Ramsey theory—lecture notes

Yuval Wigderson

Spring semester 2024
Last updated: April 19, 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 [84] 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 [84])**.** *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 [84])**.** *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 $[\![N]\!] \coloneqq \{1, \dots, N\}$, and the terminology of *coloring*. By a coloring of $[\![N]\!]$ with q colors, we just mean a partition of $[\![N]\!]$ into q sets $A_1, \dots, A_q$, 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 [25] 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. □

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 $\hat{\chi}$. □

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 $\Gamma := \{x^q : 0 \neq x \in \mathbb{Z}/p\}$ 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 [74] 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 [74])**.** *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 [39], 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 $r(k_1,\dots,k_q)$ 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 $k_1,\dots,k_q$, we have*

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

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

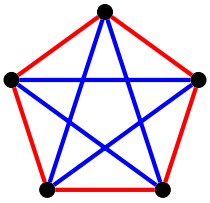
The Erdős–Szekeres bound, Theorem 2.1.4, gives us the upper bound $r(k) < 4^k$, which improves on Ramsey's earlier bound of $r(k) \leqslant k!$. To understand how good this bound is, we would like to obtain some *lower bounds* on $r(k)$.

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 $k = 2$, Proposition 2.2.1 implies that $r(2) > 4$, and it is not hard to show that in fact $r(2) > 5$, 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 [38])**.** *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 [65]). *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.

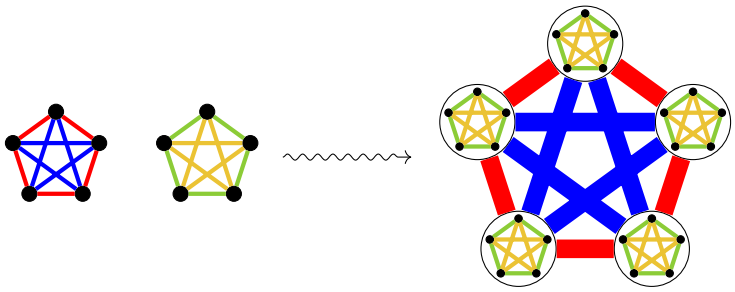
Proposition 2.2.5 (Abbott [1]). *For all positive integers k, q_1, q_2, we have*

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

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. □

[3] An earlier version of these notes, as well as many papers on the topic, attribute this result to Lefmann [64], and note that the trick goes back at least to work of Chung [13]. 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 [15, 81, 88], 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 [18], followed shortly thereafter by improvements of myself [97] and Sawin [83]. The current state of the art, due to Sawin, is the following result.

Theorem 2.3.1 (Sawin [83])**.** *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 [11] obtained the first improvement to the exponential constant in Theorem 2.1.4.

Theorem 2.3.3 (Campos–Griffiths–Morris–Sahasrabudhe [11])**.** *$r(k) < 3.9999^k$ 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 [67] 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 [28] 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 [89] using a different (and simpler) probabilistic technique, but the logarithmic gap remained for a long time until Kim [55] 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 [8, 40, 85] 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 [89] (with polylogarithmic improvements due to Bohman–Keevash [7]) 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 [67].

Theorem 2.3.4 (Mattheus–Verstraëte [67])**.** *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, 16, 68], 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 [18, 53, 68, 83, 97].

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. □

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 $f_1, \dots, f_r : [\![n]\!] \to V(G)$. 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

$$\binom{n}{k}\left(\frac{kM}{N}\right)^{rk} \leqslant \left(\frac{pk^rM^r}{N^{r-1}}\right)^k. \tag{3.1}$$

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*

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

Proof. We set $p = N^{r-1}/(2k^rM^r)$, 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 [53], 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 [53] 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 [18, 97] and explicit in [83], 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*

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

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

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

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, 68]). In their work improving the lower bound on $r(k;q)$, Conlon and Ferber [18] used an ingeneous 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 [83], 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

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

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 [97], 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 [83] 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 [83])**.** *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 [68]). 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 [49]), although we will follow a somewhat more streamlined proof due to Shearer [85]. 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 [85], is the content of the following lemma.

Lemma 4.1.3 (Shearer [85])**.** *Define*

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

extended continuously to $f(0) \coloneqq 1, f(1) \coloneqq \frac{1}{2}$. 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

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

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

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

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 [85])**.** *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 [66] 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 [86].

Theorem 4.1.5 (Shearer [86])**.** *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

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

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 [52], via an extremely long and heavily computer-assisted proof; more recently, the proof was fully formalized in a

proof assistant [51]. Even more recently, Viazovska [96] 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 [79] 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 [12], who improved Rogers' bound by a factor of $\Omega(\log d)$.

Theorem 4.2.1 (Campos–Jenssen–Michelen–Sahasrabudhe [12])**.**

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

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 [12])**.** *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 [57], although the proof technique goes back to work of Kleitman and Winston [56], 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 [82].

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 Kletiman–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 $w_{i+1}, \dots, w_k \in C_{i+1}$ such that $\{v_1, \dots, v_i, w_{i+1}, \dots, w_k\}$ forms an independent set.

We claim that we can run this process only up to step r, that is, once we select $v_1, \dots, v_r$, 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 $v_1, \dots, v_i$. At that point, as the candidate set has shrunk to size at most R, we have at most $\binom{R}{k-i}$ choices for $w_{i+1}, \dots, w_k$. 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. □

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 [16, 17], but is not quite the same as his, and the analysis also uses ideas from [23, 67]. Several alternative constructions are presented in [6].

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. □

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. □

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 [67, 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 $|X| \geqslant R \coloneqq 200q^2 \ln q$, we have that*

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

where $\beta \coloneqq 1/(10q)$.

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 [54, 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

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

From the proof of Lemma 4.3.4, we see that $S \coloneqq \sum_{\ell,p} L_{\ell,p}^2$ 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. □

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

Theorem 4.3.7. *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 Lemmas 4.3.3 and 4.3.5, we have the existence of a graph G_q on $n \coloneqq q^3$ vertices with the properties that (a) G_q is triangle-free, and (b), G_q satisfies the conditions of Lemma 4.3.1 with $R = 200q^2\ln q$ and $\beta = 1/(10q)$. Let $r \coloneqq 10q\ln q$, 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 of order k is at most

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

We may therefore apply Lemma 3.1.1 with $N = n = q^3$ and $M = 50q/\ln q$, and conclude that

$$r(3,k) \geqslant \frac{N}{4M} = \frac{q^3}{200q/\ln q} = \frac{q^2 \ln q}{200} \geqslant 10^{-6} \frac{k^2}{(\ln k)^3}. \qquad \square$$

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 $\ell_1, \dots, \ell_4$ 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 [72], which is why O'Nan configurations are so named.

Lemma 4.3.8. *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 [67, 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 4.3.8, it is not hard to prove that H_q is K_4-free with probability 1, just as in Lemma 4.3.3. The final ingredient we need, analogously to Lemma 4.3.5, is the following statement.

Lemma 4.3.9. *With positive probability, H_q has the following property. For every $X \subseteq L$ with $|X| \geqslant R := 10^7 q^2$, we have that*

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

where $\beta := 1/(300q)$.

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 4.3.9 (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 [67, Theorem 3]. A somewhat more general (and somewhat simpler) result, with a logarithmic loss in the value of R, is proved in [23, 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 4.3.9 into Lemma 4.3.1 to bound the number of independent sets of order k there are in H_q, and then plug that result into Lemma 3.1.1. Doing this yields the theorem of Mattheus and Verstraëte [67].

Theorem 4.3.10 (Mattheus–Verstraëte [67])**.** *We have*

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

for an absolute constant $C > 0$.

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 $H_1, \dots, H_q$, their *Ramsey number* $r(H_1, \dots, H_q)$ 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 $r(H_1, \dots, H_q) \leqslant r(H_1', \dots, H_q')$, 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 [31]; 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 [60], 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 [60])**.** *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

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

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. □

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 $v_1, \dots, v_n$ 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 [10].

Conjecture 5.4.5 (Burr–Erdős [10])**.** *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 [14], which established the Burr–Erdős conjecture under the stronger assumption that H has bounded maximum degree.

Theorem 5.4.6 (Chvatál–Rödl–Szemerédi–Trotter [14])**.** *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 [93]. 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 [47, 48], although it goes back in some form at least to work of Erdős and Hajnal [34]. 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 $v_1, \dots, v_n$. 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 $B_1, \dots, B_{\Delta+1}$.

Since H has maximum degree at most Δ, it is $(\Delta + 1)$-colorable, i.e. it can be partitioned into $\Delta + 1$ independent sets $C_1, \dots, C_{\Delta+1}$. 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 [47, 48])**.** *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 [48])**.** *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 $V_1, \dots, V_A$, 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 [20] 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 [44], and a much less in-depth, but hopefully gentler, introduction can be found in [98]. 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 [63], who completely resolved Conjecture 5.4.5.

Theorem 5.4.9 (Lee [63])**.** *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 [59] and Kostochka–Sudakov [58].

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 $u_1, \dots, u_t$ 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 $u_1, \dots, u_t$ 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

$$|T| \geqslant \frac{N}{4^t} = \frac{n2^{5\Delta\sqrt{\log n}}}{2^{2\sqrt{\log n}}} \geqslant n = |A|.$$

Let us arbitrarily embed A into T; that is, if we write $A = \{a_1, \dots, a_n\}$, then we pick arbitrary $x_1, \dots, x_n \in T$, 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 $b_1, \dots, b_n$. We embed them one by one, in this order. For a given vertex b_i, we know that it has at most Δ neighbors $a_{i_1}, \dots, a_{i_\Delta}$ in A. By construction, the set $\{x_{i_1}, \dots, x_{i_\Delta}\} \subseteq T$ is friendly, and thus these vertices have at least $2n$ common red neighbors. Fewer than $2n$ of these have been used in embedding $a_1, \dots, a_n, b_1, \dots, b_{i-1}$, 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. □

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 $u_1, \dots, u_t$ and considering their common neighborhood; the details are left for the homework.

Using the same ideas, but with one extra trick, Kostochka and Sudakov [58] were able to prove that if H is a d-degenerate bipartite graph on n vertices, then $r(H) \leqslant n2^{C(\log n)^{2/3}}$,

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 [59] and Kostochka–Sudakov [58], there were a series of improvements on the dependent random choice technique, culminating in Lee's theorem [63], 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 [22].

Theorem 5.4.12 (Conlon–Fox–Sudakov [22])**.** *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 [21] conjectured that in fact, such an upper bound holds in general.

Conjecture 5.4.13 (Conlon–Fox–Sudakov [21])**.** *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 [43, 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 [10] 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 [94] 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 [91, 92], 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 [14] 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. [99, 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 [14], although in some form it goes back at least to earlier work of Ruzsa–Szemerédi [80].

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 [99, 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 [93], 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 [19].

Lemma 6.1.4 (Quasirandom subset lemma; Conlon–Fox [19])**.** *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, [99, 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 [99, 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 [19] relied on an auxiliary result, the *cylinder regularity lemma* of Duke, Lefmann, and Rödl [26], 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 $d(A, B) \coloneqq e(A, B)/(|A||B|)$ 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. □

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 42, 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 [27].

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 [33], and states that induced-H-free graphs contain very large cliques or independent sets.

Conjecture 6.3.1 (Erdős–Hajnal [33])**.** *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 [9], which gives the best known result for general H. Their result improves on that of Erdős and Hajnal [34], who proved the same bound without the $\log \log N$ term.

Theorem 6.3.2 (Bucić–Nguyen–Scott–Seymour [9])**.** *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 [77], 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 [77]). *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 $G_1, \dots, G_q$ 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 $G_1 \cup \dots \cup G_q$ 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. [99, 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. □

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 [32], and was rapidly answered in the

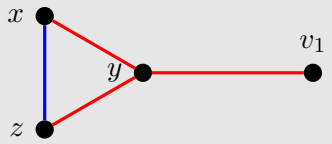
negative independently by Cherlin (unpublished), Graham [46], and van Lint (unpublished). The following slick construction is due independently to Galluccio–Simonovits–Simonyi [45] and to Szabó [90], and generalizes Graham's original argument. Given two graphs G_1, G_2, their *join*, denoted $G_1 * G_2$, 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 [45], Szabó [90])**.** *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 $K_6 \not\subseteq G$.*

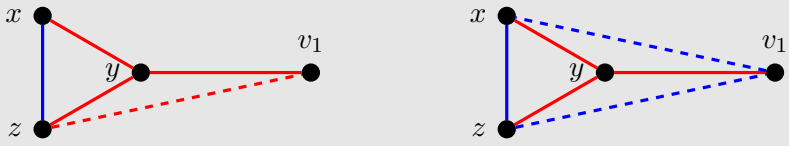
Proof. Let the vertices of G be $x, y, z, v_1, \dots, v_\ell$, where x, y, z form a triangle, $v_1, \dots, v_\ell$ form a cycle C_ℓ, and all edges between $\{x, y, z\}$ and $\{v_1, \dots, v_\ell\}$ are present. Note that if $K_6 \subseteq G$, then at least three of the vertices of this K_6 must come from $v_1, \dots, v_\ell$ (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.

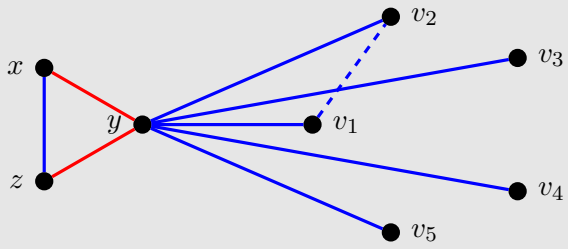
Now consider the edges between $\{x, y, z\}$ and v_1. First suppose yv_1 is red.



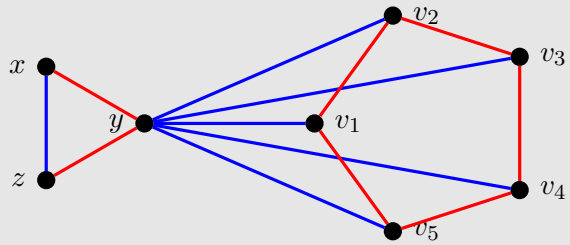
If xv_1 or zv_1 is red, then we close a red triangle xyv_1 or zyv_1, so we may assume that both these edges are blue. But that also creates a blue triangle, xzv_1.



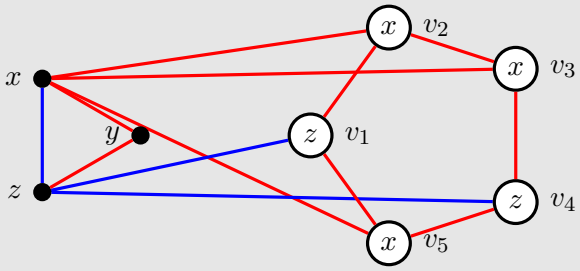
So we may assume that yv_1 is blue. By the same logic, yv_i is blue for all $i \in [\![\ell]\!]$. Note that if any of the edges v_iv_{i+1} in the cycle is blue, then we create a blue triangle yv_iv_{i+1}.



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. □

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 [32]). 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 [41].

Theorem 7.1.3 (Folkman [41])**.** *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 [69], 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 [69])**.** *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 [62].

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 [78]. 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 [78])**.** *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 $p \coloneqq CN^{-\frac{2}{k+1}}$, 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. □

Before ending this section, let us briefly discuss one further recent breakthrough on the structure of restricted Ramsey graphs, due to Reiher and Rödl [75].

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 $F \subseteq G$.

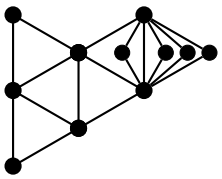
In this language, we can restate Proposition 7.1.6 as saying that K_{k+3} is not Ramsey obligatory for K_k, and Theorem 7.1.3 (or more preicsely its multicolor extension, which follows from Theorem 7.1.4) states that K_{k+1} is not Ramsey obligatory for H. On the other hand, we can easily show that certain 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 [75] is that this sufficient condition is also necessary.

Theorem 7.1.10 (Reiher–Rödl [75])**.** *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 [24], Erdős–Hajnal–Pósa [35], and Rödl [76].

Theorem 7.2.2 (Deuber [24], Erdős–Hajnal–Pósa [35], Rödl [76])**.** *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 [70] 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 [71], and which is more closely inspired by [42].

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 $G_1, \dots, G_q$. By Lemma 6.2.1, there exists $Q \subseteq V(G)$ such that $G_1[Q], \dots, G_q[Q]$ 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. □

Let us define the *induced Ramsey number* $r_{\text{ind}}(H; q)$ 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 $r_{\text{ind}}(H; 2) \leqslant 2^{2^{2^{10n}}}$ 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 [29] made the following conjecture.

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

Note that this result, if true, is best possible, since $r_{\text{ind}}(K_n; q) = r(K_n; q)$, 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 [20] proved that $r_{\text{ind}}(H; 2) \leqslant 2^{Cn \log n}$ 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 $r_{\text{ind}}(H; q) \leqslant 2^{C_q n^2}$ 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 [36], and was brought to prominence by an important result of Erdős and Szemerédi [37], 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 $r_o(k;q) > N$. □

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 [37].

Theorem 8.1.3 (Erdős–Szemerédi [37])**.** *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 [37])**.** *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

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

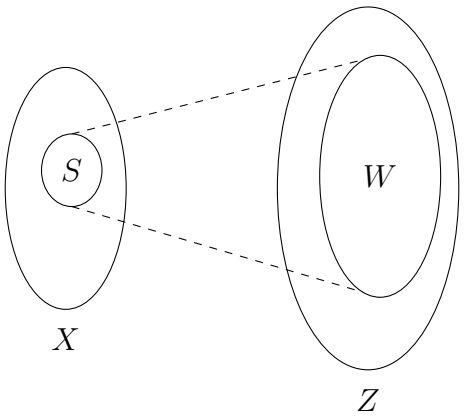
Let $Z = Y \setminus B$, and note that $|Z| \geqslant \frac{N}{4} - \frac{N}{5} = \frac{N}{20}$. Moreover, each vertex in Z has fewer than s neighbors in X. However, the number of subsets of X of size less than s is

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

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 [37], 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 [33]. 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 [33])**.** *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 $\sigma \leqslant d(S) \leqslant 1 - \sigma$. □

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 [73].

Theorem 8.2.5 (Prömel–Rödl [73])**.** *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 [42, 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 [61], which in particular proved an old conjecture of Erdős–McKay [30].

Theorem 8.2.6 (Kwan–Sah–Sauermann–Sawhney [61])**.** *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 [87]).

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 [95]. Recall that a *k-term arithmetic progression* (or *k-AP* for short) is a sequence of k integers, of the form

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

Theorem 9.1.1 (van der Waerden [95])**.** *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 [50] 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 $x^{(1)}, \dots, x^{(k)} \in [\![k]\!]^d$ with the following property. For each coordinate $i \in [\![d]\!]$, either $x_i^{(1)} = \cdots = x_i^{(k)}$, or

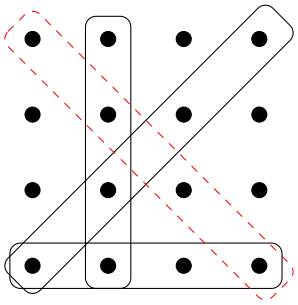
$$x_i^{(1)} = 1, x_i^{(2)} = 2, \dots, x_i^{(k)} = k. \tag{9.1}$$

Additionally, we require $x^{(1)}, \dots, x^{(k)}$ 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 $\rho \in \{1, 2, \dots, k, *\}^d$ 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

$$x^{(1)} = \rho(1), \dots, x^{(k)} = \rho(k).$$

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 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 [50])**.** *For every $k, q \geqslant 2$, 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.

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] A. Bishnoi, Finite geometry and Ramsey theory, 2021. Lecture notes available online at `https://anuragbishnoi.files.wordpress.com/2021/01/minicourse.pdf`.

[7] T. Bohman and P. Keevash, The early evolution of the H-free process, *Invent. Math.* **181** (2010), 291–336.

[8] 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.

[9] 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.

[10] 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.

[11] M. Campos, S. Griffiths, R. Morris, and J. Sahasrabudhe, An exponential improvement for diagonal Ramsey, 2023. Preprint available at arXiv:2303.09521.

[12] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe, A new lower bound for sphere packing, 2023. Preprint available at arXiv:2312.10026.

[13] F. R. K. Chung, On the Ramsey numbers $N(3, 3, \ldots, 3; 2)$, *Discrete Math.* **5** (1973), 317–321.

[14] 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.

[15] D. Conlon, A new upper bound for diagonal Ramsey numbers, *Ann. of Math. (2)* **170** (2009), 941–960.

[16] D. Conlon, A sequence of triangle-free pseudorandom graphs, *Combin. Probab. Comput.* **26** (2017), 195–200.

[17] D. Conlon, Extremal numbers of cycles revisited, *Amer. Math. Monthly* **128** (2021), 464–466.

[18] D. Conlon and A. Ferber, Lower bounds for multicolor Ramsey numbers, *Adv. Math.* **378** (2021), Paper No. 107528, 5pp.

[19] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, *Geom. Funct. Anal.* **22** (2012), 1191–1256.

[20] D. Conlon, J. Fox, and B. Sudakov, On two problems in graph Ramsey theory, *Combinatorica* **32** (2012), 513–535.

[21] 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.

[22] D. Conlon, J. Fox, and B. Sudakov, Short proofs of some extremal results II, *J. Combin. Theory Ser. B* **121** (2016), 173–196.

[23] D. Conlon, S. Mattheus, D. Mubayi, and J. Verstraëte, Ramsey numbers and the Zarankiewicz problem, 2023. Preprint available at arXiv:2307.08694.

[24] 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.

[25] L. E. Dickson, On the congruence $x^n + y^n + z^n \equiv 0 \pmod{p}$, *J. Reine Angew. Math.* **135** (1909), 134–141.

[26] 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.

[27] N. Eaton, Ramsey numbers for sparse graphs, *Discrete Math.* **185** (1998), 63–75.

[28] P. Erdős, Graph theory and probability. II, *Canadian J. Math.* **13** (1961), 346–352.

[29] 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).

[30] P. Erdős, Some of my favourite problems in various branches of combinatorics, *Matematiche (Catania)* **47** (1992), 231–240.

[31] 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.

[32] P. Erdős and A. Hajnal, On decomposition of graphs, *Acta Math. Acad. Sci. Hungar.* **18** (1967), 359–377.

[33] 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.

[34] P. Erdős and A. Hajnal, Ramsey-type theorems, *Discrete Appl. Math.* **25** (1989), 37–52. Combinatorics and complexity (Chicago, IL, 1987).

[35] 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.

[36] P. Erdős, A. Hajnal, and R. Rado, Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 93–196.

[37] P. Erdős and A. Szemerédi, On a Ramsey type theorem, *Period. Math. Hungar.* **2** (1972), 295–299.

[38] P. Erdös, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53** (1947), 292–294.

[39] P. Erdös and G. Szekeres, A combinatorial problem in geometry, *Compositio Math.* **2** (1935), 463–470.

[40] 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.

[41] J. Folkman, Graphs with monochromatic complete subgraphs in every edge coloring, *SIAM J. Appl. Math.* **18** (1970), 19–24.

[42] J. Fox and B. Sudakov, Induced Ramsey-type theorems, *Adv. Math.* **219** (2008), 1771–1800.

[43] J. Fox and B. Sudakov, Two remarks on the Burr-Erdős conjecture, *European J. Combin.* **30** (2009), 1630–1645.

[44] J. Fox and B. Sudakov, Dependent random choice, *Random Structures Algorithms* **38** (2011), 68–99.

[45] 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.

[46] R. L. Graham, On edgewise 2-colored graphs with monochromatic triangles and containing no complete hexagon, *J. Combinatorial Theory* **4** (1968), 300.

[47] R. L. Graham, V. Rödl, and A. Ruciński, On graphs with linear Ramsey numbers, *J. Graph Theory* **35** (2000), 176–192.

[48] R. L. Graham, V. Rödl, and A. Ruciński, On bipartite graphs with linear Ramsey numbers, *Combinatorica* **21** (2001), 199–209.

[49] J. E. Graver and J. Yackel, Some graph theoretic results associated with Ramsey's theorem, *J. Combinatorial Theory* **4** (1968), 125–175.

[50] A. W. Hales and R. I. Jewett, Regularity and positional games, *Trans. Amer. Math. Soc.* **106** (1963), 222–229.

[51] 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.

[52] T. C. Hales, A proof of the Kepler conjecture, *Ann. of Math. (2)* **162** (2005), 1065–1185.

[53] X. He and Y. Wigderson, Multicolor Ramsey numbers via pseudorandom graphs, *Electron. J. Combin.* **27** (2020), Paper No. 1.32, 8pp.

[54] S. Janson, T. Ł uczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[55] J. H. Kim, The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures Algorithms* **7** (1995), 173–207.

[56] D. J. Kleitman and K. J. Winston, On the number of graphs without 4-cycles, *Discrete Math.* **41** (1982), 167–172.

[57] 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.

[58] A. Kostochka and B. Sudakov, On Ramsey numbers of sparse graphs, *Combin. Probab. Comput.* **12** (2003), 627–641. Special issue on Ramsey theory.

[59] A. V. Kostochka and V. Rödl, On graphs with small Ramsey numbers, *J. Graph Theory* **37** (2001), 198–204.

[60] T. Kövari, V. T. Sós, and P. Turán, On a problem of K. Zarankiewicz, *Colloq. Math.* **3** (1954), 50–57.

[61] 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.

[62] I. Leader, Sparse Ramsey theory, 2003. Available online at `https://cranch.staff.shef.ac.uk/files/sparse-ramsey.pdf`.

[63] C. Lee, Ramsey numbers of degenerate graphs, *Ann. of Math. (2)* **185** (2017), 791–829.

[64] H. Lefmann, A note on Ramsey numbers, *Studia Sci. Math. Hungar.* **22** (1987), 445–446.

[65] X. Li, Two source extractors for asymptotically optimal entropy, and (many) more, 2023. Preprint available at arXiv:2303.06802.

[66] Y. Li and C. C. Rousseau, On book-complete graph Ramsey numbers, *J. Combin. Theory Ser. B* **68** (1996), 36–44.

[67] S. Mattheus and J. Verstraëte, The asymptotics of $r(4,t)$, *Ann. of Math. (2)* (2024), to appear. Preprint available at arXiv:2306.04007.

[68] D. Mubayi and J. Verstraëte, A note on pseudorandom Ramsey graphs, 2019. Preprint available at arXiv:1909.01461.

[69] J. Nešetřil and V. Rödl, The Ramsey property for graphs with forbidden complete subgraphs, *J. Combinatorial Theory Ser. B* **20** (1976), 243–249.

[70] 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.

[71] 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.

[72] M. E. O'Nan, Automorphisms of unitary block designs, *J. Algebra* **20** (1972), 495–511.

[73] H. J. Prömel and V. Rödl, Non-Ramsey graphs are $c \log n$-universal, *J. Combin. Theory Ser. A* **88** (1999), 379–384.

[74] F. P. Ramsey, On a problem of formal logic, *Proc. London Math. Soc. (2)* **30** (1929), 264–286.

[75] C. Reiher and V. Rödl, The girth Ramsey theorem, 2023. Preprint available at arXiv:2308.15589.

[76] V. Rödl, *The dimension of a graph and generalized Ramsey theorems*, Master's thesis, Charles University, 1973.

[77] V. Rödl, On universality of graphs with uniformly distributed edges, *Discrete Math.* **59** (1986), 125–134.

[78] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8** (1995), 917–942.

[79] C. A. Rogers, Existence theorems in the geometry of numbers, *Ann. of Math. (2)* **48** (1947), 994–1002.

[80] 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.

[81] A. Sah, Diagonal Ramsey via effective quasirandomness, *Duke Math. J.* **172** (2023), 545–567.

[82] W. Samotij, Counting independent sets in graphs, *European J. Combin.* **48** (2015), 5–18.

[83] 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.

[84] I. Schur, Über die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *Jahresber. Dtsch. Math.-Ver.* **25** (1917), 114–117.

[85] J. B. Shearer, A note on the independence number of triangle-free graphs, *Discrete Math.* **46** (1983), 83–87.

[86] J. B. Shearer, On the independence number of sparse graphs, *Random Structures Algorithms* **7** (1995), 269–271.

[87] 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`.

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

[89] J. Spencer, Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20** (1977/78), 69–76.

[90] T. Szabó, On nearly regular co-critical graphs, *Discrete Math.* **160** (1996), 279–281.

[91] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89–104.

[92] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

[93] 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.

[94] K. Tikhomirov, A remark on the Ramsey number of the hypercube, 2022. Preprint available at arXiv:2208.14568.

[95] B. L. Van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wiskunde* **15** (1927), 212–216.

[96] M. S. Viazovska, The sphere packing problem in dimension 8, *Ann. of Math. (2)* **185** (2017), 991–1015.

[97] Y. Wigderson, An improved lower bound on multicolor Ramsey numbers, *Proc. Amer. Math. Soc.* **149** (2021), 2371–2374.

[98] Y. Wigderson, Lecture notes on dependent random choice, 2023. Available online at `https://n.ethz.ch/~ywigderson/math/static/DependentRandomChoice.pdf`.

[99] Y. Zhao, *Graph theory and additive combinatorics—exploring structure and randomness*, Cambridge University Press, Cambridge, 2023.