Chapter 5

The symmetric Ramsey-problem

In this chapter we return to the symmetric Ramsey problem we studied in Section 1.3. We defined the symmetric Ramsey number $R(k,k)$ as the smallest integer n such that any graph on n vertices contains a clique or an independent set of size k. In order to deal with lower bounds, it will be convenient to call a graph *k-Ramsey* if it contains neither a clique nor an independent set of size k. The symmetric Ramsey number can be expressed with this notation as a maximum:

$$R(k,k) = 1 + \max\{n : \text{ there is a } k\text{-Ramsey graph on } n \text{ vertices}\}.$$

In particular, the existence of a k-Ramsey graph on n vertices implies the lower bound $n < R(k,k)$.

Recall the exponential upper and lower bounds

$$\sqrt{2}^k \leq R(k,k) \leq 4^k, \tag{5.1}$$

that we presented in Section 1.3. Erdős' proof of the lower bound established the *existence* of a k-Ramsey graph on $\sqrt{2}^k$ vertices, but did not give any pointers as to *how* to construct such a graph explicitly, not even on significantly fewer vertices. The best *constructive lower bound* for decades was provided by the Turán graph $T_{(k-1)^2,k-1}$ on $(k-1)^2 \ll \sqrt{2}^k$ vertices.

Knowing the existence of a special combinatorial structure, like a large Ramsey graph, is of course great, but in theoretical computer science, in particular in questions related to various models of complexity, it is desirable having the structure in our hand, constructed explicitly. Furthermore, considering that the largest known k-Ramsey graph is the uniform random graph, one might also hope that explicitly constructed Ramsey graphs would be relevant to imitating randomness efficiently—another key issue in theoretical computer science. It is doubtful that Erdős had any of these motivations in mind when, in the late 60s, he had the good taste to ask for a "direct construction" of k-Ramsey graphs on exponentially many vertices. Still, as it is the case with many of his beautiful questions, this one also hit something important right on the head. Something, the importance of which turned out only later.

In the next four sections we will see how far we can get by imitating randomness using deterministic constructions. In fact we will only be able to show the beginnings,

the tip of the iceberg. The more recent exciting breakthroughs of theoretical computer science in this direction [?, ?] are unfortunately out of the scope of our lecture notes.

In the last section of the chapter we will discuss a completely different approach to constructing k-Ramsey graphs, which highlights the influence of this question of Erdős had on extremal hypergraph theory.

5.1 Initial Constructions

5.1.1 Paley graphs

In order to constructively imitate the success of the random graph $G(n, \frac{1}{2})$ as a Ramsey graph, one might try to think of graphs in which the neighborhood of each vertex is a random-like set of roughly $n/2$ vertices. To this end the realm of Cayley graphs is natural to explore, since finding just one random-like set $S \subset G$ of generators in some group G already guarantees that all neighborhoods in the Cayley graph $C(G, S)$ are random-like.

A notable candidate for such a "quasi-random" set is the set $S = QR(p) = \{z^2 : z \in \mathbb{F}_p^*\}$ of quadratic residues in the additive group $\langle \mathbb{F}_p, + \rangle$ of the p-element field. For $p > 2$ this is a set of $\frac{p-1}{2}$ elements, which is defined via the multiplicative structure of the field ("multiply each element with itself"). The intuition is that within a finite field $\mathbb{F}_p$ of prime order the additive and the multiplicative structures should thoroughly mix each other up. Indeed, one is a cyclic group of order p, the other is a cyclic group of order $p-1$. The latter is relatively prime to the former, which anyway has only trivial subgroups, so it is hard to imagine a too large subset that is "orderly" for both structures.

For an arbitrary prime power q, the Cayley graph $C(\langle \mathbb{F}_q, + \rangle, QR(q))$ is called the *Paley graph*[1] P_q of order q. By definition, the vertex set $V(P_q) = \mathbb{F}_q$ is the q-element field, and vertices x and y are adjacent if $x-y$ is a quadratic residue. In order to have this adjacency relation symmetric, like in any Cayley graph, we must assume that $S = -S$. Here this is equivalent to $-1 \in QR(q)$, which happens if and only if $q \equiv 1 \pmod 4$. Then indeed, $x - y$ is a quadratic residue if and only if $y - x$ is.

On the one hand, we show in the next exercise that Paley graphs are beautifully symmetric.

Exercise 5.1 *(i) Show that P_q is isomorphic to its complement. In particular $\alpha(P_q) = \omega(P_q)$.*

(ii) Show that P_q is edge-transitive; that is, for every pair of edges $xy, uv \in E(P_q)$, there is an isomorphism of P_q mapping x to u and y to v.

(iii) Make a conjecture about the automorphism group of P_q.

[1] These graphs appeared first in a paper by Sachs at beginning of 60s and Erdős and Rényi for prime powers a couple of years later. The name stuck only later, due to Paley's use of the quadratic character for constructing Hadamard matrices in 1933.

In part (*iii*) of the exercise you have hopefully succeeded to show that P_q has many automorphisms. In comparison, the random graph $G(q, \frac{1}{2})$, with probability tending to 1 (as $q \to \infty$), has not got a single non-trivial automorphism.

On the other hand, the next exercise shows that Paley graphs do possess some "random-like" properties. Namely in $G\left(q, \frac{1}{2}\right)$ any two vertices have roughly $v(P_q)/4$ common neighbors (with probability tending to 1), which turns out to be the case in P_q as well.

Exercise 5.2 *Let $x, y \in V(P_q)$. Show that*

$$|N(x) \cap N(y)| = \begin{cases} \frac{q-1}{4} - 1 & \text{if } x \text{ and } y \text{ are adjacent} \\ \frac{q-1}{4} & \text{otherwise} \end{cases}$$

It is a common belief that Paley graphs of prime order have much stronger quasi-random properties than just the pairwise independence highlighted in Exercise 5.2, so far as that they are conjectured to provide relatively good Ramsey graphs. Unfortunately to show we are only able that P_q is $(\sqrt{q}+1)$-Ramsey, which is no better than what holds for Turán's construction.

Exercise 5.3 *Show that $\omega(P_q) \leq \sqrt{q}$ for any prime power $q \equiv 1 \pmod 4$. Conclude that P_q is $(\sqrt{q}+1)$-Ramsey.*

The next exercise shows that this upper bound cannot be improved for general prime powers.

Exercise 5.4 *Show that if q is an odd square, then $\omega(P_q) = \sqrt{q}$.*

For prime orders however, the situation looks much more encouraging. In Figure 5.1.1 we plotted the results of computer calculations of Shearer and Exoo about the clique number (and hence independence number) of Paley graphs of prime order, up to 10000. The figure seems to indicate that for primes p the clique number is much smaller than the $\sqrt{p}$ upper bound we were able to prove in general. In fact the growth rate looks more like polylogarithmic. For example $\lfloor\sqrt{9533}\rfloor = 97$, while the clique number of P_{9533} is only 18. Despite this convincing numerical evidence, proving $\omega(P_p) \leq p^{1/2-\epsilon}$ merely for some tiny constant $\epsilon > 0$ would already be a major number theoretic advance.

Number theorists for long studied the related classical function n_p denoting the smallest quadratic non-residue modulo p. Since the numbers $0, 1, 2, \ldots, n_p - 1$ form a clique in the Paley graph P_p, one always has $n_p \leq \omega(P_p)$. The best known upper bound on n_p is $c_\varepsilon p^{1/4\sqrt{e}+\varepsilon}$, so polynomial in p.

Assuming the generalized Riemann hypothesis (GRH), it was proven by Montgomery that there is some constant $c > 0$, such that the first $c \log p \log\log p$ integers form a clique in the Paley graph P_p for infinitely many primes p. This means that Paley graphs are *not* anticipated to provide constructive k-Ramsey graphs on $p = 2^{C\frac{k}{\log k}}$ vertices for every p. So in this regard Paley graphs do differ from the truly random graph $G(p, \frac{1}{2})$, which is k-Ramsey on exponentially many vertices in k.

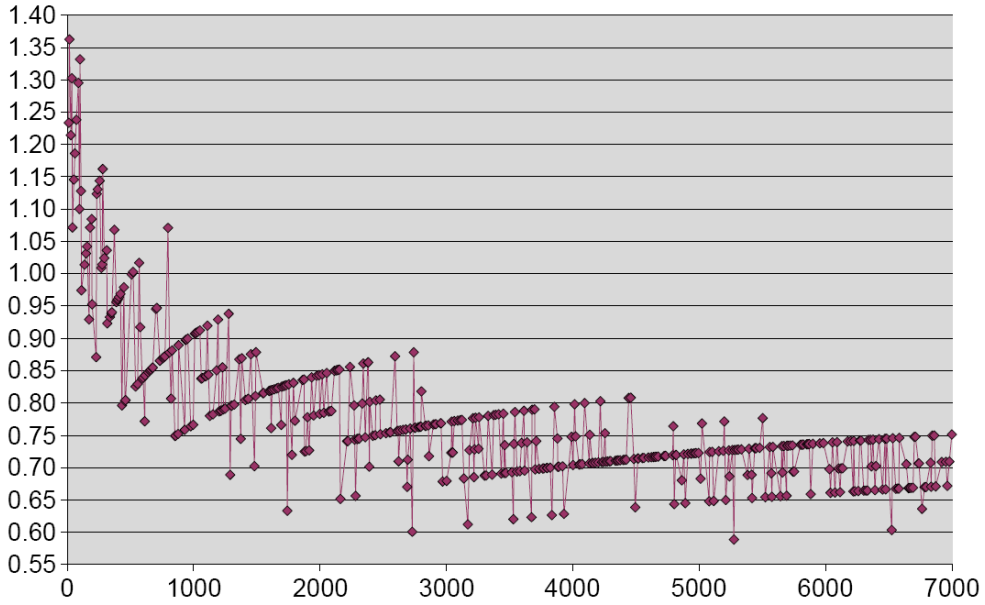

Figure 5.1: The quotient $\frac{\log n(P_p)}{\omega(P_p)}$ in the Payley-graph P_p for the primes $p \leq 7000$

From the other side, it is also true modulo the GRH that the first $\log^2 p$ integers do *not* form a clique. This might make it plausible that there is no $(\log^2 p)$-clique *anywhere* in the Paley graph, and hence they *are* a family of k-Ramsey graphs on $p = 2^{\sqrt{k}}$ vertices. Bollobás [?] speculates that Paley graphs might be k-Ramsey graph on $2^{c\frac{k}{\log k}}$ vertices, and for special primes p they might even have exponentially many vertices in terms of their clique number. It is worthwhile to compare here the exponent $c\frac{k}{\log k}$ with the best known probabilistic construction where the main term in the exponent is $\frac{k}{2}$, and the constructive lower bound of the Turán graph where the exponent is only $\log_2(k-1)^2 \approx 2\log k$.

Paley graphs provide the the tight lower bound for $R(3,3)$ and $R(4,4)$, which represent all the known exact values of symmetric Ramsey numbers.

Exercise 5.5 *Verify that P_5 is the cycle C_5 of length five. Prove that P_{17} does not contain a clique or independent set of order 4 and conclude that $R(4,4) = 18$.*

The largest Paley graph which is 5-Ramsey is P_{37}, but there exist larger 5-Ramsey graphs. The largest known has 42 vertices, proving $R(5,5) \geq 43$. An upper bound $R(5,5) \leq 48$ was recently announced. Both bounds invoke significant computer assistance.

For all other small constants, $6 \leq k \leq 20$, the best known lower bound on $R(k,k)$ is also provided by a Paley graph or the following doubling trick of Shearer [?] applied to a Paley graph.

Exercise 5.6 *Given a graph G on n vertices, we define a new graph $D = D(G)$ on $2n+2$ vertices as follows. We take two disjoint copies of G on vertex sets $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$, such that $x_i \to y_i$ is an isomorphism. Between X and Y, we add all edges $x_i y_j$ such that $x_i x_j \notin E(G)$. In particular $x_i y_i \in E(D)$ for every i. Finally, we add two new adjacent vertices u_X and u_Y, connecting u_X to every vertex in X and connecting u_Y to every vertex in Y.*
Prove that $\alpha(D(P_q)) = \alpha(P_q) + 1$ and $\omega(D(P_q)) = \omega(P_q) + 1$.

5.1.2 Beating the Turán Construction

When Erdős [?] was asking to recover his exponential lower bound by a "direct construction", he generously also admitted that he cannot even construct $(\epsilon\sqrt{n})$-Ramsey graphs on n vertices. This innocent comment of the great Master provided motivation for many, and shortly after two entirely different approaches emerged. Both of them superseded the Turán graph in that they provided constructions of graphs with no clique and independent set of size k, but having super-quadratically many, k^{2+c} vertices. And later on both of these approaches lead to even better constructions and towards important connections to theoretical computer science and extremal set theory, respectively.

The Abbott product

First, Abbott generalized the block-structure idea of the Turán construction. The Turán graph $T_{(k-1)^2,k-1}$ can be thought of as a $(k-1)$-clique, the vertices of which were blown up into independent sets of size $k-1$. Abbott realized that it is better to do a more symmetric blow-up, that is, if both what we blow up and what we place on the blown-up vertices have both small clique number and small independence number. For example, one can beat the 5-Ramsey graph $T_{16,4}$ on 16 vertices, by blowing up the vertices of a C_5 (instead of a K_4) into five vertices and placing a copy of C_5 (instead of a $\overline{K}_4$) onto each. This is also a 5-Ramsey graph, but has 25 vertices.

Beating just one concrete Turán graph is of course not what Erdős meant. But we can carry on by blowing up the above C_5-blow-up with C_5 again and again, and thus obtain an infinite sequence of constructions. To formalize, given two graphs G and H we define define their Abbott-product $G \otimes H$ by

$$\begin{aligned} V(G \otimes H) &= V(G) \times V(H), \text{ and} \\ E(G \otimes H) &= \{(g_1,h_1)(g_2,h_2) : g_1g_2 \in E(G) \text{ or } g_1 = g_2 \text{ and } h_1h_2 \in E(H)\}. \end{aligned}$$

Informally speaking, one can imagine taking $v(G)$ disjoint copies of the graph H and then include all edges between two such copies if the vertices of G corresponding to the copies are adjacent in G. The Turán graph $T_{(k-1)^2,k-1}$ is just $K_{k-1} \otimes \overline{K}_{k-1}$. One can easily check (please do!) the following properties.

Exercise 5.7 *(i) For any two graphs G and H we have $v(G \otimes H) = v(G) \cdot v(H)$, $\alpha(G \otimes H) = \alpha(G) \cdot \alpha(H)$, and $\omega(G \otimes H) = \omega(G) \cdot \omega(H)$.*

(ii) Prove that the Abbott product is associative, that is, $(G_1 \otimes G_2) \otimes G_3 \cong G_1 \otimes (G_2 \otimes G_3)$.

By the exercise, the repeated blow-up $C_5^{\otimes k}$ of the 3-Ramsey graph C_5 has clique and independence number 2^ℓ, so it represents a k-Ramsey graph construction on $5^\ell = (k-1)^{\log_2 5} \gg k^{2.32}$ vertices.

Set families with intersection restrictions

The second idea to break through the quadratic constructive lower bound of the Turán graph appeared in the same year, 1972, as the Abbott product.[2] It is due to Zsigmond Nagy, who defined a graph G_{Nagy} on the vertex set $V(G_{\text{Nagy}}) = \binom{[z]}{3}$ of triples of a z-element set. Two triples A and B are adjacent in G_{Nagy} if they intersect in exactly one element.

Exercise 5.8 *Prove that* $\alpha(G_{\text{Nagy}}) \leq z$ *and* $\omega(G_{\text{Nagy}}) \leq z$. *Conclude that* G_{Nagy} *provides an infinite sequence of construction of* k*-Ramsey graphs on* $\Omega(k^3)$ *vertices.*

5.2 What sort of explicit?

Already Abbott noted that if instead of the 3-Ramsey graph C_5, we take the powers of the Paley graph P_{17}, which is a 4-Ramsey graph on the largest possible number of vertices, we improve the construction from the previous section. Indeed, $\omega(P_{17}^{\otimes\ell}) = \alpha(P_{17}^{\otimes\ell}) = 3^\ell$, so $P_{17}^{\otimes\ell}$ is a k-Ramsey graph on $17^\ell = (k-1)^{\log_3 17} \gg k^{2.57}$ vertices.

Of course if we took the powers of P_{9533} instead, about which Exoo's computer calculated that its clique and independence number is 18, then we obtained a construction of a k-Ramsey graph on $9533^\ell = (k-1)^{\log_{18} 9533} \gg k^{3.17}$ vertices. This is now already better than the construction of Nagy from Exercise 5.8.

But the power of computer stops here. How can we construct k-Ramsey graphs on k^4 vertices? Or even larger powers of k? From the above, it is clear that we could immediately improve the construction, were we able to get our hands on just one "starter graph" G_0 which is c_0-Ramsey on c_0^m vertices for some $m > \log_{18} 9533$. The Abbott powers $G_0^{\otimes\ell}$ of such a graph have clique and independence number at most $(c_0-1)^\ell$, so they provide us with a construction k-Ramsey graphs on $v(G_0)^\ell = c_0^{m\ell} > k^m$ vertices for arbitrarily large k.

How should we get a hold of a c_0-Ramsey graph for *some* c_0 with, say, c_0^{10} vertices? Well, thanks to Erdős we know that k-Ramsey graphs *do exist* if the number of vertices is not more than $\sqrt{2}^k$. At one point the function $\sqrt{2}^k$ certainly takes over k^{10}, so let c_0 be the smallest integer such that $\sqrt{2}^{c_0} \geq c_0^{10}$. If we check the graphs on c_0^{10} vertices, one of them certainly will be c_0-Ramsey. How long would this take? Nothing ... only

[2] ... and *American Pie* by Don McLean.

constant time ... Never mind that $c_0 = 144$, so you might have to calculate the clique number and independence number of possibly $2^{\binom{144^{10}}{2}}$ graphs on 144^{10} vertices.

This is certainly a *method* that, for any exponent m, constructs an infinite sequence of k-Ramsey graphs on k^m vertices. But is this now something we want to call "explicit construction"? We can definitely agree that $C_5^{\otimes \ell}$ and $P_{17}^{\otimes \ell}$ are explicit constructions. Even $P_{9533}^{\otimes \ell}$ is one, the fact that a computer had to check the clique number and independence number of a graph on 9533 vertices does not seem relevant to that.

And once we accept the use of a computer to aid our construction, it would be hard to argue why it should matter what exactly the computer is allowed to calculate during this constant amount of time and why it should be relevant whether it actually performed that calculation already. It seems sufficient to just *know* that after the computer *did* check those $2^{\binom{144^{10}}{2}}$ graphs, it would surely hand us our appropriate starter graph, and we can proceed with our construction of k-Ramsey graphs on k^{10} vertices for arbitrary k. The disturbing fact, that these computer calculations would last longer than the age of our universe, does not feel like should play a role in whether we want to call this an "explicit construction".

It is time to stop procrastinating, face the inconvenient imprecision lurking in the background, and decide already what exactly we wanna call an explicit construction. The discussion above made it clear that our definition must deal with the possibility of computer calculations and it should definitely include an apprpriate limiting of them. After all, we certainly do NOT want to call explicit construction for example the computer checking of all graphs on n vertices, and then outputting the best Ramsey graph there is, which we know is $(2\log n)$-Ramsey. This procedure of course might require checking $2^{\binom{n}{2}}$, i.e. superexponentially many graphs, just to produce one on n vertices. It is reasonable to expect that for an explicit construction of a graph on n vertices we should be able to produce the n^2 entries of the adjacency matrix much faster, say in time polynomial in n, the customary computer scientific measure of "fast".

Definition: A family of graphs $\mathcal{G} = \{G_n : n \in S\}$, where $S \subseteq \mathbb{N}$ is an infinite subset and $v(G_n) = n$, is called *(efficiently) explicit* if there is an algorithm that on input $n \in S$ runs in time $\text{poly}(n)$ and outputs the adjacency matrix of G_n.

This is the definition adopted by the theoretical computer science community, who went on to great length extending and strengthening the randomness required of k-Ramsey graphs. They use their constructs for efficient generation of psedorandom bits, with the eventual main goal of efficiently derandomizing every randomized algorithm in mind. In the next exercise we show that the definition also caters to our wish to be able to call all of the above constructions explicit.

Exercise 5.9 *(a) Verify that the Turán graph provides an explicit family of* $(1 + o(1))\sqrt{n}$*-Ramsey graphs on* n *vertices for every* $n \in \mathbb{N}$.

(b) Verify that the graphs G_{Nagy} *provide an explicit family of* $O\left(\sqrt[3]{n}\right)$*-Ramsey graphs on* n *vertices for every* $n \in \mathbb{N}$.

(c) Verify that for every $m \in \mathbb{N}$, the above Abbott procedure can be used to create an explicit family of $O\left(\sqrt[m]{n}\right)$-Ramsey graphs on n vertices for every $n \in \mathbb{N}$.

The Abbott product argument, at least in its current form, won't give us anything *superpolynomial*, i.e., no infinite sequence of k-Ramsey graph on $k^{f(k)}$ vertices, where $f(k) \to \infty$. Even if we took a starter r_0-Ramsey-graph on $r_0^{\log\log\log r_0}$ vertices (which we certainly could), taking its Abbott-powers takes away the superpolynomial relation between the order n and the clique number ω: already for the square of the starter we would not have $n \geq \omega^{\log\log\log\omega}$.

How could we construct something truly superpolynomial? So far we have not used the full power of our definition of explicit construction for the finding of the starter graph. We spent only constant amount of time to find G_0, when we could have spent $poly(n)$. Given an integer n, we will fix integers $v = v(n)$ and $\ell = \ell(n)$ such that $v^{\ell-1} < n \leq v^\ell$. We plan to find a starter graph on v vertices and raise it to the ℓth Abbott power to obtain a graph on n vertices. We know that among the $2^{\binom{v}{2}}$graphs on v vertices at least one of them is $(2\log v)$-Ramsey. We find such a G_0 by checking for each of the $2^{\binom{v}{2}}$ graphs on v vertices, whether any of the $\binom{v}{2\log v}$ subsets of size $2\log v$ forms a clique or an independent set. This takes at most

$$2^{\binom{v}{2}} \cdot \binom{v}{2\log v} \cdot \binom{2\log v}{2} \leq 2^{\frac{v^2}{2}-\frac{v}{2}} \cdot v^{2\log v} \leq 2^{\frac{v^2}{2}} \leq n,$$

i.e. polynomially many steps, provided $v \leq \sqrt{2\log n}$ and v is large enough. We set $v = \left\lceil\sqrt{\log n}\right\rceil = v(G_0)$ and consequently choose $\ell = \left\lceil\frac{\log n}{\log v}\right\rceil$. Then $v^{\ell-1} < v(G_0^{\otimes\ell}) = n \leq v^\ell$ and

$$\alpha(G_0^{\otimes\ell}), \omega\left(G_0^{\otimes\ell}\right) \leq (2\log v)^\ell \leq (\log\log n)^{\frac{\log n}{\frac{1}{2}\log\log n}+1} = n^{(2+o(1))\frac{\log\log\log n}{\log\log n}}.$$

This is an explicit construction of a k-Ramsey graph on superpolynomially many, $k^{\Omega\left(\frac{\log\log k}{\log\log\log k}\right)}$, vertices. The exponent is quite small, but does go to infinity with k.

In the age of computer, speed, and efficiency, our definition of explicit construction sounds completely satisfactory: there is an explicit deterministic algorithm, telling us in a short time which vertices are adjacent and which vertices are not adjacent. What else would one want to call explicit?

Erdős did not specify in his question what he wants to mean by explicit construction. While his subconscious understanding of the concept might have included the limiting of computation power one way or another, he nevertheless did not accept the superpolynomial Abbott construction as explicit.[3] Intuitively it is clear what Erdős would not like: in its first phase the construction uses brute force in finding the object it knows to exist. It is not using any kind of clever idea or structure to pull out the "hay from the haystack", but rather goes in there, picks up every single object from a haystack, studies it carefully, and finds a hay eventually. This feels like cheating, even though there is a

[3]He declined to pay the "bounty prize" he set for the problem on the merit of the Abbott construction.

significant difference between doing this search in the whole haystack (of all graphs on n vertices) or just in a much smaller haystack and then using the found small hay to produce the promised "pseudo-hay" (with still more features of a needle) for arbitrary sized haystacks.

Let us take another crack at a computer scientific definition, more to Erdős' liking. An evident drawback of the Abbott construction is that the adjacency status of any particular pair of vertices cannot be decided before finding the whole starter graph first, which already takes time $poly(n)$. In our earlier constructions (Turán, Paley, Nagy) the adjacency relation was defined directly and could be decided independently for each pair of vertices. Describing two vertices in an n-vertex graph takes only $\lceil 2\log n \rceil$ bits, so ideally one would wish to decide whether they are adjacent within time $poly(\log n)$. This motivates the following definition.

Definition: A family of graphs $\mathcal{G} = \{G_n : n \in S\}$, where $S \subseteq \mathbb{N}$ and $v(G_n) = n$, is called *strongly explicit* if there is an algorithm that on inputs $u, v \in V(G_n)$ runs in time $\text{poly}(\log n)$ and decides whether $uv \in E(G_n)$.

As expected, all our constructions are strongly explicit.

Exercise 5.10 *Prove that the Turán and Nagy construction is strongly explicit.*

Exercise 5.11 *Suppose that addition and multiplication in $\mathbb{F}_q$ can be carried out in constant time. Show that P_q is strongly explicit; that is, there is some constant C such that one can decide if two given vertices u and v are adjacent in $O(\log^C(q))$ time.*
How long does it take to construct the adjacency matrix of the entire graph P_q? Provide an algorithm whose running time is best possible up to a constant factor.

Unfortunately the above superpolynomial Abbott product construction can also easily be modified to be strongly explicit: simply reduce the time spent on finding the starter graph from n to $\log n$. Carrying out the calculation like this shows that the constructed graph becomes a strongly explicit k-Ramsey graph on $k^{\Omega\left(\frac{\log\log\log k}{\log\log\log\log k}\right)}$ vertices. This is smaller than the one above, but still superpolynomial, and the fundamental flaw Erdős saw in the brute force search for the starter remains.

At the time Erdős posed his question about a "constructive" lower bound for the Ramsey function, the computer scientific notion of "efficient" was just about to be developed. So even though one could suspect that his idea of explicit would be closer to the definition of strongly explicit,he could not honestly care much about efficient computability. And there is an even more important philosophical distinction. The motivation behind Erdős' question must have rather been the desire to encounter disorder in a *concrete* large structure and thus have a much better understanding of its nature. Erdős would not care about polynomial computability of the adjacency relation, because a computer can calculate many things where the human mind is not able to see anything. For Erdős the Paley graph was an explicit construction *not* because one could compute the adjacency relation in polylogarithmic time, but because the definition of an edge is

through a concrete, mathematically described structure (a finite field and its operations), the disorder of which would also be understandable (should number theorists finally be able to prove it ...).

To draw attention to this underlying issue, we feel obliged to introduce the following definition, not quite up to the usual stuck-up standards of mathematics, but leaving open a somewhat subjective interpretation of the concept of explicit

Definition: We call a sequence of graphs *morally explicit* if Erdős would have called it explicit.

In morally explicit constructions the adjacency relation should be given directly, using only objects/structures/concepts that are precisely known already at the time of describing the construction (and not only *known to exist*, so to be found by some hypothetical search in some space, however small that space might be.) Even if we acknowledged the reality of a computer performing the actual construction, in a morally explicit construction we do care for what the computer's time is used for, and a search is disallowed.

The notion of morally explicite does *not include* any reference to fast computability, though morally explicit constructions tend to be efficiently explicit and even strongly explicit. But the Paley graph for example is morally explicit, *not* because we can decide the edge relation fast, but because the definition of an edge is direct and involves only known structures and concepts (operations within finite fields).

Abbott powers of Paley-graphs are also morally explicit, since both the product operation and the graphs we take the power of are defined directly. In particular we consider $P_{9533}^{\otimes \ell}$ a morally explicit construction, despite the fact that humans are not able to check its clique and independence number. This just makes the *proof* of its properties computer-assisted, but that does not influence the fact the constructed graph is given completely directly using the well-known adjacecy matrix of P_{9533}.

A morally explicit superpolynomial Ramsey graph.

In the following we describe yet another superpolynomial variant using the Abbott product due to Naor [?], that we would *not* be able to call *not* morally explicit. The construction will be strongly explicit and all part of the definition of an edge is known from the beginning and no part of the decision is based on search.

So far we have only made use of the existence of an incredidibly good Ramsey graph and just picked any one to be our starter. Now we will utilize that *most* of the graphs on n vertices are so, namely that the random graph $G(n, 1/2)$ has clique number and independence number that are both at most $2\log_2 n$ with extremely high probability. Hence it looks to be a good idea to take the Abbott-product of *all* graphs on n vertices, since *most* of them have very small clique- and independence-numbers.

To be more precise, let $K \subseteq [n]$ be a subset of k vertices. One can easily calculate

the probability that K induces a clique (or an independent set) in $G(n, 1/2)$:

$$\mathbb{P}[K \text{ is a clique}] = \frac{1}{2^{\binom{k}{2}}} \tag{5.2}$$

Then by the union bound

$$\mathbb{P}[\exists \text{ clique of order } k] \leq \binom{n}{k} 2^{-\binom{k}{2}} < \left(\frac{ne}{k2^{(k-1)/2}}\right)^k, \tag{5.3}$$

which is at most $\left(\frac{e}{\sqrt{2}\log_2 n}\right)^{2\log_2 n} < \frac{1}{\log_2 n}$ for $k = 2\log_2 n$. In other words, less than $\epsilon := \frac{1}{\log_2 n}$-fraction of the family $\mathcal{G} = \mathcal{G}_n$ of all labeled graphs on n vertices contains a clique of order $2\log_2 n$.

Let G be the Abbott-product of all graphs from $\mathcal{G}$. Then

$$v(G) = n^{|\mathcal{G}|},$$

where $|\mathcal{G}| = 2^{\binom{n}{2}}$. By the above one can estimate the clique number of G using (5.7) as follows:

$$\omega(G) \leq (2\log_2 n)^{(1-\epsilon)|\mathcal{G}|} n^{\epsilon|\mathcal{G}|} < (2\log_2 n)^{|\mathcal{G}|} n^{\epsilon|\mathcal{G}|} = (4\log_2 n)^{|\mathcal{G}|}. \tag{5.4}$$

Remark. Here we estimated the clique number of $(1-\epsilon)|\mathcal{G}|$ graphs by $2\log_2 n$, but were seemingly pretty generous when we estimated the clique number of the rest of the graphs by n. Nevertheless our estimate is precise enough for our purposes, since random graph theory tells us that almost all graphs *do* have clique number at least $\log_2 n$, so $\omega(G) > (\log_2 n)^{(1-o(1))|\mathcal{G}|}$.

Since the independence number can be estimated analogously by (5.7), G is an infinite sequence of k-Ramsey graphs with

$$n = k^{\Omega\left(\frac{\log\log\log k}{\log\log\log\log k}\right)}$$

vertices. (Check the calculation!) Moreover G is clearly an explicit construction, which can be constructed in polynomial time. G is finally a construction of superpolynomial order: the exponent $\frac{\log\log\log k}{\log\log\log\log k}$ does tend to infinity, though pretty slowly, it reaches the value 3 for example only when $k > 2^{256}$.

Looking at the formulas for the number $n^{|\mathcal{G}|}$ of vertices and the clique number $(4\log n)^{|\mathcal{G}|}$ of G, it is apparent what ruins an initially paradisiac clique-number/vertex-set-size relationship from logarithmic $\omega = 4\log n$ to barely subpolynomial $N^{\Omega\left(\frac{\log\log\log N}{\log\log\log N}\right)}$: the huge size of the family $\mathcal{G}$. The more times we take the factors n and $4\log n$ in the

formulas, the more the Abbott-product loses from the excellent Ramsey properties that most of the members of $\mathcal{G}$ have.

Doing a bit of (our customary) wishful thinking: wouldn't it be wonderful if there was a much smaller family $\mathcal{D}$ instead of $\mathcal{G}$, so that we could still perform the same calculations as in (5.4)? And hence we obtained a clique number bound of $(4\log n)^{|\mathcal{D}|}$ on a vertex set of size $n^{|\mathcal{D}|}$ with a much smaller exponent $|\mathcal{D}|$? Of course, *there is* such a family of size 1, we know this by the probabilistic method, even with clique and independence number bound of $2\log n$. But the point here now is that we want this family to be *explicitely constructible*.

For (5.4) we only needed that less than $\epsilon := \frac{1}{\log_2 n}$-fraction of the graphs in the family contains a clique of order $2\log_2 n$. Let us take a closer look at what property was really necessary in order to be able to infer this for the family $\mathcal{G}$. Well, first we calculated in (5.2) the probability that a particular set of k vertices forms a clique in a uniformly random member of $\mathcal{G}$ and then just used the union bound. And why did we know that the probability that a particular k-set forms a clique is equal to $2^{-\binom{k}{2}}$? Because when we selected a member of $\mathcal{G}$ uniformly at random, the appearance of each edge was mutually independent from the appearence of all other edges. The crucial observation now is that in order to guarantee (5.2) for $k = 2\log_2 n$, we do *not* need the full power of independence of all the coordinates in the family $\mathcal{G}$. The independence of any set of $2\log_2^2 n > \binom{k}{2}$ coordinates is enough. It turns out that we will be able to ensure this constructively with much fewer than $2^{\binom{n}{2}}$ graphs.

Remark. In fact the independence of all $\binom{\binom{n}{2}}{\binom{k}{2}} \sim \left(\frac{n}{k}\right)^{k^2(1+o(1))}$ subsets of $\binom{k}{2}$ coordinates is not necessary—it would be enough to have it for the $\binom{n}{k} \sim \left(\frac{n}{k}\right)^{k(1+o(1))}$ subsets corresponding to k-cliques. But we do not know how to pinpoint only those.

5.3 Limiting the randomness

5.3.1 d-wise independent sample spaces

Let us make the previous wishful thinking more precise. Our general plan is to construct a (hopefully) small multiset S of 0/1-vectors of dimension N, such that any d subset of the coordinates are mutually independent. Then, choosing $N = \binom{n}{2}$ and $d = 2\log^2 n$, and interpreting the constructed 0/1-vectors as graphs on n labeled vertices, we obtain the desired family $\mathcal{D}$, for which (5.2) is valid when $k = 2\log n$.

Definition: A *sample space* is a probability space $(S, \mathbb{P})$, where $S = S(M)$ is the multiset of the column vectors of a 0/1-matrix M, and $\mathbb{P}$ is the uniform distribution on S.

Remark: 1. If N is the length of the vectors in a sample space S, then we will often refer to S as an *N-dimensional sample space*. This does *not* in any way refer to the dimension of the linear space these vectors span over $\mathbb{F}_2$. 2. The matrix M can of course have identical columns and hence the sample space might contain vectors with multi-

plicity larger than one. For ease of notation, we chose to avoid formally describing S as a multiset of vectors. For an N-dimensional sample space $S \subseteq \{0,1\}^N$ and vector $a \in \{0,1\}^N$, the quantity $\mathbb{P}[s = a]$ represents the probability that a uniformly chosen element s of S is equal to the vector a. In other words it is equal to the number of times a appears as a column vector of M, divided by the number $|S|$ of columns.
3. The concept of a sample space is a convenient way to approximate an *arbitrary* probability space: first one approximates the probabilities of the vectors with rational numbers having a common denominator D and then takes a sample space of cardinality D where each vector has multiplicity of the numerator of its probability.

Definition: A sample space $S \subseteq \{0,1\}^N$ is called *independent* if for any vector $a \in \{0,1\}^N$, we have

$$\mathbb{P}[s = a] = \frac{1}{2^N}.$$

Remark: Independent sample spaces are in fact pretty boring: all vectors in $\{0,1\}^N$ must have the same multiplicity. We denote by G_N the $(N \times 2^N)$-matrix whose columns are the different 0/1-vectors of length N (in some arbitrary fixed order). The sample space $S(G_N) = \{0,1\}^N$ is the unique independent sample space with vectors of multiplicity one.

The problem with the full independence of independent sample spaces is their exponential size. The following is the key definition of this section.

Definition: For an $(N \times m)$-matrix M and a subset $J \subseteq [N]$ of the rows we denote by $M|_J$ the matrix obtained from M by deleting all rows indexed by elements from $[N] \setminus J$. Let $d \leq N$ be positive integers. The N-dimensional sample space $S = S(M) \subseteq \{0,1\}^N$ is called *d-wise independent* if for any subset $J \subseteq [N]$, $|J| = d$, the d-dimensional sample space $S|_J := S(M|_J) \subseteq \{0,1\}^d$, called the *projection of S on J*, is independent.

Note that being N-wise independent in dimension N is equivalent to being independent.

Remark: For a sample space $S \subseteq \{0,1\}^N$ and a subset $J = \{i_1 < \ldots < i_d\} \subseteq [N]$ of the coordinates, we denote by $s|_J := (s_{i_1}, \ldots, s_{i_d})$ the element of $S|_J$ corresponding to the element $s \in S$. Spelling out the definition with this notation: the restriction of S on J is the d-dimensional sample space

$$S|_J = \{s|_J : s \in S\} \subseteq \{0,1\}^d$$

of size $|S|$, and the sample space $S = S(M) \subseteq \{0,1\}^N$ is d-wise independent if and only if for every $J \subseteq [N]$, $|J| = d$, and vector $a \in \{0,1\}^d$, we have that

$$\mathbb{P}[s|_J = a] = \frac{1}{2^d}.$$

Exercise 5.12 *Let $S = S(M) \subseteq \{0,1\}^N$ be the sample space corresponding to the columns of a matrix M.*

(a) Show that if S is d-wise independent then it is d'-independent for every $d' \leq d$.

(b) Show that S is d-wise independent if and only if the row vectors of M, interpreted as $0/1$-valued random variables on S, are d-wise independent and uniformly distributed.

In the main theorem of this section we show that if one is content with just d-wise independence one can have a sample space of size significantly smaller than 2^N. Even more importantly, the solution is constructive.

Theorem 5.1 (Alon, Babai, Itai) *For every integer d and $N = 2^t$ with $t \in \mathbb{N}$, we can construct a d-wise independent sample space $S \subseteq \{0,1\}^N$ of size $|S| = 2N^{\lfloor \frac{d}{2} \rfloor}$.*

Proof. We will show the theorem for odd d. For even d, we just take the $(d+1)$-independent sample space of size $2N^{\frac{d}{2}}$ and use Exercise 5.12 to conclude its d-wise independence.

Independence requires that every vector occurs with the same multiplicity. Our main concern here is to ensure this *efficiently*. To this end we plan to use a linear map to generate the elements of the sample space, because in the image of a linear map every vector occurs as the image of vectors from the domain the same number of times.

Namely, applying a $(d \times m)$-matrix L to the elements of $\mathbb{F}_2^m$ produces a d-dimensional sample space $\{Lx : x \in \mathbb{F}_2^m\} \subseteq \mathbb{F}_2^d$ of size 2^m, which we denoted by $S(LG_m)$. For every $a \in \mathrm{Im}(L) \subseteq \mathbb{F}_2^d$ the inverse image $L^{-1}(a)$ is a coset of the kernel of the linear map L, and hence has the same size $|\mathbb{F}_2|^{m-rank(L)}$ as an $(m - rank(L))$-dimensional linear space.

Consequently the sample space $S(LG_m)$ is independent if and only if L is surjective, that is if the rows of L are linearly independent. In combination with Exercise 5.12, we can distill the following connection between probabilistic and linear independence.

Proposition 5.2 *Let L be a $(d \times m)$-matrix with $0/1$ entries. The sample space $S(LG_m)$ being independent is equivalent to each of the following.*

- *The rows of L, interpreted as vectors in $\mathbb{F}_2^m$, are linearly independent.*
- *The rows of LG_m, interpreted as $0/1$-valued random variables, are uniformly distributed and independent.*

Independence is just d-wise independence in dimension d. We can easily generalize the above characterization of d-wise independence to *arbitrary* dimension $N \geq d$. Let L be an $(N \times m)$-matrix. By definition, the N-dimensional sample space $S(LG_m) \subseteq \{0,1\}^N$ is d-wise independent if for every d-element subset $J \subseteq [N]$ of the rows the d-dimensional projection $S(LG_m)|_J = S((LG_m)|_J)$ is independent. The matrix $(LG_m)|_J$, obtained from LG_m by keeping the rows indexed by elements of J, is equal to the matrix $L|_J G_m$. By Proposition 5.2 the sample space $S(L|_J G_m)$ is independent if and only if the $(d \times m)$-matrix $L|_J$ is of rank d. So we have infered the following.

Corollary 5.3 *Let L be an $(N \times m)$-matrix. The sample space $S(LG_m) \subseteq \{0,1\}^N$ is d-wise independent if and only if any d rows of L are linearly independent over $\mathbb{F}_2$.*

How to obtain the magic matrix from Corollary 5.3 for the construction of our d-wise independent sample space? When we hear the condition that any d rows of a matrix should be linearly independent, it immediately rings the bell: "moment curve" (recall Wenger's construction of C_6- and C_{10}-free graphs with many edges from Section 3.4). We saw there that for every field $\mathbb{F}$ and every $d \leq |\mathbb{F}|$, any d distinct vectors from the set $M_d = \{(1, \alpha, \alpha^2, \ldots, \alpha^{d-1}) : \alpha \in \mathbb{F}\} \subseteq \mathbb{F}^d$ are linearly independent. This gives rise to an $(|\mathbb{F}| \times d)$-matrix L with the required property and we could choose $\mathbb{F}$ to be a however large finite field. Hence the sample space $S(LG_d)$ would be d-wise independent and of size 2^d, which is independent of the dimension N. Wow! At the same time this sounds very suspicious, too good to be true ...

Indeed, first of all we ignored that for a sample space we need 0/1-vectors and not coordinates from an arbitrary finite field $\mathbb{F}$. Secondly, the linear independence of the rows should be over $\mathbb{F}_2$ and not over $\mathbb{F}$. In order to fix this we need an encoding of the elements of the finite field as bit-vectors, which maintains the linear independence property. For example when we add the bit-vector of α^i and the bit-vector of β^i (over $\mathbb{F}_2$) the result should be the bit-vector of their sum (in $\mathbb{F}$).

This is how the field $\mathbb{F}_{2^t}$ comes into play. The elements of $\mathbb{F}_{2^t}$ have a canonical encoding with elements of $\mathbb{F}_2^t$, which is a linear space over $\mathbb{F}_2$, such that addition in the field $\mathbb{F}_{2^t}$ is just usual addition of vectors.[4]

Set $N = 2^t$. The dimensions of our matrix A will be $N \times (t(d-1)+1)$, where $d \leq N$ is an arbitrary integer. Let $\alpha_1, \ldots, \alpha_N$ be an arbitrary ordering of the elements of $\mathbb{F}_{2^t}$. We define the i^{th} row vector as the concatenation of an entry 1 and all the powers of the element α_i, up to the $(d-1)$th power. In fact the first coordinate 1 just represents the 0th power, which is the same for every α_i. More precisely, labelling the coordinates from 0 up to $t(d-1)$, the row vector r_i of A between coordinates $(j-1)t+1$ and jt is α_i^j (where the power is computed in $\mathbb{F}_{2^t}$ but the result is written as an element of $\mathbb{F}_2^t$).

Example. To continue our example of $t = 3$, let $N = 2^3 = 8$ and let, say, $d = 4$. The matrix we define will have dimension 8×10. The rows are labelled by the binary vectors

[4]The elements of $\mathbb{F}_{2^t}$ are polynomials of degree at most $t-1$ over $\mathbb{F}_2$, factored with a polynomial of degree t which is irreducible over $\mathbb{F}_2$. So once the irreducible polynomial is fixed, such a representation can be given as the coefficients of the terms of degree at most $t-1$.

Example. To give an example for a finite field, let $t = 3$. We fix the polynomial $f(x) = x^3 + x + 1$ of degree 3; one can check that f is irreducible over $\mathbb{F}_2$ by checking that neither of the two elements of $\mathbb{F}_2$ are roots. The elements of the field $\mathbb{F}_8$ are the polynomials $0, 1, x, x+1, x^2, x^2+1, x^2+x, x^2+x+1$. These elements can of course be denoted by 0/1 vectors of length 3, the coefficient of the monomials x^2, x, and 1 giving the three coordinates. This is completely meaningful when talking about *addition in* $\mathbb{F}_8$ as that is defined exactly as it would happen in the linear space $\mathbb{F}_2^3$. For multiplaction, however, we need the fixed polynomial $f(x)$. The product of two field elements is their usual product as polynomials modulo the equation $x^3 + x + 1 = 0$; that is, whenever we see a power larger than 2, we simplify by substituting $x^3 = -x - 1 = x + 1$. To take an example, consider
$(x^2 + x)(x + 1) = x^4 + 2x^2 + x + 1 = x^3 \cdot x + x + 1 = (x+1)x + x + 1 = x^2 + 2x + 1 = x^2 + 1$.

of length 3. Let us look at what is in the fifth row (labelled by the field element x^2+1). The first element is a 1. The next three are $1,0,1$, which are just the coordinates of x^2+1 when written in $\mathbb{F}_2^3$. For the next three entry we must calculate that $(x^2+1)^2 = x^2+x$ in the field $\mathbb{F}_8$ and for the last three we calculate that $(x^2+1)^3 = x+1$. Hence the fifth row is $1,1,0,1,1,1,0,0,1,1$.

Let us now take d arbitrary rows of the matrix A, for notational simplicity we denote them by $r_1,\dots,r_d$, defined by elements $\alpha_1,\dots,\alpha_d \in \mathbb{F}_{2^t}$. How could a linear combination $x_1r_1+\cdots+x_dr_d$ be the zero vector for some $x=(x_1,\dots,x_d)\in\mathbb{F}_2^d$? For that to happen first we would need $\sum_{i=1}^d x_i = 0$ to hold, because of the first column and then also that $\sum_{i=1}^d x_i\alpha_i^j = 0$ holds for every $j=1,\dots,d-1$, because of the columns from $(j-1)t+1$ to jt. Note that here we started to interpret these equations over $\mathbb{F}_{2^t}$, instead of just in $\mathbb{F}_2^t$.

Hence we have the following system of d equations in $\mathbb{F}_{2^t}$.

$$\begin{array}{ccccccc} x_1 & + & \cdots & + & x_d & = & 0 \\ x_1\alpha_1 & + & \cdots & + & x_d\alpha_d & = & 0 \\ x_1\alpha_1^2 & + & \cdots & + & x_d\alpha_d^2 & = & 0 \\ \vdots & & \vdots & & \vdots & & \\ x_1\alpha_1^{d-1} & + & \cdots & + & x_d\alpha_d^{d-1} & = & 0 \end{array} \tag{5.5}$$

The matrix of this system is the Vandermonde matrix, which is non-singular, since the α_i are distinct elements of $\mathbb{F}_{2^t}$. So the unique solution $x\in\mathbb{F}_{2^t}^d$ of the system is the 0-vector, and thus the d rows $r_1,\dots,r_d$ of A are linearly independent over $\mathbb{F}_2$.

Concluding, we constructed a $N\times(t(d-1)+1)$-matrix A with $0/1$ entries such that every d of its rows are linearly independent over $\mathbb{F}_2$. Corollary 5.3 then implies that the N-dimensional linear sample space $S(AG_{(d-1)t+1}) \subseteq \{0,1\}^N$of size $2^{(d-1)t+1} = 2N^{d-1}$ is d-wise independent.

This is roughly the square of the size we promised in the theorem. In order to improve, we must pinpoint what was wasted in the previous argument. The clear candidate is that even though we *do not care* whether there are coefficients $x_1,\dots,x_d$ satisfying (5.5) that are *not all* 0s or 1s, our argument still did show that there are none such. How can we make use of that the coefficients x_i of the linear combination of the rows are not just arbitrary elements from $\mathbb{F}_{2^t}$, but either 0 or 1?

What special about 0 and 1 is that squaring does not change them, so there is no point in raising them to higher powers. We can use this to show that the equation for the squares of the α_i in (5.5) is a consequence of the equation for the first powers. Indeed, simply squaring the first powers, we obtain

$$0 = (x_1\alpha_1+\cdots+x_d\alpha_d)^2 = x_1^2\alpha_1^2+\cdots+x_d^2\alpha_d^2 = x_1\alpha_1^2+\cdots+x_d\alpha_d^2.$$

In breaking up the parathesis of the square of the sum we used that in characteristic 2 the mixed terms fall out since they contain a factor 2. In the last equality we did use that $x_i = 0$ or 1.

The same squaring trick applies to the equation for the bth powers for arbitrary b. The mixed terms fall out as they have coefficient 2, and x_i^2 can be replaced with x_i because $x_i \in \mathbb{F}_2$ and thus we obtain the equation for the $(2b)$th powers:

$$0 = (x_1\alpha_1^b + \cdots + x_d\alpha_d^b)^2 = x_1^2\alpha_1^{2b} + \cdots + x_d^2\alpha_d^{2b} + \sum_{i<j} 2x_ix_j\alpha_i^b\alpha_j^b = x_1\alpha_1^{2b} + \cdots + \alpha_d^{2b}.$$

Hence the equation $0 = x_1\alpha_1^s + \cdots + \alpha_d^s$ for any even power $s = b \cdot 2^r \leq 2^t - 1$, where $r \geq 1$ and b is odd, can be obtained from the equation $0 = x_1\alpha_1^b + \cdots + \alpha_d^b$ by squaring it r times.

Motivated by this we construct a shorter matrix B using only the odd powers as follows. Let $N = 2^t$. The dimensions of our matrix B will be $N \times (t\ell+1)$, where $\ell = \frac{d-1}{2}$. Recall that $\alpha_1, \ldots, \alpha_N$ is an arbitrary ordering of the nonzero elements of $\mathbb{F}_{2^t}$. The i^{th} row vector is the concatanation of a 1 and all the odd powers of the element α_i up to $\alpha_i^{2\ell-1}$. More precisely, labeling the coordinates from 0 up to $t\ell$, for $j = 0, \ldots, \ell - 1$ the vector r_i between coordinates $jt+1$ and $(j+1)t$ is α_i^{2j+1} (where the power is computed in $\mathbb{F}_{2^t}$ but the result is written as an element of $\mathbb{F}_2^t$).

Let us take $d = 2\ell + 1$ rows $r_1, \ldots, r_d$ of the matrix, defined by elements $\alpha_1, \ldots, \alpha_d$. How could a linear combination $x_1r_1 + \cdots + x_dr_d$ be the zero vector for some $x \in \mathbb{F}_2^d$? For that we would need $\sum_{i=1}^d x_i = 0$, because of the first column and $\sum_{i=1}^d x_i\alpha_i^{2j+1} = 0$, because of the rows from $jt + 1$ to $(j + 1)t$. These are $\ell + 1$ equations and $2\ell + 1$ variables. We obtain however the remaining ℓ equations for the even powers by squaring, as described above, and end up with the the same equation system (5.5) and the same conclusion as above: there is only the trivial $x = 0$ solution. Consequently the d rows of the matrix B are linearly independet over $\mathbb{F}_2$.

Corollary 5.3 now implies that the N-dimensional linear sample space $S(BG_{t\ell+1}) \subseteq \{0,1\}^N$of size $2^{t\ell+1} = 2N^{\frac{d-1}{2}}$ is d-wise independent. This concludes the proof. □

Remark: The matrix B constructed above is well-known in classical coding theory: it is essentially the parity check matrix of the famous binary BCH-codes discovered by Hocquenghem (1959) and independently by Bose and Ray-Chaudhuri (1960). BCH-codes and their extensions are widely used in satellite communications and computer drives to correct errors in messages. The idea of error correction is the following. If a (say binary) message is sent through a "noisy channel", then it could arrive distorted, as the noise might flip some of the bits. To circumvent this, the message is encoded somehow into a longer message, with the intention that even if some bits are flipped by the noise, the message could be reconstructed. The plan is to encode the message before transmission into a sequence of *code words*, that are elements of a carefully selected set $C \subseteq \{0,1\}^N$ of vectors. The set C is refered to as a *binary code*. To measure how good a code $C \subseteq \{0,1\}^N$ is in terms of fixing the errors caused by the noise, we say that C *corrects up to d errors*, if for any vector $a \in \{0,1\}^N$, there is *at most one* code word which differs from a in at most d bits. This is a sensible definition, because then no matter what vector a is received at the end of a transmission through a noisy channel which does not introduce more than d errors to a code word, we can determine uniquely

which message (i.e. code word) was originally sent.

Obviously the more error a code can correct, the better. But one also feels that the more error one would like to be able to correct, the longer the code words will have to be, and consequently the longer it will take to communicate the same message. The latter property is measured by the *rate* of the code, i.e. the amount of useful information divided by the actual information sent. In our system we choose to send one of $|C|$ different code words of length N, that is $\log|C|$ bits of information, using N bits. So we can define the rate of the code to be the quantity $\frac{\log|C|}{N}$. One compares this number to the largest number d, such that the code corrects up to d errors. There will be no big surprises: the larger the error correction, the smaller the rate has to be. The exact dependence of these quantities on each other (together with the speed of encoding/decoding) is crucial in practical applications. In fact for BCH-codes one ignores the row of B corresponding to 0, because the cyclic nature of the multiplicative group of $\mathbb{F}_{2^t}$ makes it possible to devise very fast encoding and decoding algorithms.

A set $C \subseteq \{0,1\}^N$ of vectors is called a *binary code* and its elements are called *code words*. To measure how good a code $C \subseteq \{0,1\}^N$ is in terms of fixing the errors caused by the noise, we say that C *corrects up to d errors*, if for any vector $a \in \{0,1\}^N$, there is *at most one* code word which differs from a in at most d bits.

Exercise 5.13 *Let M be a matrix whose columns are the elements of a d-wise independent N-dimensional linear sample space S of size $|S| = m$, and let*

$$C = \{x \in \mathbb{F}_2^N : x^T M = 0\}$$

be the subset defined by the vectors orthogonal to all members of S. Show that C corrects up to $d/2$ errors.

Exercise 5.14 *A random variable is called* almost constant *if there exists a single value that it takes with probability* 1*.*

(a) Show that if the N random variables $r_1, \ldots r_N : \Omega \to \mathbb{R}$ are mutually independent and not almost constant, then the 2^N functions of the form $f_J = \prod_{j\in J}(r_j - \mathbb{E}[r_j])$, $J \subseteq [N]$, are linearly independent in the vector space $\mathbb{R}^\Omega$.

(b) In Theorem 5.1 we have constructed $2N^{\lfloor \frac{d}{2} \rfloor}$ d-wise independent 0/1-valued random variables having the uniform distribution. Here we show that this is best possible up to a constant factor depending only on d.
Let $m(N,d)$ be the sum of the following binomial coefficients:

$$m(N,d) = \begin{cases} \sum_{j=0}^{d/2} \binom{N}{j} & \text{if } d \text{ is even} \\ \sum_{j=0}^{(d-1)/2} \binom{N}{j} + \binom{n-1}{(d-1)/2} & \text{if } d \text{ is odd.} \end{cases}$$

Show that if the (not necessarily $0/1$*-valued) random variables* $r_1, \ldots, r_N$ *over the sample space* Ω *are* d*-wise independent and not almost constant, then the size* $|\Omega|$ *of the sample space is at least* $m(N,d)$ *(which is of the order* $n^{\lfloor \frac{d}{2} \rfloor}$*).*

Let us now return to our original problem of constructing Ramsey graphs. We define $N = \binom{n}{2}$, $d = 2\log_2^2 n$, and take our d-wise independent sample space of size $2(N+1)^{(d-1)/2}$ we have just constructed. We interpret the members of this sample space as graphs on n vertices and denote their family by $\mathcal{D}$. If we take the Abbott product of all graphs in $\mathcal{D}$, we have a graph G with $n^{|\mathcal{D}|}$ vertices and clique- and independence number at most $(4\log_2 n)^{|\mathcal{D}|}$. After doing the math we obtain that we constructed a k-Ramsey graph of order $k^{\Omega\left(\frac{\sqrt{\log\log k}}{\log\log\log k}\right)}$.

Exercise 5.15 *Establish that the construction we gave using the Abbott-product of graphs from the* d*-wise independent sample space we described is indeed strongly explicit. That is, for any* n *define a graph* G_n *that is* $2^{\frac{\log\log\log n}{\log\log n}}$*-Ramsey, and describe an algorithm that outputs in poly*$(\log n)$*-time whether two input vertices* u *and* $v \in [n]$ *are adjacent in* G_n *or not. Argue that* G_n *is also morally explicit.*

This is alright: we improved from three times iterated logarithm in the exponent to two-times iterated logarithm.

Can we carry the idea of sample spaces even further? Not, if we insist on d-wise independence: Exercise 5.14 above combined with Exercise 5.12 shows that, up to constant factor, the size of our d-wise independent sample space is as small as it could be.

In order to proceed, we simply have to give up on perfect $2\log_2 n$-wise independence. In fact, not insisting anymore that in calculation (5.2) the probability $\mathbb{P}[K \text{ is a clique}]$ of a k-set K hosting a clique is *exactly* $\frac{1}{2^{\binom{k}{2}}}$, but being content with it being *at most*, say, twice as large, would not have any serious effect on the rest of the proof. It turns that this idea of being "lenient" with independence is a good one—it will allow us to further significantly reduce the size of our sample space, leading us to the construction of even larger k-Ramsey graphs.

5.4 Approximating randomness

5.4.1 Almost independent sample spaces

In this section we relax the independence requirement of sample spaces that each bit-vector should appear with the exact same probability, and allow that they appear with *roughly* the same probability, up to an error of ϵ.

Definition: A sample space $S \subseteq \mathbb{F}_2^N$ is called ϵ*-close to independent* if for any vector $a \in \{0,1\}^N$, we have

$$\left| \mathbb{P}[s = a] - \frac{1}{2^N} \right| \leq \epsilon.$$

Note that being 0-close to independent is equivalent to being independent. For our application we of course need the extension of the definition to almost d-wise independence.

Definition: The sample space $S \subseteq \{0,1\}^N$ is called *ϵ-close to d-wise independent* if for any subset $J \in \binom{[N]}{d}$ of the coordinates, the sample space $S|_J \subseteq \{0,1\}^d$ is ϵ-close to independent, that is, for any vector $a \in \{0,1\}^d$, we have

$$\left| \mathbb{P}[s|_J = a] - \frac{1}{2^d} \right| \leq \epsilon.$$

In the main result of this section we show that allowing a bit of imperfectness in d-wise independence enables one to reduce the size of the sample space from the polynomial of Theorem 5.1 to a *polylogarithmic* function of N. More precisely, we will construct sample spaces that are ϵ-close to d-wise independent, and their size is only polylogarithmic in N and polynomial in their imperfectness measurements, i.e., in d and $\frac{1}{\epsilon}$.

Theorem 5.4 (Naor and Naor) *Let $N = 2^t$ with $t \in \mathbb{N}$, let $d \geq 1$ be an odd integer, and let $\epsilon > 0$. Then there is a sample space $R \subseteq \{0,1\}^N$ of size at most*

$$\frac{2\left(t^{\frac{d-1}{2}}+1\right)^2}{\epsilon^2} \sim \frac{d^2}{2\epsilon^2}\log^2 N,$$

which is ϵ-close to d-wise independent.

The main idea of the proof is to take the d-wise independent sample space $S(BG_m)$ we constructed in the last section and somehow reduce its size. The columns of the matrix BG_m are the 2^m linear combinations of the columns of B: each column of G_m is responsible for one. The plan is to take only an appropriately selected few of these, such that the d-wise independence is not ruined too much. It is tempting to select a few columns of G_m randomly, but we must remain sober and resist—we want an explicit construction. We will instead construct an m-dimensional sample space $S(Q)$ of quadratic size $p \sim \frac{m^2}{\epsilon^2}$, as opposed to 2^m, which is ϵ-close to independent. Then we will show that taking only this p linear combinations of the columns of B, as opposed to all 2^m, is enough to maintain the d-wise independence with an error ϵ. Namely, we will show that $S(BQ)$ is ϵ-close to d-wise independent.

In the next two subsections we work out the ingredients of this plan and then the proof of Theorem 5.4 will follow easily.

Linear tests

The property of being ϵ-close to d-wise independent is quite difficult to work with, let alone to show directly. Hence we develop a more effective way to establish it, a way which is much more apt to our plan to create our sample space via linear combinations.

If a sample space $S = S(M) \subseteq \{0,1\}^N$ is independent then we have seen in Exercise 5.12 that it is 1-independent, that is, the number of 0 and 1 in every row of M is the same. In the next exercise we generalize this to give yet another characterization of independent sample spaces.

Exercise 5.16 *A sample space $S \subseteq \{0,1\}^N$ is independent if and only if for every vector $a \in \{0,1\}^N \setminus \{0^N\}$,*

$$\mathbb{P}[s \cdot a = 0] = \mathbb{P}[s \cdot a = 1].$$

Here 0^N denotes the vector of length N having only 0 coordinates, while $s \cdot a = \sum_{i=1}^{N} s_i a_i$ represents the usual dot-product of vectors over $\mathbb{F}_2$.

The exercise involves $2^N - 1$ "linear test"s one performs on the sample space to verify its independence, each of which should produce a halving of the sample space. We will relax on the perfectness of these halvings to approach the concept of almost independence.

Definition: A sample space $S \subseteq \{0,1\}^N$ is called *ϵ-unbiased with respect to linear tests* if for any $a \in \{0,1\}^N \setminus \{0^N\}$,

$$|\mathbb{P}[s \cdot a = 0] - \mathbb{P}[s \cdot a = 1]| \leq \epsilon.$$

Note that S is ϵ-unbiased with respect to linear tests if and only if for any $a \in \{0,1\}^N \setminus \{0^N\}$, the 1-dimensional sample space $\{s \cdot a : s \in S\} \subseteq \{0,1\}$ is $\epsilon/2$-close to independent.

The equivalence of being ϵ-unbiased with respect to linear tests and being ϵ-close to independent, which was established in Exercise 5.16 for $\epsilon = 0$, does not hold for $\epsilon > 0$. This is shown in the next exercise.

Exercise 5.17 *Show that if a sample space $S \subseteq \{0,1\}^N$ is ϵ-close to independent then it is also $\epsilon 2^N$-unbiased with respect to linear tests. Construct a sample space that shows the statement being best possible (for all sensible values of the parameters N and ϵ).*

The following lemma states that one direction of Exercise 5.16 remains valid even if $\epsilon > 0$ and thus establishes linear tests as a method to prove ϵ-closeness to independence.

Lemma 5.4.1 (Vazirani) *Let $S \subseteq \{0,1\}^N$ be a sample space that is ϵ-unbiased with respect to linear tests. Then S is ϵ-close to independent.*

Proof. We introduce the probability distribution function p on $\mathbb{Z}_2^N$ by setting $p(x) := \mathbb{P}[s = x]$ for the probability of a vector $x \in \{0,1\}^N$ in the sample space S. We need to show that this function $p : \mathbb{Z}_2^N \to \mathbb{C}$ does not deviate more than ϵ from its average $\frac{1}{|\mathbb{Z}_2^N|} \sum_{x \in \mathbb{Z}_2^N} p(x) = \frac{1}{2^N}$. We make use of the basic properties of the discrete Fourier

transform of p on the group $\langle H, +\rangle = \langle \mathbb{Z}_2^N, +\rangle$. In particular, applying Proposition A.35 we obtain that

$$\left| p(a) - \frac{1}{2^N} \right| \leq \Phi(p) |\mathbb{Z}_2^N| \tag{5.6}$$

for every $a \in \mathbb{Z}_2^N$, where

$$\Phi(p) = \max\{|\langle \chi, p\rangle| : \chi \in \widehat{\mathbb{Z}_2^N}, \chi \neq \chi_0\}$$

is the largest absolute value among the non-principal Fourier coeffiecients of p.

Recall that the characters of $\mathbb{Z}_2^N$ are defined by $\chi_b(a) = (-1)^{b \cdot a}$, for every $b \in \mathbb{Z}_2^N$ and $a \in \mathbb{Z}_2^N$. The key observation is that the probability difference between the occurrence of 0 and 1 upon making a linear test with some test vector $b \in \mathbb{Z}_2^N$ is precisely the (non-normalized) Fourier coefficient of p corresponding to character $\overline{\chi}_b$. The test vector $b = 0^N$ corresponds then to the principal character χ_0 and therefore our assumption on S implies that all, but the principal, non-normalized Fourier coefficients of p are at most ϵ. And that, via (5.6), implies that S is ϵ-close to independent.

Indeed, for any $b \in \mathbb{Z}_2^N \setminus \{0^N\}$ we have

$$\begin{aligned} \epsilon &\geq \mathbb{P}[s \cdot b = 0] - \mathbb{P}[s \cdot b = 1] = \sum_{\substack{a \in \mathbb{Z}_2^N \\ a \cdot b = 0}} \mathbb{P}[s = a] - \sum_{\substack{a \in \mathbb{Z}_2^N \\ a \cdot b = 1}} \mathbb{P}[s = a] \\ &= \sum_{a \in \mathbb{Z}_2^N} (-1)^{a \cdot b} p(a) = \sum_{a \in \mathbb{Z}_2^N} \chi_b(a) p(a) = \left|\mathbb{Z}_2^N\right| \langle \overline{\chi}_b, p\rangle, \end{aligned}$$

and the lemma is proved. □

Our eventual goal is the construction of a small sample space that is ϵ-close to d-wise independent. The next lemma describes an easy way to combine the generator matrix L of a d-wise independent linear sample space $S(LG_m)$ with a sample space $S(Q)$ which is ϵ-close to independent and obtain a sample space that is ϵ-close to d-wise independent.

We plan to use Lemma 5.4.1 to each d-dimensional restriction of the constructed sample space, in order to establish that they are all ϵ-close to independent, and then conclude that the sample space itself is ϵ-close to d-wise independent.

Lemma 5.4.2 (Naor and Naor) *Let B be an $(N \times m)$-matrix over $\mathbb{F}_2$ such that any d-rows are linearly independent and let Q be a $(m \times p)$-matrix over $\mathbb{F}_2$ such that the sample space $S(Q) \subseteq \{0,1\}^m$ of size p is ϵ-unbiased with respect to linear tests. Then the sample space $S(BQ) \subseteq \{0,1\}^N$ of size p is ϵ-close to d-wise independent.*

Proof. We have to check that for every subset $J \subseteq [N]$ of size d the rows, the restriction of the sample space $S(BQ)$ to these d rows is ϵ-close to independent. To this end we would like to use Lemma 5.4.1 and hence verify that the d-dimensional restriction $S(BQ)|_J = S(B|_J Q)$ is ϵ-unbiased with respect to linear tests. Let $a \in \{0,1\}^d \setminus \{0^d\}$

be a d-dimensional test vector. Since $a^T(B|_J Q) = (a^T B|_J)Q$, the linear test of $S(B|_J Q)$ with test vector a and the linear test of $S(Q)$ with test vector $a^T B_J$ are the same. Note that the test vector $a^T B|_J \in \{0,1\}^m$ is non-zero, since $a \neq 0^d$ and any d rows of B are linearly independent. By assumption the sample space $S(Q)$ is ϵ-unbiased with respect to linear tests, so the probability of 0 and the probability of 1 differ by at most ϵ in the 1-dimensional sample space $S((a^T B|_J)Q)$, and hence also in $S(a^T(B|_J Q))$. □

The first ingredient of Lemma 5.4.2, a matrix B with any d of its rows being linearly independent, was constructed in Theorem 5.1. In the next subsection we construct the second ingredient: a small sample space $S(Q) \subseteq \{0,1\}^m$ which is ϵ-unbiased with respect to linear tests.

Almost independent sample spaces via the quadratic character

A field has two operations: addition and multiplication. There are many examples of the vague phenomenon that being a regular structure in some additive sense and being a regular structure in some multiplicative sense are mutually exclusive, or at least very limited in size. As a simplest example one can think of are arithmetic and geometric progressions: the largest set that is both is of size two. Recall the Paley graph we discussed in the first section of this part: for a prime $p \cong 1 \pmod 4$, the Paley graph P_p was just the Cayley graph defined on the additive group of $\mathbb{F}_p$ by the generating set $S = QR_p$ of the quadratic residues. That is, the Paley graph is defined on the additive structure of a field by a generating set that is multiplicative in nature. While we know, modulo the Generalized Rieman Hypothesis, that the Paley graph is not a perfect source of randomness, we also know that it might be a pretty good imitation, in fact waay better than anything we are able to construct today.

We use this intuition, the quadratic residues being a pseudorandom random subset within the additive structure of the finite field $\mathbb{F}_p$. Recall that the value of the quadratic character $\varrho_p : \mathbb{F}_p^* \to \{-1,1\}$ is 1 for quadratic residues and -1 for quadratic non-residues. Convert these values to bits: let $r(x) = 0$ for quadratic residues and 1 for non-residues. Expressed with a formula, we have $\varrho_p(x) = (-1)^{r(x)}$. In other words, $r = \mathbb{1}_{NQR_p}$ is just the characteristic function of the quadratic non-residues modulo p. Imagine these values in the cyclic additive order of the field, that is $r(1), r(2), (3), \ldots, r(p-1), r(0)$. For 0 let us just extend r arbitrarily, say let us have $r(0) = 1$.

Our sample space will consits of the p bit-vectors that form an interval of length m in this cyclic ordering of length p. Since intervals are very regular additive structures, we hope that the multiplicatively defined values will be quite random. Naturally, we will have to assume that m is small enough compared to p. Formally, we define a $(m \times p)$-matrix $Q = Q_m^p$, whose colums $q^{(x)} \in \mathbb{F}_2^m$ are labeled by elements $x \in \mathbb{F}_p$ and $q_i^{(x)} := r(x+i)$ for every $i = 1, 2, \ldots, m$.

Proposition 5.5 (Alon, Goldreich, Hastad, and Peralta) *For every $m \leq \sqrt{p}$, the sample space $S(Q_m^p) = \{q^{(x)} : x \in \mathbb{F}_p\}$ is $\frac{m}{\sqrt{p}}$-unbiased with respect to linear tests.*

Note that for this proposition to have any power, we better have $m \leq \epsilon\sqrt{p}$ with some $\epsilon < 1$; the smaller the ϵ, the better.

Proof. Let us fix our "linear tester" $a \in \{0,1\}^m$. As we saw in the proof of Lemma 5.4.1, the probability difference in the definition of almost independence can be expressed as follows.

$$\mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)}\cdot a = 0\right] - \mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)}\cdot a = 1\right] = \sum_{\substack{b\in\mathbb{F}_p \\ q^{(b)}\cdot a = 0}} \mathbb{P}_{x\in\mathbb{F}_p}[x=b] - \sum_{\substack{b\in\mathbb{F}_p \\ r^{(b)}\cdot a = 1}} \mathbb{P}_{x\in\mathbb{F}_p}[x=b]$$
$$= \frac{1}{p}\sum_{b\in\mathbb{F}_p}(-1)^{q^{(b)}\cdot a} = \frac{1}{p}\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(-1)^{r(b+i)a_i}$$

We want to replace each product $\prod_{i=1}^m(-1)^{r(b+i)a_i}$ with $\prod_{i=1}^m(\varrho_p(b+i))^{a_i} = \varrho_p(\prod_{i=1}^m(b+i)^{a_i})$ and then use Weil's Theorem for the quadratic character ϱ_p and the polynomial $f(x) = \prod_{i=1}^m(x+i)^{a_i}$. We can certainly do this whenever $b \in \mathbb{F}_p$ is not in the interval $[p-b, p-1]$, because then $b+i \neq 0$ and hence $(-1)^{r(b+i)} = \varrho_p(b+i)$ for every $i = 1,2,\ldots,m$, by the definition of r. These are most of the $b \in \mathbb{F}_p$; only those in the interval $[p-m, p-1]$ of length $m \leq \sqrt{p}$ are problematic. Whenever $b \in [p-m, p-1]$ the corresponding product contains a factor $(-1)^{r(b+i)a_i}$ with $b+i=0$. Considering that for the sake of Weil's Theorem $\varrho_p(0)$ is defined to be 0, whenever $b+i=0$, we have that $\left|(-1)^{r(b+i)a_i} - \varrho_p(b+i)^{a_i}\right|$ is either 0 or 1 (depending on whether $a_i = 0$ or 1).

$$\left|\mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)}\cdot a = 0\right] - \mathbb{P}_{x\in\mathbb{F}_p}\left[q^{(x)}\cdot a = 1\right]\right| = \left|\frac{1}{p}\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(-1)^{r(b+i)a_i}\right| \leq$$
$$\leq \frac{1}{p}\left|\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(\varrho_p(b+i))^{a_i}\right| + \frac{1}{p}\sum_{b\in[p-m,p-1]}\left|\prod_{i=1}^{m}(-1)^{r(b+i)a_i} - \prod_{i=1}^{m}(\varrho_p(b+i))^{a_i}\right|$$
$$\leq \frac{1}{p}\left|\sum_{b\in\mathbb{F}_p}\varrho_p\left(\prod_{i=1}^{m}(b+i)^{a_i}\right)\right| + \frac{m}{p}$$
$$\leq \frac{m-1}{\sqrt{p}} + \frac{m}{p} \leq \frac{m}{\sqrt{p}}.$$

In the last step we used $m \leq \sqrt{p}$ and in the next to last we applied Weil's theorem for the quadratic character ϱ_p which has order 2 and the polynomial $f(x) = \prod_{i=1}^m(x+i)^{a_i}$ which has at most m distinct roots and is certainly not a square. □

Proof. We can now put together the proof of Theorem 5.4 by using Lemma 5.4.2 with the almost independent independent sample space of Proposition 5.5 and the d-wise independent linear sample space of Theorem 5.1.

Let $m = t^{\frac{d-1}{2}} + 1$. First construct the $(N \times m)$-matrix B with the property that any d rows are linearly independent. Then, after choosing a prime p between $\frac{\left(t^{\frac{d-1}{2}}+1\right)^2}{\epsilon^2}$

and its double, construct the above sample space $S(Q_m^p) \subseteq \{0,1\}^p$ with $m = t\frac{d-1}{2} + 1$. By Proposition 5.5 $S(Q_m^p)$ is $\frac{m}{\sqrt{p}}$-unbiased with respect to linear tests. Note that $\frac{m}{\sqrt{p}} \leq \epsilon$. According to Lemma 5.4.2 the sample space $S(BQ_m^p)$ of size p is ϵ-close to d-wise independent. This concludes the proof of the theorem. □

Better Ramsey-graphs

Let us now try to use our sample spaces from Theorem 5.4 which are ϵ-close to d-wise independent in our quest for explicit Ramsey graphs.

We could again take our constructive sample space, like we did earlier, interpret it as graphs on $N = \binom{n}{2}$ vertices and take the Abbott product of all of them. But in fact, since our sample space is now so small, we can do even better. We can return to the original idea of the Abbott construction: checking for the perfect "starter graph" with brute force in polynomial time, and then taking the Abbott-powers of this single graph with good Ramsey properties.

Our goal in this section is the construction of a graph G on n vertices in time polynomial in n with $\omega(G), \alpha(G) < 2^{\sqrt{\log n} \log\log n}$. In the solitude of your home you should check that it is equivalent to constructing a k-Ramsey graph with $k^{\frac{\log k}{(\log\log k)^2}}$ vertices. Recall that this will be a further improvement in the line of our constructive lower bounds: the exponent of the order of the construction in Subsection 5.3.1 was twice iterated logarithm and now we have essentially a single $\log k$ in the exponent (disregarding the lower order $(\log\log k)^2$ in the denominator.)

This construction was apparently folklore, here we follow the description of Baraz. Let us fix the number of vertices n and define the integer $k = 2^{\sqrt{\log n}}$.

We aim to find our "good starter" graph H on k vertices. What is special about the selection of k. We will see that on the one hand we can choose a sample space of size polynomial in n of graphs on k vertices, which γ-close to d-wise independent, where γ is small enough and d is large enough. On the other hand it is possible to check for small enough cliques on k vertices.

We take a sample space $S \subseteq \{0,1\}^{\binom{k}{2}}$ which is $2^{-5\log^2 k}$-close to being $4.5\log^2 k$-wise independent. By Theorem 5.4 there exists such a space of size

$$\approx 20.25 \log^4 k 2^{10\log^2 k} \log^2 \binom{k}{2} = k^{O(\log k)} = n^{O(1)},$$

i.e., the size of this space is polynomial in n.

Note that for any graph on k vertices we can check, just by brute force, whether the clique number and the independence number of it is at most $3\log k$, in time

$$\binom{k}{3\log k}\binom{3\log k}{2} = k^{O(\log k)} = n^{O(1)},$$

which is polynomial in n.

Hence in polynomial time we can check for each member of this sample space, whether its clique number and independence number is at most $3\log k$. What is left to prove is that in S, there exist such a graph. This follows from the almost d-wise independence of the space. Fix a subset L of the vertices, $|L| = 3\log k$. Then by the almost $4.5\log^2 k$-independence of the sample space,

$$\mathbb{P}[L \text{ is a clique or independent set}] = 2 \cdot \left(\frac{1}{2^{\binom{|L|}{2}}} + \frac{1}{2^{5\log^2 k}} \right) \ll \frac{1}{\binom{k}{3\log k}}.$$

That is *there exists* a member of the sample space S for which no set of size $3\log k$ is a clique or an independent set. This will be our starter graph H and our brute force search will certainly find it in polynomial time in n.

Now take the $\sqrt{\log n}$th Abbott-power of H. This product graph has $k^{\sqrt{\log n}} = n$ vertices and can be constructed in time polynomial in n (Exercise ??). By (5.7), its clique number and independence numnber is certainly upper bounded by

$$(3\log k)^{\sqrt{\log n}} = (3\sqrt{\log n})^{\sqrt{\log n}} = 2^{\sqrt{\log n}\log\log n\left(\frac{1}{2}+\frac{\log 3}{\log\log n}\right)}.$$

The extra factor in the exponent is smaller than 1 for large enough n and hence we are done.

Note however a crucial difference in the construction of this last example and the rest of this section. When we took the Abbott-product of all graphs in Subsection 5.2 or when we took the Abbott-product of all graphs from the d-wise independent sample space in Subsection 5.3.1, these were strongly explicit, even morally explicit constructions. In our current construction one needs to construct the starter graph first before being able to answer adjacency queries about its Abbott-power and this alone already takes time polynomial in n, and not in $\log n$. So we have "only" an efficiently explicit construction.

The best strongly explicit construction by the Abbott-product (from Subsection 5.3.1) has a twice iterated logarithm in the exponent. In the next section we discuss a surprisingly simple strongly and morally explicit construction, which beats slightly even the efficiently explicit Abbott-type construction above.

5.5 Ramsey graphs via intersection theorems

In 1977 Frankl extended the construction of Nagy using the theory of *sunflowers* to obtain a constructive superpolynomial lower bound $k^{f(k)}$, with $f(k) = \Omega\left(\frac{\log k}{\log\log k}\right) \to \infty$. Later Frankl and Wilson (1981) gave a simpler proof through the linear algebra method. This is what we will discuss here. Let p be a prime and define the graph G by

$$V(G) = \binom{[p^3]}{p^2-1}, \quad A \text{ and } B \text{ are adjacent if } |A \cap B| \equiv -1 \pmod{p}.$$

Observe that for $p = 2$ we get back Nagy's construction with $k = 8$.

Theorem 5.6 *Graph G contains no clique and no independent set of size*

$$\sum_{i=0}^{p-1} \binom{p^3}{i} + 1.$$

Provided that the theorem holds, we have a $\sim p^{2p}$-Ramsey graph on $\sim p^{p^2}$ vertices.

Exercise 5.18 *Check (precisely!) that for* every k *we have a k-Ramsey graph with $k^{\Omega\left(\frac{\log k}{\log\log k}\right)}$ vertices.*

The proof of Theorem 5.6 is again a wonderful application of the linear algebra method, which goes one step further than the proof of the theorem of Nagy. Now characteristic vectors do not suffice; we need a simple technical lemma about *function spaces*. Let F be a field and $\Omega \subseteq F^n$. Then the set $F^\Omega = \{f : \Omega \to F\}$ of functions is a *vector space over F*.

Lemma 5.6.1 *If $f_1, \ldots, f_m \in F^\Omega$ and $v_1, \ldots, v_m \in \Omega$ such that*

- *$f_i(v_i) \neq 0$, and*
- *$f_i(v_j) = 0$ for all $j < i$,*

then $f_1, \ldots, f_m$ are linearly independent in F^Ω.

Proof. (of Lemma 5.6.1) Suppose $\lambda_1 f_1 + \cdots + \lambda_m f_m = 0$, and let j be the smallest index j with $\lambda_j \neq 0$. Substituting v_j into this function equation we have

$$\underbrace{\lambda_1 f_1(v_j) + \cdots + \lambda_{j-1} f_{j-1}(v_j)}_{=0,\text{ since } \lambda_i = 0,\ i<j} + \underbrace{\lambda_j f_j(v_j)}_{\neq 0} + \underbrace{\lambda_{j+1} f_{j+1}(v_j) + \cdots + \lambda_m f_m(v_j)}_{=0,\text{ since } f_i(v_j) = 0,\ j<i} = 0,$$

a contradiction. □

Proof. (of Theorem 5.6) For a set $A \in 2^{[p^3]}$ let $v_A \in \{0,1\}^{p^3}$ be the characteristic vector of A. The linear algebra method is based on a simple, but crucial identity connecting the size of the intersection of two sets to the inner product of their characteristic vectors, namely that $|A \cap B| = \langle v_A, v_B \rangle$.

Independent sets. Let $A_1, \ldots, A_s$ be an independent set in G, so $|A_i \cap A_j| \not\equiv -1 \pmod p$ for every $i \neq j$. For each i let $v_i = v_{A_i}$ be the characteristic vector of A_i. Our plan is to define a function $f_i : \{0,1\}^{p^3} \to \mathbb{F}_p$ for every $i = 1, \ldots s$, prove that they are linearly independent and bound the dimension of the vector space they span — giving us an upper bound on s. Let

$$\tilde{f}_i(x) = \prod_{l=0}^{p-2} (\langle x, v_i \rangle - l),$$

for all i. Obviously we have $\tilde{f}_i(v_i) \neq 0$, since $|A_i| \equiv -1 \pmod{p}$. On the other hand, we have $\tilde{f}_i(v_j) = 0$ for all $j \neq i$, since $\{A_1, \ldots, A_s\}$ is an independent set. Our technical lemma then implies that $\tilde{f}_1, \ldots, \tilde{f}_s$ are linearly independent. The dimension of the space these functions span could be quite large, since each variable x_j, $j = 1, \ldots, p^3$ could appear with powers ranging from 0 to $p-1$. To reduce the dimension of the space, we apply a "multilinearization trick" and define $f_i(x)$ from $\tilde{f}_i(x)$ by replacing each occurrence of a large power x_i^l $(l > 1)$ with x_i. Observe that $f_i \equiv \tilde{f}_i$ on $\{0,1\}^{p^3}$. Since all the f_i are multilinear polynomials, the dimension of the space spanned by them is the number of monomials of degree at most $p-1$,

$$1 + p^3 + \binom{p^3}{2} + \cdots + \binom{p^3}{p-1}.$$

Cliques. To bound the clique number of G we proceed similarly, but we will work over $\mathbb{R}$ instead of $\mathbb{F}_p$. Let $B_1, \ldots, B_t$ be a clique in G, so $|B_i \cap B_j| \equiv -1 \pmod{p}$ for every $i \neq j$. Let $L = \{p-1, 2p-1, \ldots, p^2-p-1\}$ be the set of possible intersection sizes. Note that $|L| = p-1$. For each i let $w_i = v_{B_i}$ be the characteristic vector of B_i and let

$$\tilde{f}_i(x) = \prod_{l \in L} (\langle x, w_i \rangle - l)$$

be functions $\{0,1\}^{p^3} \to \mathbb{R}$ for all i. Since $|B_i| = p^2 - 1 \notin L$, we have $\tilde{f}_i(w_i) \neq 0$. On the other hand, $\tilde{f}_i(w_j) = 0$ for all $j \neq i$. Lemma ?? then implies that $\tilde{f}_1, \ldots, \tilde{f}_t$ are linearly independent. Again, we multilinearize the functions and define $f_i(x)$ from $\tilde{f}_i(x)$ by replacing each occurrence of a large power x_i^l $(l > 1)$ with x_i. Since $|L| = p-1$, all the f_i are multilinear polynomials of degree at most $p-1$. Thus the dimension of the space spanned by them is at most

$$1 + p^3 + \binom{p^3}{2} + \cdots + \binom{p^3}{p-1}.$$

□

Exercise 5.19 *The proof of the following theorem is an immediate generalization of the claim we had about the clique number of the Frankl-Wilson graph. (Think this over!)*

Theorem *Let L be a set of integers with $|L| = s$. Let $B_1, \ldots, B_t \in 2^{[n]}$ be a uniform L-intersecting family, i.e. all $|B_i|$ have the same size and $|B_i \cap B_j| \in L$ for every $i \neq j$. Then $t \leq \sum_{i=0}^{s} \binom{n}{i}$.* □

Generalize this statement further to arbitrary *L-intersecting families, i.e. derive the same conclusion when the $|B_i|$ are not necessarily all equal. (Hint: Select the functions $\tilde{f}_i$ more carefully and use Lemma 5.6.1 in its full power.)*

Bipartite Ramsey problem The bipartite Ramsey number $BR(k,k)$ denotes the smallest integer N such that every two-coloring of $K_{N,N}$ contains a monochromatic $K_{k,k}$. The story of bipartite Ramsey numbers is very similar to that of the ordinary Ramsey numbers in the sense that we *know* that $BR(k,k)$ is exponential: the uniform random two-coloring shows that $BR(k,k) > \sqrt{2}^k$. The parallels stop right there though, as comparable constructive lower bounds are much harder to obtain. Abbott's product, Nagy's set intersection construction, and even the simple Turán's construction has no obvious analogue. Even for a construction of quadratic order, we have to work! There are a couple of different constructios yielding a quadratic lower bound; we treat these in the exercises.

Exercise 5.20 *A square matrix H with entries $+1$ and -1 is called an* Hadamard matrix *if the rows are pairwise orthogonal.*

(a) Show that that the colums of an Hadamard matrix are pairwise orthogonal.

(b) Show that the order of an Hadamard matrix is 1 or 2 or divisible by 4.[5]

(c) Construct $2^n \times 2^n$ Hadamard matrices for every integer $n \geq 1$.

(d) Given an Hadamard matrix $H = (h_{ij})$, define a two-coloring of $K_{N,N}$ by coloring the edge xy red if the entry $h_{x,y}$ is $+1$ and blue otherwise. Prove that for arbitrary integers s,t, $1 \leq s,t \leq N$, we have

$$\left| \sum_{i=1}^{r} \sum_{j=1}^{s} h_{i,j} \right| \leq \sqrt{rsN}.$$

Conclude that this coloring is $\sqrt{N}$-Ramsey.

Exercise 5.21 *Let $q \cong 3 \pmod 4$ be a prime power and let $\rho_q : \mathbb{F}_q \to \{1,-1\}$ denote the quadratic chracter, extended to 0 by $\rho_q(0) = 1$. The rows and columns of the matrix $Q = (r_{a,b})$ are labeled by the elements of $\mathbb{F}_q$ and its entries are defined by $r_{a,b} = \rho_q(a-b)$. Let H be the $(q+1) \times (q+1)$-matrix we obtain by adding to Q first a column of -1 (of length q) and then a row of 1s (of length $q+1$. Prove that H is an Hadamard matrix.*

Using the projective norm graphs on $n = q^t - q^{t-1}$ vertices we know even more in some sense. These graphs do not contain $K_{t,t!+1}$ and one can also prove that their bipartite complement does not contain $\overline{K}_{n^{1/2+1/t},n^{1/2+1/t}}$. Selecting $t = c \ln n / \ln \ln n$ we have that there is no $\overline{K}_{Cn^{1/2}\ln n, Cn^{1/2}\ln n}$ and no $K_{c\ln n/\ln\ln n, n^{\epsilon(c)}}$, where $\epsilon(c) \to 0$. Despite having such asymmetric construction, with much better parameters in the forbidden red bi-clique, it was a long-standing open problem to go below $\sqrt{n}$ by *any* infinite factor for *both* the red and the blue bi-clique.

[5] It is a notorious conjecture of design theory that Hadamard matrices exist for all N divisible by 4.

The quadratic constructive lower bound for the bipartite Ramsey number was broken through by a factor tending to infinity only in 2004. The first superpolynomial construction was given in 2010. In 2012 the Frankl-Wilson barrier was surpassed for bipartite graphs. The problem was the subject of vigorous research and further milestones were achieved in the past few years. The current record is due to Gil Cohen, which is a family of strongly explicit $O\left((\log n)^{(\log\log\log n)^c}\right)$-Ramsey graphs. Recall that in the random graph the largest clique and independent set is logarithmic and in this construction it is already "almost" polylogarithmic ...[6] Unfortunately all these constructions are complicated and lengthy to be presented here, but it is good to know: we are almost there.

[6]We discussed earlier the rate of speed by which the function $\log\log\log n$ tends to infinity...