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

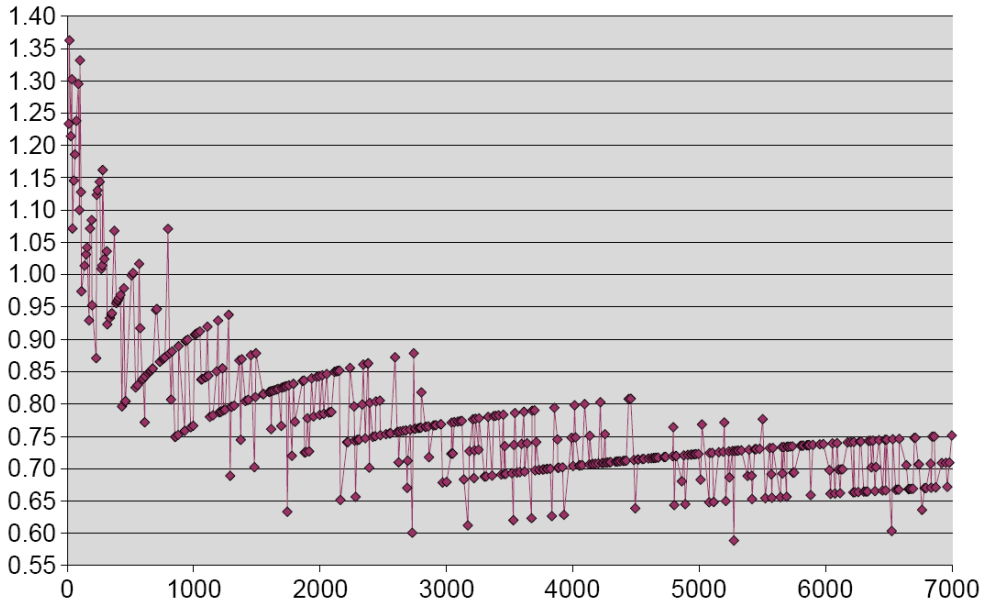

Figure 5.1: The quotient $\frac{\log n(P_p)}{\omega(P_p)}$ in the Payley-graph P_p for the primes $p \leq 7000$

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 $\log_2(k-1)^2 \approx 2 \log k$.)

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

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

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

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

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 $\omega\left(\mathbb{G}\right) > (\log_2 n)^{(1-o(1))|\mathcal{D}|}$.

Since the independence number can be estimated analogously by (5.1), 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 G is clearly an explicit construction, it 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 number of vertices $n^{|\mathcal{D}|}$ and the clique number $(4\log n)^{|\mathcal{D}|}$ of 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 5 *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 $c_1, \ldots, c_m \in \{0,1\}^N$ is d-wise independent if and only if any d rows of the matrix L with columns $c_1, \ldots c_m$ 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 $M_d = \{(1, \alpha, \alpha^2, \ldots, \alpha^{d-1}) : \alpha \in \mathbb{F}\} \subseteq \mathbb{F}^d$ 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 β^i (mod 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.