induced Ramsey theorem for *bipartite* H.

Theorem 9.5.1. *For every bipartite graph H and every $q \geqslant 2$, there exists a bipartite graph G such that G is q-color induced Ramsey for H.*

Note that Theorem 9.5.1 does not follow from Theorem 7.2.2, since we are also requiring G to be bipartite. However, one can adapt the technique we used to prove Theorem 7.2.2 to prove Theorem 9.5.1 as well—one simply sets G to be a random bipartite graph[4]. One can also prove Theorem 9.5.1 directly from Ramsey's theorem (see e.g. [33, Lemma 9.3.3]). However, as observed by Nešetřil and Rödl [99], there is also a short proof of Theorem 9.5.1 from the Hales–Jewett theorem.

Proof of Theorem 9.5.1. Let H have vertex parts A, B and edge set $E \subseteq A \times B$. Let $k = |E|$, and let $d = \mathrm{HJ}(k; q)$.

We define the bipartite graph G as follows. Its vertex parts are A^d, B^d, and two vertices are adjacent if and only if they are adjacent in every coordinate, that is

$$((a_1, \ldots, a_d), (b_1, \ldots, b_d)) \in E(G) \qquad \Longleftrightarrow \qquad (a_i, b_i) \in E(H) \text{ for all } i \in [\![d]\!].$$

Note that we can naturally identify $E(G)$ with E^d, namely by saying that the edge $(e_1, \ldots, e_d)$ joins $(a_1, \ldots, a_d)$ to $(b_1, \ldots, b_d)$ in G if and only if e_i joins a_i to b_i in H for all $i \in [\![d]\!]$. Equivalently, if we let $a(e), b(e)$ denote the endpoints of e in A, B, respectively, then the edge $(e_1, \ldots, e_d)$ joins $(a(e_1), \ldots, a(e_d))$ to $(b(e_1), \ldots, b(e_d))$.

Now, fix a q-coloring of $\chi : E(G) \to [\![q]\!]$. By the above, we can view this as a q-coloring of E^d. By our choice of $d = \mathrm{HJ}(k; q)$, we know that this coloring contains a monochromatic

[4]One also needs a bipartite version of the quasirandom subset lemma—simply applying Lemma 6.1.4 directly to a coloring of G will not work (exercise: why not?). Luckily, proving such a bipartite statement turns out to be substantially easier than proving Lemma 6.1.4 [103].

combinatorial line $E' = \{e^{(1)}, \dots, e^{(k)}\}$; it remains to understand what this combinatorial line "means", and to see that it corresponds to a monochromatic induced copy of H.

Let $M \subseteq [\![d]\!]$ be the set of moving coordinates of the combinatorial line, and let e_i, for $i \notin M$, be the edge of H such that $e_i^{(j)} = e_i$ for all j. Let $e^{(j)}$ join the vertices $a^{(j)} \in A^d, b^{(j)} \in B^d$. Note that for every $i \notin M$, all vectors $a^{(j)}$ (resp. $b^{(j)}$) have the same ith coordinate, namely $a(e_i)$ (resp. $b(e_i)$). On the other hand, $a^{(j)}$ has the same entry in all coordinates in M, namely the A-endpoint of the jth edge of H, since all coordinates $i \in M$ of $e^{(j)}$ have this edge as their ith entry.

Therefore, if we set $A' = \{a^{(j)} : j \in [\![k]\!]\}, B' \coloneqq \{b^{(j)} : j \in [\![k]\!]\}$, we find that (A', B', E') is a copy of H in G. Moreover, by construction, all edges of this copy receive the same color under χ, since E' is a monochromatic combinatorial line. All that remains is to note that this copy is induced. Indeed, suppose that $f = (f_1, \dots, f_d) \in E(G)$ joins some $a^{(j)} \in A'$ to some $b^{(\ell)} \in B'$. Then for every $i \notin M$, we must have that f_i joins a_i to b_i, hence $f_i = e_i$. On the other hand, all M coordinates of $a^{(j)}$ are equal, as are all M coordinates of $b^{(\ell)}$. Hence, all M coordinates of f must be equal. But these properties characterize the elements of E'—the elements of E' are precisely those elements of E^d whose M coordinates are equal, and whose ith coordinate equals e_i for all $i \notin M$. This implies that $f \in E'$, and therefore the H-copy is indeed induced in G, as $A' \times B'$ contains no edges of G besides those already in E'. □

We remark that this proof demonstrates the power of the abstract setting of the Hales–Jewett theorem. Indeed, the restriction to combinatorial lines may seem arbitrary at first—if we're working in $[\![k]\!]^d$, which is naturally contained in $\mathbb{R}^d$, why not consider all *geometric* lines? The reason is that the greater abstraction, and the restriction to combinatorial lines, allows us to apply Theorem 9.1.3 to settings such as this one, where there are no natural geometric lines.

9.6 Proof of the Hales–Jewett theorem

Finally, we turn to the proof of Theorem 9.1.3. We use the following notation in the proof (and nowhere else): given sets X, Y and elements $x \in X, y \in Y$, we denote an element of $X \times Y$ by $x \times y$, rather than the more common (x, y). The reason is that we will really be considering products of multiple spaces, and it will be more convenient to use the $\times$ notation than to nest several layers of parentheses.

Proof of Theorem 9.1.3. We proceed by induction on k, where for every fixed k we prove the statement for all q. The base case[5] is $k = 1$, which is easy since we may take $\mathrm{HJ}(1; q) = 1$ for any q. We now assume that the statement is proven for $k - 1$, and wish to prove it for k.

We let $d_1, d_2, \dots, d_q$ be an extremely rapidly increasing sequence of numbers, where each d_i is chosen to be very large relative to $d_1, \dots, d_{i-1}$. Concretely, we set $d_1 = \mathrm{HJ}(k - 1; q)$,

[5] If you don't like this base case, on the homework you will show by an elementary argument that $\mathrm{HJ}(2; q)$ is finite for all q, and thus we can use that as our base case.

and let

$$d_i = \mathrm{HJ}\left(k-1; q^{k^{d_1+\cdots+d_{i-1}}}\right)$$

for every $2 \leqslant i \leqslant q$. Let $d = d_1 + \cdots + d_q$, and fix a coloring $\chi : [\![k]\!]^d \to [\![q]\!]$.

Throughout the proof, we think of $[\![k]\!]^d$ as a product set $[\![k]\!]^{d_1} \times \cdots \times [\![k]\!]^{d_q}$. Note that each $z \in [\![k]\!]^{d_q}$ is the "suffix" of $k^{d_1+\cdots+d_{q-1}}$ elements of $[\![k]\!]^d$, namely each $y \in [\![k]\!]^{d_1+\cdots+d_{q-1}}$ yields an element $y \times z \in [\![k]\!]^d$ with suffix z. We now define a coloring χ_q of $[\![k]\!]^{d_q}$ with $q^{k^{d_1+\cdots+d_{q-1}}}$ colors by coloring z with the list of all colors of all vectors in $[\![k]\!]^d$ whose suffix is z. That is, we define

$$\chi_q(z) \coloneqq \left(\chi(y \times z) : y \in [\![k]\!]^{d_1+\cdots+d_{q-1}}\right).$$

Note that we color z by a tuple of length $k^{d_1+\cdots+d_{q-1}}$, and each entry of this tuple is one of the q original colors, hence we use $q^{k^{d_1+\cdots+d_{q-1}}}$ colors in total.

Now, consider the subgrid $[\![k-1]\!]^{d_q} \subseteq [\![k]\!]^{d_q}$. We picked $d_q = \mathrm{HJ}(k-1; q^{k^{d_1+\cdots+d_{q-1}}})$, and χ_q gives us a coloring of this subgrid, so we conclude that there is a combinatorial line in $[\![k-1]\!]^{d_q}$ which is monochromatic under χ_q. Let ρ_q be the root of this combinatorial line. Let $L_q \subseteq [\![k]\!]^{d_q}$ be the combinatorial line corresponding to ρ_q; that is, we extend the original combinatorial line in $[\![k-1]\!]^{d_q}$ by adding to it one more element, namely the element $\rho_q(k)$ obtained by substituting k for every $*$ in ρ_q.

Note that L_q is not necessarily monochromatic under χ_q; all we know is that its first $k-1$ points $\rho_q(1), \ldots, \rho_q(k-1)$ receive the same color under χ_q. However, we do know the following property: for every $y \in [\![k]\!]^{d_1+\cdots+d_{q-1}}$,

$$\text{the color } \chi(y \times \rho_q(j)) \text{ is independent of } j, \text{ unless } j = k. \tag{9.4}$$

Indeed, by the definition of χ_q and the fact that $\rho_q(1), \ldots, \rho_q(k-1)$ have the same color under χ_q, we have that

$$\chi(y \times \rho_q(1)) = \cdots = \chi(y \times \rho_q(k-1))$$

for all $y \in [\![k]\!]^{d_1+\cdots+d_{q-1}}$. This is precisely the statement (9.4), saying that the only way j can influence $\chi(y \times \rho_q(j))$ is by whether $j = k$ or not.

We now restrict our attention to the subspace $[\![k]\!]^{d_1} \times \cdots \times [\![k]\!]^{d_{q-1}} \times L_q$. We define a coloring χ_{q-1} of $[\![k]\!]^{d_{q-1}}$ with $q^{k^{d_1+\cdots+d_{q-2}}}$ colors by setting

$$\chi_{q-1}(z) = \left(\chi(y \times z \times \rho_q(1)) : y \in [\![k]\!]^{d_1+\cdots+d_{q-2}}\right).$$

Again by our choice of $d_{q-1} = \mathrm{HJ}(k-1; q^{k^{d_1+\cdots+d_{q-2}}})$, we can find a combinatorial line in $[\![k-1]\!]^{d_{q-1}}$ which is monochromatic under χ_{q-1}. Let ρ_{q-1} be its root, and let $L_{q-1} \subseteq [\![k]\!]^{d_{q-1}}$ be obtained by extending it to $\rho_{q-1}(k)$. Arguing as above, and using (9.4), we conclude that for every $y \in [\![k]\!]^{d_1+\cdots+d_{q-2}}$,

$$\text{the color } \chi(y \times \rho_{q-1}(j) \times \rho_q(j')) \text{ is} \left\{ \begin{array}{ll} \text{independent of } j, j', & \text{if } j, j' < k, \text{ and} \\ \text{independent of } j', & \text{if } j = k \text{ and } j' < k. \end{array} \right\}$$

We now continue in this fashion all the way down. We eventually end up with combinatorial lines $L_1, \dots, L_q$, with roots $\rho_1, \dots, \rho_q$, respectively, such that the following holds. For any point $\rho_1(j_1) \times \cdots \times \rho_q(j_q) \in L_1 \times \cdots \times L_q \subseteq [\![k]\!]^d$, its color

$$\chi(\rho_1(j_1) \times \cdots \times \rho_q(j_q)) \text{ is } \left\{ \begin{array}{ll} \text{independent of } j_1, \dots, j_q, & \text{if } j_1, \dots, j_q < k, \\ \text{independent of } j_2, \dots, j_q, & \text{if } j_1 = k \text{ and } j_2, \dots, j_q < k, \\ \text{independent of } j_3, \dots, j_q, & \text{if } j_1 = j_2 = k \text{ and } j_3, \dots, j_q < k, \\ & \vdots \\ \text{independent of } j_q, & \text{if } j_1 = \cdots = j_{q-1} = k \text{ and } j_q < k, \\ \text{some color}, & \text{if } j_1 = \cdots = j_q = k. \end{array} \right\} \tag{9.5}$$

An equivalent way of saying this condition is the following. Let us say that two sequences in $[\![k]\!]^q$ are *friendly* if they begin with some sequence of k's of the same length, and afterwards never use the symbol k. That is, $(j_1, \dots, j_q), (j'_1, \dots, j'_q) \in [\![k]\!]^q$ are friendly if there exists $0 \leqslant t \leqslant q$ such that $j_1 = j'_1 = \cdots = j_t = j'_t = k$, but $j_\ell, j'_\ell < k$ for all $\ell > t$. Thus, for example, the sequences $(k, k, 1, 3)$ and $(k, k, 7, 11)$ are friendly, whereas the sequences $(k, k, 1, 3)$ and $(k, k, k, 4)$ are not. Sequences like $(1, 3, k, 4)$ or $(k, k, 2, k)$—which have an instance of k after an inital segment—are not friendly with any sequence. Then an equivalent way of saying the condition in (9.5) is that

$$\text{if } (j_1, \dots, j_q), (j'_1, \dots, j'_q) \text{ are friendly, then } \chi(\rho_1(j_1) \times \cdots \times \rho_q(j_q)) = \chi(\rho_1(j'_1) \times \cdots \times \rho_q(j'_q)). \tag{9.6}$$

Having found our combinatorial lines $L_1, \dots, L_q$ with this key property, the rest of the proof follows from a simple pigeonhole argument. We consider the $q + 1$ points

$$\left\{ \begin{array}{c} \rho_1(1) \times \rho_2(1) \times \cdots \times \rho_{q-1}(1) \times \rho_q(1), \\ \rho_1(k) \times \rho_2(1) \times \cdots \times \rho_{q-1}(1) \times \rho_q(1), \\ \rho_1(k) \times \rho_2(k) \times \cdots \times \rho_{q-1}(1) \times \rho_q(1), \\ \vdots \\ \rho_1(k) \times \rho_2(k) \times \cdots \times \rho_{q-1}(k) \times \rho_q(1), \\ \rho_1(k) \times \rho_2(k) \times \cdots \times \rho_{q-1}(k) \times \rho_q(k). \end{array} \right\} \tag{9.7}$$

As these $q + 1$ points receive only q colors, by the pigeonhole principle two of them must be equal. Say that these two equally-colored points have their final k in positions $a - 1$ and b, respectively, where $a \leqslant b$. Then consider the combinatorial line whose elements are

$$\left\{ \begin{array}{c} \rho_1(k) \times \cdots \times \rho_{a-1}(k) \times \rho_a(1) \times \cdots \times \rho_b(1) \times \rho_{b+1}(1) \times \cdots \times \rho_q(1), \\ \rho_1(k) \times \cdots \times \rho_{a-1}(k) \times \rho_a(2) \times \cdots \times \rho_b(2) \times \rho_{b+1}(1) \times \cdots \times \rho_q(1), \\ \vdots \\ \rho_1(k) \times \cdots \times \rho_{a-1}(k) \times \rho_a(k) \times \cdots \times \rho_b(k) \times \rho_{b+1}(1) \times \cdots \times \rho_q(1). \end{array} \right\}$$

Equivalently, this is the combinatorial line whose root is $\rho_1(k) \times \cdots \times \rho_{a-1}(k) \times \rho_a \times \cdots \times \rho_b \times \rho_{b+1}(1) \times \cdots \times \rho_q(1)$. We claim that this line is monochromatic under χ.

Indeed, its first $k-1$ points are all friendly with one another, so these first $k-1$ points all receive the same color by (9.6). On the other hand, the first and the last point receive the same color, since that is how we picked these two points from (9.7). Hence all k points receive the same color. $\square$

9.7 Bounds and density theorems

The bound on $\mathrm{HJ}(k;q)$ given by the proof above is absolutely enormous. Indeed, we proved that

$$\mathrm{HJ}(k;q) \leqslant d = d_1 + \cdots + d_q,$$

and each d_i is itself $\mathrm{HJ}(k-1; q^{k^{d_1+\cdots+d_{i-1}}})$. Essentially, this means that in order to get $\mathrm{HJ}(k;q)$, we need to iterate the function $x \mapsto \mathrm{HJ}(x;q')$, for some large q', roughly q times. Already for $\mathrm{HJ}(3;q)$, this gives an upper bound which is roughly of the form

$$\mathrm{HJ}(3;q) \leqslant \left.3^{3^{\cdot^{\cdot^{\cdot^{3^q}}}}}\right\}{\scriptstyle q \text{ times}}.$$

Then, to get a bound on $\mathrm{HJ}(4;q)$, we'd need to iterate this tower function roughly q times, yielding a so-called wowzer bound, where *wowzer function* $w(x)$ is defined by

$$w(x) = \underbrace{\left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\}\left.2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}\right\}\cdots}_{x \text{ times}}.$$

In general, the bound we obtain for $\mathrm{HJ}(k;q)$ is in the $(k-1)$st level of the so-called *Ackermann hierarchy.*

For a long time, these (abysmal) bounds were the best known. Moreover, these also yielded the best known bounds on various consequences of the Hales–Jewett theorem, such as van der Waerden's theorem and the Gallai–Witt theorem. A major breakthrough was obtained by Shelah [127], who proved that

$$\mathrm{HJ}(k;q) \leqslant w(k+q+2).$$

In particular, this remains the only known *primitive recursive* bound, which roughly means a bound that stays at a fixed level of the Ackermann hierarchy. Good expositions of Shelah's proof are given in [100] and [75, Chapter 26], and an extremely short exposition is given in [104]. While Shelah's bound remains the best known bound on $\mathrm{HJ}(k;q)$ in general, let us mention a result of Conlon [23], who showed that $\mathrm{HJ}(3;q) \leqslant 2^{2^{2^{Cq}}}$ for some constant $C > 0$.

Of course, the main reason we care about the Hales–Jewett theorem is because of its applications, so it is natural to ask whether we can obtain better bounds than these for van der Waerden's theorem or the Gallai–Witt theorem. In the latter case, the answer seems to essentially be "no". For many years, the only non-one-dimensional configuration for which the Gallai–Witt theorem was known to have better bounds was the set

$$S = \{(0,0),(1,0),(0,1)\} \subseteq \mathbb{Z}^2.$$

Indeed, Graham and Solymosi [64] proved that any q-coloring of $[\![N]\!]^2$, where $N = 2^{2^{Cq}}$ contains a monochromatic copy of S; however, even this result is now known to follow from bounds on Hales–Jewett numbers, thanks to the result of Conlon [23] mentioned above.

For van der Waerden's theorem, much more is known. However, to disucss the improved bounds for van der Waerden's theorem, we first need to discuss *density results*.

Recall that some of the Ramsey-theoretic theorems we proved, such as Theorems 5.2.1, 5.3.1 and 5.4.10 were proved by restricting to the densest color class in any coloring of $E(K_N)$. In fact, in some cases, such as Theorem 5.3.2 and Lemma 5.4.11, we explicitly extracted a lemma saying that any graph with many edges contains a certain structure, and such a lemma immediately implies the Ramsey-theoretic result by restricting to the densest color class. We thus say that results like Theorem 5.3.2 are density Ramsey theorems: they state that, to find a certain structure, it is not necessary to color some object, but merely to restrict to an arbitrary large subset of it.

Already in 1936, Erdős and Turán asked whether there is a density version of van der Waerden's theorem. Namely, is it the case that for every $\delta > 0$ and $k \geqslant 3$, there exists some N such that every subset $A \subseteq [\![N]\!]$ with $|A| \geqslant \delta N$ contains a k-AP? If true, this result immediately implies van der Waerden's theorem, since any q-coloring of $[\![N]\!]$ contains a color class of size at least δN, where $\delta = \frac{1}{q}$.

The $k = 3$ case of the Erdős–Turán conjecture was proved by Roth [116, 117], in an extremely influential paper that is arguably the origin of modern additive combinatorics. Moreover, Roth's proof gave fairly good bounds on this N—he showed that if $N \geqslant 2^{2^{C/\delta}}$, for some constant $C > 0$, then every subset $A \subseteq [\![N]\!]$ with $|A| \geqslant \delta N$ contains a 3-AP. In particular, this implies the bound $W(3; q) \leqslant 2^{2^{Cq}}$, which was much better than any other bound on van der Waerden numbers known at the time. Over the years, there have been very many papers improving on Roth's result; see [102] for an excellent survey on the topic. Very recently, a major breakthrough of Kelley and Meka [76], slightly improved by Bloom–Sisask [8], showed that the same conclusion holds if $N \geqslant 2^{C/\delta^9}$, nearly matching an old construction of Behrend [6], who showed that such a result is not true if $N \leqslant 2^{c/\delta^2}$ for some $c > 0$. In particular, we now know that $W(3; q) \leqslant 2^{Cq^9}$.

However, for $k \geqslant 4$, the Erdős–Turán conjecture turns out to be substantially harder. It was finally proved by Szemerédi [135, 136].

Theorem 9.7.1 (Roth [116, 117], Szemerédi [135, 136])**.** *For every $\delta > 0$ and $k \geqslant 3$, there exists some N such that every $A \subseteq [\![N]\!]$ with $|A| \geqslant \delta N$ contains a k-AP.*

The original proof of Szemerédi was an extremely complicated and involved inductive argument. Important ingredients in his proof include both van der Waerden's theorem (applied to find very long progressions), and an early version of the regularity method that we discussed in Chapter 6. Because of this, Szemerédi's proof gave essentially no bounds on how large N has to be, and in particular gives even worse bounds on $W(k; q)$ than those arising from the Hales–Jewett theorem.

Since then, Szemerédi's theorem has been re-proved a number of times, which I will briefly discuss in a highly non-chronological order. In the early 2000s, Gowers [62] found an analytic proof of Szemerédi's theorem, which substantially improved the bounds; in particular, he proved what remains the best general bound on van der Waerden numbers, namely

$$W(k;q) \leqslant 2^{2^{q^{2^{2^{k+9}}}}}.$$

His proof is, in a certain appropriate sense, a generalization of Roth's original proof for the $k = 3$ case, but proving such a generalization required creating an entire theory of *higher order Fourier analysis*.

Shortly after Szemerédi's original proof, Furstenberg [58] found an alternative proof, using ergodic theory. This proof gives absolutely no estimates on how large N has to be, but has a great number of other advantages. In particular, the same approach allowed Furstenberg and Katznelson [59] to prove a density version of the Gallai–Witt theorem, now known as the *multidimensional Szemerédi theorem*: for every $S \subseteq \mathbb{Z}^t$ and every $\delta > 0$, there exists some N such that if $A \subseteq [\![N]\!]^t$ satisfies $|A| \geqslant \delta N^t$, then A contains a homothetic copy of S.

These ergodic-theoretic proofs were non-quantitative, but we now do know quantitative versions of the multidimensional Szemerédi theorem, thanks to independent work of Gowers [63] and of (subsets of) Frankl, Kohayakwa, Nagle, Rödl, Schacht, and Skokan (see e.g. [113]). These authors managed to set up an appropriate analogue of the regularity method to hypergraphs, and were able to use this machinery to prove the so-called *hypergraph removal lemma*, a purely combinatorial statement which easily implies the multidimensional Szemerédi theorem. Unfortunately, the bounds given by the hypergraph regularity method are of Ackermann type, and thus this approach does not give improved bounds for the Gallai–Witt theorem.

Of course, the natural remaining question is whether there is a density version of the Hales–Jewett theorem itself. The answer is yes, as was first proved by Furstenberg and Katznelson [60] via ergodic-theoretic tools.

Theorem 9.7.2 (The Density Hales–Jewett theorem; Furstenberg–Katznelson [60])**.** *For every $\delta > 0$ and $k \geqslant 2$, there exists some d such that every $A \subseteq [\![k]\!]^d$ with $|A| \geqslant \delta k^d$ contains a combinatorial line.*

In a major breakthrough, the first-ever Polymath project [105] was able to give a new proof of the Density Hales–Jewett theorem, yielding quantitative bounds; unfortunately, these bounds are roughly comparable to those arising from the proof we gave of Theorem 9.1.3, and in particular do not recover anything like Shelah's bound. Nonetheless, this was an important breakthrough, giving yet another new proof of Szemerédi's theorem and the multidimensional Szemerédi theorem (and thus of van der Waerden's theorem and the Gallai–Witt theorem). The Polymath proof was subsequently simplified by Dodos, Kanellopoulos, and Tyros [34].

Finally, let me just mention a recent breakthrough in this area, which is yet another new proof of Szemerédi's theorem, due to Leng–Sah–Sawhney [89]. Their proof is based on that of Gowers mentioned above, but thanks to a substantially improved version of one of the main tools (the so-called *inverse theorem for the Gowers uniformity norms*), they are able to show that Theorem 9.7.1 holds already when $N \geqslant 2^{2^{(\log \frac{1}{\delta})^{C_k}}}$, for some constant $C_k > 0$ depending only on k. This is the best known bound for any $k \geqslant 5$.

Chapter 10

Hypergraph Ramsey numbers

10.1 The hypergraph Ramsey theorem

A graph G consists of a vertex set $V(G)$, as well as an edge set $E(G)$, which is a collection of pairs of elements from $V(G)$. A natural generalization of this is a *hypergraph*, where, rather than taking pairs of vertices, we take larger tuples.

Definition 10.1.1. Let $t \geqslant 2$ be an integer. A *t-uniform hypergraph* $\mathcal{H}$ consists of a pair $(V(\mathcal{H}), E(\mathcal{H}))$, where $V(\mathcal{H})$ is a finite set, and $E(\mathcal{H})$ is a collection of t-tuples of distinct elements of $V(\mathcal{H})$. Elements of $V(\mathcal{H})$ are called *vertices*, and elements of $E(\mathcal{H})$ are called *hyperedges* (or *t-edges*, or sometimes simply *edges*).

Definition 10.1.2. For integers $k \geqslant t \geqslant 2$, the *complete t-uniform hypergraph on k vertices*, denoted $K_k^{(t)}$, is the t-uniform hypergraph with k vertices in which each of the $\binom{k}{t}$ t-tuples of vertices are hyperedges. Equivalently, if we denote by $\binom{V}{t}$ the set of all t-subsets of a set V, then $K_k^{(t)}$ is the hypergraph with vertex set V of order k, and edge set $E(K_k^{(t)}) = \binom{V}{t}$.

For $t = 2$, this definition of $K_k^{(2)}$ precisely agrees with the usual definition of the complete graph K_k. Perhaps unsurprisingly, there is a version of Ramsey's theorem for hypergraphs, which was also proved by Ramsey [109].

Theorem 10.1.3 (Ramsey [109])**.** *For all integers $k \geqslant t \geqslant 2, q \geqslant 2$, there exists some N such that the following holds. In any q-coloring $\chi : E(K_N^{(t)}) \to [\![q]\!]$, there is a monochromatic copy of $K_k^{(t)}$. In other words, there exist k vertices such that each of the $\binom{k}{t}$ t-tuples among them receive the same color under χ.*

Continuing our earlier practice, we define the *t-uniform Ramsey number* $r_t(k; q)$ to be the least N for which Theorem 10.1.3 is true, and we use the shorthand $r_t(k)$ when $q = 2$. We also define the *off-diagonal t-uniform Ramsey number* $r_t(k_1, \dots, k_q)$ to be the least N so that in any q-coloring of $E(K_N^{(t)})$, there is a monochromatic copy of $K_{k_i}^{(t)}$ in color i, for some $i \in [\![q]\!]$. Similarly, for any t-uniform hypergraphs $\mathcal{H}_1, \dots, \mathcal{H}_q$, we denote by $r_t(\mathcal{H}_1, \dots, \mathcal{H}_q)$

the least N such that any q-coloring of $K_N^{(t)}$ contains a monochromatic copy of $\mathcal{H}_i$ in color i, for some $i \in [\![q]\!]$, and write $r_t(\mathcal{H}; q)$ for shorthand if $\mathcal{H}_1 = \cdots = \mathcal{H}_q = \mathcal{H}$.

Probably the most natural way to prove Theorem 10.1.3 is via the following argument, directly mimicking the proof of Theorem 2.1.4.

Proof of Theorem 10.1.3. Let us only deal with the case $q = 2$. We prove by induction on t the statement that $r_t(k, \ell)$ exists for all $k, \ell \geqslant t$, and for any fixed t we prove this statement by induction on $k + \ell$. Note that the base case $t = 2$ is already done by Theorem 2.1.1, so we fix some $t \geqslant 3$ and assume the result has been proved for $t - 1$. For this fixed t, the base case $k + \ell = 2t$ is trivial, so we may assume the result has been proved for the pairs $(k - 1, \ell)$ and $(k, \ell - 1)$.

The key claim is that the following recursive bound holds, analogously to (2.1):

$$r_t(k, \ell) \leqslant r_{t-1}(r_t(k - 1, \ell), r_t(k, \ell - 1)) + 1. \tag{10.1}$$

Note that we are done if we prove (10.1), since by induction, we know that the numbers $a \coloneqq r_t(k - 1, \ell)$ and $b \coloneqq r_t(k, \ell - 1)$ are finite, as is the number $r_{t-1}(a, b)$. Thus, (10.1) implies Theorem 10.1.3, at least in the case $q = 2$.

To prove (10.1), let $N = r_{t-1}(r_t(k - 1, \ell), r_t(k, \ell - 1)) + 1$, and consider any 2-coloring $\chi : E(K_N^{(t)}) \to \{\text{red}, \text{blue}\}$. Fix a vertex $v \in V(K_N^{(t)})$. There is a bijection between hyperedges containing v and $(t - 1)$-tuples of vertices in $V(K_N^{(t)}) \setminus \{v\}$. That is, we can use χ to define a coloring $\psi : E(K_{N-1}^{(t-1)}) \to \{\text{red}, \text{blue}\}$, by setting

$$\psi(\{w_1, \ldots, w_{t-1}\}) \coloneqq \chi(\{w_1, \ldots, w_{t-1}, v\}).$$

By the definition of N, we know that ψ contains a monochromatic red clique of order $r_t(k-1, \ell)$, or a monochromatic blue clique of order $r_t(k, \ell-1)$. The two cases are symmetric, so let us assume we are in the first. Looking at χ on these $r_t(k - 1, \ell)$ vertices, we can either find a monochromatic blue $K_\ell^{(t)}$, or a monochromatic red $K_{k-1}^{(t)}$. In the first case we are done. In the second case, we have $k - 1$ vertices, such that each of the t-tuples among them are colored red. Moreover, by the definition of ψ, if we combine any $(t - 1)$-tuple from this set with v, we obtain another t-tuple that is colored red by χ. That is, we have found a monochromatic red $K_k^{(t)}$, showing that we are done in this case as well. □

Remark. While this proof is clearly reminiscient of the proof of Theorem 2.1.4, you might think that some things are different. For example, (10.1) is a bit different from (2.1), in that the former has this strange r_{t-1} term, whereas the latter simply has a sum. It is worth pondering what a 1-*uniform hypergraph* should be, and what the 1-uniform version of Theorem 10.1.3 should say. If you think about this enough, you'll come to realize that the proof above really is nothing more than a generalization of the proof of Theorem 2.1.4.

The proof above shows that $r_t(k, \ell)$ is finite for all t, k, ℓ. However, the bound it gives is absolutely enormous. For example, just trying to upper-bound $r_3(k, k)$, we find from (10.1) that

$$r_3(k) \leqslant r_2(r_3(k - 1, k), r_3(k, k - 1)) + 1.$$

Plugging in our bound $r_2(a) < 4^a$, this implies that

$$r_3(k) \leqslant 4^{r_3(k-1,k)}.$$

That is, a single step of the recursion has cost us an exponential! Continuing in this way, this proof yields a bound roughly of the form

$$r_3(k) \leqslant \left.4^{4^{\cdot^{\cdot^{\cdot^{4}}}}}\right\} 2k \text{ times}.$$

But then the bound in uniformity 4 is then much worse—a single step of the recursion (10.1) for $t = 4$ shows that $r_4(k)$ is bounded as a tower-type function of $r_4(k-1,k)$. That is, this proof yields a wowzer-type bound on $r_4(k)$, and in general, the bounds it gives for uniformity t are at the $(t-1)$th level of the Ackermann hierarchy.

Are such abysmal bounds necessary? At first glance, one might suspect that they are—exponential bounds really are the truth for $r_2(k)$, so the argument above is not particularly wasteful for uniformity 2. However, Erdős and Rado [48] discovered an alternative proof of Theorem 10.1.3, which gives a much stronger bound.

Theorem 10.1.4 (Erdős–Rado [48])**.** *For all integers $t \geqslant 3, q \geqslant 2$, and $k_1, \ldots, k_q > t$, we have*

$$r_t(k_1, \ldots, k_q) \leqslant q^{1+\binom{r_{t-1}(k_1-1,\ldots,k_q-1)}{t-1}}.$$

In particular,

$$r_t(k;q) \leqslant q^{1+\binom{r_{t-1}(k-1)}{t-1}}.$$

Theorem 10.1.4 is sometimes called the *stepping-down* argument; it shows that we can bound a t-uniform Ramsey number by (an exponential function of) a $(t-1)$-uniform Ramsey number, that is, we step down one level in the uniformity. As an immediate consequence, we obtain much stronger bounds on hypergraph Ramsey numbers: for any fixed t, the bound is a fixed tower of 2s.

Corollary 10.1.5. *We have*

$$r_3(k;q) \leqslant 2^{2^{(Cq\log q)k}}$$

for some absolute constant $C > 0$. Similarly,

$$r_4(k;q) \leqslant 2^{2^{2^{(C'q\log q)k}}},$$

and in general,

$$r_t(k;q) \leqslant \left.2^{2^{\cdot^{\cdot^{\cdot^{2^{(C_t q\log q)k}}}}}}\right\} t-1 \text{ twos},$$

for some constant C_t depending only on t.

Proof. By Theorem 10.1.4, we know that

$$r_3(k; q) \leqslant q^{1+\binom{r_2(k-1;q)}{2}}.$$

Plugging in the bound $r_2(k-1; q) \leqslant r(k; q) \leqslant q^{qk}$ from Theorem 2.1.5, we find that

$$1 + \binom{r_2(k-1; q)}{2} \leqslant 1 + \binom{q^{qk}}{2} \leqslant q^{2qk}.$$

Note that

$$\log\left(q^{q^{2qk}}\right) = q^{2qk} \log q = 2^{(2q \log q)k + \log \log q} \leqslant 2^{(3q \log q)k},$$

since $\log \log q \leqslant q \log q$ for all $q \geqslant 2$. Therefore,

$$r_3(k; q) \leqslant q^{q^{2qk}} \leqslant 2^{2^{(3q \log q)k}}.$$

The general bound is proved in exactly the same way. □

We now turn to the proof of the stepping-down lemma. To keep the notation manageable, we first present a proof in the case $t = 3, q = 2, k_1 = k_2 = k$, which nonetheless captures all of the ideas of the general proof. We will then give the full proof, which will be a copy-pasted version of the special case, with appropriate modifications.

Proof of Theorem 10.1.4 for $t = 3, q = 2, k_1 = k_2 = k$. Let $r = r_2(k-1)$, let $N = 2^{1+\binom{r}{2}}$, and fix a coloring $\chi : E(K_N^{(3)}) \to \{\text{red}, \text{blue}\}$. Our goal is to find a set $W = \{w_1, \ldots, w_{r+1}\} \subseteq V(K_N^{(3)})$ with the following property: for every $1 \leqslant i < j < r + 1$, the triples $\{w_i, w_j, w_\ell\}$ receive the same color, for all $j < \ell \leqslant r + 1$. Said differently, the color of a triple of vertices in W depends only on the first two vertices in the tuple, with the last vertex being irrelevant. Said differently again, there exists a coloring $\psi : \binom{W}{2} \to \{\text{red}, \text{blue}\}$ with the property that for all $1 \leqslant i < j < \ell \leqslant r + 1$, we have

$$\chi(\{w_i, w_j, w_\ell\}) = \psi(\{w_i, w_j\}). \tag{10.2}$$

Before seeing how to find such a W, let's see why it suffices for our purposes. Recall that we chose $r = r_2(k-1)$. Therefore, $\{w_1, \ldots, w_r\}$ contains a monochromatic K_{k-1} under the coloring ψ. Let the vertex set of this monochromatic clique be $\{v_1, \ldots, v_{k-1}\} \subseteq \{w_1, \ldots, w_r\}$. Then we claim that $\{v_1, \ldots, v_{k-1}, w_{r+1}\}$ is a monochromatic $K_k^{(3)}$ under χ. Indeed, this is immediate from the fact that $\{v_1, \ldots, v_{k-1}\}$ is monochromatic under ψ, as well as the key property (10.2) of the set W.

To find W, we proceed as follows. We let w_1, w_2 be two arbitrary vertices of $K_N^{(3)}$. Consider the remaining vertices. For every $w \in V(K_N^{(3)}) \setminus \{w_1, w_2\}$, the hyperedge $\{w_1, w_2, w\}$ is either red or blue under χ. Therefore, by the pigeonhole principle, there exists some set $S_3 \subseteq V(K_N^{(3)}) \setminus \{w_1, w_2\}$ with

$$|S_3| \geqslant \left\lceil \frac{N-2}{2} \right\rceil$$

and with all tuples $\{w_1, w_2, w\}$, for $w \in S_3$, receiving the same color under χ. That is, each $w \in S_3$ is a valid choice for all future vertices of W, as all vertices in S_3 satisfy the key property we want W to satisfy.

Inductively, suppose we have picked out $w_1, \dots, w_{m-1}$, and we have a set $S_m \subseteq V(K_N^{(3)}) \setminus \{w_1, \dots, w_{m-1}\}$ with the property that all vertices in S_m are valid choices for all future $w \in W$. To continue this process, we first pick an arbitrary $w_m \in S_m$. Having made this choice, we now need to shrink S_m to some subset S_{m+1}, while maintaining the key property we want, namely that all vertices in S_{m+1} are valid choices for future $w \in W$. Notice that for this to work, the only tuples we need to worry about are those involving the newly-chosen w_m—all tuples not involving this vertex are already OK by the fact that $S_{m+1} \subseteq S_m$, as well as our inductive assumption on S_m.

For a vertex in $w \in S_m \setminus \{w_m\}$, consider all hyperedges of the form $\{w_i, w_m, w\}$, where $1 \leqslant i \leqslant m-1$. There are $m-1$ such hyperedges, and each one receives one of two colors under χ. So the total number of lists of colors for all such hyperedges is 2^{m-1}. Therefore, by the pigeonhole principle, we can find such an S_{m+1}, with the key property we want to maintain, with

$$|S_{m+1}| \geqslant \left\lceil \frac{|S_m| - 1}{2^{m-1}} \right\rceil. \tag{10.3}$$

We continue in this way. As long as we ensure that $S_{r+1} \neq \varnothing$, then we can pick some $w_{r+1} \in S_{r+1}$, and this will yield our desired set W.

So all that remains to prove is that by our choice of N, we have that $S_{r+1} \neq \varnothing$. To show this, we claim that for all $m \geqslant 3$, we have

$$|S_m| \geqslant \frac{N}{2^{1+\binom{m-1}{2}}}.$$

We prove this claim by induction. For the base case $m = 3$, we have that

$$|S_3| \geqslant \left\lceil \frac{N-2}{2} \right\rceil \geqslant \frac{N}{4} = \frac{N}{2^{1+\binom{3-1}{2}}},$$

where the first inequality $r \geqslant 2$, hence $N \geqslant 4$. Inductively, assuming we have proved the claim for S_m, we know by (10.3) that

$$|S_{m+1}| \geqslant \left\lceil \frac{|S_m| - 1}{2^{m-1}} \right\rceil = \frac{|S_m|}{2^{m-1}} \geqslant \frac{N/2^{1+\binom{m-1}{2}}}{2^{m-1}} = \frac{N}{2^{1+\binom{m}{2}}},$$

where the final equality uses Pascal's identity $\binom{m-1}{2} + (m-1) = \binom{m}{2}$. This completes the proof of the claim, and applying it with $m = r+1$ we conclude that

$$|S_{r+1}| \geqslant \frac{N}{2^{1+\binom{r}{2}}} = 1,$$

and thus we indeed have that $S_{r+1} \neq \varnothing$. $\square$

Proof of Theorem 10.1.4. Let $r = r_{t-1}(k_1-1,\dots,k_q-1)$, let $N = q^{1+\binom{r}{t-1}}$, and fix a q-coloring $\chi : E(K_N^{(t)}) \to [\![q]\!]$. Our goal is to find a set $W = \{w_1,\dots,w_{r+1}\} \subseteq V(K_N^{(t)})$ with the following property: for every $1 \leqslant i_1 < \cdots < i_{t-1} < r+1$, the tuples $\{w_{i_1},\dots,w_{i_{t-1}},w_i\}$ receive the same color, for all $i_{t-1} < i \leqslant r+1$. Said differently, the color of a t-tuple of vertices in W depends only on the first $t-1$ vertices in the tuple, with the last vertex being irrelevant. Said differently again, there exists a coloring $\psi : \binom{W}{t-1} \to [\![q]\!]$ with the property that for all $1 \leqslant i_1 < \cdots < i_t \leqslant r+1$, we have

$$\chi(\{w_{i_1},\dots,w_{i_t}\}) = \psi(\{w_{i_1},\dots,w_{i_{t-1}}\}). \tag{10.4}$$

Before seeing how to find such a W, let's see why it suffices for our purposes. Recall that we chose $r = r_{t-1}(k_1-1,\dots,k_q-1)$. Therefore, there exists some color $i \in [\![q]\!]$ such that $\{w_1,\dots,w_r\}$ contains a monochromatic $K_{k_i-1}^{(t-1)}$ in color i, under the coloring ψ. Let the vertex set of this monochromatic clique be $\{v_1,\dots,v_{k_i-1}\} \subseteq \{w_1,\dots,w_r\}$. Then we claim that $\{v_1,\dots,v_{k_i-1},w_{r+1}\}$ is a monochromatic $K_{k_i}^{(t)}$ in color i under χ. Indeed, this is immediate from the fact that $\{v_1,\dots,v_{k_i-1}\}$ is monochromatic under ψ, as well as the key property (10.4) of the set W.

To find W, we proceed as follows. We let $w_1,\dots,w_{t-1}$ be $t-1$ arbitrary vertices of $K_N^{(t)}$. Consider the remaining vertices. For every $w \in V(K_N^{(t)}) \setminus \{w_1,\dots,w_{t-1}\}$, the hyperedge $\{w_1,\dots,w_{t-1},w\}$ receives one of q colors under χ. Therefore, by the pigeonhole principle, there exists some set $S_t \subseteq V(K_N^{(t)}) \setminus \{w_1,\dots,w_{t-1}\}$ with

$$|S_t| \geqslant \left\lceil \frac{N-(t-1)}{q} \right\rceil$$

and with all tuples $\{w_1,\dots,w_{t-1},w\}$, for $w \in S_t$, receiving the same color under χ. That is, each $w \in S_t$ is a valid choice for all future vertices of W, as all vertices in S_t satisfy the key property we want W to satisfy.

Inductively, suppose we have picked out $w_1,\dots,w_{m-1}$, and we have a set $S_m \subseteq V(K_N^{(t)}) \setminus \{w_1,\dots,w_{m-1}\}$ with the property that all vertices in S_m are valid choices for all future $w \in W$. To continue this process, we first pick an arbitrary $w_m \in S_m$. Having made this choice, we now need to shrink S_m to some subset S_{m+1}, while maintaining the key property we want, namely that all vertices in S_{m+1} are valid choices for future $w \in W$. Notice that for this to work, the only tuples we need to worry about are those involving the newly-chosen w_m—all tuples not involving this vertex are already OK by the fact that $S_{m+1} \subseteq S_m$, as well as our inductive assumption on S_m.

For a vertex in $w \in S_m \setminus \{w_m\}$, consider all hyperedges of the form $\{w_{i_1},\dots,w_{i_{t-2}},w_m,w\}$, where $1 \leqslant i_1 < \cdots < i_{t_2} \leqslant m-1$. There are $\binom{m-1}{t-2}$ such hyperedges, and each one receives one of q colors under χ. So the total number of lists of colors for all such hyperedges is $q^{\binom{m-1}{t-2}}$. Therefore, by the pigeonhole principle, we can find such an S_{m+1}, with the key property we want to maintain, with

$$|S_{m+1}| \geqslant \left\lceil \frac{|S_m|-1}{q^{\binom{m-1}{t-2}}} \right\rceil. \tag{10.5}$$

We continue in this way. As long as we ensure that $S_{r+1} \neq \varnothing$, then we can pick some $w_{r+1} \in S_{r+1}$, and this will yield our desired set W.

So all that remains to prove is that by our choice of N, we have that $S_{r+1} \neq \varnothing$. To show this, we claim that for all $m \geqslant t$, we have

$$|S_m| \geqslant \frac{N}{q^{1+\binom{m-1}{t-1}}}.$$

We prove this claim by induction. For the base case $m = t$, we have that

$$S_t \geqslant \left\lceil \frac{N-(t-1)}{q} \right\rceil \geqslant \frac{N}{q^2} = \frac{N}{q^{1+\binom{t-1}{t-1}}},$$

where the first inequality uses that $r \geqslant \max_i(k_i - 1) > t - 1$, hence $N \geqslant q^t \geqslant 2t$. Inductively, assuming we have proved the claim for S_m, we know by (10.5) that

$$|S_{m+1}| \geqslant \left\lceil \frac{|S_m|-1}{q^{\binom{m-1}{t-2}}} \right\rceil = \frac{|S_m|}{q^{\binom{m-1}{t-2}}} \geqslant \frac{N/q^{1+\binom{m-1}{t-1}}}{q^{\binom{m-1}{t-2}}} = \frac{N}{q^{1+\binom{m}{t-1}}},$$

where the final equality uses Pascal's identity $\binom{m-1}{t-2} + \binom{m-1}{t-1} = \binom{m}{t-1}$. This completes the proof of the claim, and applying it with $m = r+1$ we conclude that

$$|S_{r+1}| \geqslant \frac{N}{q^{1+\binom{r}{t-1}}} = 1,$$

and thus we indeed have that $S_{r+1} \neq \varnothing$. □

10.2 Lower bounds on hypergraph Ramsey numbers

We have improved our original Ackermann-type bounds on hypergraph Ramsey numbers, but are our new upper bounds close to the truth? Given everything we've already seen, clearly the first thing we should try for lower-bounding $r_t(k;q)$ is to use a random coloring.

Proposition 10.2.1. *For all $t, q \geqslant 2$ and $k \geqslant \max\{t, q\}$, we have*

$$r_t(k;q) > q^{k^{t-1}/t^t} = 2^{ck^{t-1}},$$

where $c > 0$ is a constant depending only on q and t.

Proof. Let $N = q^{k^{t-1}/t^t}$, and consider a random q-coloring of $K_N^{(t)}$, where each of the $\binom{N}{t}$ hyperedges receives one of the q possible colors uniformly at random. For a fixed k-set, the probability that it is monochromatic is exactly $q^{1-\binom{k}{t}}$, since there are q options for the color, and then all $\binom{k}{t}$ hyperedges in the set must receive this color. Therefore, by the union bound, the probability that there is a monochromatic k-set is at most

$$\binom{N}{k} q^{1-\binom{k}{t}} < N^k q^{-(k/t)^t} = 1,$$

where we use the fact that $k \geqslant \max\{t, q\}$ to conclude that $q\binom{N}{k} < N^k$ and that $\binom{k}{t} \geqslant (k/t)^t$. Since this probability is strictly less than 1, we conclude that $r_t(k; q) > N$. □

For $t = 2$, as we know, this gives the correct dependence, namely exponential in k. But for any $t \geqslant 3$, there is a substantial gap relative to Corollary 10.1.5. For example, for $t = 3$ and $q = 2$, we have

$$2^{ck^2} < r_3(k) < 2^{2^{Ck}},$$

for some absolute constants $c, C > 0$. For larger values of t, the gap is even worse—the lower bound is exponential in a power of k, whereas the upper bound is a tower of height t.

Luckily, there is a beautiful argument, called the *stepping-up lemma* of Erdős–Hajnal–Rado[1] [47], which yields much better lower bounds. At a high level, it allows us to convert a lower bound for $r_{t-1}(k/2; q)$ into a lower bound for $r_t(k; q)$ which is *exponentially* larger. In particular, it "should" allow us to close the gap above, by acting in concert with the stepping-down argument Theorem 10.1.4, as the two yield upper and lower bounds on $r_t(k; q)$ which are exponential in the $(t-1)$-uniform Ramsey number. However, there is an important catch: the stepping-up lemma only works if we start with a construction in uniformity 3 or above.

Theorem 10.2.2 (Erdős–Hajnal–Rado [47])**.** *For every $k \geqslant t \geqslant 3, q \geqslant 2$, we have*

$$r_{t+1}(2k + t - 4; q) > 2^{r_t(k;q)-1}.$$

As a corollary, we get a lower bound which "almost" matches Corollary 10.1.5, but there is a gap of 1 in the height of the tower.

Corollary 10.2.3. *We have*

$$r_4(k) \geqslant 2^{2^{ck^2}},$$

for some absolute constant $c > 0$. In general, for every $t \geqslant 4$, there is a constant $c_t > 0$ such that

$$r_t(k) \geqslant \left.2^{2^{\cdot^{\cdot^{\cdot^{2^{c_t k^2}}}}}}\right\} t-2 \text{ twos}.$$

Proof. Applying Theorem 10.2.2, we find that

$$r_4(k) > 2^{r_3(k/2)-1}.$$

By Proposition 10.2.1, we have that $r_3(k/2) > 2^{ck^2}$, for an appropriate constant $c > 0$, which implies the claimed bound. The general case follows by induction, where at each step the constant c_t is roughly four times smaller than c_{t-1}. □

[1] Some sources, such as [27, 68] attribute the stepping-up lemma to Erdős and Hajnal, but it appears to have first appeared in the seminal paper of Erdős, Hajnal, and Rado [47], and was attributed by Erdős [40] to Erdős–Hajnal–Rado.

The most important open problem about hypergraph Ramsey numbers is to close this exponential gap. Note that if one closes this gap for any uniformity $t \geqslant 3$, then one automatically closes it for all higher uniformities, thanks to the stepping-down and stepping-up lemmas, Theorems 10.1.4 and 10.2.2. In particular, closing the gap for uniformity 3 would close it for all uniformities. It is generally believed that the upper bound is closer to the truth.

Conjecture 10.2.4 (Erdős–Hajnal–Rado [47])**.** *There exists an absolute constant $c > 0$ such that $r_3(k) \geqslant 2^{2^{ck}}$. As a consequence, for every $t \geqslant 3$, there exist constants $c_t, C_t > 0$ such that*

$$t-1 \text{ twos}\left\{ 2^{2^{\cdot^{\cdot^{2^{c_t k}}}}} \right. \leqslant r_t(k) \leqslant \left. 2^{2^{\cdot^{\cdot^{2^{C_t k}}}}} \right\} t-1 \text{ twos}.$$

One important reason to believe this conjecture is that it is known to be true once the number of colors is at least four, via a variant of the stepping-up lemma due to Hajnal[2].

Theorem 10.2.5 (Hajnal)**.** *For every $k, q \geqslant 2$, we have*

$$r_3(k; 2q) > 2^{r_2(k-1;q)-1}.$$

In particular,

$$r_3(k; 4) > 2^{2^{ck}}$$

for some absolute constant $c > 0$.

10.2.1 The stepping-up argument

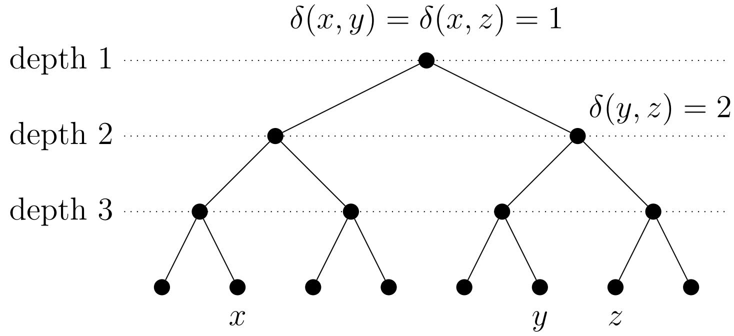
It remains to prove Theorems 10.2.2 and 10.2.5. We begin with Theorem 10.2.5, since it is somewhat simpler (both conceptually and notationally), and captures several of the key ideas of the proof of Theorem 10.2.2.

Proof of Theorem 10.2.5. Let $M = r_2(k-1; q) - 1$, and fix a coloring $\chi : E(K_M) \to [\![q]\!]$ with no monochromatic K_{k-1}; such a coloring exists by the definition of M. Let $N = 2^M$. Our goal is to construct a $(2q)$-coloring of $E(K_N^{(3)})$ with no monochromatic $K_k^{(3)}$.

We think of the vertices of $K_N^{(3)}$ as being the leaves of a binary tree of depth M. For $x, y \in V(K_N^{(3)})$, we denote by $\delta(x, y)$ the depth of the nearest common ancestor[3] of x and y.

[2]It is again not 100% clear to whom this result should correctly be attributed. Graham, Rothschild, and Spencer [68] attribute it to Erdős–Hajnal, but Erdős [40] attributes it to Hajnal.

[3]If you'd prefer, you can think of $V(K_N^{(3)})$ as the set of all binary strings of length M, and then $\delta(x, y)$ is simply the first coordinate in which x and y disagree.



Let us label the vertices of $K_N^{(3)}$ as $v_1, \dots, v_N$, where they come in this order when read left to right along the leaves of the tree. The key fact we need about the quantity δ, which is evident from the picture above, is that whenever $i < j < \ell$, we have that

$$\delta(v_i, v_j) \neq \delta(v_j, v_\ell) \qquad \text{and} \qquad \delta(v_i, v_\ell) = \min\{\delta(v_i, v_j), \delta(v_j, v_\ell)\}. \tag{10.6}$$

In other words, the common ancestor of v_i and v_ℓ is either the common ancestor off v_i, v_j or the common ancestor of v_j, v_ℓ, whichever of the two is higher in the tree. Actually, we will use a more general version of this fact, which is that for all $i_1 < \cdots < i_m$, we have that

$$\delta(v_{i_1}, v_{i_m}) = \min\{\delta(v_{i_1}, v_{i_2}), \delta(v_{i_2}, v_{i_3}), \dots, \delta(v_{i_{m-1}}, v_{i_m})\}. \tag{10.7}$$

Recall that we have fixed a coloring $\chi : E(K_M) \to [\![q]\!]$, and let us identify $V(K_M)$ with $[\![M]\!]$. We are now ready to define the coloring ψ of $E(K_N^{(3)})$. ψ will use $2q$ colors, which we think of as pairs in $[\![q]\!] \times \{\text{up}, \text{down}\}$. For a triple $\{v_i, v_j, v_\ell\} \subseteq V(K_N^{(3)})$, where $i < j < \ell$, we define

$$\psi(\{v_i, v_j, v_\ell\}) \coloneqq \begin{cases} (A, \text{up}) & \text{if } \delta(v_i, v_j) < \delta(v_j, v_\ell) \text{ and } \chi(\delta(v_i, v_j), \delta(v_j, v_\ell)) = A, \\ (A, \text{down}) & \text{if } \delta(v_i, v_j) > \delta(v_j, v_\ell) \text{ and } \chi(\delta(v_i, v_j), \delta(v_j, v_\ell)) = A. \end{cases}$$

Unpacking the definition, we are doing the following. First, we compute $\delta_1 \coloneqq \delta(v_i, v_j)$ and $\delta_2 \coloneqq \delta(v_j, v_\ell)$. By (10.6), these are two distinct integers in $[\![M]\!]$, so we may view them as distinct vertices of K_M; thus, we obtain a color $A \coloneqq \chi(\delta_1, \delta_2)$. In defining ψ, we write down this color A, as well as recording the information of whether $\delta_1 < \delta_2$ or $\delta_1 > \delta_2$; "up" is the former and "down" is the latter. This certainly gives us a $(2q)$-coloring of $E(K_N^{(3)})$; it remains to prove that there is no monochromatic $K_k^{(3)}$ under ψ.

Suppose for contradiction that $S = \{v_{i_1}, \dots, v_{i_k}\}$ form a monochromatic clique under ψ, where $i_1 < \cdots < i_k$. Let us assume first that this clique is monochromatic in color (A, up), for some $A \in [\![q]\!]$. Define $\delta_1 \coloneqq \delta(v_{i_1}, v_{i_2}), \dots, \delta_{k-1} \coloneqq \delta(v_{i_{k-1}}, v_{i_k})$. We claim that $\{\delta_1, \dots, \delta_{k-1}\} \subseteq V(K_M)$ forms a monochromatic K_{k-1} in color A under χ, which is a contradiction since we assumed that χ contains no monochromatic K_{k-1}.

We now turn to proving this claim, that is, that for all $1 \leqslant a < b \leqslant k-1$ we have that $\chi(\delta_a, \delta_b) = A$. First, the fact that $\{v_{i_1}, v_{i_2}, v_{i_3}\}$ is colored (A, up) by ψ implies that $\delta_1 < \delta_2$. Similarly, the color of $\{v_{i_2}, v_{i_3}, v_{i_4}\}$ implies that $\delta_2 < \delta_3$. Continuing in this fasion, we conclude that $\delta_1 < \delta_2 < \cdots < \delta_{k-1}$. Now, consider the hyperedge $\{v_{i_a}, v_{i_b}, v_{i_{b+1}}\}$, which is also colored (A, up). By (10.7), we have that

$$\delta(v_{i_a}, v_{i_b}) = \min\{\delta(v_{i_a}, v_{i_{a+1}}), \ldots, \delta(v_{i_{b-1}}, v_{i_b})\} = \min\{\delta_a, \ldots, \delta_{b-1}\} = \delta_a,$$

where the final step uses the monotonicity $\delta_1 < \cdots < \delta_{k-1}$. Additionally, $\delta(v_{i_b}, v_{i_{b+1}}) = \delta_b$ by definition of δ_b. But now examining the definition of ψ, we see that

$$\psi(\{v_{i_a}, v_{i_b}, v_{i_{b+1}}\}) = (A, \text{up}) \qquad \text{implies} \qquad \chi(\delta_a, \delta_b) = A.$$

Since this holds for arbitrary $1 \leqslant a < b \leqslant k-1$, we conclude that $\{\delta_1, \ldots, \delta_{k-1}\}$ is indeed a monochromatic K_{k-1} under χ, a contradiction.

The case where S is monochromatic under ψ with color (A, down) yields a contradiction via a nearly identical argument, completing the proof. □

We now turn to the proof of Theorem 10.2.2. As in the proof of Theorem 10.1.4, we begin by showing the argument in case $t = 3$ and $q = 2$, which captures all of the key ideas with fewer notational difficulties; we will then present the proof in full generality.

Proof of Theorem 10.2.2 in case $t = 3, q = 2$. We wish to prove that for every $k \geqslant 3$, we have that

$$r_4(2k-1) > 2^{r_3(k)-1}.$$

Let $M = r_3(k) - 1$, and fix a coloring $\chi : E(K_M^{(3)}) \to \{\text{red}, \text{blue}\}$ with no monochromatic $K_k^{(3)}$; such a coloring exists by the definition of M. Let $N = 2^M$. Our goal is to construct a coloring $\psi : E(K_N^{(4)}) \to \{\text{red}, \text{blue}\}$ with no monochromatic $K_{2k-1}^{(4)}$.

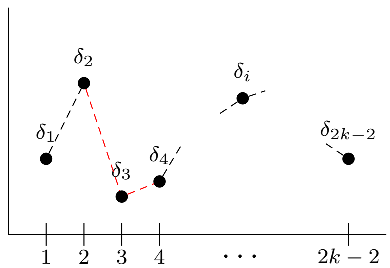
As in the proof of Theorem 10.2.5, we identify $V(K_M^{(3)})$ with $[\![M]\!]$, we think of the vertices $v_1, \ldots, v_N$ of $K_N^{(4)}$ as being the roots of a binary tree of depth M, and we define the function $\delta(x_i, x_j)$ as before. For a 4-tuple $\{v_a, v_b, v_c, v_d\}$ with $1 \leqslant a < b < c < d \leqslant N$, we define $\psi(\{v_a, v_b, v_c, v_d\})$ as follows. Let $\delta_1 = \delta(v_a, v_b), \delta_2 = \delta(v_b, v_c), \delta_3 = \delta(v_c, v_d)$. Recall that by (10.6), we have that $\delta_1 \neq \delta_2$ and $\delta_2 \neq \delta_3$. We then set

$$\psi(\{v_a, v_b, v_c, v_d\}) \coloneqq \begin{cases} \chi(\{\delta_1, \delta_2, \delta_3\}) & \text{if } \delta_1 < \delta_2 < \delta_3, \\ \chi(\{\delta_1, \delta_2, \delta_3\}) & \text{if } \delta_1 > \delta_2 > \delta_3, \\ \text{red} & \text{if } \delta_1 < \delta_2 > \delta_3 \\ \text{blue} & \text{if } \delta_1 > \delta_2 < \delta_3. \end{cases}$$

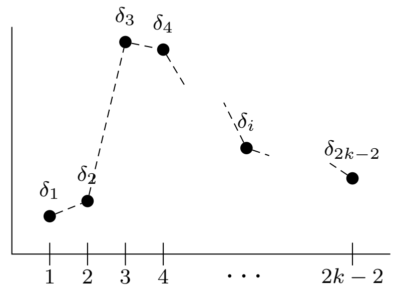
In other words, ψ is defined based on the relative order of the three integers $\delta_1, \delta_2, \delta_3$. If they form a monotonic sequence, then we define ψ to take the same value as χ does on the hyperedge $\{\delta_1, \delta_2, \delta_3\} \in E(K_M^{(3)})$. However, if this sequence is not monotonic, we completely ignore the coloring χ, and define ψ based entirely on the order—red if δ_2 is the maximum, and blue if it's the minimum. We claim that there is no monochromatic $K_{2k-1}^{(4)}$ under ψ.

Indeed, suppose for contradiction that $v_{a_1}, \ldots, v_{a_{2k-1}}$ form a monochromatic $K_{2k-1}^{(4)}$, where $1 \leqslant a_1 < \cdots < a_{2k-1} \leqslant N$. Let us assume first that this $K_{2k-1}^{(4)}$ is red under ψ. Let $\delta_i = \delta(v_{a_i}, v_{a_{i+1}})$, for $1 \leqslant i \leqslant 2k-2$.

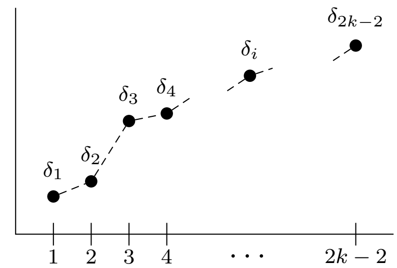
The key observation is that the sequence $\delta_1, \ldots, \delta_{2k-2}$ cannot be an arbitrary sequence of integers. First, we recall by (10.6) that any two consecutive numbers in this sequence are distinct. But more importantly, by our definition of ψ, this sequence can have no local minimum, that is, there cannot exist i such that $\delta_{i-1} > \delta_i < \delta_{i+1}$. Indeed, if we had such an i, then the hyperedge $\{v_{a_{i-1}}, v_{a_i}, v_{a_{i+1}}, v_{a_{i+2}}\}$ would be colored blue by ψ, a contradiction. This implies that the sequence $\delta_1, \ldots, \delta_{2k-2}$ must be *unimodal*—it increases for a while, then reaches a maximum, then decreases for a while[4].



an illegal sequence



a legal sequence (unimodular)



a legal sequence (maximum at one end)

We now simply observe that in any unimodular sequence of $2k-2$ numbers (with every consecutive pair distinct), we can find k of them (in fact, either the first k or the last k) which are either strictly increasing or strictly decreasing. Indeed, if the maximum value occurs at index $k-1$ or before, then the last k terms are strictly decreasing, whereas if the index occurs at value k or later, then the first k terms are strictly increasing.

Let us suppose first that $\delta_1, \ldots, \delta_k$ are strictly increasing. Then the definition of the coloring ψ, as well as the fact that $\{v_{a_1}, \ldots, v_{a_{2k-1}}\}$ is a monochromatic clique under ψ, implies that $\{\delta_1, \ldots, \delta_k\}$ must form a monochromatic red clique under χ, thanks to the same argument we used to finish the proof of Theorem 10.2.5. This is a contradiction to our choice of χ. Similarly, if $\delta_{k-1}, \ldots, \delta_{2k-2}$ form a decreasing sequence, then we again find that they span a monochromatic red $K_k^{(3)}$ under χ, another contradiction.

This completes the proof of the case where $\{v_{a_1}, \ldots, v_{a_{2k-1}}\}$ form a monochromatic red clique under ψ. The remaining case, in which they form a blue clique, is essentially identical; the only difference is that now, the sequence $\delta_1, \ldots, \delta_{2k-2}$ contains no local *maximum*, as such a local maximum would yield a red hyperedge in our putative blue clique. As such, the sequence $\delta_1, \ldots, \delta_{2k-2}$ must either start with a decreasing sequence of order k, or end with an increasing sequence of order k, and in either case we obtain a contradiction. □

Proof of Theorem 10.2.2 in generality. Let $M = r_t(k; q) - 1$, and fix a coloring $\chi : E(K_M^{(t)}) \to [\![q]\!]$ with no monochromatic $K_k^{(t)}$; such a coloring exists by the definition of M. Let $N = 2^M$.

[4]It is of course possible that the maximum is at one of the two ends, such that there is actually no increasing or decreasing portion; that is, "for a while" includes the possibility of increasing or decreasing for zero terms.

Our goal is to construct a coloring $\psi : E(K_N^{(t+1)}) \to [\![q]\!]$ with no monochromatic $K_{2k+t-4}^{(t+1)}$.

As in the proof of Theorem 10.2.5, we identify $V(K_M^{(t)})$ with $[\![M]\!]$, we think of the vertices $v_1, \dots, v_N$ of $K_N^{(t+1)}$ as being the roots of a binary tree of depth M, and we define the function $\delta(x_i, x_j)$ as before. For a $(t+1)$-tuple $\{v_{a_1}, \dots, v_{a_{t+1}}\}$ with $1 \leqslant a_1 < \dots < a_{t+1} \leqslant N$, we define $\psi(\{v_{a_1}, \dots, v_{a_{t+1}}\})$ as follows. Let $\delta_1 = \delta(v_{a_1}, v_{a_2}), \delta_t = \delta(v_{a_t}, v_{a_{t+1}})$. We then set

$$\psi(\{v_{a_1}, \dots, v_{a_{t+1}}\}) \coloneqq \begin{cases} \chi(\{\delta_1, \dots, \delta_t\}) & \text{if } \delta_1, \dots, \delta_t \text{ is a monotonic sequence,} \\ 1 & \text{if } \delta_1 < \delta_2 > \delta_3, \\ 2 & \text{if } \delta_1 > \delta_2 < \delta_3, \\ 1 & \text{in all other cases.} \end{cases}$$

In other words, ψ is defined based on the relative order of the integers $\delta_1, \dots, \delta_t$. If they form a monotonic sequence, then we define ψ to take the same value as χ does on the hyperedge $\{\delta_1, \dots, \delta_t\} \in E(K_M^{(t)})$. However, if this sequence is not monotonic, we completely ignore the coloring χ, and define ψ based entirely on the order—by color 1 if it starts with a local maximum $\delta_1 < \delta_2 > \delta_3$, by color 2 if it starts with a local minimum, and again by color 1 in all other cases†. We claim that there is no monochromatic $K_{2k+t-4}^{(t+1)}$ under ψ.

Indeed, suppose for contradiction that $v_{a_1}, \dots, v_{a_{2k+t-4}}$ form a monochromatic $K_{2k+t-4}^{(4)}$, where $1 \leqslant a_1 < \dots < a_{2k+t-4} \leqslant N$. Let us assume first that this $K_{2k+t-4}^{(4)}$ is colored 1 under ψ. Let $\delta_i = \delta(v_{a_i}, v_{a_{i+1}})$, for $1 \leqslant i \leqslant 2k + t - 5$.

The key observation is that the sequence $\delta_1, \dots, \delta_{2k+t-5}$ cannot be an arbitrary sequence of integers. First, we recall by (10.6) that any two consecutive numbers in this sequence are distinct. But more importantly, by our definition of ψ, this sequence can have no local minimum in the range $[2, 2k-3]$, that is, there cannot exist $2 \leqslant i \leqslant 2k-3$ such that $\delta_{i-1} > \delta_i < \delta_{i+1}$. Indeed, if we had such an i, then the hyperedge $\{v_{a_{i-1}}, v_{a_i}, \dots, v_{a_{i+t-2}}\}$ would be colored 2 by ψ, a contradiction. This implies that the sequence $\delta_1, \dots, \delta_{2k-2}$ must be *unimodal*—it increases for a while, then reaches a maximum, then decreases for a while‡.

We now simply observe that in any unimodular sequence of $2k-2$ numbers (with every consecutive pair distinct), we can find k of them (in fact, either the first k or the last k) which are either strictly increasing or strictly decreasing. Indeed, if the maximum value occurs at index $k-1$ or before, then the last k terms are strictly decreasing, whereas if the index occurs at value k or later, then the first k terms are strictly increasing.

Let us suppose first that $\delta_1, \dots, \delta_k$ are strictly increasing. Then the definition of the coloring ψ, as well as the fact that $\{v_{a_1}, \dots, v_{a_{2k+t-4}}\}$ is a monochromatic clique under ψ, implies that $\{\delta_1, \dots, \delta_k\}$ must form a monochromatic clique in color 1 under χ, thanks to the same argument we used to finish the proof of Theorem 10.2.5. This is a contradiction to our choice of χ. Similarly, if $\delta_{k-1}, \dots, \delta_{2k-2}$ form a decreasing sequence, then we again find that they span a monochromatic $K_k^{(t)}$ under χ, another contradiction.

This completes the proof of the case where $\{v_{a_1}, \dots, v_{a_{2k+t-4}}\}$ form a monochromatic clique in color 1 under ψ. The next case, in which they form a clique in color 2, is essentially identical; the only difference is that now, the sequence $\delta_1, \dots, \delta_{2k-2}$ contains no local *maximum*, as such a local maximum would yield a hyperedge of color 1 in our putative clique of color 2. As such, the sequence $\delta_1, \dots, \delta_{2k-2}$ must either start with a decreasing sequence of order k, or end with an increasing sequence of order k, and in either case we obtain a contradiction.

The final case (which is only relevant when $q \geqslant 3$), is if $\{v_{a_1}, \dots, v_{a_{2k+t-4}}\}$ form a monochromatic clique in color $c \geqslant 3$ under ψ. But this case is even easier than the previous ones—in this case, the sequence $\delta_1, \dots, \delta_{2k-2}$ can contain no local maximum *or* local minimum, implying that $\{\delta_1, \dots, \delta_{2k-2}\}$ form a monochromatic $K^{(t)}_{2k-2}$ under χ, and even more resounding contradiction. □

[†] It doesn't actually matter how we color these final hyperedges; as we will shortly see, our proof will not use them at all. We color them 1 only for concreteness.

[‡] It is of course possible that the maximum is at one of the two ends, such that there is actually no increasing or decreasing portion; that is, "for a while" includes the possibility of increasing or decreasing for zero terms.

10.3 Points in convex position

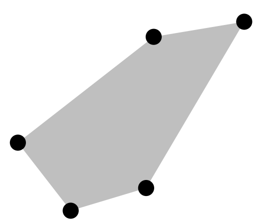
The paper of Erdős and Szekeres [52] in which they proved Theorem 2.1.4—one of the most influential and foundational papers in the field—was titled "A combinatorial problem in geometry". We will now study this geometric problem, and see how it relates to Ramsey theory.

Definition 10.3.1. Let $p_1, \dots, p_k$ be points in $\mathbb{R}^d$. A point $p \in \mathbb{R}^d$ is *in their convex hull* if there exist numbers $\lambda_1, \dots, \lambda_k \geqslant 0$ with $\sum_{i=1}^k \lambda_i = 1$ such that

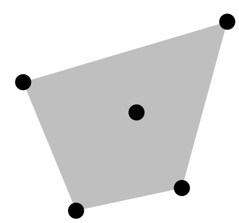
$$p = \sum_{i=1}^k \lambda_i p_i.$$

That is, p is in the convex hull of $p_1, \dots, p_k$ if p is a weighted average of them.

Definition 10.3.2. A collection $p_1, \dots, p_k$ of points in $\mathbb{R}^d$ is *in convex position* if no p_i is in the convex hull of $p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_k$.



Five points in convex position
(the gray region is their convex hull)

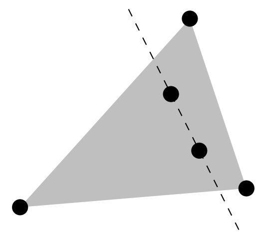


Five points *not* in convex position
(the gray region is their convex hull)

The question studied by Erdős and Szekeres begins with a simple observation of Esther Klein.

Proposition 10.3.3 (Klein)**.** *Among any five points in $\mathbb{R}^2$, no three of them collinear, there are four points in convex position.*

Proof. Consider the convex hull of the five points. It is a polygon with at most five vertices. If it has four or five vertices, then four of these vertices yield our four desired points in convex position. So we may assume that the convex hull is a triangle, meaning that the final two points lie inside the triangle, as shown in the following picture.



Consider the line through the two interior points. Since no three points are collinear, two of the vertices of the triangle must lie on one side of this line. But then these two vertices, plus the two interior points, yield four points in convex position. □

Although this was before Ramsey theory really existed, Klein realized that there was a Ramsey-theoretic flavor to this result. She asked Erdős and Szekeres whether Proposition 10.3.3 could be generalized to finding arbitrarily large collections of points in convex position. Erdős and Szekeres proved that the answer is yes.

Theorem 10.3.4 (Erdős–Szekeres [52]). *For every $k \geqslant 4$, there exists some N such that the following holds. Among any N points in $\mathbb{R}^2$, no three of them collinear, there are k points in convex position.*

We will see three proofs of this theorem (and a fourth proof is in the homework); the first is the original proof of Erdős and Szekeres [52].

Erdős and Szekeres's proof of Theorem 10.3.4. We will show that the theorem holds with $N = r_4(5, k)$. Fix N points $p_1, \dots, p_N$ in $\mathbb{R}^2$, no three of them collinear. We identify $V(K_N^{(4)})$ with $\{p_1, \dots, p_N\}$, and define a two-coloring of $E(K_N^{(4)})$ as follows. Given a 4-tuple $\{p_a, p_b, p_c, p_d\}$, we color it blue if these four points are in convex position, and red otherwise.

The first observation is that we cannot have a monochromatic red $K_5^{(4)}$. Indeed, this would correspond to five points in the plane, no three collinear, such that *every* 4-tuple among them is not in convex position. Proposition 10.3.3 says that such a configuration cannot exist.

Therefore, by the choice of N, there must exist k points, say $p_1, \dots, p_k$, such that each hyperedge among them is colored blue. That is, every 4-tuple among them *is* in convex position. To complete the proof, we require the following simple lemma.

Lemma 10.3.5 (Carathéodory's theorem). *Let $p_1, \dots, p_k$ be a collection of points in $\mathbb{R}^2$, such that each 4-tuple among them is in convex position. Then $p_1, \dots, p_k$ are in convex position.*

In a moment, we will give a formal proof of Lemma 10.3.5, but the intuitive proof is the following. Suppose for contradiction that $p_1, \ldots, p_k$ are not in convex position, and say without loss of generality that p_k is in the convex hull of $p_1, \ldots, p_{k-1}$, and call this convex hull P. Then P is a convex polygon, whose vertices are (some subset of) $p_1, \ldots, p_{k-1}$. Pick an arbitrary triangulation of P, that is, a partition of P into triangles whose vertices are vertices of P itself. Since $p_k \in P$, we must have that p_k is contained in one of the triangles of the triangulation. But that means that p_k is in the convex hull of three vertices of P; this yields four points out of $p_1, \ldots, p_k$ which are not in convex position.

Given Lemma 10.3.5, the proof is complete: we have found k points from our original collection that are in convex position. $\square$

While the geometric proof sketch presented above can be made rigorous, there is also a fairly simple linear-algebraic proof of Lemma 10.3.5, which we now present.

Proof of Lemma 10.3.5. We may assume that $k \geqslant 5$, for otherwise there is nothing to prove. Suppose for contradiction that one of the points, say p_k, is in the convex hull of of the remaining points. This means that there exist numbers $\lambda_1, \ldots, \lambda_{k-1} \geqslant 0$ with $\sum \lambda_i = 1$ and

$$p_k = \sum_{i=1}^{k-1} \lambda_i p_i.$$

Let us fix such a collection $\lambda_1, \ldots, \lambda_{k-1}$ with the fewest number of non-zero elements. That is, we may assume by renaming the points that $\lambda_1, \ldots, \lambda_t > 0$, that $\lambda_{t+1}, \ldots, \lambda_{k-1} = 0$, and that no such representation is possible with fewer than t non-zero coefficients.

If $t \leqslant 3$, then we have shown that the points p_1, p_2, p_3, p_k are not in convex position (since p_k is in the convex hull of p_1, p_2, p_3), contradicting our assumption that all 4-tuples are in convex position. Therefore we may assume that $t \geqslant 4$. Consider the vectors

$$v_1 \coloneqq p_1 - p_t, \qquad v_2 \coloneqq p_2 - p_t, \qquad \ldots, \qquad v_{t-1} \coloneqq p_{t-1} - p_t.$$

These are $t - 1 \geqslant 3$ vectors in $\mathbb{R}^2$, so they must be linearly dependent. That is, there exist $\alpha_1, \ldots, \alpha_{t-1} \in \mathbb{R}$, at least one of which is non-zero, such that $\sum_{i=1}^{t-1} \alpha_i v_i = 0$. Now note that for any $\varepsilon \geqslant 0$, we have

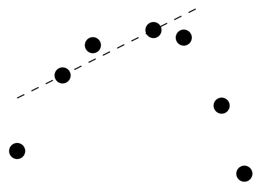
$$\begin{aligned} p_k &= \sum_{i=1}^{t} \lambda_i p_i \\ &= \lambda_t p_t + \sum_{i=1}^{t-1} \lambda_i p_i + \sum_{i=1}^{t-1} \varepsilon \alpha_i v_i \\ &= \lambda_t p_t + \sum_{i=1}^{t-1} [(\lambda_i + \varepsilon\alpha_i) p_i - \varepsilon \alpha_i p_t] \\ &= \sum_{i=1}^{t-1} (\lambda_i + \varepsilon \alpha_i) p_i + \left(\lambda_t - \varepsilon \sum_{i=1}^{t-1} \alpha_i \right) p_t \\ &=: \sum_{i=1}^{t-1} \mu_i(\varepsilon) p_i + \mu_t(\varepsilon) p_t. \end{aligned}$$

Notice that each $\mu_i(\varepsilon)$ is a continuous (in fact, linear) function of ε. Also, by assumption, we have that $\mu_i(0) > 0$ for all $i \in [\![t]\!]$. Also, by construction, we have that $\sum_i \mu_i(\varepsilon) = 1$ for all ε. However, since one of the α_i is non-zero, we see that in the limit $\varepsilon \to \infty$, at least one of the $\mu_i(\varepsilon)$ must become negative. Therefore, there is some smallest value ε^* such that $\mu_i(\varepsilon^*) = 0$ for at least one i, and $\mu_j(\varepsilon^*) \geqslant 0$ for all $j \neq i$. However, this gives us a new representation of p_k as a convex combination of $p_1, \ldots, p_{k-1}$ with fewer non-zero coefficients, contradicting our choice of $\lambda_1, \ldots, \lambda_{k-1}$. □

An alternative proof of Theorem 10.3.4 was found by Tarsi, who showed how to obtain the same result by using a diagonal 3-uniform Ramsey theorem, rather than the off-diagonal 4-uniform Ramsey theorem used by Erdős and Szekeres.

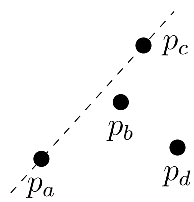
Tarsi's proof of Theorem 10.3.4. Let $N = r_3(k)$, and fix points $p_1, \ldots, p_N$ in $\mathbb{R}^2$. By rotating the plane if necessary, we may assume that all the points $p_1, \ldots, p_N$ have distinct x-coordinates. Let us also relabel them so that they are sorted by x-coordinate, that is, so that p_1 is to the left of p_2, which is to the left of p_3, and so on. We identify $V(K_N^{(3)})$ with $\{p_1, \ldots, p_N\}$, and color $E(K_N^{(3)})$ as follows. For $1 \leqslant i < j < \ell \leqslant N$, we color the hyperedge $\{p_i, p_j, p_\ell\}$ red if p_j lies above the line $p_i p_\ell$, and blue if p_j lies below the line $p_i p_\ell$.

By the choice of N, there is a monochromatic $K_k^{(3)}$, say $p_{i_1}, \ldots, p_{i_k}$, where $i_1 < \cdots < i_k$. Let us suppose this $K_k^{(3)}$ is red. This means that every point in this set lies above the line between its neighbors on the left and right; intuitively, this means that the points need to look like this:



In particular, the points $p_{i_1}, \ldots, p_{i_k}$ are in convex position, as is hopefully intuitive from the picture. This is in fact true, and is a discrete version of the well-known fact that a function with non-positive second derivative is concave.

To prove that $p_{i_1}, \ldots, p_{i_k}$ are in convex position, it suffices by Lemma 10.3.5 to show that any four of them are in convex position. So let p_a, p_b, p_c, p_d be four points, ordered from left to right, with the property that each of the triples they define is red, that is, that each point lies above the line connecting its two neighbors. If they are not in convex position, then one of them must be in the convex hull of the other three, and it is not hard to see that the interior point must be either p_b or p_c (p_a and p_d are necessarily extreme points because they minimize and maximize, respectively, the x-coordinate among these four points). If, say, p_b is in the convex hull of p_a, p_c, p_d, then we see that p_b lies below the line between p_a and p_c, a contradiction.



Similarly, if p_c is an interior point, it lies below the line joining p_b, p_d, another contradiction. This shows that all 4-tuples are indeed in convex position, and thus we have found our desired k-set in convex position by Lemma 10.3.5. In case $\{p_{i_1}, \ldots, p_{i_k}\}$ form a blue clique, the same argument works after vertically reflecting the whole picture. □

Let us define the *Klein number* $\mathrm{Kl}(k)$ to be the least integer N such that every collection of N points in the plane, no three collinear, contains k points in convex position. For $k = 3$ we see that $\mathrm{Kl}(3) = 3$. Proposition 10.3.3 shows that $\mathrm{Kl}(4) \leqslant 5$, and by considering a triangle with an interior point we conclude that $\mathrm{Kl}(4) = 5$. Makai and Turán proved that $\mathrm{Kl}(5) = 9$, via an elementary but involved case analysis. More recently, Szekeres and Peters [134] used a computer search to verify that $\mathrm{Kl}(6) = 17$. Observing the pattern, you may be tempted to make the following conjecture.

Conjecture 10.3.6 (Erdős–Szekeres [52]). *We have* $\mathrm{Kl}(k) = 2^{k-2} + 1$ *for every* $k \geqslant 3$.

In a later paper, Erdős and Szekeres [49] proved that $\mathrm{Kl}(k) > 2^{k-2}$ for every k, thus proving the lower bound in this conjecture. You will see this construction on the homework. What about the upper bound?

Our first proof showed that $\mathrm{Kl}(k) \leqslant r_4(5, k)$. Applying Theorem 10.1.4 twice, we find that

$$\mathrm{Kl}(k) \leqslant r_4(5,k) \leqslant 2^{1+\binom{r_3(4,k-1)}{3}} \leqslant 2^{r_3(4,k-1)^3} \leqslant 2^{2^{3\binom{r_2(3,k-2)}{2}}} \leqslant 2^{2^{O(k^4/(\log k)^2)}},$$

where the final step uses Theorem 4.1.4. The second proof gives a slightly better bound,

$$\mathrm{Kl}(k) \leqslant r_3(k) \leqslant 2^{1+\binom{r_2(k-1)}{2}} \leqslant 2^{2^{O(k)}}.$$

However, already in their original paper, Erdős and Szekeres proved a much stronger bound.

Theorem 10.3.7 (Erdős–Szekeres [52])**.** *For every $k \geqslant 3$, we have*

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

We will prove Theorem 10.3.7 in Chapter 11. The geometric part of the proof is essentially the same as what we saw in Tarsi's proof of Theorem 10.3.4 above, but there is a different combinatorial argument that is much more quantitatively efficient than simply reducing to the hypergraph Ramsey theorem.

Thus, in 1960, Erdős and Szekeres knew that $2^{k-2} < \mathrm{Kl}(k) < 4^{k-2}$. These bounds stood almost unchanged for over 50 years, until a recent breakthrough of Suk [132] proved that the lower bound is indeed close to the truth, i.e. that 2 is the correct base of the exponent.

Theorem 10.3.8 (Suk [132])**.** *We have*

$$\mathrm{Kl}(k) \leqslant 2^{k+o(k)}$$

as $k \to \infty$.

Suk's proof is quite short and simple, but uses a number of very clever ideas coming from both geometry and combinatorics, and we will not discuss it in this course.

Chapter 11

Canonical Ramsey theorems

This chapter contains several rather disparate topics, which nonetheless share some thematic connection. The *extremely* high-level idea is the following. Most mathematical objects are endowed with a notion of sub-objects (e.g. subsets, subgraphs, subgroups, subspaces, subschemes, subterfuges…). Certain objects are *canonical*, in the sense that all of their sub-objects "look like" the original object. For example, an elementary result in group theory is that all subgroups of a cyclic group are cyclic; a more pronounced version of the same fact is that any subgroup of $\mathbb{Z}$ is isomorphic to $\mathbb{Z}$. A substantially deeper and more difficult statement along the same lines is that any subgroup of a free group is again free.

One question we are interested in is a full classification of such examples: for any given notion of mathematical object, what is a complete list of the canonical ones? Having accomplished this task (which requires formalizing what we mean by "looking like" the original object), one can turn to proving a Ramsey-theoretic statement, along the lines of "any sufficiently large object must contain an arbitrarily large canonical sub-object".

We can view Ramsey's theorem as an instance of this general philosophy. Indeed, consider the class of graphs, endowed with the sub-object relation of induced subgraphs. Then complete graphs and empty graphs are examples of canonical objects, since any induced subgraph of a complete graph is again complete, and any induced subgraph of an empty graph is empty. Moreover, Ramsey's theorem implies that every sufficiently large graph contains an arbitrarily large complete or empty induced subgraph.

11.1 Monotone sequences

Consider a sequence $a_1, \ldots, a_k$ of distinct real numbers. A natural definition for a "canonical" sequence is a monotone sequence (that is, a sequence which either strictly increasing or strictly decreasing), since any subsequence of an increasing sequence is again increasing, and the same holds for decreasing sequences.

As you might expect, there is a Ramsey-theoretic statement, asserting that every sequence of distinct real numbers contains a long monotone subsequence; this was proved in the same seminal paper of Erdős and Szekeres [52].

Theorem 11.1.1 (Erdős–Szekeres [52])**.** *Given $k \geqslant 2$, let $N = (k-1)^2 + 1$. Then any sequence $a_1, \ldots, a_N$ of distinct real numbers contains a monotone subsequence of length k. That is, there exist indices $1 \leqslant i_1 < \cdots < i_k \leqslant N$ such that*

$$a_{i_1} < a_{i_2} < \cdots < a_{i_k} \qquad \text{or} \qquad a_{i_1} > a_{i_2} > \cdots > a_{i_k}.$$

There are many known proofs of this theorem (see [131] for an exposition of several different proofs); we will show a particularly elegant proof discovered by Seidenberg [124].

Proof of Theorem 11.1.1 (Seidenberg [124]). For an index $m \in [\![N]\!]$, let $\delta(m)$ denote the length of the longest decreasing subsequence ending at a_m, and let $\iota(m)$ denote the length of the longest increasing sequence ending at a_m. We wish to prove that $\delta(m) \geqslant k$ or $\iota(m) \geqslant k$ for some $m \in [\![N]\!]$. So suppose for contradiction that this is not the case, that is, that $1 \leqslant \delta(m), \iota(m) \leqslant k-1$; note that we have a lower bound of 1 on both functions, since we can always view a_m itself as both an increasing and a decreasing subsequence ending at a_m.

This means that there are at most $(k-1)^2$ possible values for the pair $(\delta(m), \iota(m))$. Since $N = (k-1)^2+1$, the pigeonhole principle implies that there exists two indices $1 \leqslant \ell < m \leqslant N$ such that $(\delta(\ell), \iota(\ell)) = (\delta(m), \iota(m))$. Since the elements of our sequence are distinct, we have $a_\ell < a_m$ or $a_\ell > a_m$. Suppose first that $a_\ell < a_m$. Then any increasing sequence ending in a_ℓ can be extended by one to obtain an increasing sequence ending at a_m, implying that $\iota(m) > \iota(\ell)$, a contradiction. Similarly, if $a_\ell > a_m$, then $\delta(m) > \delta(\ell)$, another contradiction. In either case we are done. $\square$

It is not hard to show (as you will do on the homework) that this bound is tight, in that there exist sequences of $(k-1)^2$ distinct real numbers with no monotone subsequence of length k.

An equivalent way of viewing Theorem 11.1.1 is as saying that in any function $a : [\![N]\!] \to \mathbb{R}$, where $N = (k-1)^2 + 1$, there are k vertices on which a is monotone. If we identify $[\![N]\!]$ with $V(K_N)$, then this becomes a statement about functions $a : V(K_N) \to \mathbb{R}$. The following generalization to functions $E(K_N) \to \mathbb{R}$ was first proved by Chvátal and Komlós [18].

Theorem 11.1.2 (Chvátal–Komlós [18])**.** *Let $k \geqslant 3$, let $N = \binom{2k-2}{k-1} + 1$, and identify $V(K_N)$ with $[\![N]\!]$. Given any function $a : E(K_N) \to \mathbb{R}$, there is a monotone path of length k, that is, there exist indices $1 \leqslant i_0 < \cdots < i_k \leqslant N$ such that*

$$a((i_0, i_1)) \leqslant a((i_1, i_2)) \leqslant \cdots \leqslant a((i_{k-1}, i_k)) \quad \text{or} \quad a((i_0, i_1)) \geqslant a((i_1, i_2)) \geqslant \cdots \geqslant a((i_{k-1}, i_k)).$$

Note that by *length*, we mean the number of edges, so we are finding a path with $k+1$ vertices. Note too that the path we find is monotone in two senses—the vertices come in the same order as the given order of $V(K_N) \cong [\![N]\!]$, and the value of a on the edges is monotone. This is in direct analogy with Theorem 11.1.1, where we pass to a subset of the indices, where the order is maintained, such that the value of a becomes monotone. The proof we present of Theorem 11.1.2 is due to Lovász [92, Solution 14.27] and Moshkovitz–Shapira [94].

Proof of Theorem 11.1.2. For an edge uv, where $u < v$, we write $\delta(uv)$ for the length (that is, number of edges) of the longest decreasing path ending in the edge uv, and $\iota(uv)$ for the length of the longest increasing path ending in uv. It suffices to prove that $\delta(uv) \geqslant k$ or $\iota(uv) \geqslant k$ for some edge uv, so we assume for contradiction that $1 \leqslant \delta(uv), \iota(uv) \leqslant k-1$ for all $uv \in E(K_N)$.

We define a partial order $\preccurlyeq$ on $\mathbb{Z}^2$ by setting $(x, y) \preccurlyeq (x', y')$ if $x \leqslant x'$ and $y \leqslant y'$. For a vertex $v \in V(K_N)$, we now set

$$\Phi(v) \coloneqq \{(x, y) \in [\![k-1]\!]^2 : (x, y) \preccurlyeq (\delta(uv), \iota(uv)) \text{ for some } u \in V(K_N) \text{ with } u < v\}.$$

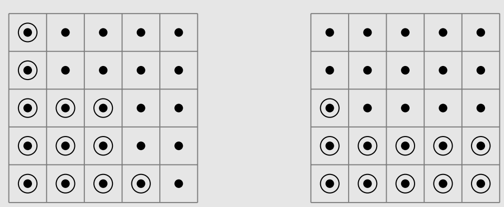
That is, we first write down all pairs $(\delta(uv), \iota(uv))$ over all $u < v$, and then we let $\Phi(v)$ be the "down-set" generated by these points, that is, the collection of all points down or to the left of these points. Note that $\Phi(v) \subseteq [\![k-1]\!]^2$.

The key claim is that $\Phi(v) \neq \Phi(w)$ for all $v \neq w$. Indeed, suppose that $\Phi(v) = \Phi(w)$ for some v, w, and assume without loss of generality that $v < w$. By definition, we have that $(\delta(vw), \iota(vw)) \in \Phi(w)$, hence $(\delta(vw), \iota(vw)) \in \Phi(v)$ as well. Again by the definition of $\Phi(v)$, this implies that there exists some $u < v$ such that

$$(\delta(vw), \iota(vw)) \preccurlyeq (\delta(uv), \iota(uv)).$$

In other words, $\delta(vw) \leqslant \delta(uv)$ and $\iota(vw) \leqslant \iota(uv)$. If $a((u, v)) \geqslant a((v, w))$, then any decreasing path ending at uv can be extended to an decreasing path ending at vw, contradicting the first inequality. Similarly, if $a((u, v)) \leqslant a((v, w))$, then any increasing path ending at uv can be extended to an increasing path ending at vw, another contradiction. We conclude that, as claimed, $\Phi(v) \neq \Phi(w)$ for all $v \neq w$.

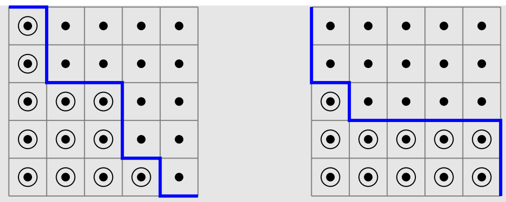
Notice that for every v, the set $\Phi(v) \subseteq [\![k-1]\!]^2$ is a *Ferrers diagram*, that is, a collection of points in $[\![k-1]\!]^2$ such that each row is left-aligned and each column is down-aligned.



Two Ferrers diagrams in $[\![5]\!]^2$

Indeed, by definition, $\Phi(v)$ consists of all points preceding some set in the order $\preccurlyeq$, and such a set is necessarily a Ferrers diagram.

Therefore, Φ gives us an injective function from $V(K_N)$ to the set of all Ferrers diagrams in $[\![k-1]\!]^2$. So to obtain the desired contradiction, it suffices to prove that the number of Ferrers diagrams in $[\![k-1]\!]^2$ is exactly $\binom{2k-2}{k-1}$, since our choice of N would then contradict the existence of such an injection. In order to count the number of Ferrers diagrams, we can identify each Ferrers diagram with its *boundary*, namely the set of edges separating the Ferrers diagram from its complement in $[\![k-1]\!]^2$.



Boundaries of Ferrers diagrams (in blue)

Note that we can recover a Ferrers diagram from its boundary, and vice versa, so the number of possible Ferrers diagrams is the same as the number of possible boundaries. As we can see from these pictures, the boundary of a Ferrers diagram in $[\![k-1]\!]^2$ can be viewed as a path from $(0, k-1)$ to $(k-1, 0)$ only going in the directions right and down. The set of such paths is in bijection with the set of words in the symbols R and D (for *right* and *down*), in which both symbols appear exactly $k-1$ times. The number of such words is exactly $\binom{2k-2}{k-1}$, hence this is the number of possible boundaries, and thus, the number of Ferrers diagrams, as claimed. □

As a corollary of Theorem 11.1.2, we can prove Theorem 10.3.7, giving a much better estimate on the number of points needed to guarantee k points in convex position.

Proof of Theorem 10.3.7. Let $N = \binom{2k-4}{k-2}+1$, and let $p_1, \ldots, p_N$ be N points in the plane, with no three collinear. By rotating the plane, we may assume that the x-coordinates of $p_1, \ldots, p_N$ are distinct, and that the points are sorted by x-coordinate.

For $1 \leqslant u < v \leqslant N$, define $a((u, v))$ to be the slope[†] of the line between p_u and p_v. By Theorem 11.1.2 (applied with parameter $k-1$), we can find indices $1 \leqslant i_1 < \cdots < i_k \leqslant N$ such that

$$a((i_1, i_2)) \leqslant \cdots \leqslant a((i_{k-1}, i_k)) \qquad \text{or} \qquad a((i_1, i_2)) \geqslant \cdots \geqslant a((i_{k-1}, i_k)).$$

Note that in fact, all of these inequalities are strict. Indeed, the fact that $p_{i_{j-1}}, p_{i_j}, p_{i_{j+1}}$ are not collinear precisely means that $a((i_{j-1}, i_j)) \neq a((i_j, i_{j+1}))$.

If we have $a((i_1, i_2)) < \cdots < a((i_{k-1}, i_k))$, then the points $p_{i_1}, \ldots, p_{i_k}$ form a *cup*, that is, a sequence of points where each point lies below the line between its two neighbors. On the other hand, if $a((i_1, i_2)) > \cdots > a((i_{k-1}, i_k))$, then these points form a *cap*, where each point lies above the line between its two neighbors. In either case, as argued in the second proof of Theorem 10.3.4, these k points are in convex position by Lemma 10.3.5. □

We remark that one can construct examples of $\binom{2k-4}{k-2}$ points in the plane with no cup or cap of size k. This shows that the proof above cannot be further improved, and also implies that Theorem 11.1.2 is best possible.

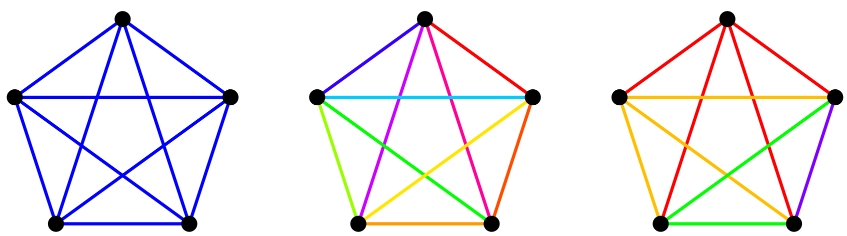
[†] Since we assume the points have distinct x-coordinates, the slope of this line is a well-defined real number.

11.2 The canonical Ramsey theorem

We now turn to the canonical Ramsey theorem for edge-colorings of the complete graph. Of course, as discussed above, Ramsey's theorem itself is such a statement—any coloring of a complete graph with a fixed number of colors must contain an arbitrarily large monochromatic clique, and monochromatic cliques are clearly canonical, as any subset of a monochromatic clique is another monochromatic clique. However, what if we remove the restriction that the number of colors is fixed?

That is, the question we are asking is the following: we color $E(K_N)$, for a large N, with an arbitrary number of colors. What kinds of subcolorings are canonical, in the sense that all of their induced subgraphs yield colorings of the same type? Certainly, monochromatic cliques are still canonical. On the other hand, once the number of colors is unbounded, we get a new type of canonical coloring: a *rainbow* coloring of K_N, in which each of the edges receives a different color (so $\binom{N}{2}$ colors are used in total).

It is tempting to conjecture that these are the only ones, but this turns out to not be the case. There is a third type of coloring, which we will call *starry.* A coloring of $E(K_N)$ is called starry if there are distinct colors $c_1, \ldots, c_{N-1}$ and if one can sort the vertices as $v_1, \ldots, v_N$, such that the color of the edge v_iv_j, where $i < j$, is c_i. In other words, each color class is a star, with the first star centered at v_1, the second centered at v_2 (and not containing v_1), and so on. Note that this is a canonical coloring, as any subset of vertices induces another starry coloring.



As it turns out, these really are the only canonical colorings, in the sense that a canonical Ramsey theorem holds: every sufficiently large edge-colored clique contains an arbitrarily large clique which is monochromatic, rainbow, or starry. This was proved by Erdős and Rado [48], in a result that is now usually called *the* canonical Ramsey theorem.

Theorem 11.2.1 (Erdős–Rado [48])**.** *For every $k \geqslant 2$, there exists some N such that if $E(K_N)$ is colored (with an arbitrary number of colors), there is a K_k which is monochromatic, rainbow, or starry.*

The original proof of Erdős and Rado used a clever reduction to the hypergraph Ramsey theorem in uniformity 4. Namely, for every 4-tuple of vertices, they considered the equivalence relation of colors on the $\binom{4}{2} = 6$ edges. That is, rather than remembering the actual colors on each of these 6 edges, they only record which pairs of edges receive the same color. As it turns out, there are 203 equivalence relations[1] on a set of size 6, so they obtain a 203-coloring of $E(K_N^{(4)})$. By Theorem 10.1.3, there is a monochromatic $K_k^{(4)}$ in this coloring (assuming N is sufficiently large), and an elementary argument (involving some casework)

[1]The number of equivalence relations on a set of size n is given by the *Bell number* B_n, and $B_6 = 203$.

shows that in each of the 203 cases[2], this monochromatic $K_k^{(4)}$ yields a monochromatic, rainbow, or starry K_k in the original coloring. A nice exposition of this proof can be found in [68, Section 5.5].

However, from a quantitative perspective, the proof of Erdős and Rado is not very good. Letting $\mathrm{ER}(k)$ denote the least N such that Theorem 11.2.1 holds, the proof of Erdős–Rado only shows that $\mathrm{ER}(k) \leqslant r_4(k; 203) \leqslant 2^{2^{2^{O(k)}}}$, thanks to the bounds on hypergraph Ramsey numbers. A much better bound, with an alternative proof that is also extremely elegant, was found by Lefmann and Rödl [88].

Theorem 11.2.2 (Lefmann–Rödl [88])**.** *We have* $\mathrm{ER}(k) \leqslant k^{4k^2}$ *for all* $k \geqslant 2$.

In particular, Theorem 11.2.2 gives a finite bound on $\mathrm{ER}(k)$, thus proving Theorem 11.2.1. In the course of the proof of Theorem 11.2.2, we will need the following extremely useful lemma, which allows us to find rainbow cliques in edge-colored graphs where every color class is a graph with bounded maximum degree.

Lemma 11.2.3. *Let* $k, M \geqslant 2$ *be integers, and suppose that* $E(K_M)$ *is colored so that every vertex is incident to at most* M/k^4 *edges in every color. Then there is a rainbow* K_k *in this coloring.*

Proof. Every vertex must be incident to at least one edge of some color, hence no such coloring can exist if $M < k^4$. Thus the statement is vacuously true in these cases, and we may assume henceforth that $M \geqslant k^4$. Also, since every coloring of $E(K_2)$ is rainbow, we may assume henceforth that $k \geqslant 3$. Let χ be the coloring of $E(K_M)$.

Let $v_1, \ldots, v_k$ be a uniformly random sequence of k distinct vertices from K_M. That is, we pick a set of k distinct vertices uniformly at random among the $\binom{M}{k}$ options, and then pick a random ordering of that set and label it $v_1, \ldots, v_k$. Equivalently, we let v_1 be a uniformly random vertex, v_2 a uniformly random vertex among the remaining vertices, and so on. The key property that we need about this distribution is that if we condition on the outcome of any subset of these vertices, the marginal distribution of any remaining vertex is that of a uniformly random vertex of K_M, apart the ones already picked. Thus, for example, if x, y are two distinct vertices of K_M, and we condition on $v_3 = x, v_7 = y$, the marginal distribution of v_4 is uniformly random on the set $V(K_M) \setminus \{x, y\}$.

For distinct indices $i, j, \ell \in [\![k]\!]$, let $\mathcal{E}_{i,j,\ell}$ be the event that the edges v_iv_j and v_iv_ℓ receive the same color. We wish to estimate $\Pr(\mathcal{E}_{i,j,\ell})$. Given two distinct vertices $x, y \in V(K_M)$, we begin by estimating $\Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y)$. Given $v_i = x, v_j = y$, the event $\mathcal{E}_{i,j,\ell}$ is simply the event that $\chi(xv_\ell) = \chi(xy)$, where the only randomness remaining is in the choice of v_ℓ. By assumption, x is incident to at most M/k^4 edges in color $\chi(xy)$, and v_ℓ is a uniformly random vertex in the set $V(K_M) \setminus \{x, y\}$, hence

$$\Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y) \leqslant \frac{1}{M-2} \cdot \frac{M}{k^4} \leqslant \frac{2}{k^4}.$$

[2] In fact, it is not hard to show that most of the 203 cases are actually impossible, so the true number of cases is much smaller.

Since the same upper bound holds for $\Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y)$ for all x, y, the same bound holds for $\Pr(\mathcal{E}_{i,j,\ell})$. More formally, by the law of total probability, we have

$$\Pr(\mathcal{E}_{i,j,\ell}) = \sum_{x,y} \Pr(\mathcal{E}_{i,j,\ell} \mid v_i = x, v_j = y) \Pr(v_i = x, v_j = y) \leqslant \frac{2}{k^4} \sum_{x,y} \Pr(v_i = x, v_j = y) = \frac{2}{k^4}.$$

Since the events $\mathcal{E}_{i,j,\ell}$ and $\mathcal{E}_{i,\ell,j}$ are the same, there are at most $k^3/2$ such events we need to consider. Hence, by the union bound, the probability that $\mathcal{E}_{i,j,\ell}$ occurs for some triple i, j, ℓ is at most $\frac{k^3}{2} \cdot \frac{2}{k^4} = \frac{1}{k} \leqslant \frac{1}{3}$.

Similarly, for four distinct indices i, j, ℓ, m, let $\mathcal{E}_{i,j,\ell,m}$ be the event that the edges $v_i v_j$ and $v_\ell v_m$ receive the same color. For fixed vertices x, y, z, we now condition on the outcome $v_i = x, v_j = y, v_\ell = z$. By assumption, z has at most M/k^4 neighbors in color $\chi(xy)$. Once we condition, the event $\mathcal{E}_{i,j,\ell,m}$ is just the event that $\chi(zv_m) = \chi(xy)$, where the only randomness is in the choice of v_m, which is uniform on a set of size $M - 3$. So we have

$$\Pr(\mathcal{E}_{i,j,\ell,m} \mid v_i = x, v_j = y, v_\ell = z) \leqslant \frac{1}{M-3} \cdot \frac{M}{k^4} \leqslant \frac{2}{k^4}.$$

Again applying the law of total probability, we conclude that $\Pr(\mathcal{E}_{i,j,\ell,m}) \leqslant \frac{2}{k^4}$. The total number of such events is at most $k^4/4$, since we obtain the same event if we swap i, j or ℓ, m. So by the union bound, the probability that $\mathcal{E}_{i,j,\ell,m}$ happens for some 4-tuple (i, j, ℓ, m) is at most $\frac{k^4}{4} \cdot \frac{2}{k^4} = \frac{1}{2}$.

In total, we find that the probability that $v_1, \dots, v_k$ span a rainbow K_k is at least $1 - \frac{1}{3} - \frac{1}{2} > 0$, hence there is a rainbow K_k in the coloring. □

Now that we have Lemma 11.2.3, we can proceed with the proof of Theorem 11.2.2. Before doing so, it's worth thinking of an alternative way of presenting the proof of Theorem 2.1.4. To show that $r(k) \leqslant 4^k$, let us fix a 2-coloring of $E(K_N)$, where $N = 4^k = 2^{2k}$. We pick an arbitrary vertex v_1. At least half of its incident edges are of the same color, which we call c_1. We now restrict to the c_1-colored neighborhood S_1 of v_1, and pick from there an arbitrary vertex v_2. At least half of its incident edges in S_1 are of the same color, say c_2. We let S_2 be this neighborhood, and proceed in this fashion. Since

$$|S_{i+1}| \geqslant \left\lceil \frac{|S_i| - 1}{2} \right\rceil$$

for all i, we conclude that $|S_i| \geqslant 2^{2k-i}$ for all i. Hence we can continue this process for at least $2k$ steps, to produce vertices $v_1, \dots, v_{2k}$ and colors $c_1, \dots, c_{2k}$. Again by the pigeonhole principle, at least k of these colors must be the same, say $c_{i_1}, \dots, c_{i_k}$ are all red. But by the way we constructed this sequence, this shows that $v_{i_1}, \dots, v_{i_k}$ form a red K_k.

The proof of Theorem 11.2.2 uses a very similar argument, which we will now see.

Proof of Theorem 11.2.2. Let $N = k^{4k^2}$, and fix an arbitrary coloring of $E(K_N)$. We let $S_0 = V(K_N)$. We now run the following process, for all $i \geqslant 1$.

1. If $|S_{i-1}| < 2$, stop the process.
2. If every vertex in S_{i-1} is incident to at most $|S_{i-1}|/k^4$ edges in each color, we apply Lemma 11.2.3 to S_{i-1} with $M = |S_{i-1}| \geqslant 2$. We conclude that S_{i-1} contains a rainbow K_k, completing the proof.
3. If not, there is some vertex $v_i \in S_{i-1}$ and some color c_i such that v_i is incident to at least $|S_{i-1}|/k^4$ edges of color c_i in S_{i-1}. We let S_i be the c_i-colored neighborhood of v_i in S_{i-1}.
4. Increment i by 1 and return to step 1.

If we ever find a rainbow K_k in this process, we are done, so we may assume that that never happens. Note that as long as the process continues, we have that $|S_i| \geqslant |S_{i-1}|/k^4$, so by induction we have that $|S_i| \geqslant k^{4(k^2-i)}$. Hence we can continue this process at least until step $i-1 = k^2-1$. In other words, this process produces a sequence $v_1, \dots, v_{k^2}$ of vertices and $c_1, \dots, c_{k^2}$ of colors, with the property that each v_i is adjacent in color c_i to all v_j with $j > i$.

Suppose first that k of the colors $c_1, \dots, c_{k^2}$ are equal, say $c_{i_1}, \dots, c_{i_k}$ are all red. Then $v_{i_1}, \dots, v_{i_k}$ form a monochromatic red K_k, and we are done. But if this does not happen, then at least k different colors must appear in the list $c_1, \dots, c_{k^2}$, say $c_{j_1}, \dots, c_{j_k}$ are all distinct. Then $v_{j_1}, \dots, v_{j_k}$ form a starry K_k, and we are again done. □

Theorem 11.2.2 states that $\mathrm{ER}(k) \leqslant k^{4k^2} = 2^{4k^2 \log k}$. How good is this bound? The best known lower bound, which is off by a logarithmic factor in the exponent, is given by the following simple proposition.

Proposition 11.2.4. *We have*

$$\mathrm{ER}(k) \geqslant r(k; k-2).$$

In particular, using Proposition 2.2.5, we find that

$$\mathrm{ER}(k) \geqslant 2^{\frac{1}{4}k^2 - k}.$$

Actually, using the techniques of Chapter 3, one can improve the constant factor $\frac{1}{4}$, since (3.2) states that

$$r(k; k-2) \geqslant \frac{(2^{(k-3)/2})^k}{4(2k)^{k-4}(2^{(k-4)/8})^k} = \frac{2^{\frac{3}{8}k^2}}{4(2k)^{k-4}2^k} = 2^{(\frac{3}{8}-o(1))k^2}.$$

Similarly, using Lemma 3.3.2 one can get a further small improvement on the constant $\frac{3}{8}$. But it remains a major open problem to close the logarithmic gap in the estimates $2^{\Omega(k^2)} \leqslant \mathrm{ER}(k) \leqslant 2^{O(k^2 \log k)}$.

Proof of Proposition 11.2.4. Let $N = r(k; k-2) - 1$, and consider a $(k-2)$-coloring χ of $E(K_N)$ with no monochromatic K_k. Note that a starry coloring of K_k must use $k-1$ colors, so there is also no starry K_k in χ, since χ only uses $k-2$ colors. Similarly, a rainbow coloring of K_k must use $\binom{k}{2} > k-2$ colors, hence there is no rainbow K_k in χ either. This shows that $\mathrm{ER}(k) > N$, proving the proposition. □

Note that this construction rules out the existence of starry or rainbow K_k in a pretty silly fashion, by simply using too few colors to allow these structures to appear. However, as far as I know, this is the only technique that anyone has ever found for lower-bounding $\mathrm{ER}(k)$; in particular, no one knows of a "smarter" way of excluding rainbow or starry K_k. It seems quite possible that if one could come up with such a technique, and thus color with, say, $\Theta(k \log k)$ colors while still avoiding a starry K_k, then one could close the logarithmic gap between the lower and upper bounds on $\mathrm{ER}(k)$.

Chapter 12

The book algorithm

12.1 What are books, and why do they matter?

In this chapter, we will see a proof sketch of Theorem 2.3.3, the breakthrough result of Campos–Griffiths–Morris–Sahasrabudhe [13] showing that $r(k) \leqslant (4-\delta)^k$ for some absolute constant $\delta > 0$. The key idea in the proof is a process for finding large monochromatic cliques in colorings of $E(K_N)$, which they termed the *book algorithm*.

To start with, we need to define book graphs.

Definition 12.1.1. 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. Equivalently, $B_{t,m}$ is obtained from $K_{t,m}$ 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_k.

Note that two important special cases are $t = 1$, where $B_{1,m}$ is simply the star graph $K_{1,m}$, and $m = 1$, where $B_{t,1}$ is simply the clique K_{t+1}. The "book" terminology comes from the case $t = 2$, in which case $B_{2,m}$ consists of m triangles sharing an edge, which kind of looks 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*.

As it turns out, book graphs are used (implicitly or explicitly) in essentially every known proof of Ramsey's theorem. In particular, even Ramsey's [109] original proof proved the finiteness of $r(k)$ by an auxiliary double induction, which showed that $r(B_{t,m})$ is finite for all t, m. However, for our purposes, the relation between books and $r(k)$ is given by the following simple lemma.

Lemma 12.1.2. *Suppose that a two-coloring χ of $E(K_N)$ contains a monochromatic red copy of $B_{t,m}$, where $m \geqslant r(k-t, \ell)$. Then χ 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. $\square$

If this proof looks familiar, it should! In fact, if you think about it, the $t = 1$ case of this lemma is exactly the proof of Theorem 2.1.4. Indeed, in that proof, we showed that if a coloring contains (say) 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 12.1.2 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 12.1.2 basically goes back to the work of Erdős and Szekeres [52], it was first formulated in this language by Thomason [138]. In particular, Thomason noted that Lemma 12.1.2 immediately implies that

$$r(k) \leqslant r(B_{t,m}), \qquad \text{where } m = r(k-t, k). \tag{12.1}$$

Indeed, by the definition of $r(B_{t,m})$, every two-coloring of $E(K_N)$, where $N = r(B_{t,m})$, contains a monochromatic copy of $B_{t,m}$, which then yields a monochromatic copy of K_k by Lemma 12.1.2. As such, Thomason[1] [138] raised the question of determining $r(B_{t,m})$. In particular, Thomason made the following bold conjecture.

Conjecture 12.1.3 (Thomason [138])**.** *$r(B_{t,m}) \leqslant 2^t(m + t - 2) + 2$ for all $m, t \geqslant 1$.*

At the moment, Conjecture 12.1.3 is wide open (and there does not seem to be strong evidence either in favor or against it). Note that this conjecture is extremely powerful, since even the $m = 1$ case implies that $r(t+1) = r(B_{t,1}) \leqslant t2^t$, a much stronger upper bound than anything we currently know. However, an asymptotic version of the conjecture was recently proved by Conlon [21].

Theorem 12.1.4 (Conlon [21])**.** *For every fixed $t \geqslant 1$, we have $r(B_{t,m}) = (2^t + o(1))m$ as $m \to \infty$.*

It is fairly straightforward to show that a random coloring on $(2^t - o(1))m$ vertices contains no monochromatic $B_{t,m}$ with positive probability, hence the main result in Theorem 12.1.4 is the upper bound. In particular, Theorem 12.1.4 implies that $r(B_{t,m}) \leqslant 2^{t+1}m$ for any m which is sufficiently large in terms of t. Plugging this into (12.1), we find that if k is sufficiently large in terms of t (so that $m = r(k-t, k)$ is sufficiently large), then

$$r(k) \leqslant 2^{t+1}m = 2^{t+1}r(k-t, k) \leqslant 2^{t+1}\binom{2k-t}{k}, \tag{12.2}$$

where the final inequality uses Theorem 2.1.4. Unfortunately, Conlon's proof of Theorem 12.1.4 uses Szemerédi's regularity lemma, and therefore requires that m is of tower type in t for this to hold. This in turn means that k must also be of tower type in t, and hence (12.2) gives no meaningful improvement to the Erdős–Szekeres bound of $r(k) \leqslant \binom{2k}{k}$. Conlon's proof was subsequently improved in [29], so that m is "only" required to be of order roughly $2^{2^{2^{t^{25}}}}$ for (12.2) to hold, but this is still far too weak to meaningfully improve Theorem 2.1.4.

[1] At roughly the same time, the same question was also raised by Erdős, Faudree, Rousseau, and Schelp [41], although their motivation was somewhat different.

Nonetheless, Lemma 12.1.2 was used in what were, prior to the work of Campos–Griffiths–Morris–Sahasrabudhe [13], the best known upper bounds on $r(k)$, due to Conlon [19] and Sah [119], who showed[2] that $r(k) \leqslant 4^{k-\Omega((\log k)^2)}$. I will not discuss these techniques in any detail, as they are complex and no longer the state of the art. But the basic idea is to derive, from the failure of the Erdős–Szekeres argument to yield a stronger bound, good quasirandomness properties of a given coloring of $E(K_N)$. By then applying something like the embedding lemma, Lemma 6.1.3, one can use this quasirandomness to find a monochromatic $B_{t,m}$ in the coloring, and then apply Lemma 12.1.2 to complete the proof.

12.2 The algorithms

One of the many new ideas introduced by Campos–Griffiths–Morris–Sahasrabudhe [13] is to use Lemma 12.1.2 directly, rather than its consequences (12.1) and (12.2). Namely, 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 12.1.2. We now describe this algorithm.

We henceforth fix a two-coloring χ of $E(K_N)$. We assume that χ has no monochromatic K_k, and our goal is to eventually obtain a contradiction if N is sufficiently large, namely at least $(4-\delta)^k$. Before describing the book algorithm of Campos–Griffiths–Morris–Sahasrabudhe, we describe for comparison the "Erdős–Szekeres algorithm", which is essentially an alternative way of presenting the proof of Theorem 2.1.4.

12.2.1 The Erdős–Szekeres algorithm

In the Erdős–Szekeres algorithm, we maintain three sets A, B, X; 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 a red clique, B is a blue clique, all edges between A and X are red, and all edges between B and X are blue. To initialize the process, we set $A = B = \varnothing$, and $X = V(K_N)$. We now repeatedly run the following steps.

1. If $|X| \leqslant 1$, $|A| \geqslant k$, or $|B| \geqslant k$, stop the process.
2. Pick a vertex $v \in X$, and check whether v has at least $\frac{1}{2}(|X|-1)$ red neighbors in X.
3. If yes, move v to A and shrink X to the red neighborhood of v. That is, update $A \to A \cup \{v\}$ and $X \to X \cap N_R(v)$, and keep B the same. Call this a *red step.*
4. If not, then v has at least $\frac{1}{2}(|X|-1)$ blue neighbors in X. We now move v to B, and shrink X to the blue neighborhood of v. That is, 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.

[2]This bound is due to Sah [119], who managed to remove an extra $\log\log k$ factor from the exponent in the result of Conlon [19] by proving an optimized version of the embedding lemma in this setting.

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. The basic observation driving the Erdős–Szekeres argument is that if $N \geqslant 4^{k+o(k)}$, then when this process stops, we necessarily produce a monochromatic K_k. Indeed, if we ever stop because $|A| \geqslant k$ or $|B| \geqslant k$, then we definitely have such a monochromatic K_k. However, since $|A| + |B|$ increases by 1 through every iteration of the process, we can do at most $2(k-1)$ steps if we never reach $|A| \geqslant k$ or $|B| \geqslant k$. This in turn means that X shrinks by at most a factor[3] of $2^{2(k-1)+o(k)}$, since it shrinks by a factor of $2 + o(1)$ at every step. Thus, if we start with $N = 2^{2k+o(k)}$, then we will never terminate the process because X becomes too small, and thus can only terminate when we find a monochromatic K_k.

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

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 $A \to A \cup \{v\}$ 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[4] $\gamma|X|$ blue neighbors in X. We now move v to B, and shrink X to the blue neighborhood of v. That is, 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.

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 $1 - \gamma + o(1)$ at every red step, and by a factor of $\gamma + o(1)$ at every blue step, so at the end of the process we have

$$|X| \geqslant 2^{-o(k)}(1-\gamma)^k \gamma^\ell N.$$

On the other hand, the process only terminates if $|X| \leqslant 1$, so this implies $N \leqslant 2^{o(k)}(1-\gamma)^{-k}\gamma^{-\ell}$. One can check, by Stirling's approximation, that

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

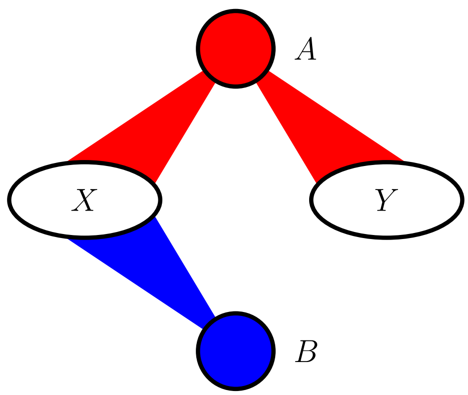
for all $\ell \leqslant k$, and hence this is a contradiction if we choose N sufficiently large, namely of the form $2^{o(k)}\binom{k+\ell}{\ell}$. This recovers Theorem 2.1.4 up to the subexponential error term.

[3]For the rest of this chapter, we will start ignoring the additive -1 terms arising from removing a vertex from X. Of course they need to be carefully dealt with to obtain a correct proof, but they will always contribute a sub-exponential error, and we will simply absorb them in the $o(k)$ in the exponent.

[4]Again, we have started dropping the -1 terms, and will now stop commenting on this.

12.2.2 The book algorithm

We are now ready to describe the book algorithm of Campos–Griffiths–Morris–Sahasrabudhe. As before, we fix a coloring χ of $E(K_N)$, and assume that χ contains 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. 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 12.1.2 to obtain a contradiction. We initialize the process with $A = B = \varnothing$, and $X \sqcup Y$ an arbitrary partition of $V(K_N)$. 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.



Note that there is 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

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

and recall that at the beginning of the process we have $p \geqslant \frac{1}{2}$. 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 2.3.3, we will make the following (completely unjustified) assumption.

Assumption 12.2.1. *Throughout the entire 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, this assumption says that 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 12.2.1 asserts that whenever such a change happens, we magically end up back with the same bi-regularity.

While Assumption 12.2.1 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. As we've seen in a few places (e.g. Lemma 5.2.2 and Theorem 8.1.4), 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 2.3.3, one must repeatedly "clean" X by removing such outliers, and thus one can indeed maintain an approximate version of Assumption 12.2.1, at least ensuring that all vertices in X have roughly the same degree[5]. However, for our exposition, we ignore these important technicalities, and stick with Assumption 12.2.1.

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 12.2.1, 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

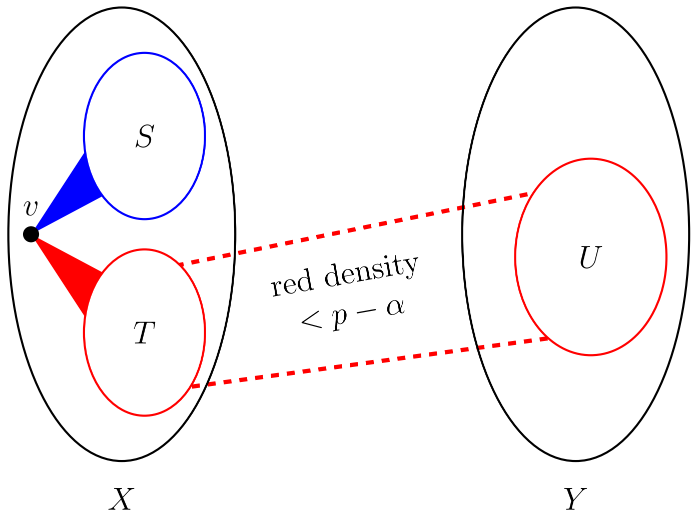
$$p' := d_R(X \cap N_R(v), Y \cap N_R(v)).$$

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 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 12.2.1), and update p to at least $p - \alpha$.

In the Erdős–Szekeres algorithm, 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 at 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–Sahasrabudhe. Pick a vertex $v \in X$, and consider the following picture.

[5]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.



Since v is not prosperous, the red edge density between $T \coloneqq N_R(v) \cap X$ and $U \coloneqq N_R(v) \cap Y$ must be less than $p - \alpha$. However, by Assumption 12.2.1, every vertex in U has $p|X|$ red neighbors in X. Therefore,

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

Rearranging, we find that

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

Let $\beta \coloneqq |S|/|X|$, 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

$$d_R(S, U) > p + \alpha\frac{1 - \beta}{\beta}. \tag{12.3}$$

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 anyway 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 analyzing 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 $\alpha \leqslant \varepsilon/k$, 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. A density-boost step only *increases* p by α, so if we pick $\alpha \leqslant \varepsilon/k$, then even if we do k density-boost steps (the maximum possible number), we will only increase the density by ε, which we just argued is some insignificant amount. In particular, we can just pretend that the density p stays fixed at $\frac{1}{2}$ throughout the entire process. But in this case, we are basically back to the Erdős–Szekeres setting: in the worst case we will do k red steps and k blue or density-boost steps, each time shrinking $|X|$ by roughly a factor of 2. As such, we will not be able to prove any bound better than $r(k) \leqslant 4^k$. The place where the book algorithm wins over the Erdős–Szekeres argument is in obtaining a stronger upper bound on the number of density-boost steps.

The way to ensure this 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 $\alpha = \varepsilon/k$—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

$$\alpha(p) := \begin{cases} \varepsilon/k & \text{if } p \leqslant \frac{1}{2} + \frac{1}{k}, \\ \varepsilon(p - \frac{1}{2}) & \text{otherwise.} \end{cases} \tag{12.4}$$

Again, the point of this is that, if we are at some step of the process where $\alpha > \frac{1}{2}$, 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.

With all of these preliminaries, we are finally able to define the book algorithm[6].

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 (12.4), where ε is some fixed parameter throughout the process.

[6] Actually, the algorithm described here is still incomplete, and a substantial simplification of the actual algorithm defined and analyzed in [13]. For technical reasons I will not go into, things need to be set up somewhat differently to actually deal with the issues arising from the fact that Assumption 12.2.1, as well as the further simplifying assumptions we will shortly make, are not actually true.

3. Check if 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 if 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, every vertex $v \in X$ 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.

12.3 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 12.3.1. *We have $p \geqslant \frac{1}{2} - \varepsilon$ throughout the entire process.*

Proof. As discussed above, every blue step keeps p constant (by Assumption 12.2.1), 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 $p \geqslant \frac{1}{2} + \frac{1}{k}$, whereas $p - \alpha(p) = \varepsilon/k$ whenever $p \leqslant \frac{1}{2} + \frac{1}{k}$. Since we do $t \leqslant k$ red steps, p can never drop below $\frac{1}{2} - t(\varepsilon/k) \geqslant \frac{1}{2} - \varepsilon$. □

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 12.3.2. *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 12.2.1. By Lemma 12.3.1, 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 $|Y| = N/2$. □

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 $\beta_1, \ldots, \beta_s$ be the sequence of values of β for each of the s blue steps. Define β by

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

Lemma 12.3.3. *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 AM-GM inequality 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 $|X| = N/2$, 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 becuase of our "dynamic" choice of α; this is the content of the next lemma, which is called the *zig-zag lemma* in [13].

Lemma 12.3.4. *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 12.3.4[7], but the following sketch captures the main ideas.

[7] In fact, given all of the places we cheated in setting up the book algorithm, Lemma 12.3.4 is probably not even true as stated here.

Proof sketch of Lemma 12.3.4. 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 $q \coloneqq p - \frac{1}{2}$. By our choice of α in (12.4), 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 $q' = q - \frac{1}{2}$). If the step we do is a blue step, then by Assumption 12.2.1, 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 (12.3) 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

$$\frac{q'}{q} \geqslant \begin{cases} 1 & \text{when we do a blue step,} \\ 1-\varepsilon & \text{when we do a red step,} \\ 1+\varepsilon\frac{1-\beta_i}{\beta_i} & \text{when we do the } i\text{th density-boost step.} \end{cases} \tag{12.5}$$

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 (12.5) over all steps of the algorithm, we find that

$$\frac{q_{\text{final}}}{q_{\text{initial}}} \geqslant (1-\varepsilon)^t \prod_{i=1}^{s}\left(1+\varepsilon\frac{1-\beta_i}{\beta_i}\right) \approx e^{-\varepsilon t}\exp\left(\varepsilon\sum_{i=1}^{s}\frac{1-\beta_i}{\beta_i}\right), \tag{12.6}$$

where we approximate $1+x$ and $1-x$ as e^x and e^{-x}, respectively, an approximation that is valid for sufficiently small[8] x. 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 $p_{\text{initial}} \geqslant \frac{1}{k}$. Therefore, $q_{\text{final}}/q_{\text{initial}} \leqslant \frac{k}{2} \leqslant k$. Plugging this into (12.6) and taking logarithms, we find that

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

implying that

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

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

[8] This approxiamtion can be made rigorous (especially for ε, which is sufficiently small for this to be OK), but we're still cheating here. We have no guarantee that $\varepsilon\frac{1-\beta_i}{\beta_i}$ is small, since we have no control over β_i. A real argument would actually need to separate out the contribution from the steps where β_i is very small, and thus where such an approximation is not valid.

As an immediate consequence of Lemma 12.3.4, we obtain our upper bound on the number s of density-boost steps.

Lemma 12.3.5. *We have*

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

Equivalently,

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

Proof. We have that

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

Plugging this into Lemma 12.3.4 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 $\beta/(1-\beta) \leqslant 1$, yielding the first claimed bound. The second bound follows by solving for[9] β. $\square$

We are now ready to put everything together. The process ends when $|X| \leqslant 1$, which by Lemma 12.3.3 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 12.3.5 to find that

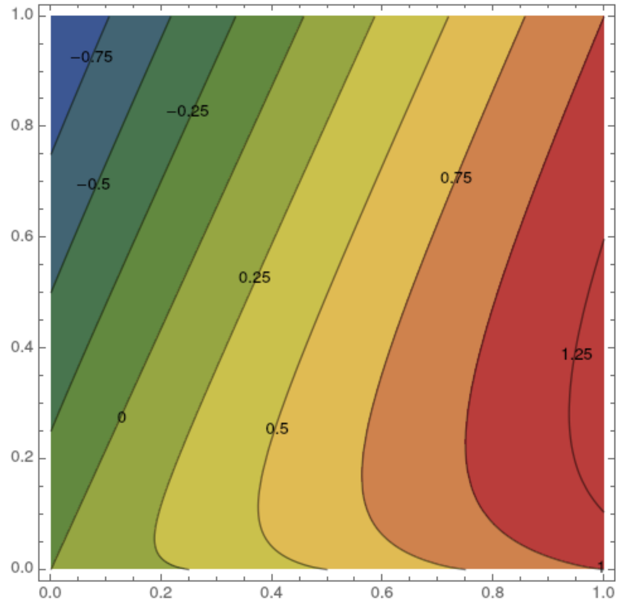
$$N \leqslant \left(\frac{t+s}{s}\right)^s 2^{k+t-s+o(k)}. \tag{12.7}$$

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 $x \coloneqq t/k, y \coloneqq s/k$, and $C \coloneqq \frac{\log N}{k} - 1$, so that $N = 2^{(1+C)k}$; our goal is to be able to pick C a little smaller than 1 and still obtain a contradiction. In terms of these parameters, we can rewrite (12.7) as

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

In other words, we would be done if we could prove that the maximum value of $G(x,y)$ over the square $[0,1]^2$ is strictly less than 1. However, this is not true, as shown on the following contour plot; the maximum value of G is roughly 1.33.

[9] Again, there is some cheating going on here—one can only obtain the claimed asymptotic if s is not too small as a function of k, in order to absorb the error terms. We will continue ignoring this issue.



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 12.1.2.

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 12.3.2. Moreover, by Lemma 12.1.2, we win if $m \geqslant r(k-t, k)$. In other words, we obtain a contradiction unless

$$N \leqslant 2^{t+s+o(k)} r(k-t, k). \tag{12.8}$$

By Theorem 2.1.4, we know that

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

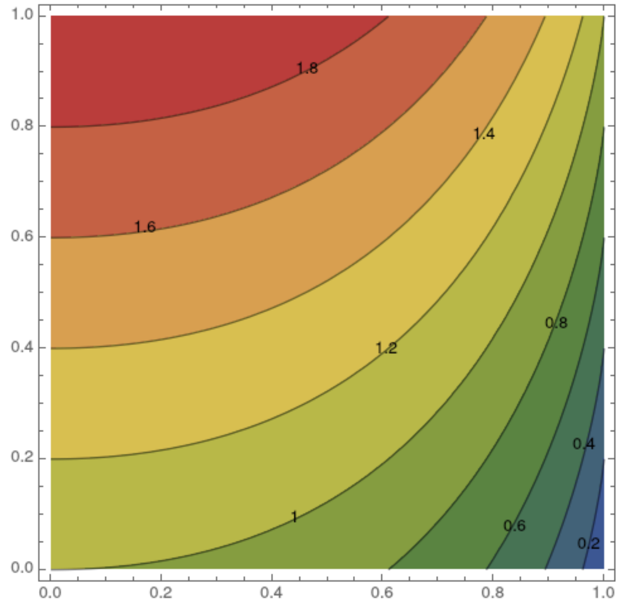
A useful upper bound on binomial coefficients is that $\binom{a}{b} \leqslant 2^{aH(b/a)}$, where $H(x) := -x \log x - (1-x)\log(1-x)$ is the *binary entropy function*. Plugging this in, we find that

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

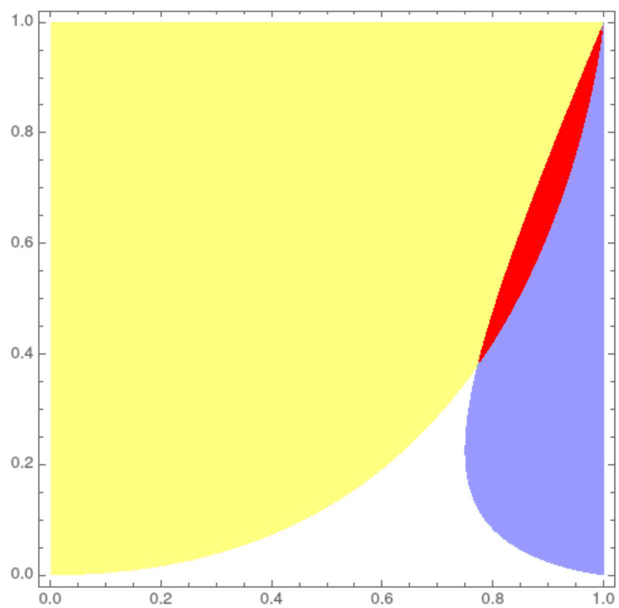
Taking logarithms of (12.8) and dividing by k shows that

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

Putting all of this together, we 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 $G > 1$.



Uhhhhhh... that's not good! 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, so 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 simple argument in Theorem 2.1.4.

12.4 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 problem. In order to solve it, we will use two tricks, both of which involve tweaking the book algorithm. The first is to examine our criterion for deciding whether to do red or blue steps. Recall that, as in the Erdős–Szekeres algorithm, 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 12.3.2. The fact that $|Y|$ decays like $2^{-t}N$ is unavoidable (and best possible) by Theorem 12.1.4, 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 12.3.5 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 fix a number $\mu \in [0,1]$, which will be fixed throughout the argument. In step 3 of the book algorithm, 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 now 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 tradeoff—if μ is very small then every blue step decreases $|X|$ by a lot, and thus the process will terminate quickly, and we need to balance these two effects.

In this modified book algorithm, Lemmas 12.3.1, 12.3.2, 12.3.4 and 12.3.5 remain true; the only change is that Lemma 12.3.3 needs to be modified to the following statement, reflecting the fact that each blue (resp. red) step shrinks X by a factor of μ (resp. $1-\mu$) in the worst case. The proof is otherwise identical to that of Lemma 12.3.3.

Lemma 12.4.1 (Modified Lemma 12.3.3)**.** *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

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

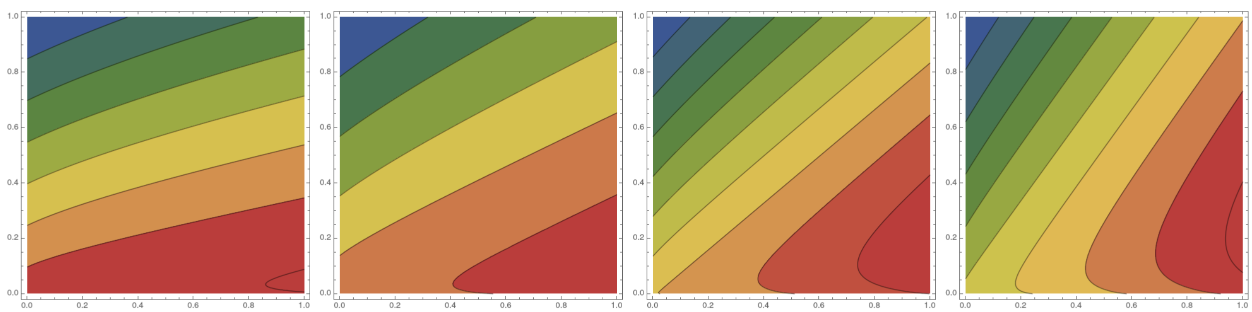
Since the process terminates when $|X| \leqslant 1$, we conclude from Lemma 12.4.1 that

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

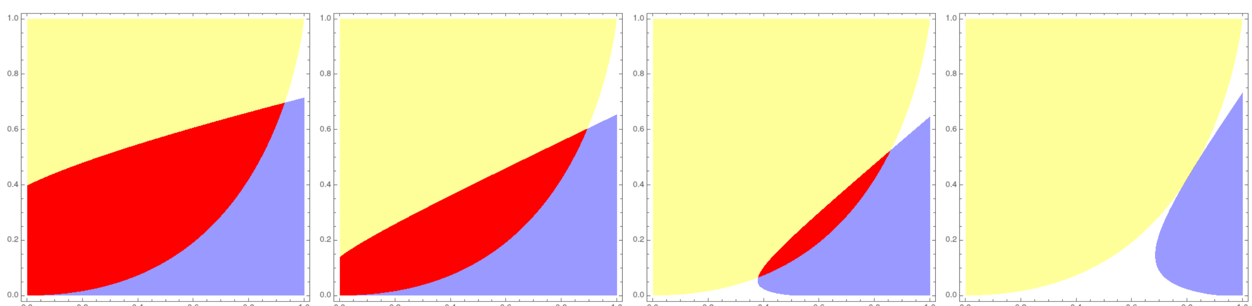
where the final inequality follows from the lower bound on β in Lemma 12.3.5. Taking logarithms and dividing by k, we conclude that

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

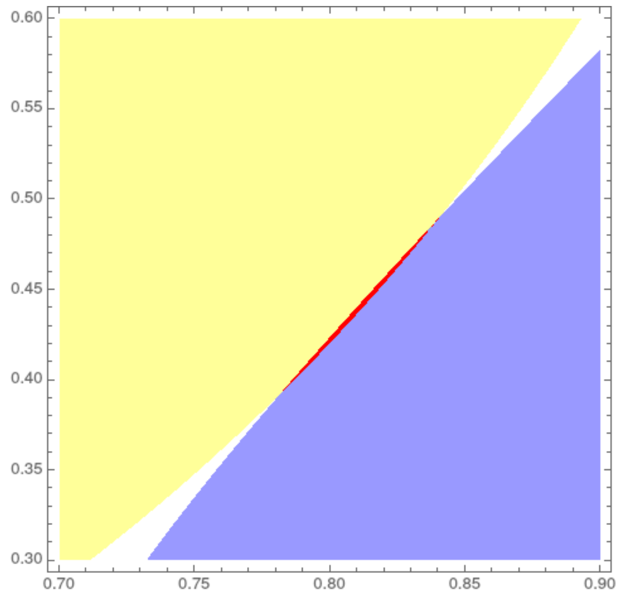
Note that in the case $\mu = \frac{1}{2}$, we precisely recover the previous function G, which of course makes sense as we are then recovering the previous book algorithm. Here are contour plots of G_μ for $\mu \in \{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\}$, respectively.



And here are pictures of the regions where $F > 1$ and $G_\mu > 1$, for $\mu \in \{\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\}$.



It looks like we're already done at $\mu = \frac{2}{5}$, but unfortunately we're not: one can check that $\min\{F(x,y), G_{2/5}(x,y)\}$ 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.

12.4.1 Off-diagonal Ramsey numbers

So far, we have played with the parameter μ in order to move around 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 $G_{2/5}$ 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 $\ell \approx k/4$.

There is actually a good reason to expect this to work. Remember that in the Erdős–Szekeres algorithm, as presented in Section 12.2.1, 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 $\ell \approx k/4$, 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 $\mu^\ell \approx (1/5)^{k/4} \approx 0.67^k$; in contrast, in the argument above, the blue steps shrink X by $(2/5)^k = 0.4^k$, which is much smaller. Hence we may expect the trade-offs to work well for us.

For completeness, here is our modified 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 $p_0 = d_R(X,Y)$ be the density of red edges between X and Y at the beginning of the process, and define

$$\alpha(p) := \begin{cases} \varepsilon/k & \text{if } p \leqslant p_0 + \frac{1}{k}, \\ \varepsilon(p-p_0) & \text{otherwise.} \end{cases} \tag{12.9}$$

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 (12.4).
3. Check if 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 if 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, every vertex $v \in X$ 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.

Of course, Lemmas 12.3.4, 12.3.5 and 12.4.1 remain true in this setting. Unfortunately, there is an additional complication introduced by moving to the diagonal setting. Before, when we seeked to upper-bound $r(k)$, we could assume that the initial red density p_0 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_0. Let us make another completely unjustified assumption.

Assumption 12.4.2. *At the beginning of the process, we have $p_0 \geqslant \frac{k}{k+\ell} = 1 - \mu$.*

Note that this is a natural assumption, since if the red edge density were substantially smaller $1-\mu$ "everywhere", then the simple Erdős–Szekeres algorithm should already be able to prove a stronger upper bound than $r(k,\ell) \leqslant \binom{k+\ell}{\ell}$. In fact, one can essentially force Assumption 12.4.2 to hold because of such an argument; if we start with $p_0 < \frac{k}{k+\ell}$, 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 the Erdős–Szekeres algorithm itself will prove that $r(k,\ell) \ll \binom{k+\ell}{\ell}$.

Given Assumption 12.4.2, we obtain the following modified versions of Lemmas 12.3.1 and 12.3.2. The proof is identical (recall the modified definition of α in (12.9)), and the only other thing to note is that Y shrinks by a factor of $p \geqslant 1-\mu-\varepsilon$ during every red or density-boost step.

Lemma 12.4.3. *We have $p \geqslant p_0 - \varepsilon \geqslant (1-\mu) - \varepsilon$ 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 (modulo Assumptions 12.2.1 and 12.4.2, and the sketchiness in the proof of Lemma 12.3.4) an exponentially-improved upper bound on $r(k,\ell)$. The bound claimed below is actually stronger than anything proved in [13]; it is very possible that this bound is stronger than anything that can actually be rigorously proved without dozens of pages of computation, so you should take the theorem statement with a grain of salt.

Theorem 12.4.4. *We have $r(k,\ell) \leqslant 2^{-\frac{1}{3}k+o(k)}\binom{k+\ell}{\ell}$ for all $\ell \leqslant k/4$.*

Proof. Let $N = 2^{(1+C+o(1))k}$, and fix a 2-coloring of $E(K_N)$. Let us assume for contradiction that there is no red K_k or blue K_ℓ in this coloring. We apply the off-diagonal book algorithm above, 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 never output that A is a red K_k or B is a blue K_ℓ, then the process only terminates when $|X| \leqslant 1$, and we also have that $b + s \leqslant \ell$. Plugging this into Lemma 12.4.1, we find that

$$N \leqslant 2^{o(k)}(1-\mu)^{-t}\mu^{s-\ell}\beta^{-s} \leqslant 2^{o(k)}(1-\mu)^{-t}\mu^{s-\ell}\left(\frac{s+t}{s}\right)^s. \tag{12.10}$$

Note that we have plugged in the assumption $b+s \leqslant \ell$, which gives us the better exponent $s-\ell$ on μ. Taking logarithms and dividing by k shows that

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

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 12.4.3, we have

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

If $|Y| \geqslant r(k-t,\ell)$, then we are done by Lemma 12.1.2, so we may assume that $|Y| < r(k-t,\ell)$. Taking logarithms and dividing by k again shows that

$$C - o(1) \leqslant -1 + (x+y)\log\left(\frac{1}{1-\mu}\right) + \frac{1}{k}\log r(k-t,\ell). \tag{12.11}$$

By Theorem 2.1.4, we have

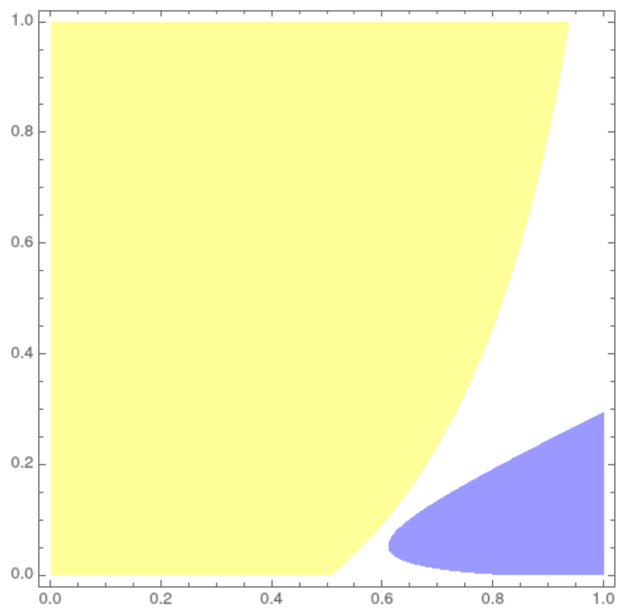
$$\begin{aligned}\log r(k-t,\ell) &\leqslant \log \binom{k-t+\ell}{k-t} \\ &\leqslant (k-t+\ell) H\left(\frac{k-t}{k-t+\ell}\right) \\ &= k \cdot \left(1-x+\frac{\mu}{1-\mu}\right) H\left(\frac{1-x}{1-x+\mu/(1-\mu)}\right).\end{aligned}$$

Plugging this into (12.11) shows that

$$C-o(1) \leqslant -1+(x+y)\log\left(\frac{1}{1-\mu}\right)+\left(1-x+\frac{\mu}{1-\mu}\right) H\left(\frac{\mu/(1-\mu)}{1-x+\mu/(1-\mu)}\right) =: \widetilde{F}_\mu(x,y).$$

We are no longer trying to beat the bound $r(k) \leqslant 4^k$, so our goal is no longer obtaining a contradiction for some $C < 1$. Instead, we are comparing to $\frac{1}{k}\log\binom{k+\ell}{\ell}$, which equals $(1+\frac{\mu}{1-\mu})H(\mu)+o(1)$. So what we would like to show is that for all $\mu \leqslant \frac{1}{5}$, we have $\min\{\widetilde{F}_\mu(x,y), \widetilde{G}_\mu(x,y)\} < \frac{\mu}{1-\mu}H(\mu) - \delta$ for all $x, y \in [0,1]$, where $\delta > 0$ is some absolute constant that will end up in the exponent of N.

In fact, one can check that for $\mu \leqslant \frac{1}{5}$, we may take δ as large as $\frac{1}{3}$. Indeed, here is a plot of the regions where $\widetilde{F}_\mu > \frac{\mu}{1-\mu}H(\mu) - \frac{1}{3}$ and $\widetilde{G}_\mu > \frac{\mu}{1-\mu}H(\mu) - \frac{1}{3}$, respectivey, 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 whenever $C > \frac{\mu}{1-\mu}H(\mu) - \frac{1}{3}$, proving the bound

$$r(k,\ell) \leqslant 2^{\left(1+\frac{\mu}{1-\mu}H(\mu)-\frac{1}{3}+o(1)\right)k} = 2^{-\frac{1}{3}k+o(k)}\binom{k+\ell}{\ell}$$

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

12.4.2 Back to diagonal

Now that we have an upper bound on $r(k,\ell)$ for $\ell \leqslant k/4$, we can finally complete the proof of Theorem 2.3.3. We will actually prove the following bound; again, the exact statement is slightly stronger than what is in [13], and should not be taken too literally, since there are many parts of the proof that need to be made formal, and would likely lead to a worse bound.

Theorem 12.4.5. *We have* $r(k) \leqslant 2^{(2-\frac{1}{200}+o(1))k} \approx 3.986^k$.

Proof. Let $N = 2^{(1+C+o(1))k}$, and fix a two-coloring χ of $E(K_N)$, which we may assume has no monochromatic K_k. We run the book algorithm with $k = \ell$ and $\mu = \frac{2}{5}$. Thanks to Theorem 12.4.4 (plus Theorem 2.1.4), we know that

$$r(k-t,k) \leqslant \begin{cases} \binom{2k-t}{k-t} & \text{if } t < \frac{3}{4}k, \\ 2^{-\frac{1}{3}k+o(k)}\binom{2k-t}{k-t} & \text{if } t \geqslant \frac{3}{4}k. \end{cases}$$

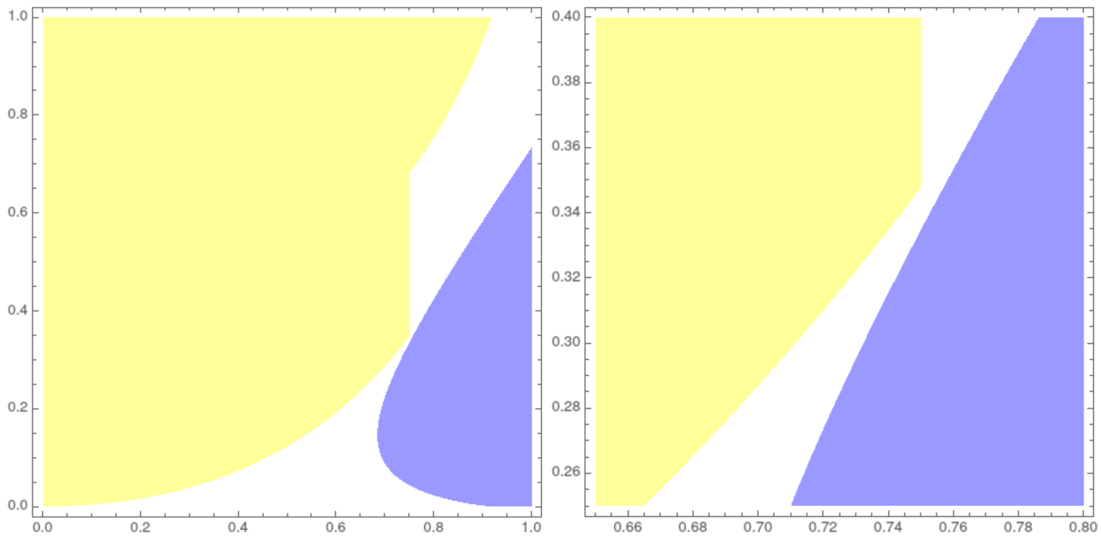
Recall that we obtain a contradiction if $|Y| \geqslant r(k-t,k)$ at the end of the process, hence we may assume that $|Y| \leqslant r(k-t,k)$. Combining this with Lemma 12.3.2[10], we see that we get a contradiction if

$$\begin{aligned} C - o(1) &\leqslant -1 + (x+y) + \frac{1}{k}\log r(k-t,k) \\ &\leqslant \begin{cases} -1 + (x+y) + (2-x)H(\frac{1-x}{2-x}) & \text{if } x < \frac{3}{4}, \\ -1 + (x+y) - \frac{1}{3} + (2-x)H(\frac{1-x}{2-x}) & \text{if } x \geqslant \frac{3}{4} \end{cases} \\ &= F(x,y) - \frac{1}{3}\mathbf{1}_{x \geqslant \frac{4}{5}} \\ &=: \widehat{F}(x,y). \end{aligned}$$

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.

[10] We are back to the diagonal setting, so we may assume that $p_0 \geqslant \frac{1}{2}$. Therefore Lemmas 12.3.1 and 12.3.2 are again valid.



In fact, one can check that $\max_{x,y\in[0,1]} \min\{\widehat{F}(x, y), G_{\frac{2}{5}}(x, y)\} < 0.995$. Therefore, we obtain a contradiction if $C \geqslant .995 = 1 - \frac{1}{200}$, proving that $r(k) \leqslant 2^{(2-\frac{1}{200}+o(1))k}$, as claimed. $\square$

Acknowledgements: I am grateful to Yu-Cheng (Daniel) Chiu, Zach Hunter, Zhihan Jin, Aleksa Milojević, Immanuel Quarch, and Yanbo Zhang for correcting a large number of errors in earlier drafts of these lecture notes.

Bibliography

[1] H. L. Abbott, A note on Ramsey's theorem, *Canad. Math. Bull.* **15** (1972), 9–10.

[2] M. Ajtai, J. Komlós, and E. Szemerédi, A note on Ramsey numbers, *J. Combin. Theory Ser. A* **29** (1980), 354–360.

[3] N. Alon and V. Rödl, Sharp bounds for some multicolor Ramsey numbers, *Combinatorica* **25** (2005), 125–141.

[4] N. Alon and J. H. Spencer, *The probabilistic method*, Wiley Series in Discrete Mathematics and Optimization, fourth ed., John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[5] J. Balogh and W. Samotij, An efficient container lemma, *Discrete Anal.* (2020), Paper No. 17, 56.

[6] F. A. Behrend, On sets of integers which contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. U.S.A.* **32** (1946), 331–332.

[7] A. Bishnoi, Finite geometry and Ramsey theory, 2021. Lecture notes available online at `https://anuragbishnoi.files.wordpress.com/2021/01/minicourse.pdf`.

[8] T. F. Bloom and O. Sisask, An improvement to the Kelley–Meka bounds on three-term arithmetic progressions, 2023. Preprint available at arXiv:2309.02353.

[9] T. Bohman and P. Keevash, The early evolution of the H-free process, *Invent. Math.* **181** (2010), 291–336.

[10] T. Bohman and P. Keevash, Dynamic concentration of the triangle-free process, in *The Seventh European Conference on Combinatorics, Graph Theory and Applications*, *CRM Series*, vol. 16, Ed. Norm., Pisa, 2013, 489–495.

[11] M. Bucić, T. Nguyen, A. Scott, and P. Seymour, Induced subgraph density. I. A loglog step towards Erdős–Hajnal, 2023. Preprint available at arXiv:2301.10147.

[12] S. A. Burr and P. Erdős, 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*, *Colloq. Math. Soc. János Bolyai*, vol. Vol. 10, North-Holland, Amsterdam-London, 1975, 215–240.

[13] M. Campos, S. Griffiths, R. Morris, and J. Sahasrabudhe, An exponential improvement for diagonal Ramsey, 2023. Preprint available at arXiv:2303.09521.

[14] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe, A new lower bound for sphere packing, 2023. Preprint available at arXiv:2312.10026.

[15] A. K. Chandra, M. L. Furst, and R. J. Lipton, Multi-party protocols, in *Proceedings of the fifteenth annual ACM Symposium on Theory of Computing*, 1983, 94–99.

[16] F. R. K. Chung, On the Ramsey numbers $N(3, 3, \ldots, 3; 2)$, *Discrete Math.* **5** (1973), 317–321.

[17] C. Chvatál, V. Rödl, E. Szemerédi, and W. T. Trotter, Jr., The Ramsey number of a graph with bounded maximum degree, *J. Combin. Theory Ser. B* **34** (1983), 239–243.

[18] V. Chvátal and J. Komlós, Some combinatorial theorems on monotonicity, *Canad. Math. Bull.* **14** (1971), 151–157.

[19] D. Conlon, A new upper bound for diagonal Ramsey numbers, *Ann. of Math. (2)* **170** (2009), 941–960.

[20] D. Conlon, A sequence of triangle-free pseudorandom graphs, *Combin. Probab. Comput.* **26** (2017), 195–200.

[21] D. Conlon, The Ramsey number of books, *Adv. Comb.* (2019), Paper No. 3, 12pp.

[22] D. Conlon, Extremal numbers of cycles revisited, *Amer. Math. Monthly* **128** (2021), 464–466.

[23] D. Conlon, Monochromatic combinatorial lines of length three, *Proc. Amer. Math. Soc.* **150** (2022), 1–4.

[24] D. Conlon and A. Ferber, Lower bounds for multicolor Ramsey numbers, *Adv. Math.* **378** (2021), Paper No. 107528, 5pp.

[25] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, *Geom. Funct. Anal.* **22** (2012), 1191–1256.

[26] D. Conlon, J. Fox, and B. Sudakov, On two problems in graph Ramsey theory, *Combinatorica* **32** (2012), 513–535.

[27] D. Conlon, J. Fox, and B. Sudakov, Recent developments in graph Ramsey theory, in *Surveys in combinatorics 2015*, *London Math. Soc. Lecture Note Ser.*, vol. 424, Cambridge Univ. Press, Cambridge, 2015, 49–118.

[28] D. Conlon, J. Fox, and B. Sudakov, Short proofs of some extremal results II, *J. Combin. Theory Ser. B* **121** (2016), 173–196.

[29] D. Conlon, J. Fox, and Y. Wigderson, Ramsey numbers of books and quasirandomness, *Combinatorica* **42** (2022), 309–363.

[30] D. Conlon, S. Mattheus, D. Mubayi, and J. Verstraëte, Ramsey numbers and the Zarankiewicz problem, 2023. Preprint available at arXiv:2307.08694.

[31] W. Deuber, Generalizations of Ramsey's theorem, in *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vols. I, II, III*, *Colloq. Math. Soc. János Bolyai*, vol. Vol. 10, North-Holland, Amsterdam-London, 1975, 323–332.

[32] L. E. Dickson, On the congruence $x^n + y^n + z^n \equiv 0 \pmod p$, *J. Reine Angew. Math.* **135** (1909), 134–141.

[33] R. Diestel, *Graph theory*, *Graduate Texts in Mathematics*, vol. 173, fifth ed., Springer, Berlin, 2017.

[34] P. Dodos, V. Kanellopoulos, and K. Tyros, A simple proof of the density Hales-Jewett theorem, *Int. Math. Res. Not. IMRN* (2014), 3340–3352.

[35] R. A. Duke, H. Lefmann, and V. Rödl, A fast approximation algorithm for computing the frequencies of subgraphs in a given graph, *SIAM J. Comput.* **24** (1995), 598–620.

[36] N. Eaton, Ramsey numbers for sparse graphs, *Discrete Math.* **185** (1998), 63–75.

[37] P. Erdős, Graph theory and probability. II, *Canadian J. Math.* **13** (1961), 346–352.

[38] P. Erdős, Problems and results on finite and infinite graphs, in *Recent advances in graph theory (Proc. Second Czechoslovak Sympos., Prague, 1974)*, Academia, Prague, 1975, 183–192. (loose errata).

[39] P. Erdős, Some of my favourite problems in various branches of combinatorics, *Matematiche (Catania)* **47** (1992), 231–240.

[40] P. Erdős, Problems and results in discrete mathematics, *Discrete Math.* **136** (1994), 53–73.

[41] P. Erdős, R. J. Faudree, C. C. Rousseau, and R. H. Schelp, The size Ramsey number, *Period. Math. Hungar.* **9** (1978), 145–161.

[42] P. Erdős and R. L. Graham, On partition theorems for finite graphs, in *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vols. I, II, III*, *Colloq. Math. Soc. János Bolyai*, vol. Vol. 10, North-Holland, Amsterdam-London, 1975, 515–527.

[43] P. Erdős and A. Hajnal, On decomposition of graphs, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 359–377.

[44] P. Erdős and A. Hajnal, On spanned subgraphs of graphs, in *Contributions to graph theory and its applications (Internat. Colloq., Oberhof, 1977) (German)*, Tech. Hochschule Ilmenau, Ilmenau, 1977, 80–96.

[45] P. Erdős and A. Hajnal, Ramsey-type theorems, *Discrete Appl. Math.* **25** (1989), 37–52. Combinatorics and complexity (Chicago, IL, 1987).

[46] P. Erdős, A. Hajnal, and L. Pósa, Strong embeddings of graphs into colored graphs, in *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vols. I, II, III*, *Colloq. Math. Soc. János Bolyai*, vol. Vol. 10, North-Holland, Amsterdam-London, 1975, 585–595.

[47] P. Erdős, A. Hajnal, and R. Rado, Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 93–196.

[48] P. Erdős and R. Rado, Combinatorial theorems on classifications of subsets of a given set, *Proc. London Math. Soc. (3)* **2** (1952), 417–439.

[49] P. Erdős and G. Szekeres, On some extremum problems in elementary geometry, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **3/4** (1960/61), 53–62.

[50] P. Erdős and A. Szemerédi, On a Ramsey type theorem, *Period. Math. Hungar.* **2** (1972), 295–299.

[51] P. Erdös, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292–294.

[52] P. Erdös and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* **2** (1935), 463–470.

[53] G. Fiz Pontiveros, S. Griffiths, and R. Morris, The triangle-free process and the Ramsey number $R(3, k)$, *Mem. Amer. Math. Soc.* **263** (2020), v+125.

[54] J. Folkman, Graphs with monochromatic complete subgraphs in every edge coloring, *SIAM J. Appl. Math.* **18** (1970), 19–24.

[55] J. Fox and B. Sudakov, Induced Ramsey-type theorems, *Adv. Math.* **219** (2008), 1771–1800.

[56] J. Fox and B. Sudakov, Two remarks on the Burr-Erdős conjecture, *European J. Combin.* **30** (2009), 1630–1645.

[57] J. Fox and B. Sudakov, Dependent random choice, *Random Structures Algorithms* **38** (2011), 68–99.

[58] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256.

[59] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.* **34** (1978), 275–291.

[60] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *J. Anal. Math.* **57** (1991), 64–119.

[61] A. Galluccio, M. Simonovits, and G. Simonyi, On the structure of co-critical graphs, in *Graph theory, combinatorics, and algorithms, Vol. 1, 2 (Kalamazoo, MI, 1992)*, Wiley-Intersci. Publ., Wiley, New York, 1995, 1053–1071.

[62] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), 465–588.

[63] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math. (2)* **166** (2007), 897–946.

[64] R. Graham and J. Solymosi, Monochromatic equilateral right triangles on the integer grid, in *Topics in discrete mathematics*, *Algorithms Combin.*, vol. 26, Springer, Berlin, 2006, 129–132.

[65] R. L. Graham, On edgewise 2-colored graphs with monochromatic triangles and containing no complete hexagon, *J. Combinatorial Theory* **4** (1968), 300.

[66] R. L. Graham, V. Rödl, and A. Ruciński, On graphs with linear Ramsey numbers, *J. Graph Theory* **35** (2000), 176–192.

[67] R. L. Graham, V. Rödl, and A. Ruciński, On bipartite graphs with linear Ramsey numbers, *Combinatorica* **21** (2001), 199–209.

[68] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, Wiley-Interscience Series in Discrete Mathematics and Optimization, second ed., John Wiley & Sons, Inc., New York, 1990. A Wiley-Interscience Publication.

[69] J. E. Graver and J. Yackel, Some graph theoretic results associated with Ramsey's theorem, *J. Combinatorial Theory* **4** (1968), 125–175.

[70] A. W. Hales and R. I. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222–229.

[71] T. Hales, M. Adams, G. Bauer, T. D. Dang, J. Harrison, L. T. Hoang, C. Kaliszyk, V. Magron, S. McLaughlin, T. T. Nguyen, Q. T. Nguyen, T. Nipkow, S. Obua, J. Pleso, J. Rute, A. Solovyev, T. H. A. Ta, N. T. Tran, T. D. Trieu, J. Urban, K. Vu, and R. Zumkeller, A formal proof of the Kepler conjecture, *Forum Math. Pi* **5** (2017), e2, 29pp.

[72] T. C. Hales, A proof of the Kepler conjecture, *Ann. of Math. (2)* **162** (2005), 1065–1185.

[73] X. He and Y. Wigderson, Multicolor Ramsey numbers via pseudorandom graphs, *Electron. J. Combin.* **27** (2020), Paper No. 1.32, 8pp.

[74] S. Janson, T. Ł uczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[75] S. Jukna, *Extremal combinatorics: With applications in computer science*, Texts in Theoretical Computer Science. An EATCS Series, second ed., Springer, Heidelberg, 2011.

[76] Z. Kelley and R. Meka, Strong bounds for 3-progressions, in *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE Computer Soc., Los Alamitos, CA, 2023, 933–973.

[77] J. H. Kim, The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures Algorithms* **7** (1995), 173–207.

[78] D. J. Kleitman and K. J. Winston, On the number of graphs without 4-cycles, *Discrete Math.* **41** (1982), 167–172.

[79] Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers, *Random Structures Algorithms* **46** (2015), 1–25.

[80] A. Kostochka and B. Sudakov, On Ramsey numbers of sparse graphs, *Combin. Probab. Comput.* **12** (2003), 627–641. Special issue on Ramsey theory.

[81] A. V. Kostochka and V. Rödl, On graphs with small Ramsey numbers, *J. Graph Theory* **37** (2001), 198–204.

[82] T. Kövari, V. T. Sós, and P. Turán, On a problem of K. Zarankiewicz, *Colloq. Math.* **3** (1954), 50–57.

[83] E. Kushilevitz and N. Nisan, *Communication complexity*, Cambridge University Press, Cambridge, 1997.

[84] M. Kwan, A. Sah, L. Sauermann, and M. Sawhney, Anticoncentration in Ramsey graphs and a proof of the Erdős-McKay conjecture, *Forum Math. Pi* **11** (2023), Paper No. e21, 74pp.

[85] I. Leader, Sparse Ramsey theory, 2003. Available online at `https://web.archive.org/web/20240418132306/https://cranch.staff.shef.ac.uk/files/sparse-ramsey.pdf`.

[86] C. Lee, Ramsey numbers of degenerate graphs, *Ann. of Math. (2)* **185** (2017), 791–829.

[87] H. Lefmann, A note on Ramsey numbers, *Studia Sci. Math. Hungar.* **22** (1987), 445–446.

[88] H. Lefmann and V. Rödl, On Erdős-Rado numbers, *Combinatorica* **15** (1995), 85–104.

[89] J. Leng, A. Sah, and M. Sawhney, Improved bounds for Szemerédi's theorem, 2024. Preprint available at arXiv:2402.17995.

[90] X. Li, Two source extractors for asymptotically optimal entropy, and (many) more, 2023. Preprint available at arXiv:2303.06802.

[91] Y. Li and C. C. Rousseau, On book-complete graph Ramsey numbers, *J. Combin. Theory Ser. B* **68** (1996), 36–44.

[92] L. Lovász, *Combinatorial problems and exercises*, second ed., North-Holland Publishing Co., Amsterdam, 1993.

[93] S. Mattheus and J. Verstraëte, The asymptotics of $r(4,t)$, *Ann. of Math. (2)* (2024), to appear. Preprint available at arXiv:2306.04007.

[94] G. Moshkovitz and A. Shapira, Ramsey theory, integer partitions and a new proof of the Erdős-Szekeres theorem, *Adv. Math.* **262** (2014), 1107–1129.

[95] D. Mubayi and J. Verstraëte, A note on pseudorandom Ramsey graphs, 2019. Preprint available at arXiv:1909.01461.

[96] J. Nešetřil and V. Rödl, The Ramsey property for graphs with forbidden complete subgraphs, *J. Combinatorial Theory Ser. B* **20** (1976), 243–249.

[97] J. Nešetřil and V. Rödl, A simple proof of the Galvin-Ramsey property of the class of all finite graphs and a dimension of a graph, *Discrete Math.* **23** (1978), 49–55.

[98] J. Nešetřil and V. Rödl, Partition theory and its applications, in B. Bollobás (ed.), *Surveys in Combinatorics*, London Mathematical Society Lecture Note Series, Cambridge University Press, 1979, 96–156.

[99] J. Nešetřil and V. Rödl, The partite construction and Ramsey set systems, *Discrete Math.* **75** (1989), 327–334.

[100] A. Nilli, Shelah's proof of the Hales-Jewett theorem, in *Mathematics of Ramsey theory*, *Algorithms Combin.*, vol. 5, Springer, Berlin, 1990, 150–151.

[101] M. E. O'Nan, Automorphisms of unitary block designs, *J. Algebra* **20** (1972), 495–511.

[102] S. Peluse, Recent progress on bounds for sets with no three terms in arithmetic progression, *Astérisque* (2022), No. 1196, 581.

[103] Y. Peng, V. Rödl, and A. Ruciński, Holes in graphs, *Electron. J. Combin.* **9** (2002), Research Paper 1, 18.

[104] D. H. J. Polymath, Coloring Hales–Jewett theorem, 2009. Avilable online at `http://web.archive.org/web/20210606051518/https://asone.ai/polymath/index.php?title=Coloring_Hales-Jewett_theorem`.

[105] D. H. J. Polymath, A new proof of the density Hales-Jewett theorem, *Ann. of Math. (2)* **175** (2012), 1283–1327.

[106] H. J. Prömel and V. Rödl, Non-Ramsey graphs are $c \log n$-universal, *J. Combin. Theory Ser. A* **88** (1999), 379–384.

[107] R. Rado, Note on combinatorial analysis, *Proc. London Math. Soc. (2)* **48** (1943), 122–160.

[108] R. Rado, Some partition theorems, in *Combinatorial theory and its applications, I-III (Proc. Colloq., Balatonfüred, 1969)*, *Colloq. Math. Soc. János Bolyai*, vol. 4, North-Holland, Amsterdam-London, 1970, 929–936.

[109] F. P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc. (2)* **30** (1929), 264–286.

[110] C. Reiher and V. Rödl, The girth Ramsey theorem, 2023. Preprint available at arXiv:2308.15589.

[111] V. Rödl, *The dimension of a graph and generalized Ramsey theorems*, Master's thesis, Charles University, 1973.

[112] V. Rödl, On universality of graphs with uniformly distributed edges, *Discrete Math.* **59** (1986), 125–134.

[113] V. Rödl, B. Nagle, J. Skokan, M. Schacht, and Y. Kohayakawa, The hypergraph regularity method and its applications, *Proc. Natl. Acad. Sci. USA* **102** (2005), 8109–8113.

[114] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8** (1995), 917–942.

[115] C. A. Rogers, Existence theorems in the geometry of numbers, *Ann. of Math. (2)* **48** (1947), 994–1002.

[116] K. Roth, Sur quelques ensembles d'entiers, *C. R. Acad. Sci. Paris* **234** (1952), 388–390.

[117] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.

[118] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, *Colloq. Math. Soc. János Bolyai*, vol. 18, North-Holland, Amsterdam-New York, 1978, 939–945.

[119] A. Sah, Diagonal Ramsey via effective quasirandomness, *Duke Math. J.* **172** (2023), 545–567.

[120] W. Samotij, Counting independent sets in graphs, *European J. Combin.* **48** (2015), 5–18.

[121] J. H. Sanders, *A generalization of Schur's theorem*, Ph.D. thesis, Yale University, 1968.

[122] W. Sawin, An improved lower bound for multicolor Ramsey numbers and a problem of Erdős, *J. Combin. Theory Ser. A* **188** (2022), Paper No. 105579, 11.

[123] I. Schur, Über die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *Jahresber. Dtsch. Math.-Ver.* **25** (1917), 114–117.

[124] A. Seidenberg, A simple proof of a theorem of Erdős and Szekeres, *J. London Math. Soc.* **34** (1959), 352.

[125] J. B. Shearer, A note on the independence number of triangle-free graphs, *Discrete Math.* **46** (1983), 83–87.

[126] J. B. Shearer, On the independence number of sparse graphs, *Random Structures Algorithms* **7** (1995), 269–271.

[127] S. Shelah, Primitive recursive bounds for van der Waerden numbers, *J. Amer. Math. Soc.* **1** (1988), 683–697.

[128] V. T. Sós, Induced subgraphs and Ramsey colorings, 2013. Presented at the 16th International Conference on Random Structures and Algorithms. `https://web.archive.org/web/20150910071523/http://rsa2013.amu.edu.pl/abstracts/Sos.Vera.pdf`.

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

[130] J. Spencer, Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20** (1977/78), 69–76.

[131] J. M. Steele, Variations on the monotone subsequence theme of Erdős and Szekeres, in *Discrete probability and algorithms (Minneapolis, MN, 1993)*, *IMA Vol. Math. Appl.*, vol. 72, Springer, New York, 1995, 111–131.

[132] A. Suk, On the Erdős-Szekeres convex polygon problem, *J. Amer. Math. Soc.* **30** (2017), 1047–1053.

[133] T. Szabó, On nearly regular co-critical graphs, *Discrete Math.* **160** (1996), 279–281.

[134] G. Szekeres and L. Peters, Computer solution to the 17-point Erdős-Szekeres problem, *ANZIAM J.* **48** (2006), 151–164.

[135] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89–104.

[136] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

[137] E. Szemerédi, Regular partitions of graphs, in *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, *Colloq. Internat. CNRS*, vol. 260, CNRS, Paris, 1978, 399–401.

[138] A. Thomason, On finite Ramsey numbers, *European J. Combin.* **3** (1982), 263–273.

[139] K. Tikhomirov, A remark on the Ramsey number of the hypercube, 2022. Preprint available at arXiv:2208.14568.

[140] B. L. Van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wiskunde* **15** (1927), 212–216.

[141] M. S. Viazovska, The sphere packing problem in dimension 8, *Ann. of Math. (2)* **185** (2017), 991–1015.

[142] Y. Wigderson, An improved lower bound on multicolor Ramsey numbers, *Proc. Amer. Math. Soc.* **149** (2021), 2371–2374.

[143] Y. Wigderson, Lecture notes on dependent random choice, 2023. Available online at `https://n.ethz.ch/~ywigderson/math/static/DependentRandomChoice.pdf`.

[144] E. Witt, Ein kombinatorischer Satz der Elementargeometrie, *Math. Nachr.* **6** (1952), 261–262.

[145] Y. Zhao, *Graph theory and additive combinatorics—exploring structure and randomness*, Cambridge University Press, Cambridge, 2023.