Chapter 5

The symmetric Ramsey-problem

5.1 What sort of explicit?

Let us recall that the Ramsey number

$$R(k,l) = \min\{n : \forall \text{ graph on } n \text{ vertices contains either } K_k \text{ or } \overline{K_l}\}.$$

The most interesting question concerns the symmetric case, i.e. when $k = l$. We call a graph k-*Ramsey* if both the largest independent set and clique are of order less than k. $R(2,2) = 2$ is a triviality, while $R(3,3) = 6$ is a standard first year combinatorics exercise. It is already a nontrivial task to construct a 4-Ramsey graph of order 17 and prove that it is the best possible, i.e. that $R(4,4) = 18$. About $R(5,5)$ we only know that it is between 43 and 49.

In 1935 Erdős and Szekeres showed that $R(k,l) \leq \binom{k+l-2}{k-1}$, so in particular $R(k,k) < 4^k$. For a while the Turán graph (1941) on $(k-1)^2$ vertices provided the best lower bound. In fact Turán believed this to be the truth, i.e. that $R(k,k) = (k-1)^2$. It came as a great surprise in 1947 when Erdős, using non-constructive methods proved that $R(k,k)$ is of exponential order. His paper, showing the *existence* of k-Ramsey graphs of order $\sqrt{2}^k$, is often considered the starting point of the Probabilistic Method in combinatorics.

It is a frustrating fact that today, these two ingenious but relatively simple arguments provide more or less the best known bounds. Some small improvements came along later, but only by a polynomial factor for the upper bound and a constant factor for the lower bound, requiring more and more advanced methods. The upper bound improvements culminated in the recent work of Conlon who managed to slice down a factor slightly larger than polynomial from the upper bound, though his bound is still way below an exponential improvement. The 70-year-old lower bound of Erdős and the 80-year-old upper bound of Erdős and Szekeres still stand rock solid, noone can show $R(k,k) \geq 1.42^k$ or $R(k,k) \leq 3.99^k$. It is one of the great open problems of combinatorics to prove that $\lim_{k\to\infty} \frac{\log R(k,k)}{k}$ exists and if it does to determine its value.

The lower bound of $\sqrt{2}^k$ obtained by Erdős was using the probabilistic method, and did not give any pointers *how* to construct a good Ramsey graph explicitly, not even with significantly worse parameters. The best *constructive lower bound* for decades was provided by the Turán graph on $(k-1)^2$ vertices.

A notable candidate for good Ramsey-graphs are the *Paley graphs.* The Paley graph P_p is defined on $V(P_p) = \mathbb{F}_p$ for every prime p for that -1 is a quadratic residue modulo p (i.e., $p \equiv 1 \pmod 4$). Vertices x and y are adjacent if $x - y$ is a quadratic residue. Observe that, because of our assumption on p, $x - y$ is a quadratic residue if and only if $y - x$ is a quadratic residue; that is adjacency is well-defined. It is a common beleif that the Paley graphs provide good k-Ramsey graphs — except noone can prove it. In fact, to prove that $\omega(P_p) \leq p^{1/2-\epsilon}$ for some positive ϵ would be a major number theoretic advance. Modulo 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 the Paley graphs *cannot* be expected to provide constructive k-Ramsey graphs on $p = 2^{\frac{k}{c \log k}}$ vertices in general. However, it is also true modulo the GRH that there is a constant C such that the first $C \log p \log\log p$ integers do *not* form a clique. This might be a good indication to believe that the Paley graphs *are* k-Ramsey graphs on $p = 2^{\frac{k}{C \log k}}$ vertices for arbitrary prime p? (It is worth to compare the exponent $\frac{k}{C \log k}$ with the best known probabilistic lower bound where the main term in the exponent is $\frac{k}{2}$ and the constructive lower bound of the Turán graph where the exponent would be $\log_2(k-1)^2 \approx 2 \log k$.)

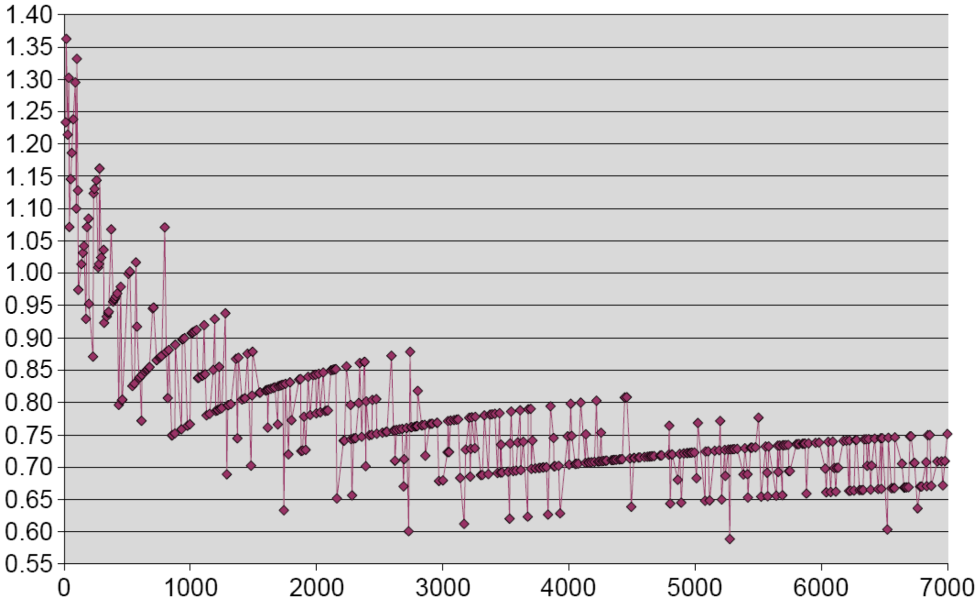

Figure 5.1: The quotient $\frac{\log n(P_p)}{\omega(P_p)}$ in the Payley-graph P_p for the primes $p \leq 7000$

These results show that even though Paley graphs are k-Ramsey graphs of not exponential order in general, there is indication that they are pretty close to that. Not to mention that for sporadic values of p, they could still be reaching exponential order. Figure 5.1 is based on computer calculations made by Shearer about the clique number (and hence independence number) of Paley graphs for primes up to $p < 7000$. One can always observe some irregularly small values, like the clique number of P_{5501} is only 16, it is remarkable to compare this with the upper bound that one can actually prove in general, which is $\lfloor\sqrt{5501}\rfloor = 74$.

Exercise 5.1 *Show that the Paley graph P_p is self-complementary and edge-transitive (that is for each pair of edges xy and $uv \in E(P_p)$ there is a graph automorphism $\phi : V(P_p) \to V(P_p)$ such that $\phi(\{x,y\}) = f(\{u,v\})$.*

Exercise 5.2 *Observe that P_5 provides the construction for $R(3,3) = 6$. Prove that P_{17} does not contain a clique or independent set of order 4 and show that $R(4,4) = 18$.*

Exercise 5.3 *One can define the Paley graph P_q analogously for prime powers q. Show that if q itself is an odd square, then $\omega(P_q) = \sqrt{q}$.*

Knowing the existence of certain combinatorial structures is great, however in theoretical computer science, in particular in questions related to various models of complexity, it is desirable having the the structure in our hand, constructed explicitly. Moreover, as the best known "construction" of a k-Ramsey graph is the random graph $G(n, 1/2)$, good explicit constructions for the Ramsey problem might also be useful in imitating randomness efficiently, another key feature in theoretical computer science. I doubt Erdős had any of these motivations in mind, when in the late 60s he had the the good taste to ask for an explicit construction of k-Ramsey graphs on 1.01^k vertices. Still, as it is the case with many of his beautiful questions, this one also hit something important right on the head; something whose importance turned out only later. In the last section of this chapter we will see that besides the above connections to computer science, the question of explicit constructions had a great influence in motivating extremal hypergraph theory; a completely unexpected development.

5.1.1 The Abbott-product

Answering the challenge of Erdős, in 1972 Abbott gave a curious super-quadratic constructive lower bound. For any integer t, he gave a method to construct an infinite sequence of k-Ramsey graphs on k^t vertices "efficiently". Given two graphs G and H (to simplify the definition assume they contain one loop at each vertex) let us define their 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, one can imagine that we take $v(G)$ 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. One can easily check (please do!) that

$$\begin{aligned} v(G\otimes H) &= v(G)\cdot v(H), \\ \omega(G\otimes H) &= \omega(G)\cdot\omega(H) \text{ and} \\ \alpha(G\otimes H) &= \alpha(G)\cdot\alpha(H) \end{aligned} \tag{5.1}$$

Exercise 5.4 *Prove the properties in* (5.1).

Suppose that we got for birthday a graph G with $n(G) \geq \max\{\omega(G), \alpha(G)\}^{10}$. Then by the multiplicativity of these parameters, for $G \otimes G$ we have a similar inequality:

$$n(G \otimes G) = n(G)^2 \geq \max\{\omega(G), \alpha(G)\}^{20} = \max\{\omega(G \otimes G), \alpha(G \otimes G)\}^{10},$$

The same is true for any Abbott-power of G, which gives us the infinite sequence of explicit Ramsey graphs — provided that we have the graph to start from.

How can we get a hold of just one k-Ramsey graph for *some* k with, say, k^{10} vertices? Well, we know k-Ramsey graphs *do exist* if the number of vertices is not more than $\sqrt{2}^k$. Certainly, at one point $\sqrt{2}^k$ overtakes k^{10}, so let k_0 be the smallest integer such that $\sqrt{2}^{k_0} \geq k_0^{10}$. Check the graphs on k_0^{10} vertices, one of them certainly will be k_0-Ramsey. How long will this take? Nothing... only constant time... Never mind that $k_0 = 144$ so you might have to calculate the clique number and independence number of possibly $2^{\binom{144^{10}}{2}}$ graphs on 144^{10} vertices.

Is this now an "explicit construction"? Apparently Erdős did not think so and was not too content with it. Today, one would disagree with him (not about being non-content). In the age of computer and efficiency, it sounds completely reasonable to call the above an explicit construction: there is a fast (that is, polynomial time) algorithm telling us which vertices are adjacent and which vertices are not, i.e., the graph is constructable in polynomial time. What else would you want to call explicit?

Exercise 5.5 *Prove that the Abbott product is an explicit construction in the "efficient", computer scientific sense. That is, show that for any n you are able to construct the adjacency matrix of a $\sqrt[10]{n}$-Ramsey graph G_n on n vertices, in time polynomial in n. Give a concrete upper bound, bounded by a polynomial in n, on the number of steps this takes.*
Even more, show that given any two vertices i and j from the vertex set $[n]$ of G_n, you can tell whether they are adjacent in time polynomial in just $\log n$. *This question is motivated by the fact that describing i and j only takes* $\log n$ *bits.*

Intuitively it is clear what Erdős didn't like about the Abbott construction: it is "cheating" to look at that many graphs to find our starter. In the 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 the haystack, studies it carefully, and finds the hay eventually (which is BTW not real hay, more like a pseudo-hay with still more features similar to a needle...). On the other hand, one must also not forget that such brute force is used only in a very small (constant size) haystack, which will eventually be negligible compared to the graphs constructed from it.

Before going on to study constructions more to Erdős' liking in the next section, we further explore the Abbott-product in particular to enhance our definition of an explicit construction.

One problem with the above argument in its current form is that it won't give us anything superpolynomial, that is no k-Ramsey graph on $k^{f(k)}$ vertices with $f(k) \to \infty$. Even if we had a starter k_0-Ramsey-graph with $k_0^{\log\log\log k_0}$ vertices, by taking its Abbott-powers we don't get an infinite sequence with the same parameters. The Abbott-product takes away the superpolinomial relation between the order and the clique number: already for the square of the starter we would not have $n \geq \omega^{\log\log\log\omega}$.

How can we get something really superpolynomial? Well, we know that *most* of the graphs on n vertices are incredibly good Ramsey graphs: in other words 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 $G(n, 1/2)$:

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

Then by the union bound

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

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 $\mathcal{D} = \mathcal{D}_n$ of all labeled graphs on n vertices contains a clique of order $2\log_2 n$.

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

$$v\left(\boldsymbol{G}\right) = n^{|\mathcal{D}|},$$

where $|\mathcal{D}| = 2^{\binom{n}{2}}$. By the above one can estimate the clique number of $\boldsymbol{G}$ using (5.1) as follows:

$$\omega\left(\boldsymbol{G}\right) \leq (2\log_2 n)^{(1-\epsilon)|\mathcal{D}|} n^{\epsilon|\mathcal{D}|} < (2\log_2 n)^{|\mathcal{D}|} n^{\epsilon|\mathcal{D}|} = (4\log_2 n)^{|\mathcal{D}|}.$$

Remark. Here we estimated the clique number of $(1-\epsilon)|\mathcal{D}|$ 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 relatively precise since random graph theory tells us that almost all graphs do have clique number at least $\log_2 n$, so $\omega\left(\boldsymbol{G}\right) > (\log_2 n)^{(1-o(1))|\mathcal{D}|}$.

Since the independence number can be estimated analogously by (5.1), $\mathcal{G}$ is an infinite sequence of k-Ramsey graphs with

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

vertices. (Check the calculation!) Moreover $\mathcal{G}$ is clearly an explicit construction, it can be constructed in polynomial time. $\mathcal{G}$ 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 number of vertices $n^{|\mathcal{D}|}$ and the clique number $(4\log n)^{|\mathcal{D}|}$ of $\mathcal{G}$ it becomes apparent that the larger the family $\mathcal{D}$ the more we lose from the Ramsey properties of the majority of its members by the product. So it would be nice if we could guarantee the same calculations, properties with a smaller family. For this we need to look closer what we really did need about the family $\mathcal{D}$ in order to carry out the critical calculations? Well, we needed to know the probability that a particular set of k vertices forms a clique and then just used the union bound. Why did we know that the probability that a particular k-set forms a clique is $2^{-\binom{k}{2}}$? Because when we select a member of $\mathcal{D}$ uniformly at random the appearance of each edge is independent from the appearence of all other edges. The crucial observation is now that we do *not* need the full power of independence of the coordinates in the family $\mathcal{D}$. We use this calculation for $k = 2\log_2 n$ so the independence of any set of $2\log_2^2 n > \binom{k}{2}$ variables is enough to guarantee (5.2). And then, everything else follows.

5.1.2 d-wise independent sample spaces

Let us make the previous wishful thinking more precise.
Definition: A *sample space* $S \subseteq \{0,1\}^N$ is a multiset of vectors endowed with the uniform distribution.
Remark: 1. We rather choose to avoid using the formal notation of a multiset. For example when we talk about the cardinality of a sample space S and write $|S|$, we mean the cardinality as a multiset, where each element is counted with multiplicity.
2. The concept of a multiset with the uniform distribution is a convenient way to approximate a probability space on the *set* of vectors $\{0,1\}^N$ with an *arbitrary* distribution: first we approximate the probabilities of the vectors with rational numbers having a common denominator D and then we take the sample space of cardinality D where each vector has multiplicity of the numerator of its probability.
3. We adopt the usual convention and think of vectors written vertically, i.e., members of the sample space are $N \times 1$-matrices. Then a sample space can be thought of as a $N \times |S_N|$-matrix whose columns are endowed with the uniform distribution.
Definition: A sample space $S \subseteq \{0,1\}^N$ is *independent* if for any $a \in \{0,1\}^N$, we have

$$Pr_{s\in S}[s = a] = \frac{1}{2^N}.$$

Remark: In fact independent sample spaces are pretty boring. The sample space $S = 2^{[N]}$ is independent and all independent sample spaces are essentially of this form: members of $2^{[N]}$ must have the same multiplicity.

The problem with the perfect independence of independent sample spaces is their size 2^N. The following is the key definition of this subsection.

Definition: For a sample space $S \subseteq \{0,1\}^N$ and a subset $J = \{i_1, \ldots, i_d\} \subseteq [N]$ of the coordinates, let

$$S|_J := \{(s_{i_1}, \ldots, s_{i_d}) : s \in S\} \subseteq \{0,1\}^d$$

be the sample space in dimension d with cardinality $|S|_J| = |S|$. The sample space $S|_J$ is called the *projection* of S onto J.

A sample space $S \subseteq \mathbb{F}_2^N$ is called *d-wise independent* if for any $J \subseteq [N]$, $|J| = d$, the projection $S|_J \subseteq \{0,1\}^d$ is independent.

Remark: 1. The d-wise independence of a sample space S_N is equivalent to the (well-established) notion of d-wise independence of the set of N *uniform* random variables obtained from the rows of the matrix whose columns are the elements of S_N.

Exercise 5.6 *Show that d-wise independence of a sample space implies its d'-independence for every $d' \leq d$.*

The following theorem claims 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 odd integer d and $N = 2^t - 1$ with $t \in \mathbb{N}$, we can construct a d-wise independent linear sample space $S \subseteq \{0,1\}^N$ of size $|S| = 2(N+1)^{\frac{d-1}{2}}$.*

Remark: The restriction of d being odd is not significant one. For an even d, one could take the $(d+1)$-independent sample space of size $2(N+1)^{\frac{d}{2}}$ from the theorem and use Exercise 5.6 to conclude its d-wise independence.

Note the word *linear* in the statement. It means that we willingly restrict our search for a d-wise independent sample space to those ones that are closed under addition (and constant multiplication, which, in characteristic 2, is not saying too much). In particular, we focus on finding a *generating set of vectors* whose span possesses the d-wise independence property.

We will use the following simple fact about linear maps: if $L : \mathbb{F}^m \to \mathbb{F}^d$ is a linear map, where $m \geq d$ and $\mathbb{F}$ is a (finite) field, then the number of solutions $x \in \mathbb{F}^m$ to $Lx = a$ is either $|\mathbb{F}|^{m-rank(L)}$ or 0, depending on whether a is in the image of L. In particular, to prove that the number of solution is *the same* for each a, it is enough to check that the linear map L is surjective, that is its matrix of L has rank d. Consequently, if the d rows of the matrix L with entries from $\mathbb{F}_2$ are linearly independent, then the *multi*set

$$S^L := \{Lx : x \in \mathbb{F}_2^m\} \subseteq \{0,1\}^d$$

is an independent sample space of size 2^m in dimension d. Note that S^L can also be written as the sample space generated by the columns $c_i \in \mathbb{F}_2^d$ of L, that is,

$$S^L = \left\{ \sum_{i=1}^{m} x_i c_i : x \in \mathbb{F}_2^m \right\}.$$

Hence we observed the following connection between linear and "probabilistic" independence.

Claim 6 *Let L be a $d \times m$-matrix with 0 or 1 entries, where $d \leq m$. The following are equivalent*

- *the rows of L are linearly independent over $\mathbb{F}_2$*
- *the sample space S^L generated by the columns of L is independent.*

The whole point of the above simple training with basic linear algebra was to formulate the following immediate consequence for d-wise independence.

Corollary 5.2 *The linear sample space $S^L \subseteq \{0,1\}^N$ generated by vectors $c_1, \dots, c_m \in \{0,1\}^N$ is d-wise independent if and only if any d rows of the matrix L with columns $c_1, \dots c_m$ are linearly independent.*

Proof. (of Theorem 5.1) Now how to get the magic matrix expressed in Corollary 5.2? When we hear the condition of Corollary 5.2 that any d rows of a matrix are 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 3.3). We saw there that for any field $\mathbb{F}$ and any $d \leq |\mathbb{F}|$ vectors from the set $M_d = \{(1, \alpha, \alpha^2, \dots, \alpha^{d-1}) : \alpha \in \mathbb{F}\} \subseteq \mathbb{F}^d$ is linearly independent. This gives rise to an $(|\mathbb{F}| \times d)$-matrix with the required property and we could choose $\mathbb{F}$ to be a however large finite field. Hence we would have a linear sample space of size 2^d (independent of the length N) and keep the d-wise independence property. Wow! At the same time this also sounds suspicious, too good to be true ...

Yes, first of all we ignored that for a sample space we need 0/1-vectors and not coordinates from an arbitrary finite field. Let us try to fix this and start with a a bit of wishful thinking. If we could just encode the elements of the finite field as bit-vectors, but still keep the linear independence property ... In principle the elements of, say, $\mathbb{F}_{37}$ can be encoded with bit-vectors of length $\lceil \log_2 37 \rceil = 6$. But then, to keep the linear independence, we would need somehow that when we add the bit-vector of α^i and the bit-vector of $\beta^i \pmod 2$ the result would be the bit-vector of their sum in the field $\mathbb{F}_{37}$. Furthermore linear independence of the vectors in M_{37} is over $\mathbb{F}_{37}$, while the independence of the bit-vectors should be over $\mathbb{F}_2$. So just an arbitrary bit-vector encoding will not do.

That's 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.[1]

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 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 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 this matrix, for notational simplicity we denote them by $r_1, \ldots, r_d$, defined by elements $\alpha_1, \ldots, \alpha_d$. How could a linear combination $x_1 r_1 + \cdots + x_d r_d$ be the zero vector for some $x = (x_1, \ldots, x_d) \in \mathbb{F}_2^d$? 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, because of the columns from $(j-1)t+1$ to jt. Note that here we started to interpret the equations over $\mathbb{F}_{2^d}$, even though we would only be concerned about a nontrivial solution x from $\mathbb{F}_2^d$. But at this point it is not clear how to distinguish them from other solutions in $\mathbb{F}_{2^d}^d$.

[1] The elements of $\mathbb{F}_{2^t}$ are polynomials of degree at most $t-1$ over $\mathbb{F}_2$ factored with an irreducible degree t polynomial. 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 degree 3 polynomial $f(x) = x^3+x+1$; it is irreducible, (please belive me, I checked it...). The members 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 members 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 in fact 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 $(x^2+x)(x+1) = x^4+2x^2+x+1 = x^3 \cdot x + x + 1 = (x+1)x+x+1 = x^2+2x+1 = x^2+1$.

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

$$\begin{array}{ccccccc} x_1 & + & \cdots & + & x_d & = & 0 \\ x_1\alpha_1 & + & \cdots & + & x_d\alpha_d & = & 0 \\ x_1\alpha_1^2 & + & \cdots & + & x_d\alpha_d^2 & = & 0 \\ & \vdots & & \vdots & & \vdots & \\ x_1\alpha_1^d & + & \cdots & + & x_d\alpha_d^d & = & 0 \end{array} \tag{5.4}$$

The matrix of this system is the Vandermonde matrix, which is non-singular, so the unique solution $x \in \mathbb{F}_{2^d}^d$ is the 0-vector: the d rows $r_1, \ldots, r_d$ are linearly independent.

Concluding, we constructed a $N \times (t(d-1)+1)$-matrix A with every d of its rows linearly independent over $\mathbb{F}_2$. Then Corollary 5.2 implies that the linear sample space $S^A \subseteq \{0,1\}^N$ generated by the columns of A is d-wise independent and its size is $2^{(d-1)t+1} = 2N^{d-1}$.

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 for this is our inability so far to use that that the coefficients x_i of the linear combination of the rows are not just arbitrary elements from $\mathbb{F}_{2^d}$, but either 0 or 1. How can we make use of that? Squares of sums in characteristic 2 are very simple to handle, because the mixed terms fall out, so let us consider the square of equation of the first powers of the α_i:

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

We just derived that the equation for the squares of the α_i is a consequence of the equation for the first powers. In the last equality we did use that $x_i = 0$ or 1, because we replaced x_i^2 with x_i.

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.

Hence we can construct a shorter matrix B using only the odd powers as follows. Let $N = 2^t - 1$. The dimensions of our matrix B will be N times $t\ell + 1$, where $\ell < N/2$ is an arbitrary integer, and our $d = 2\ell + 1$.

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. More precisely, labeling the coordinates from 0 up to tl, 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 $x_1 r_1 + \cdots + x_d r_d$ 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 $\sum_{i=1}^d x_i \alpha_j^{2i-1} = 0$, because of the rows from $jt+1$ to $(j+1)t$. These are ℓ equations and $2\ell+1$ variables. We obtain however the equations for the even powers as described above and end with the thesame equation system (5.4) and the same conclusion as above: there is only the trivial $x = 0$ solution. The d rows are independet.

The dimensions of our matrix B is $N \times t\ell + 1$, whose columns generate a d-independent sample space of size $2^{t\ell+1} = 2(N+1)^\ell$.

□

Remark: The matrix constructed above is well-known in classical coding theory: it is essentially the parity check matrix of the famous BCH-codes discovered by Hocquenghem (1959) and independently by Bose and Ray-Chaudhuri (1960). Matrices with our property define linear codes where the weight of each code-word is at least d, and as such these codes correct up to $d/2$ errors.

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 $\mathcal{A}$. If we take the Abbott product of all graphs in $\mathcal{A}$, we have a graph G with $n^{|\mathcal{A}|}$ vertices and clique- and independence number at most $(4\log_2 n)^{|\mathcal{A}|}$. 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 5.7 *Verify the calculation.*

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

Can we get even better? We will further reduce the size of our sample space significantly by being content with providing $2\log_2^2 n$-wise independence only *approximately.*

5.1.3 Almost independent sample spaces

We relax the requirement of independent sample spaces and *not* require any longer that each bit-vector appears with the same probability, but only that each appears 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 $a \in \{0,1\}^N$, we have

$$|Pr_{s\in S}(s = a) - 2^{-N}| \leq \epsilon.$$

Note that being 0-close to independent is equivalent to being independent.

We remark that in our applications the ϵ will be chosen to be much smaller than 2^{-N}, so the definition of ϵ-closeness does not become meaningless.

As we commented earlier, in order to get the full power of independence one needs at least 2^N vectors in the sample space. First we will see a construction of Alon, Goldreich, Hastad, and Peralta, which shows how to explicitly construct a sample space of size of only $\frac{N^2}{\epsilon^2}$ provided we are content with our sample space being only ϵ-close to independent.

Then we will combine this construction with our construction of the previous section of a d-wise independent linear sample space of size $2(N+1)^{(d-1)/2}$ and obtain a sample space of even smaller size (logarithmic in N), which is only ϵ-close to being d-wise independent.

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 (d-dimensional) projection sample space $S|_J$ is ϵ-close to independent.

Note that being 0-close to d-wise independent is equivalent to being d-wise independent.

In the main result of this section we will prove the following theorem, which constructs sample spaces whose size is only *logarithmic in N* and polynomial in their imperfectness measurements, i.e., in d and $\frac{1}{\epsilon}$.

Theorem 5.3 (Naor and Naor) *Let $N = 2^t - 1$ 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*

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

which is ϵ-close to d-wise independent.

Proof. The proof of this theorem will be carried out in three subsections.

Linear tests

The concept of being ϵ-close to independent is somewhat inconvenient/tedious to check, hence we develop a more effective way to test it.

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

$$|Pr_{s\in S}[s \cdot a = 0] - Pr_{s\in S}[s \cdot a = 1]| \leq \epsilon.$$

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 $\mathbb{F}_2$. 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.

Exercise 5.8 *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 is a sort of converse of the previous exercise and establishes the usefulness of linear tests in proving ϵ-closeness to independence.

Lemma 5.3.1 (Vazirani) *Let $S \subseteq \{0,1\}^d$ be a sample space that is ϵ-unbiased with respect to linear tests. Then S is ϵ-close to independent.*

Proof. (Alon, Goldreich, Hastad, and Peralta) We make use of the basic properties of the discrete Fourier transform on the group $\langle H, +\rangle = \langle \mathbb{F}_2^d, +\rangle$, contained in the Toolbox. Let us fix a vector $a \in \{0,1\}^d$ and let $p(a) = \Pr[s = a]$ be the probability in question. We need to show that the function $p : H \to \mathbb{C}^*$ does not deviate much from its average $1/2^d$.

The key observation is that the probability difference when we make a linear test with some test vector $b \in H$ is precisely the Fourier transform of p evaluated at the character corresponding to b. Hence by our assumption all, but the leading, Fourier coefficients of p are known to to be small (at most $\epsilon/2^d$) and this will imply that p is uniformly random looking. The leading Fourier coefficient corresponds to the vector $b = 0^d$ and hence it is $1/2^d$.

Formally, let $b \in H$. Then

$$\begin{aligned}
\Pr[s \cdot b = 0] - \Pr[c \cdot b = 1] &= \sum_{\substack{a \in H \\ a \cdot b = 0}} \Pr[s = a] - \sum_{\substack{a \in H \\ a \cdot b = 1}} \Pr[s = a] \\
&= \sum_{a \in H} (-1)^{a \cdot b} p(a) = \sum_{a \in H} \chi_b(a) p(a) = \hat{p}(\chi_b),
\end{aligned}$$

where χ_b is the character of H defined by $\chi_b(a) = (-1)^{b \cdot a}$. Using the formula of the Inverse Fourier Transform to express $p(a)$, we obtain an estimate on how much can $p(a)$ deviate from its average.

$$\begin{aligned}
\left| p(a) - \frac{1}{2^d} \right| &= \frac{1}{2^d} \left| \sum_{b \in H} \hat{p}(\overline{\chi}_b) \chi_b(a) - 1 \right| \\
&\leq \frac{1}{2^d} \left(|\hat{p}(\overline{\chi}_0)|\, |\chi_0(a)| + \sum_{\substack{b \in H \\ b \neq 0}} |\hat{p}(\overline{\chi}_b)|\, |\chi_b(a)| - 1 \right) \\
&\leq \frac{1}{2^d} \left(1 + \sum_{\substack{b \in H \\ b \neq 0}} \epsilon - 1 \right) = \frac{2^d - 1}{2^d} \epsilon,
\end{aligned}$$

and the lemma is proved. □

The previous lemma is also useful to test ϵ-closeness to d-wise independence.

Definition: The sample space $S \subseteq \{0,1\}^N$ is called *ϵ-unbiased with respect to linear tests of size at most d* if for every $a \in \{0,1\}^N \setminus \{0^N\}$, $0 < \sum_i a_i \leq d$, we have

$$|Pr_{s\in S}[s \cdot a = 0] - Pr_{s\in S}[s \cdot a = 1]| \leq \epsilon.$$

The following corollary is immediate from the definitions and applying Lemma 5.3.1 for the d-dimensional projections.

Corollary 5.4 *Let $S \subseteq \{0,1\}^N$ be a sample space that is ϵ-unbiased with respect to linear tests of size at most d. Then S is ϵ-close to d-wise independent.*

Almost independent sample spaces via the quadratic character

In this section we use the intuition that quadratic residues form a random subset within the additive structure of the finite field $\mathbb{F}_p$, where p is a prime number. This is the same heuristics why we think the Paley-graphs have relatively good Ramsey-properties. We will construct p bit vectors of length m (where m will be at most $\sqrt{p}$). For each $x \in \mathbb{F}_p$ we consider the integers $x+1$, $x+2$, etc ...., $x+m$ and the characteristic vector $r^{(x)}$ of them being not a quadratic residue.

Formally, we define the sample space $B^p_m := \{r^{(x)} : x \in \mathbb{F}_p\} \subseteq \{0,1\}^m$ of size p, where

$$r_i^{(x)} = \begin{cases} 0 & \text{if } x+i \in QR(p) \\ 1 & \text{if } x+i \in QNR(p) \text{ or } = 0 \end{cases}$$

Proposition 5.5 *(Alon, Goldreich, Hastad, and Peralta) For every $m \leq \sqrt{p}$, the sample space B^p_m is $\frac{m}{\sqrt{p}}$-unbiased respect to linear tests.*

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

Proof. The key is the use of the theorem of Weil (Theorem A.37), which states that the values of non-principal characters behave randomly in some sense. We are able to apply this powerful tool, because of the explicit connection between the vectors of the sample space and the quadratic character ϱ_p of $\mathbb{F}_p^*$. That is, we have $(-1)^{r_i^{(x)}} = \rho_p(x+i)$ for every $x \in \mathbb{F}_p$ and $i \in [m]$ with $x+i \neq 0$.

Let us fix our "linear tester" $a \in \{0,1\}^m$. As we saw in the proof of Lemma 5.3.1, the probability difference in question can be expressed as follows.

$$\begin{aligned}\Pr\left[r^{(x)} \cdot a = 0\right] - \Pr\left[r^{(x)} \cdot a = 1\right] &= \sum_{\substack{b\in\mathbb{F}_p \\ r^{(b)}\cdot a=0}} Pr_{x\in\mathbb{F}_p}(x=b) - \sum_{\substack{b\in\mathbb{F}_p \\ r^{(b)}\cdot a=1}} Pr_{x\in\mathbb{F}_p}(x=b) \\ &= \frac{1}{p}\sum_{b\in\mathbb{F}_p}(-1)^{r^{(b)}\cdot a} = \frac{1}{p}\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(-1)^{r_i^{(b)}a_i}\end{aligned}$$

We want to replace $\prod_{i=1}^m (-1)^{r_i^{(b)} a_i}$ 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. Since for the sake of Weil's Theorem ϱ_p is extended to the whole $\mathbb{F}_p$ by defining $\varrho_p(0) := 0$, we can make the replacement only if $b+i \neq 0$ or, if $b+i=0$ but the corresponding power $a_i = a_{p-b}$ is 0. Otherwise, when $\prod_{i=1}^m (\varrho_p(b+i))^{a_i} = 0$, that is, when $b \in [p-m, p-1]$ and $a_{p-b} = 1$, then we estimate the absolute value of the terms trivially, by 1. Note that since $m \ll p$, for most of the b this does not happen.

$$
\begin{aligned}
&|\Pr\left[r^{(x)} \cdot a = 0\right] - \Pr\left[r^{(x)} \cdot a = 1\right]| \leq \\
&\qquad \leq \frac{1}{p}\left|\sum_{b\in\mathbb{F}_p}\prod_{i=1}^{m}(\varrho_p(b+i))^{a_i} + \sum_{\substack{b\in[p-m,p-1]\\ a_{p-b}=1}}\left((-1)^{r^{(x)}\cdot a} - \prod_{i=1}^{m}(\varrho_p(b+i))^{a_i}\right)\right| \\
&\qquad \leq \frac{1}{p}\left|\sum_{b\in\mathbb{F}_p}\varrho_p\left(\prod_{i=1}^{m}(b+i)^{a_i}\right)\right| + \frac{n}{p} \\
&\qquad \leq \frac{n-1}{\sqrt{p}} + \frac{m}{p} \leq \frac{m}{\sqrt{p}}.
\end{aligned}
$$

At the end 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. □

Lemma 5.5.1 (Naor and Naor) *Suppose that for some integers N, d, m, p, we can construct a d-wise independent linear sample space $L \subseteq \{0,1\}^N$ of size 2^m and a sample space $S \subseteq \{0,1\}^m$ which is ϵ-unbiased with respect to linear tests and has size p. Then we can construct a sample space $R \subseteq \{0,1\}^N$ of size p which is ϵ-close to d-wise independent.*

Proof. Let $b_1, \ldots, b_m \in \mathbb{F}_2^N$ be the basis which generates the linear sample space L. We denote by B the $m \times N$ matrix whose rows are the b_i. In terms of d-wise independence L is perfect, the problem is its size 2^m, which we would like to reduce. The idea is that instead of putting *all* linear combinations of the basis into the sample space we only put a well-selected subset of the linear combinations: those ones whose coefficients come from the sample space S, which is ϵ-unbiased with respect to linear tests. This way we get a sample space of size only p and as we will see the d-wise independence properties carry through.

Formally, let S be the multiset of the (column) vectors $r_1, \ldots, r_p \in \mathbb{F}_2^m$ and define $R = \{r_i^T B : i = 1, \ldots, p\}$ to be the multiset of the p linear combinations of the basis vectors $b_1, \ldots, b_m \in \mathbb{F}_2^N$. We will prove that R is ϵ-unbiased with respect to linear tests of size at most d and then Corollary 5.4 implies that it is also ϵ-close to d-wise independent. Let us fix a j-element subset $J \subseteq [N]$ of the coordinates for some $j \leq d$. We need to check what is the probability of 0 and 1 in the sample space $\{r_i^T B \cdot 1_J : i = 1, \dots, p\}$, where $1_J \in \mathbb{F}_2^N$ denotes the characteristic vector of the set J. Since of course

$$(r_i^T B) \cdot 1_J = r_i \cdot (B 1_J),$$

we have that $Pr_{i \in [p]}[r_i \cdot (B1_J) = \alpha] = Pr_{i \in [p]}[(r_i^T B) \cdot 1_J = \alpha]$ for $\alpha = 0, 1$. Applying that S is ϵ-unbiased with respect to linear tests, with the test vector $B1_J \in \mathbb{F}_2^m$, we conclude that

$$|\Pr\left[(r_i^T B) \cdot 1_J = 0\right] - \Pr\left[(r_i^T B) \cdot 1_J = 1\right]| = |\Pr[r_i \cdot (B1_J) = 0] - \Pr[r_i \cdot (B1_J) = 1]| < \epsilon.$$

□

Now we can finish the proof of Theorem 5.3 easily by using Lemma 5.5.1 with the almost independent independent sample spaces of Proposition 5.5 and the d-wise independent sample spaces of Theorem 5.1

First construct a d-wise independent linear sample space $L \subseteq \{0,1\}^N$ of size $2^{t^{\frac{d-1}{2}}+1}$. Then, after choosing a prime p between $\frac{\left(t^{\frac{d-1}{2}}+1\right)^2}{\epsilon^2}$ and its double, construct the sample space $S = B_m^p \subseteq \{0,1\}^p$ with $m = t^{\frac{d-1}{2}} + 1$, which is ϵ-close to independent by Proposition 5.5. Now Lemma 5.5.1 concludes the proof. □

Better Ramsey-graphs

Let us now try to use our sample spaces from Theorem 5.3 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 5.1.2 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 5.3 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,

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

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 5.5). By (5.1), 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 5.1.1 or when we took the Abbott-product of all graphs from the d-wise independent sample space in Subsection 5.1.2 we were not only constructing the adjacency matrix of the graph in time polynomial in n, but were able to answer a query quickly requesting the adjacency relation of two particular vertices. The query containing the labels of the two vertices in question has only $2\log n$ bits, so one would possibly want to have the answer in time polynomial in $\log n$. This is possible in those constructions as the Abbott-product is efficient in this sense (see Exercise 5.5).

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 $\log n$. This explains the following definition. A construction of a graph on n vertices is called *strongly explicit* if adjacency queries can be answered in time polynomial in $\log n$. A construction of a graph on n vertices is called *weakly explicit* if the adjacency matrix of the graph can be constructed in time polynomial in n.

One could suspect that the "definition" or rather "feeling" of explicit construction a'la Erdős would be closer to the definition of the strongly explicit one above. However there is an important "philosophical" distinction. 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 created. Erdős refused to pay his award to Peter Frankl, who came up to him with the Abbott product construction. His refusal was not based on a mathematically founded argument, rather by a philosophically motivated one. "I don't know what a construction is, but I will know when I see one and this is not it" he might have said. The motivation behind his original question was rather the desire to see disorder in an understandable fashion. Erdős would not care about polynomial computability of the adjacency relation; a computer can calculate many things where the human mind is not able to see anything. On the other hand, he would also not worry about the adjacency relation in the Paley graph being really computable in polylogarithmic time, before proclaiming the Paley graph a "construction". The Paley graph is not an explicit construction because of efficient computability, it is explicit because one looks at it and sees mathematically explainable disorder (should number theorists finally be able to prove that so).

The best strongly explicit construction by the Abbott-product (from Subsection 5.1.2) has a twice iterated logarithm in the exponent. In the next section we discuss a surprisingly simple strongly explicit construction, which beats slightly even the weakly explicit Abbott-type construction above.

The following exercise is good preparation for that. It was the first real breakthrough over the quadratic constructive lower bound of the Turán graph and it came in the same year (1972) as the Abbott-product. Nagy defined an infinite sequence of k-Ramsey graphs on $\Theta(k^3)$ vertices. Let G be the graph with $V(G) = \binom{[k]}{3}$, and $A \sim B$ if $|A \cap B| = 1$. The proof of correctness of the construction, i.e. that they don't contain large clique and independent set, is a beautiful application of the Linear Algebra Method.

Exercise 5.9 *Prove that the graph of Nagy contains no clique and no independent set of order $k+1$. (Hint for a proof via linear algebra: Prove that set of characteristic vectors of an independent set (or a clique) is linearly independent over an appropriately chosen field. Hint for a combinatorial proof: there is one.)*

5.2 The construction of Frankl and Wilson

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 5.10 *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 $A_1, \dots, A_s$ 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 $i = 1, \dots s$, 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 $\{A_1, \dots, A_s\}$ is an independent set. Our technical lemma then implies that $\tilde{f}_1, \dots, \tilde{f}_s$ are linearly independent. The dimension of the space these functions span could be quite large, since each variable x_j, $j = 1, \dots, p^3$ 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 $p-1$,

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

Cliques. To bound the clique number of G we proceed similarly, but we will work over $\mathbb{R}$ instead of $\mathbb{F}_p$. Let $B_1, \dots, B_t$ be a clique in G, so $|B_i \cap B_j| \equiv -1 \pmod p$ for every $i \neq j$. Let $L = \{p-1, 2p-1, \dots, p^2-p-1\}$ 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 $\tilde{f}_1, \dots, \tilde{f}_t$ 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

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

□

Exercise 5.11 *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 We formulate the Ramsey problem for bipartite graphs.

$$BR(k,l) = \min\{n : \forall \text{ subgraph of } K_{n,n} \text{ contains either } K_{k,k} \text{ or } \overline{K_{l,l}}\}.$$

Here containment of bipartite graphs is understood the natural way, respecting the specification of the two parts.

Again the most interesting case is the symmetric, when $k = l$. The story of bipartite Ramsey numbers is very similar to ordinary in the sense that we *know* that $BR(k,k)$ is exponential by the probabilistic method. Constructively the situation is very different though: until very recently there was no super-quadratic constructive lower bound. Even the product construction of Abbott does not have an obvious counterpart in the bipartite world. There are several construction yielding a quadratic lower bound, like the ones based on Hadamard matrices. In some sense we know even more: the norm graphs on n vertices do not contain $K_{t,t!+1}$ and one can prove that their 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 bi-clique, we cannot go below $\sqrt{n}$ by *any* infinite factor for *both* the bi-clique and the bi-independent set. Very recently B. Barak, G. Kindler, R. Shaltiel, B. Sudakov and A. Wigderson (Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers and Extractors, *Proc. of the 37-th ACM STOC* (2005), 1-10.) broke through the constructive quadratic barrier and exhibited a constructive lower bound of k^t on $BR(k,k)$ for arbitrary t.
Even more recently even the Frankl-Wilson barrier for bipartite graphs was broken, but only with a weakly explicit construction. (B. Barak, A. Rao, R. Shaltiel, and A. Wigderson (2-source dispersers for sub-polynomial ebtropy and Ramsey graphs beating the Frankl-Wilson construction, *Proc. of the 38-th ACM STOC* (2006), to appear)

A.5 Basic Properties of Characters

Let H be a finite abelian group. For the sake of the this exposition we mostly write the group operation additively (denoted by $+$), however later we will also use characters of multiplicative groups and even mix the two.

The homomorphisms of $(H,+)$ into the multiplicative group $(\mathbb{C}^*,\cdot)$ of the complex numbers are called *characters* of H. Formally, $\chi : H \to \mathbb{C}^*$ is a *character* of H if

$$\chi(a+b) = \chi(a)\chi(b) \text{ for every } a, b \in H.$$

The *principal character* χ_0 is defined by

$$\chi_0(a) = 1, \text{ for every } a \in H,$$

and exists for arbitrary group H.

Another important example is the *quadratic residue character* ρ_q of the multiplicative group $(\mathbb{F}_q^*,\cdot)$ of a finite field: $\rho_q(x) = 1$ whenever $x \in \mathbb{F}_q^*$ is a quadratic residue and $\rho_q(x) = -1$ otherwise. The map ρ_q is a homomorphism because as we saw earlier in Appendix A.2, a square times a square or a non-square times a non-square is a square, while a square times a non-square is a non-square.

The fact that the quadratic residue character has only values 1 and -1 is not an accident: all character values must be some root of unity.

Exercise A.2 *Prove that*

- $\chi(a)$ *is a* $|H|^{th}$ *root of unity.*
- $\chi(-a) = \chi(a)^{-1} = \overline{\chi(a)}$

The values of any non-principal character sum up to 0.

Proposition A.26 *For any character* $\chi \neq \chi_0$*,*

$$\sum_{a \in H} \chi(a) = 0.$$

Proof. Let $b \in H$ be such that $\chi(b) \neq 1$; such an element b exists since χ is not principal. Then, using that $a \to a+b$ is a bijection from H to H, we have that

$$\sum_{a \in H} \chi(a) = \sum_{a \in H} \chi(a+b) = (\sum_{a \in H} \chi(a))\chi(b).$$

Then the claim follows. □

Let $\hat{H}$ be the set of characters. It will turn out that H has exactly $|H|$ characters. Even more, there is a natural group structure on $\hat{H}$ and the two groups are isomorphic.

Proposition A.27 *$\hat{H}$ is an abelian group with the operation $\cdot$, defined by*

$$(\chi \cdot \psi)(a) := \chi(a)\psi(a).$$

Proof. Exercise. □

The group H and its group of characters are isomorphic.

Theorem A.28 *$H \cong \hat{H}$.*

Proof. We establish the proof in two steps. First we explicitely give the characters of the cyclic group $(\mathbb{Z}_n, +)$.

Proposition A.29 *Let ω be an arbitrary primitive n^{th} root of unity (i.e. $\omega^i = 1$ if and only if $n|i$) and define the map $\chi_j : \mathbb{Z}_n \to \mathbb{C}^*$ by $\chi_j(a) := \omega^{ja}$. Then*

- *χ_j is a character for every $j \in \mathbb{Z}_n$.*
- *the mapping sending $j \in \mathbb{Z}_n$ to $\chi_j \in \hat{\mathbb{Z}}_n$ is an isomorphism between $\mathbb{Z}_n$ and $\hat{\mathbb{Z}}_n$.*

Proof. The first statement follows easily from the definition: $\chi_j(a+b) = \omega^{j(a+b)} = \omega^{ja}\omega^{jb} = \chi_j(a)\chi_j(b)$.
For the second statement let us see first that the mapping is a homomorphism from $(\mathbb{Z}_n, +)$ to $(\hat{\mathbb{Z}}_n, \cdot)$. Indeed, $j + \ell \in \mathbb{Z}_n$ is mapped to $\chi_{j+\ell} = \chi_j \cdot \chi_\ell$. The mapping is injective, since $\chi_j(1) = \chi_\ell(1)$ would mean that $\omega^{j-\ell} = 1$ and since ω is primitive, we have n dividing $j - \ell$, so $j = \ell$. Let us see finally that the mapping is surjective. Let χ be an arbitrary character of $(\mathbb{Z}_n, +)$. Since $\chi(1)$ is an nth root of unity by Exercise ... and ω is primitive, there is a j, such that $\chi(1) = \omega^j$. Then, since χ is a character, $\chi(a) = \chi(1 + \cdots + 1) = \chi(1)^a = \omega^{ja} = \chi_j(a)$ for every $a \in \mathbb{Z}_n$, so χ is identical to χ_j. □

Secondly we show how to obtain the characters of a direct sum from the characters of its summands.

Proposition A.30 *If $H = H_1 \times H_2$, then $\hat{H} \cong \hat{H}_1 \times \hat{H}_2$*

Proof. Exercise □

To conclude the proof of Theorem A.28 note that any finite abelian group is the direct product of cyclic groups, hence by the previous two proposition

$$H \cong \mathbb{Z}_{s_1} \times \cdots \times \mathbb{Z}_{s_r} \cong \hat{\mathbb{Z}}_{s_1} \times \cdots \times \hat{\mathbb{Z}}_{s_r} \cong \hat{H}.$$

□

Example Let $H = \overbrace{\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2}^{k}$. Then $\hat{H} = \{\chi_w : w \in \{0,1\}^k\}$, where $\chi_w(a) = (-1)^{w \cdot a}$ and $w \cdot a = \sum_{i=1}^{k} w_i a_i$ is the usual scalar product of vectors.

Inner product and orthonormal basis

$\mathbb{C}^H := \{f : H \to \mathbb{C}\}$ is an n-dimensional linear space over $\mathbb{C}$. We define an inner product on $\mathbb{C}^H$:

$$\langle f, g \rangle = \frac{1}{n} \sum_{a \in H} \overline{f(a)} g(a).$$

Corollary A.31 *(First orthogonality relation) For any $\chi, \psi \in \hat{H}$,*

$$\langle \chi, \psi \rangle = \begin{cases} 1 & \textit{if } \chi = \psi \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. Exercise. □

Corollary A.32 *$\hat{H}$ forms an orthonormal basis in $\mathbb{C}^H$.*

Discrete Fourier transform

Corollary A.33 *Every $f \in \mathbb{C}^H$ can be written uniquely as the linear combination of characters:*

$$f = \sum_{\chi \in \hat{H}} c_\chi \chi,$$

where $c_\psi = \langle \psi, f \rangle$ are called the Fourier coefficients *of f.*

Proof. By the previous Corollary the characters form an orthonormal basis in $\mathbb{C}^H$, so we can express f uniquely as their linear combination $f = \sum c_\chi \chi$ with $c_\chi \in \mathbb{C}$. Taking the inner product of both sides with any fixed character ψ from the left, we see by the first orthogonality relation that all terms cancel except $\langle \psi, f \rangle$ and c_ψ. □

Definition: The *Fourier transform* of $f : H \to \mathbb{C}$ is a function $\hat{f} : \hat{H} \to \mathbb{C}$, defined by

$$\hat{f}(\chi) := nc_{\overline{\chi}} = \sum_{a \in G} \chi(a) f(a).$$

The following formula of the *Inverse Fourier transform:*

$$f = \sum_{\chi \in \hat{H}} c_\chi \chi = \sum_{\chi \in \hat{H}} \frac{1}{n} \hat{f}(\overline{\chi}) \chi.$$

Quasi-randomness of Cayley-graphs

For a subset $S \subseteq H$ let us define

$$\Phi(S) = \max\{|H| \widehat{\mathbb{1}}_S(\chi) : \chi \in \hat{H}, \chi \neq \chi_0\}.$$

Just to have an idea about how large $\Phi(S)$ is let us calculate an upper bound (why is it that??): $|H| \widehat{\mathbb{1}}_S(\chi_0) = |H| \frac{1}{|H|} \sum_{s \in S} \chi_0(s) = |S|$. For a lower bound see the following small Claim

Claim 8

$$\Phi(S) \geq \sqrt{|S|}2,$$

provided $|S| \leq \frac{n}{2}$.

Let now $S \subseteq H$ be a subset such that $S = -S$. The Cayley graph $G = G(H, S)$ is defined on the vertex set $V(G) = H$. Two vertices $u, v \in V$ are adjacent if $v - u \in S$. In other words, the neighborhood of each vertex $w \in H$ is the set $w + S$ and thus the Cayley graph is d-regular with $d = |S|$.

Exercise A.3 *Give a proof of the following on the language of characters:*
Let $\langle H, + \rangle$ *be an abelian group and* S *be a subset, such that* $S = -S$. *Let* G *be the corresponding Cayley graph. For any subsets* $B, C \subseteq V(G)$,

$$\left| e(B,C) - |B||C|\frac{|S|}{|H|} \right| \leq \Phi(S)\sqrt{|B||C|}.$$

Solution:

The following theorem shows that the closer $\Phi(S)$ is to the lower bound of the Claim the stronger pseudorandom properties the corresponding Cayley graph exhibits.

Theorem A.34 *For any subsets* $B, C \subseteq V(G(S))$,

$$\left| e(B,C) - |B||C|\frac{|S|}{|H|} \right| \leq \Phi(S)\sqrt{|B||C|},$$

where $e(B,C)$ *denotes the number of ordered pairs* $(u,v) \in B \times C$, *such that* $uv \in E(G(S))$.

Proof.

$$\begin{aligned}
e(B,C) &= \sum_{u\in B}\sum_{v\in C}\sum_{s\in S} \mathbb{1}_{\{0\}}(u+s-v) \\
&= \sum_{u\in B}\sum_{v\in C}\sum_{s\in S}\sum_{\chi\in\widehat{H}} \widehat{\mathbb{1}}_{\{0\}}(\chi)\chi(u+s-v) \\
&= \sum_{\chi\in\widehat{H}}\sum_{u\in B}\sum_{v\in C}\sum_{s\in S} \frac{1}{|H|}\chi(u)\chi(s)\chi(-v) \\
&= \sum_{\chi\in\widehat{H}} \frac{1}{|H|}(\sum_{u\in B}\chi(u))(\sum_{s\in S}\chi(s))(\sum_{z\in -C}\chi(z)) \\
&= \frac{|B||C||S|}{|H|} + \sum_{\chi\neq\chi_0}\frac{1}{|H|}(\sum_{u\in B}\chi(u))(|H|\widehat{\mathbb{1}}_S(\chi))(\sum_{z\in -C}\chi(z))
\end{aligned}$$

On the one hand $|(|H|\widehat{\mathbb{1}}_S(\chi))| \leq \Psi(S)$.

On the other hand by the Cauchy-Schwartz-inequality

$$\begin{aligned}
\left|\sum_{\chi\neq\chi_0}(\sum_{u\in B}\chi(u))(\sum_{z\in -C}\chi(z))\right| &\leq \sum_{\chi\neq\chi_0}\left|\sum_{u\in B}\chi(u)\right|\left|\sum_{z\in -C}\chi(z)\right| \\
&\leq \sum_{\chi\in\widehat{H}}\left|\sum_{u\in B}\chi(u)\right|\left|\sum_{z\in -C}\chi(z)\right| \\
&\leq \sqrt{\sum_{\chi\in\widehat{H}}\left(\sum_{u\in B}\chi(u)\right)^2}\sqrt{\sum_{\chi\in\widehat{H}}\left(\sum_{z\in -C}\chi(z)\right)^2} \\
&\leq \sqrt{\sum_{\chi\in\widehat{H}}\left(|H|\widehat{\mathbb{1}}_B(\chi)\right)^2}\sqrt{\sum_{\chi\in\widehat{H}}\left(|H|\widehat{\mathbb{1}}_{-C}(\chi)\right)^2} \\
&\leq |H|^2\sqrt{\langle\mathbb{1}_B,\mathbb{1}_B\rangle}\sqrt{\langle\mathbb{1}_{-C},\mathbb{1}_{-C}\rangle} \\
&\leq |H|^2\sqrt{\frac{|B|}{|H|}}\sqrt{\frac{|-C|}{|H|}} \\
&\leq |H|\sqrt{|B|}\sqrt{|C|}
\end{aligned}$$

and the theorem follows. □

The following is an easy corollary.

Corollary A.35 *Let $G = G(H, S)$ be a Cayley graph. Then*

$$\alpha(G) \leq \frac{\Phi(S)|H|}{|S|}.$$

Proof. Let I be an independent set of maximum size, that is $|I| = \alpha(G)$. By Theorem A.34 we have that

$$\left|e(I,I) - |I|^2\frac{|S|}{|H|}\right| \leq \Phi(S)|I|.$$

Since $e(I,I) = 0$, we have $|I|^2\frac{|S|}{|H|} \leq \Phi(S)|I|$, which implies the statement. □

The following simple proposition shows that in fact we already proved Theorem A.34 and Corollary A.35 in the previous section.

Proposition A.36 *The spectrum of the Cayley graph $G(H, S)$ is the n-element multiset $\{\sum_{s\in S}\chi(s) : \chi \in \widehat{H}\} = \{|H|\widehat{\mathbb{1}}_S(\chi) : \chi \in \widehat{H}\}$. The eigenvectors are the n characters. In particular, the eigenvectors do not depend on S.*

Proof.

$$(A\chi)_v = \sum_{w\in G w-v\in S}\chi(w) = \sum_{s\in S}\chi(v+s) = \left(\sum_{s\in S}\chi(s)\right)\chi(v).$$

Hence χ is indeed an eigenvector with eigenvalue $\sum_{s\in S}\chi(s)$ □

Character sum estimates

The following famous theorem of Weil states that the values of a polynomial substituted into a non-principal character behave uniformly (in some weak sense) .

Theorem A.37 (Weil) *Let q be a prime power and let χ be a multiplicative character of $\mathbb{F}_q^*$ of order d, extended to $\mathbb{F}_q$ by $\chi(0) = 0$. Then for any polynomial $f(x) \in \mathbb{F}_q[x]$ which has precisely m distinct zeros and is not a dth power (over the algebraic closure) we have*

$$\left|\sum_{x\in\mathbb{F}_q} \chi(f(x))\right| \le (m-1)\sqrt{q}.$$

Note that Proposition A.26 is a special case of Weil's theorem for $f(x) = x$.

In light of how hard it is to *estimate* the sum of characters (Weil's theorems about various character sums are highly non-trivial), it is refreshing to see the simple proof of the following *precise formula* involving the additive *and* multiplicative charecters of a finite field together.

Theorem A.38 *(Gaussian sums) Let $\mathbb{F}$ be a finite field and let χ be a character of the additive group of $\mathbb{F}$, while let ψ be a character of the multiplicative group of $\mathbb{F}$. Then*

$$\left|\sum_{C\in\mathbb{F} C\neq 0} \chi(C)\psi(C)\right| = \begin{cases} |\mathbb{F}|-1 & \text{if } \chi=\chi_0 \text{ and } \psi=\psi_0 \\ 0 & \text{if } \chi=\chi_0 \text{ and } \psi\neq\psi_0 \\ 1 & \text{if } \chi\neq\chi_0 \text{ and } \psi=\psi_0 \\ \sqrt{|\mathbb{F}|} & \text{if } \chi\neq\chi_0 \text{ and } \psi\neq\psi_0, \end{cases}$$

where χ_0 is the pricipal additive character and ψ_0 is the prinicipal multiplicatice character.

Proof. In fact the whole proof is just applying Proposition A.26 over and over again; the first three cases being quite straightforward. To appply Proposition A.26 for the fourth case, we need a couple of simple manipulations.

$$\begin{aligned} \left|\sum_{C\neq 0}\chi(C)\psi(C)\right|^2 &= \left(\sum_{C\neq 0}\chi(C)\psi(C)\right)\overline{\left(\sum_{C\neq 0}\chi(C)\psi(C)\right)} \\ &= \sum_{C\neq 0}\sum_{D\neq C,0}\chi(C)\psi(C)\overline{\chi(C)\psi(C)} + \sum_{C\neq 0}\chi(C)\psi(C)\overline{\chi(C)\psi(C)} \\ &= \sum_{C\neq 0}\sum_{D\neq C,0}\chi(C-D)\psi\left(\frac{C}{D}\right) + \sum_{C\neq 0}|\chi(C)|^2\,|\psi(C)|^2 \end{aligned}$$

Each character value is a root of unity, thus its norm is 1 implying that the second term consits of sum of 1s and thus equal to $|\mathbb{F}|-1$. To manipulate the first term we change variables.

$$\begin{aligned}
\sum_{C\neq 0}\sum_{D\neq C,0}\chi(C-D)\psi\left(\frac{C}{D}\right) &= \sum_{W\neq 0,1}\sum_{D\neq 0}\chi(D(W-1))\psi(W) \\
&= \sum_{W\neq 0,1}(-1)\cdot\psi(W) \\
&= 1
\end{aligned}$$

The next to last ineaquality follows from Proposition A.26 since for a fixed $W \neq 1$ the values $D(W-1)$ run through the nonzero elements of $\mathbb{F}$, while D runs through the nonzero elements of $\mathbb{F}$. The last inequality also follows from Proposition A.26; this time employed for the multiplicative character ψ.

□