Ramsey theory—lecture notes

Yuval Wigderson

Spring semester 2024
Last updated: February 21, 2024

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 [38] 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 [38])**.** *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 [38])**.** *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 [15] 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 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 [33] 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 [33])**.** *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 [18], 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:

$$r(k,\ell) \leqslant r(k-1,\ell) + r(k,\ell-1). \tag{2.1.1}$$

To prove (2.1.1), 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 (2.1.1).

We can now complete the induction. By (2.1.1) 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

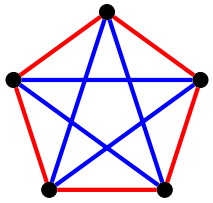
The Erdős–Szekeres bound, Theorem 2.1.4, gives us the upper bound $r(k) < 4^k$, which improves on Ramsey's earlier bound of $r(k) \leqslant k!$. To understand how good this bound is, we would like to obtain some *lower bounds* on $r(k)$.

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 [17])**.** *For any $k \geqslant 2$, we have $r(k) > 2^{k/2}$.*

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

$$\binom{N}{k} 2^{1-\binom{k}{2}} < \frac{N^k}{k!} \cdot 2^{1-\frac{k^2-k}{2}} < \frac{N^k}{2^{1+\frac{k}{2}}} \cdot 2^{1+\frac{k}{2}-\frac{k^2}{2}} = \left(N \cdot 2^{-\frac{k}{2}}\right)^k = 1, \tag{2.2.1}$$

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 [29]). *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 Lefmann [28] noticed a simple trick (which goes back at least to work of Chung [9]) that does much better.

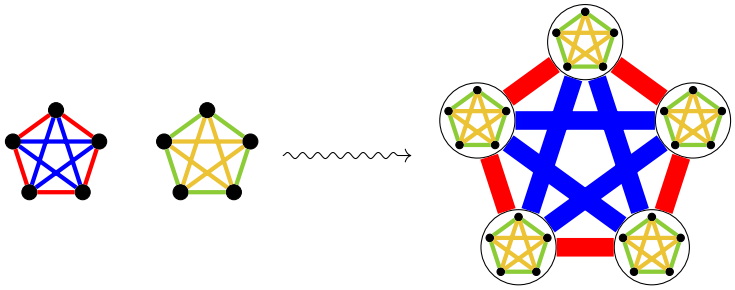
Proposition 2.2.5 (Chung [9], Lefmann [28]). *For all positive integers k, q_1, q_2, we have*

$$r(k; q_1 + q_2) - 1 \geqslant (r(k;q_1) - 1)(r(k;q_2) - 1). \tag{2.2.2}$$

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 (2.2.2).

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

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 [10, 35, 40], 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 [13], followed shortly thereafter by improvements of myself [43] and Sawin [37]. The current state of the art, due to Sawin, is the following result.

Theorem 2.3.1 (Sawin [37])**.** *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 [7] obtained the first improvement to the exponential constant in Theorem 2.1.4.

Theorem 2.3.3 (Campos–Griffiths–Morris–Sahasrabudhe [7])**.** *$r(k) < 3.9999^k$ 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 [30] 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

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

where the subscript indicates that the implicit constant may depend on 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 [1], 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 [16] used a very sophisticated and intricate probabilistic argument to obtain a nearly matching lower bound,

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

Erdős's result was re-proved by Spencer [41] using a different (and simpler) probabilistic technique, but the logarithmic gap remained for a long time until Kim [25] finally managed to prove that the upper bound is correct, that is

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

More recent improvements to the lower and upper bounds [6, 19, 39] 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.

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 [41] (with polylogarithmic improvements due to Bohman–Keevash [5]) 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 [30].

Theorem 2.3.4 (Mattheus–Verstraëte [30])**.** *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 [2, 11, 31], 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 Lefmann [28] 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 [2], and were crystallized in a series of works [13, 23, 31, 37, 43].

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 a 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 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 not, we color $\{u,v\}$ by color $r+1$. Then each of the first r colors is K_s-free, by the above.

[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.

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

$$\binom{n}{k}\left(\frac{kM}{N}\right)^{rk} \leqslant \left(\frac{pk^rM^r}{N^{r-1}}\right)^k. \tag{3.1.1}$$

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 equation (3.1.1) 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 [23], 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 gotten 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 [2], 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 [23] 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 [13, 43] and explicit in [37], 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

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

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

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

In many settings, such as the lower bounds on $r(3,k)$ and $r(4,k)$ that we will discuss shortly, finding such graphs is quite difficult, and is basically the crux of the argument (see also [2, 31]). In their work improving the lower bound on $r(k;q)$, Conlon and Ferber [13] 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 [37], 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 (2.2.1). 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

$$r(k;q) \geqslant \frac{2^{k/2}N^{q-2}}{4k^{q-2}M^{q-2}} = \frac{(2^{(q-1)/2})^k}{4(2k)^{q-2}(2^{(q-2)/8})^k} = \left(2^{\frac{q-1}{2} - \frac{q-2}{8}}\right)^{k-o(k)} = \left(2^{\frac{3}{4}q - \frac{1}{4}}\right)^{k-o(k)}, \quad (3.3.1)$$

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 (3.3.1) 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 (3.3.1) was proved in [43], 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 [37] 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 [37])**.** *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 [31]). 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

$$r(3,k) \leqslant \binom{k+1}{2} \leqslant k^2. \tag{4.1.1}$$

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 [1] (improving on earlier work of Graver–Yackel [20]), although we will follow a somewhat more streamlined proof due to Shearer [39]. 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 (4.1.1) 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 (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 [39], is the content of the following lemma.

Lemma 4.1.3 (Shearer [39])**.** *Define*

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

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

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

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

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

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

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

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

$$\begin{aligned}(n - \deg(v) - 1)(\widehat{d} - d) &= (nd - 2\deg(v)r(v)) - (nd - d\deg(v) - d) \\ &= d\deg(v) + d - 2\deg(v)r(v). \end{aligned} \tag{4.1.2}$$

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

$$\begin{aligned}\alpha(G) &\geqslant 1 + (n - \deg(v) - 1)f(\widehat{d}) \\ &\geqslant 1 + (n - \deg(v) - 1)f(d) + (n - \deg(v) - 1)(\widehat{d} - d)f'(d) \\ &= 1 + (n - \deg(v) - 1)f(d) + (d\deg(v) + d - 2\deg(v)r(v))f'(d), \end{aligned} \tag{4.1.3}$$

where the final equality uses (4.1.2).

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

$$\frac{1}{n}\sum_{v \in V} (\deg(v) + 1)f(d) = (d+1)f(d). \tag{4.1.4}$$

For the first quantity, we first compute that

$$\sum_{v \in V(G)} \deg(v)r(v) = \sum_{v \in V(G)} \sum_{u \sim v} \deg(u) = \sum_{u \in V(G)} \sum_{v \sim u} \deg(u) = \sum_{u \in V(G)} \deg(u)^2 \geqslant nd^2, \tag{4.1.5}$$

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

$$\frac{1}{n}\sum_{v\in V(G)}(d\deg(v)+d-2\deg(v)r(v))f'(d)\leqslant(d^2+d-2d^2)f'(d)=(d-d^2)f'(d), \quad (4.1.6)$$

where we *reversed* the direction of the inequality from (4.1.5) 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, (4.1.4) and (4.1.6) 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 (4.1.3) 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 [1], Shearer [39])**.** *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 (2.1.1), 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.

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 [22], via an extremely long and heavily computer-assisted proof; more recently, the proof was fully formalized in a proof assistant [21]. Even more recently, Viazovska [42] 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 [34] to $\theta(d) = \Omega(d \log d)$. 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 [8], who improved Rogers' bound by a factor of $\Omega(\log d)$.

Theorem 4.2.1 (Campos–Jenssen–Michelen–Sahasrabudhe [8])**.**

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

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 [8])**.** *Let G be an n-vertex graph with maximum degree Δ. Suppose that every pair of distinct vertices in G have 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 $(1-o(1))\ln\Delta$ 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 [27], although the proof technique goes back to work of Kleitman and Winston [26], and the same idea was first applied in this setting by Alon and Rödl [2]. An excellent survey on this topic, including a detailed proof of Lemma 4.3.1, was written by Samotij [36].

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 [11, 12], but is not quite the same as his, and the analysis also uses ideas from [14, 30]. Several alternative constructions are presented in [4].

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 [30, 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 [24, Corollary 2.27 and Remark 2.28] or [3, 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

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

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 a map $\sigma : \mathbb{F}_{q^2}^2 \times \mathbb{F}_{q^2}^2 \to \mathbb{F}_{q^2}$ by

$$\sigma((a_1,a_2),(b_1,b_2)) = a_1 b_1^q + a_2 b_2^q.$$

We define

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

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 [32], 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 [30, 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 $|X| \geqslant R \coloneqq 10^7 q^2$, we have that*

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

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

Unfortunately, for technical reasons arising from the fact that $|L| \approx q^4$ is much larger than $|P| \approx q^3$, it seems impossible to prove Lemma 4.3.9 (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 [30, Theorem 3]. A somewhat more general (and somewhat simpler) result, with a logarithmic loss in the value of R, is proved in [14, 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 [30].

Theorem 4.3.10 (Mattheus–Verstraëte [30])**.** *We have*

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

for an absolute constant $C > 0$.

Bibliography

[1] M. Ajtai, J. Komlós, and E. Szemerédi, A note on Ramsey numbers, *J. Combin. Theory Ser. A* **29** (1980), 354–360.

[2] N. Alon and V. Rödl, Sharp bounds for some multicolor Ramsey numbers, *Combinatorica* **25** (2005), 125–141.

[3] 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.

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

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

[6] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[19] 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.

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

[21] 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.

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

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

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

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

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

[27] 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.

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

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

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

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

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

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

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

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

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

[37] 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.

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

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

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

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

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

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