LET'S BE EXPLICIT!

Extremal Graph Theory with emphasis on constructions

Lecture Notes

Summer Semester 2018, Freie Universität Berlin

Tibor Szabó

2004-2018

July 27, 2024

Contents

Chapter 1

Introduction

1.1 Prologue

Extremal Graph Theory, on the most general level, investigates the extremal (maximal or minimal) value of various graph parameters over the family of graphs having a particular property. It is a lively subject with a rich history, where numerous natural questions have beautiful answers. It is a field very much driven by problems; many of the interesting ones are still wide open and stimulate an abundance of research.

Each such problem has two sides: one is the construction of an extremal structure, the other is the proof of its optimality. In this course we are putting extra emphasis on explicit constructions of extremal graphs, which do not customarily feature in standard treatments of the field. These constructions often require useful tools from algebra, geometry, or discrete Fourier analysis; the other main objective of these notes is to highlight them.

We will consider two families of problems: *Turán-type problems* and *Ramsey-type problems*. In the first lecture we discuss the underlying ideas of both areas by looking more closely at a classical question from each genre.

The study of explicit constructions is intimately connected to the notion of quasi-random graphs. On the one hand, quasi-random graphs should imitate randomness as closely as they can in some sense: this is essential when the extremal example is random-like. On the other hand, one can also exploit the imperfect randomness of quasi-random graphs to find structures that are not typical at all. The Ramsey- and Turán-type questions we study eloquently demonstrate this duality.

In Turán-type problems the optimal solution is frequently obtained not by random methods, but rather by bumping into a beautiful and unique structure. More often than not, these structures are very much quasi-random but seriously deviate from random in one regard: the one that is the focus of the particular Turán-type problem. One can say that solving a Turán-type problem is like finding a needle in a haystack. The hay represents the random objects taking up almost all the space, while the needle represents the optimal solution we search for: rare, unique, and hard to come by.

Ramsey-type problems are the opposite in some sense: an optimal or nearly optimal solution is obtained by random methods; often *most* of the solutions are provably nearly optimal. In this case however finding an *explicit construction* poses a difficult prob-

lem. The deficient randomness of quasi-randomness makes the parameters of available explicit constructions lag significantly behind the random ones. Returning to the folklore metaphor about hay and needles, when investigating a Ramsey-type problem, we can imagine we are a horse standing in front of a huge haystack. We are hungry, we are trying to eat. We reach into the stack, pull out something: it's a needle. We reach again, pull out something: again a needle. Pull again, needle again... Clearly almost anything is edible, yet we are still unable to eat for some mysterious reason. Constructing good Ramsey-graphs explicitly: it is really like finding hay in a haystack.

Notation. $\mathbb{N}$ denotes the set of positive integers and $\mathbb{N}_0$ denotes the set of non-negative integers.

For a graph G, $V(G)$ denotes the vertex set and $E(G)$ denotes the edge set. Even though formally $E(G) \subseteq \binom{V(G)}{2}$, we often write xy instead of $\{x, y\}$ for the edge connecting vertices x and y. Sometimes we write $G = (V, E)$ for a graph with vertex set V and edge set E. The number of vertices of G is denoted by $v(G)$, the number of edges by $e(G)$. For subsets $X, Y \subseteq V(G)$, we denote by $E_G(X)$ the set of edges of G with both endpoints in X and write $e_G(X) = |E_G(X)|$ for its cardinality. $e_G(X, Y)$ denotes the number of ordered pairs (x, y) such that $x \in X$, $y \in Y$ and $xy \in E(G)$. If X and Y are disjoint, then $e_G(X, Y)$ is just the number of edges between X and Y. The neighborhood of vertex v is denoted by $N_G(v)$; formally $N_G(v) := \{u \in V(G); uv \in E(G)\}$. The degree of v is denoted by $d_G(v) := |N_G(v)|$. The set of neighbors of v in the subset $X \subseteq V(G)$ is denoted by $N_G(v, X)$ and $d_G(v, X) := |N_G(v, X)|$ denotes the degree of v into X. Often, when the underlying graph is clear from the context, the subscript G is omitted. For a subset $S \subseteq V(G)$ of the vertices, $G[S]$ denotes the subgraph of G induced by the vertex set S. The *minimum degree of* G is denoted by $\delta(G) = \min_{v \in V(G)} d(v)$, while the *maximum degree of* G is denoted by $\Delta(G) = \max_{v \in V(G)} d(v)$. The *independence number* of G is the size of the largest independent set of G and it is denoted by $\alpha(G)$. The *clique number* of G is the order of the largest clique in G and it is denoted by $\omega(G)$

The asymptotic notation. Most of the time we will be interested in asymptotic behavior of the encountered quantities. Hence, we start by recalling some definitions for abbreviating asymptotics.

Definition: Let $f, g : \mathbb{N} \to \mathbb{R}$ be functions. Then we write

- $f(n) = o(g(n))$ if
$$\frac{f(n)}{g(n)} \longrightarrow 0 \qquad \text{as } n \to \infty \ .$$
Sometimes we also write $f \ll g$.
- $f(n) = O(g(n))$ if
$$\exists\ N \in \mathbb{N}\ \ \forall\ n \geq N\ :\ \left|\frac{f(n)}{g(n)}\right| \leq C$$

for some constant $C \geq 0$.

- $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$.
- $f(n) = \Theta(g(n))$ if

$$f(n) = O(g(n)) \qquad \text{and} \qquad f(n) = \Omega(g(n)) \ .$$

Sometimes we also write $f \sim g$.

- $f(n) \approx g(n)$ if

$$\frac{f(n)}{g(n)} \longrightarrow 1 \qquad \text{as } n \to \infty \ .$$

- $f(n) \lessapprox g(n)$ if

$$\limsup_{n\to\infty} \frac{f(n)}{g(n)} \leq 1 \qquad \text{as } n \to \infty \ .$$

Estimations. To estimate binomial coefficients we often use

- For integers $n \geq k \geq 0$, we have

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq e^k \left(\frac{n}{k}\right)^k .$$

This being settled, in the following sections we concentrate on describing the type of questions which are important in this course.

1.2 Turán-type problems

Let us consider five isolated vertices. At most how many edges are we able to cram onto them if we are not allowed to create a triangle? In a first try one might end up with a cycle of length five. The graph C_5 has five edges and one cannot insert another edge in it without creating a triangle. In other words, C_5 is a *maximal* triangle-free graph with respect to addition of edges. More trial reveals that C_5 does not represent the *maximum* though: the complete bipartite graph $K_{2,3}$ is triangle-free and has six edges. Shortly we will see that $K_{2,3}$ is optimal: every graph on five vertices and seven edges contains a triangle.

The general question, about the maximum number of edges in an n-vertex triangle-free graph, was posed as a problem in a Dutch journal by W. Mantel in 1907. Correct solutions were submitted by five readers, including of course the poser, hence today the statement is referred to as *Mantel's Theorem*. Mantel showed that the complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ on n vertices has the largest number of edges among all graphs on n vertices not containing a triangle. Note that

$$e(K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}) = \left\lfloor \frac{n}{2} \right\rfloor \cdot \left\lceil \frac{n}{2} \right\rceil = \left\lfloor \frac{n^2}{4} \right\rfloor \ .$$

Theorem 1.1 (Mantel, 1907) *Any triangle-free graph on n vertices has at most $\lfloor \frac{n^2}{4} \rfloor$ edges.*

Proof. Let $G = (V, E)$ be a triangle-free graph on n vertices. Observe that the neighborhood $N(v) = \{u \in V \ : \ vu \in E\}$ of *any* vertex v must be an independent set. In other words, every edge of G has at least one of its endpoints in $W(v) := V \setminus N(v)$. Hence summing up the degrees of the vertices in $W(v)$ accounts for each edge at least once, twice only those which are entirely contained in $W(v)$. That is, for any vertex $v \in V$ we have that

$$e(G) = e(W(v)) + e(N(v), W(v)) \leq 2e(W(v)) + e(N(v), W(v)) = \sum_{x \in W(v)} d(x)$$

Estimating each degree further with the maximum degree Δ of G, we have

$$e(G) \leq \sum_{x \in W(v)} d(x) \leq \sum_{x \in W(v)} \Delta = (n - d(v)) \cdot \Delta,$$

To obtain the strongest estimate, it obviously makes sense to use a vertex $v \in V$ of degree as large as possible. For a vertex v of degree Δ the above estimate reduces to $e(G) \leq (n - \Delta)\Delta$. This is a quadratic polynomial in Δ that attains its maximum over the integers for $\Delta = \lceil \frac{n}{2} \rceil$. Substituting we find that

$$e(G) \leq \left(n - \left\lceil \frac{n}{2} \right\rceil\right) \cdot \left\lceil \frac{n}{2} \right\rceil = e(K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}) \ .$$

□

Remark: Observe that the proof also establishes the uniqueness of $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ as the unique extremal construction. Indeed, the first inequality is an equality if and only if $W(v)$ is an independent set, so the graph is bipartite with $N(v)$ being the other class. The second inequality is tight if and only if all $x \in W(v)$ have maximum degree, so the graph is complete bipartite, while the third inequality is tight if and only if the parts are as equal as possible.

Mantel's Theorem says that in order to "kill" all triangles of the complete graph one must delete roughly half of its edges (recall that the complete graph on n vertices has $\binom{n}{2} \approx n^2/2$ edges). What if, instead of forbidding triangles, one forbids the presence of K_4, or more generally the presence of k-cliques? Do we have to delete significantly fewer edges to achieve that? How many fewer? How does this fraction depend on k? In the early 40's Turán, unaware of the work of Mantel, arrived at this problem and generalized Mantel's Theorem. His motivation was Ramsey theory; this relationship is discussed in the next section.

Consider the following graph on n vertices. Partition the vertex set into $k - 1$ parts $V_1, \ldots, V_{k-1}$ of equal size (if $k - 1$ does not divide n take lower and upper integer parts, hence making a partition such that the cardinalities of any two parts differ by at most 1). Leave these $k - 1$ parts independent in the graph, but introduce an edge between

any two vertices from different parts. This is the *complete* $(k-1)$*-partite graph* with parts $V_1, \ldots, V_{k-1}$, which is also called the *Turán graph* and denoted by $T_{n,k-1}$.

The Turán graph (or, for that matter, any graph whose vertex set can be covered by $k-1$ independent sets) is K_k-free: it does not contain a subgraph isomorphic to K_k. Indeed, by the pigeonhole principle, among any k vertices of a $(k-1)$-partite graph there will be two which are in the same part and hence are *not* adjacent. Turán proved that the Turán graph is *the* unique graph containing the most edges among all K_k-free graphs. So to get rid of all k-cliques, one must delete roughly a $\frac{1}{k-1}$-fraction of the edges of K_n.

Theorem 1.2 (Turán, 1941) *Let G be a graph on n vertices not containing a k-clique K_k. Then*

$$e(G) \leq e(T_{n,k-1}) \approx \left(\frac{n}{k-1}\right)^2 \binom{k-1}{2} = \left(1 - \frac{1}{k-1}\right) \frac{n^2}{2} \ .$$

The exact formula for the number of edges of the Turán graph $T_{n,k-1}$ can of course be easily calculated for every n, but the general formula is not so interesting and the asymptotic expression above (which is tight when $k-1$ divides n) says much more.

There are many ways to prove Turán's Theorem; later we will look at two different arguments.

Up to now we have considered graphs that did not contain cliques. One can of course ask questions of this type for any forbidden subgraph H. This is exactly what Turán did in his letters to Erdős and via these questions he practically initiated the field of Extremal Graph Theory. To introduce the corresponding key definition, we say that G is H*-free* and write $H \not\subseteq G$ if G does not have a subgraph isomorphic to H.

Definition: The *Turán number* (or *extremal number*) $ex(n, H)$ of a graph H is defined as

$$ex(n, H) := \max \{e \ : \exists \text{ an } H\text{-free graph } G \text{ with } v(G) = n, e(G) = e\} \ .$$

As we have seen

$$ex(n, K_3) = \left\lfloor \frac{n^2}{4} \right\rfloor \qquad \text{and} \qquad ex(n, K_k) = e(T_{n,k-1}) \ .$$

Turán asked what if, instead of forbidding K_4 (which is the graph of the tetrahedron), we forbid some other platonic polyhedra? How many edges can a graph without an octahedron, or cube, or dodecahedron or icosahedron, have? These problems could seem somewhat arbitrary, but as it turns out they do contain some of the most interesting features of the area and at first sight the results certainly come as a surprise. When one is told that, after more than sixty years, the asymptotic answer is not known for only one of the five platonic polyhedra, one tends to guess that the outstanding question might be about the dodecahedron or the icosahedron, since their graphs are more "complicated".

It comes then as a minor shock that "complicatedness", in an everyday-sense, has nothing to do with this problem being hard.

$$\begin{aligned} ex(n, \texttt{tetrahedron}) &= \frac{n^2}{3} + o(n^2) \\ ex(n, \texttt{octahedron}) &= \frac{n^2}{4} + o(n^2) \\ ex(n, \texttt{dodecahedron}) &= \frac{n^2}{4} + o(n^2) \\ ex(n, \texttt{icosahedron}) &= \frac{n^2}{3} + o(n^2) \\ ex(n, \texttt{cube}) &= o(n^2) \ . \end{aligned}$$

The only problem still open is the one about cube-free graphs! This is quite astounding considering the fact that we do know quite precisely at most how many edges a dodecahedron-free graph can contain. The above are all corollaries of the following general theorem of Erdős, Stone, and Simonovits.

Theorem 1.3 (Erdős, Stone, Simonovits) *For any graph H, we have*

$$ex(n, H) = \left(1 - \frac{1}{\chi(H) - 1}\right)\binom{n}{2} + o(n^2) \ .$$

The special feature of this theorem is that the Turán number of any non-bipartite graph H essentially depends only on the chromatic number of H. For bipartite graphs the leading term disappears and we only know that the Turán number is of subquadratic order.

The Turán number of even the simplest bipartite graphs is often not known and is the subject of vigorous research. For example the order of magnitude of the Turán numbers of the cube Q_3, or the eight-cycle C_8 or the complete bipartite graph $K_{4,4}$ are still a mystery. Here is a sample of what we do know:

$$\begin{aligned} ex(n, C_4) &= \Theta(n^{3/2}) \\ ex(n, C_6) &= \Theta(n^{4/3}) \\ ex(n, C_{10}) &= \Theta(n^{6/5}) \\ ex(n, K_{3,3}) &= \Theta(n^{5/3}) \\ ex(n, K_{4,7}) &= \Theta(n^{7/4}) \\ \Omega(n^{3/2}) \leq ex(n, Q_3) &\leq O(n^{8/5}) \\ \Omega(n^{8/7}) \leq ex(n, C_8) &\leq O(n^{5/4}) \\ \Omega(n^{5/3}) \leq ex(n, K_{4,4}) &\leq O(n^{7/4}) \end{aligned}$$

Concluding this section we present some of the classic conjectures about the Turán number of simple bipartite graphs. In a significant portion of the course we will discuss these in depth.

Conjecture 1

$$ex(n, C_{2k}) \;=\; \Theta(n^{1+\frac{1}{k}}).$$

This conjecture is verified for $k = 2, 3, 5$. Recall that we do not know the order of magnitude of $ex(n, C_8)$.

Conjecture 2 *Let $s \geq t \geq 2$ be integers.*

$$ex(n, K_{t,s}) \;=\; \Theta\left(n^{2-\frac{1}{t}}\right).$$

This conjecture is known to be true for arbitrary s if $t = 2$ or 3, and for $s > (t-1)!$ when $t \geq 4$.

Conjecture 3

$$ex(n, Q_3) \;=\; \Theta(n^{8/5})$$

One general observation about Turán-type problems is that extremal graphs in all known cases have a very strong, pronounced structure. Nothing is accidental about them, but at the same time their existence often feels like coincidence. This seems valid not only in the simplest case, where the simple-minded Turán-graph provides the unique optimal structure for K_k-free graphs, but also for the optimal C_{10}-free or $K_{4,7}$-free structures we will encounter later, which are much more complex. As a prototype of this phenomenon, let us mention that for $k \geq 3$ a "typical" n-vertex graph is K_k-free only if its number of its edges is at most $o\left(n^{2-\frac{2}{k-1}}\right)$— much much smaller than the quadratic number of edges of the Turán graph.
Anybody who tried it would agree: constructing an extremal object for a Turán-type problem, as it is said in the Prologue, is like looking for a needle in a haystack. In the first part of these notes we will be after those needles.

Exercise 1.1 *Determine $ex(n, K_{1,4})$ and $ex(n, P_4)$ for every n.*

Exercise 1.2 *Let G be the "diamond" graph on 4 vertices: $V(G) = [4]$ and $E(G) = \{12, 13, 14, 23, 24\}$. Determine $ex(n, G)$ for every n.*

Exercise 1.3 *Show that for any tree T with t edges, $\frac{(t-1)n}{2} - o(n) \leq \mathrm{ex}(n, T) \leq (t-1)n$. In the special case of the star graph, $T = K_{1,t}$, show that the lower bound is correct.*

Exercise 1.4 *Let G be a graph on n vertices with $\left\lfloor \frac{n^2}{4} \right\rfloor + 1$ edges. Show that G contains at least $\left\lfloor \frac{n}{2} \right\rfloor$ triangles.*

Exercise 1.5 *For the octahedron graph, $K_{2,2,2}$, show that for all $n \geq 3$ we have the* strict *inequality* $\mathrm{ex}(n, K_{2,2,2}) > e(T_{n,2})$. *How large an n-vertex octahedron-free graph can you find?*

Exercise 1.6 *The TV remote of George requires two working batteries to function. Opening the drawer in which he keeps the batteries, he finds eight. He remembers that four of them work and four of them do not, but there is no way of telling them apart without testing them in the remote. How quickly can George guarantee to get his remote working?*

Exercise 1.7 *Show that for any k, the n-vertex K_k-free graph with the maximum number of copies of K_3 is the Turán graph $T_{n,k-1}$.* (Hint: *You can try apply the Zykov symmetrization: for non-adjacent vertices $u, v \in V(G)$, let G' be the graph obtained from G by deleting all edges between v and $N(v)$ and adding all edges between v and $N(u)$.)*

Remark. *One can generalise this by replacing 'K_3' with 'K_t' for any t.*

Exercise 1.8 *Give a Regularity Lemma-free proof of the Erdős-Stone Theorem.*

(a) Show that for every $r, s \in \mathbb{N}$ and $\epsilon \in \mathbb{R}$, $0 < \epsilon < 1/r$ there exists $t \in \mathbb{N}$, $\delta > 0$ and N_0 such that for any graph G on $n > N_0$ vertices and with minimum degree $\delta(G) \geq \left(1 - \frac{1}{r} + \epsilon\right) n$, and any r pairwise disjoint t-element subsets $B_1, \ldots, B_r \subseteq V(G)$, the set

$$W = \{w \in V(G) \setminus (B_1, \cup \cdots \cup B_r) : |N(w, B_i)| \geq s \text{ for all } i = 1, \ldots r\}$$

has size at least δn.

(b) Use (a) to show that for any $\epsilon > 0$ and integers $r \geq 2$, $s \geq 1$ there exists an integer $N = N(r, s, \epsilon)$, such that any graph G on $n \geq N$ vertices and with $\delta(G) \geq \left(1 - \frac{1}{r-1} + \epsilon\right) n$ contains $T_{rs,r}$.

(c) For every $r \in \mathbb{N}$ and $\epsilon \in \mathbb{R}$, $0 < \epsilon < 1/r$, there exists a $\delta = \delta(r, \epsilon)$ and $M = M(r, \delta)$ such that, for all graphs G with $n \geq M$ vertices and at least $\left(1 - \frac{1}{r-1} + \epsilon\right) \frac{n^2}{2}$ edges there is a subgraph $H \subseteq G$ with at least δn vertices, where each vertex has at least $\left(1 - \frac{1}{r-1} + \frac{\epsilon}{2}\right) v(H)$ neighbors.

(d) Conclude the Erdős-Stone Theorem.

1.3 Ramsey-type problems

A standard combinatorial exercise is the following.

Proposition 1.4 *In a party of six people there always exists three who pairwise know each other or three who pairwise do not know each other.*

Proof. We establish that at least one of the two conclusions necessarily holds. Let Frank be one of the six in the party. By the pigeonhole principle, either Frank knows at least three other people in the party, or there are at least three others whom he does not know. Consider the first case — the second is handled similarly — and let Esther, George, and Paul be three people who know Frank. If two of them, say Esther and George, also know each other, then they together with Frank would be three who pairwise know each other. Otherwise, Esther, George and Paul would pairwise not know each other and we arrive at the second of the possible conclusions. □

Remark: 1. For the argument to make sense and the statement of the claim to be true we must assume that there are no movie-stars in the party, that is, the relation of "knowing" each other must be symmetric.
2. Note that the same claim is not true for five people: just consider the party of five where each person knows two others in a cyclic fashion.

One can translate the statement of the claim into graph theoretic language. To each person we assign a vertex and make two vertices adjacent (by an undirected edge) if the corresponding people know each other. In this context Proposition 1.4 says that in any graph on 6 vertices there are three vertices that are pairwise adjacent or three vertices that are pairwise non-adjacent. In other words, for any graph G on six vertices the clique number or the independence number has to be at least 3. The five-cycle C_5, whose clique number and independence number are both 2, shows that six vertices are necessary.

The generalization of this problem was investigated by the great British logician/philosopher/economist Frank Plumpton Ramsey[1] in the late 1920's. To this end, for every $k, \ell \in \mathbb{N}$ we define the *Ramsey number* $R(k,\ell)$, as the smallest integer $n \in \mathbb{N}$ such that any graph on n vertices contains a clique of size k or an independent set of size ℓ. That is,

$$R(k,\ell) := \min\{n \in \mathbb{N} \,:\, \omega(G) \geq k \text{ or } \alpha(G) \geq \ell, \quad \forall G \text{ with } v(G) = n\} \,.$$

By Proposition 1.4 $R(3,3) \leq 6$, and $\omega(C_5) = \alpha(C_5) = 2$ shows that $R(3,3) > v(C_5) = 5$. Consequently we know the exact value of $R(3,3)$: it is 6. Nevertheless, even that $R(4,4)$ would be finite is not obvious at first glance. Motivated by a problem in mathematical logic, Ramsey proved that the Ramsey numbers $R(k,\ell)$ are indeed finite for every k and ℓ.

Our knowledge about exact values of Ramsey numbers is surprisingly limited, only one more symmetric value is known: $R(4,4) = 18$. Already the calculation of $R(5,5)$ exceeds, by far, the capacities of not only mathematicians, but even computers! Considering that this value is somewhere between 43 and 48, it is pretty astounding and humbling to come to terms with the cold hard fact: humanity is unable to program its computers to decide a question about graphs on fourty-something vertices.

A few years later, unaware of the work of Ramsey, Paul Erdős and George Szekeres arrived at the same problem. Their motivation was a question in geometry, the famous

[1] Besides being a mathematician, Ramsey also published fundamental papers in philosophy and economics. He died at the age of 26.

Happy Ending Problem.[2] They[3] proved the finiteness of Ramsey numbers, with a slightly better quantitative upper bound, by generalizing the pigeonhole principle argument for Proposition 1.4.

Theorem 1.5 *For any $k, \ell \in \mathbb{N}$ we have*

$$R(k,\ell) \leq \binom{k+\ell-2}{k-1} .$$

Proof. We give a proof by induction on $k+\ell$. First let us consider the base cases. Since a set consisting of one vertex is both a clique and an independent set, $R(k,1) = 1 = R(1,k)$ for any $k \in \mathbb{N}$.

Now let $k, \ell \geq 2$, and suppose that both $R(k-1,\ell)$ and $R(k,\ell-1)$ exist by induction. We will show that

$$R(k,\ell) \leq R(k-1,\ell) + R(k,\ell-1). \tag{1.1}$$

Consider an arbitrary graph G on $v(G) = R(k-1,\ell) + R(k,\ell-1)$ vertices. We will find in G a clique of size k or an independent set of size ℓ.

Let $v \in V(G)$ be any vertex and, similarly to the proof of Proposition 1.4, let us take a look at how many neighbors and non-neighbors it has. We claim that v has either at least $R(k-1,\ell)$ neighbors or at least $R(k,\ell-1)$ non-neighbors. Indeed, otherwise the sum of the number of neighbors and the number of non-neighbors of v would be less than $R(k-1,\ell) + R(k,\ell-1) - 1 = v(G) - 1$, a contradiction.

Let us assume first that $|N(v)| \geq R(k-1,\ell)$. Then either $\alpha(G[N(v)]) \geq \ell$ and we found an independent set of size ℓ in G, or $\omega(G[N(v)]) \geq k-1$ and this $(k-1)$-clique of $G[N(v)]$ together with v forms a k-clique of G.

An analogous argument works in the second case, when the set $U := V(G) \setminus (N(v) \cup \{v\})$ of non-neighbors of v is large, i.e., if $|U| \geq R(k,\ell-1)$. Then either $\omega(G[U]) \geq k$, in which case we found a k-clique in G, or $\alpha(G[U]) \geq \ell - 1$ and then this independent set of size $\ell-1$ together with v forms an independent set of size ℓ in G.

This completes the proof of the recursion (1.1). Then by the induction we have

$$R(k,l) \leq R(k-1,l) + R(k,l-1) \leq \binom{k+l-3}{k-2} + \binom{k+l-3}{k-1} = \binom{k+l-2}{k-1}.$$

□

The most interesting special case is the symmetric one, when the two parameters are equal. The previous theorem implies an upper bound which is exponential in k.

Corollary 1.6 *For any $k \in \mathbb{N}$ we have*

$$R(k,k) \leq \binom{2k-2}{k-1} = O\left(\frac{4^k}{\sqrt{k}}\right) .$$

[2] ..., which, in fact, initiated the field of combinatorial geometry.

[3] In fact, Erdős [?] generously attributes the proof to Szekeres. Since they gave *two* different proofs of the Happy Ending Problem in their joint paper, it could be that Erdős gave the other, which gives a better bound for the geometry problem.

After their 1935 joint paper, Erdős and Szekeres did not pursue the Ramsey problem any further. In particular, they did not make any serious attempt to complement their upper bound with a lower bound. Erdős had already left Budapest and embarked on his life-long academic journey around the world, hopping between departments, institutes, meetings, and collaborators. He was extremely prolific in number theory and analysis, but not published in combinatorics at all. Szekeres, a hobby mathematician without formal training at the time, worked in Budapest as a chemist till 1939, when he fled Hungary with his wife, Esther Klein[4], to escape the persecution of Jews. He survived the war in Shanghai working all sorts of jobs, and obviously did not have too much time on his hand for mathematics.[5]

It was a third Hungarian protagonist, Paul Turán who pushed matters further eventually. Unlike his two university friends, Turán remained in Hungary throughout the second world war. As a Jew, he could not get a teaching position after graduation, supported himself from private tutoring, and pursued his research interests on the side. Despite being an analytic number theorist primarily, Turán found Ramsey's Theorem intriguing, and the theorem was to play a crucial role in his life on a couple of occasions during the 1940s. His personal account of these encounters in his *A note of welcome* in the first issue of the newly established *Journal of Graph Theory* in 1970 should be a mandatory historical read for every student of Graph Theory.

First of all, it was via Ramsey's Theorem that Turán *arrived* at the formulation of the extremal graph theory problem that eventually became Turán's Theorem in 1941. To illuminate this, we interpret Ramsey's Theorem as the study of the independence number of K_k-free graphs, or rather, how small it could be. Formally speaking, if $\ell = \ell(n,k)$ denoted the minimum value of $\alpha(G)$ among all K_k-free graphs G on n vertices, then $n < R(k,\ell+1)$ would provide a lower bound on the Ramsey number. Since small independence number is naturally associated with having a lot of edges, Turán was led to wanting to find out more about the maximum edge number of K_k-free graphs on n vertices. Eventually this has led to a whole family of problems we call today Turán-type problems, and the "discovery" of the field of extremal graph theory.

So a plausible strategy for a lower bound on the Ramsey number could be to cram as many edges as possible onto the vertices, such that the graph is still K_k-free, and just hope for the best in terms of the independence number. This is exactly the question we solved in the previous section, so let us see what the optimal construction there, the Turán graph, gives us here. What is the *largest* vertex number n such that the K_k-free Turán graph $T_{n,k-1}$ has no independent set of size k? Since the maximum independent sets of a Turán graph are exactly its classes, they must all be of size at most $k-1$. Consequently, the best lower bound we can obtain by way of the Turán graph is using $T_{(k-1)^2,k-1}$ on $(k-1)^2$ vertices, which shows

$$R(k,k) > (k-1)^2 \ .$$

This is "just" a quadratic lower bound standing opposite the exponential upper bound

[4]the poser of the Happy Ending Problem

[5]At least until he settled in Australia in 1948, finally to a proper mathematics position.

of Corollary 1.6. Which one is (closer to) the truth? The proof of both bounds are pretty straightforward. Should the simplistic inductive argument of the upper bound be improved significantly? After all, it involves throwing away potentially half of the vertices in each of the $2k$ induction step. Or should the simple symmetry of the Turán graph be broken somehow?

In October/November of 1944 Turán worked intensively on improving the upper bound on $R(k,k)$. He tried to show by various inductive approaches that $T_{(k-1)^2,k-1}$ is optimal for the Ramsey question as well. After a promising start, these attempts all broke down at some point. In [?] Turán writes:

> "I had no other support for the truth of this conjecture than the symmetry and some dim feeling of beauty; perhaps the ugly reality was what made me believe in the strong connection of beauty and truth. But this unsuccesful fight gave me strength hence, when it was necessary, I could act properly. In one of my first letters to Erdős after the war I wrote of this conjecture to him. In his answer he proved that my conjecture was utterly false;"

The strategy of Erdős was remarkably simple, yet quite unusual at the time: count those graphs on vertex set $[n]$ that contain a clique or an independent set of size k, and compare the number to the number of all graphs on vertex set $[n]$. If, for some n, the latter is strictly larger than the former, then we established the *existence* of a k-Ramsey graph on n vertices.

Clearly, there are $2^{\binom{n}{2}}$ graphs on the vertex set $[n]$, since for every one of the $\binom{n}{2}$ pairs of vertices there are two choices: the pair is either an edge or not.

Determining the number of graphs G with $V(G)=[n]$ and $\omega(G)\geq k$ or $\alpha(G)\geq k$ might look like a daunting task—were we interested in the exact value. But Erdős never lost target of the orginal target and was content with just *estimating* this number. He did not overcomplicate things. He classified graphs G with $V(G)=[n]$ and $\omega(G)\geq k$ or $\alpha(G)\geq k$ according to which of the k-subset induces a clique or an independent set and counted these separately. For any particular k-subset $K\subseteq[n]$ of the vertices, the number of those graphs which induce a clique or independent set on K is $2\cdot 2^{\binom{n}{2}-\binom{k}{2}}$. Indeed, the pairs inside K must all be edges or must all be non-edges, while each of the remaining $\binom{n}{2}-\binom{k}{2}$ pairs can be freely chosen to be an edge or a non-edge. So the number of graphs on vertex set $[n]$ containing a clique or independent set of size k is at most

$$\binom{n}{k}2^{\binom{n}{2}-\binom{k}{2}+1}<\left(\frac{ne}{k}\right)^k 2^{\binom{n}{2}-\binom{k}{2}+1}.$$

Note that this seems like a crude overestimate,[6] graphs with many k-cliques and/or independent sets are counted many times; for example the complete graph K_n is counted $\binom{n}{k}$ times.

[6] In reality it is not

Recall now that if for some integer it holds that

$$2^{\binom{n}{2}} > \left(\frac{ne}{k}\right)^k 2^{\binom{n}{2}-\binom{k}{2}+1} = 2 \cdot \left(\frac{ne\sqrt{2}}{k\sqrt{2}^k}\right)^k \cdot 2^{\binom{n}{2}},$$

then there is at least one graph on n vertices which contains neither a clique nor an independent set of order k. The inequality is equivalent to

$$\frac{1}{2^{1/k}} \cdot \frac{k\sqrt{2}^k}{e\sqrt{2}} > n,$$

so we have show that there is at least one graph on $\left\lfloor \frac{1}{2^{1/k}} \cdot \frac{k\sqrt{2}^k}{e\sqrt{2}} \right\rfloor$ vertices which contains neither a clique nor an independent set of order k, implying an exponential lower bound on the Ramsey number.

Both of the two exponential bounds we derived here,

$$\left(\frac{1}{e\sqrt{2}} + o(1)\right) k\sqrt{2}^k < R(k,k) < \left(\frac{1}{\sqrt{\pi}} + o(1)\right) \frac{4^k}{\sqrt{k}},$$

are ancient, yet they are essentially the best known today! In terms of the upper bound, Conlon [?] showed that $R(k,k)$ is smaller than 4^k by an *arbitrary* polynomial factor. Still, nobody can prove that $R(k,k) \leq 3.99^k$ for example. There is no substantial improvement in the lower bound either. All what happened is that the constant factor $\sqrt{2}$ has moved up to the numerator some forty years ago [?], but it is unclear whether a lower bound of, say, $k^{1.1}\sqrt{2}^k$ would hold.

The determination of the limit $\lim_{k\to\infty} \sqrt[k]{R(k,k)}$ is one of the most notorious open problems of combinatorics, yet we do not even know whether the limit exists. Erdős offered a prize of $500 for deciding either questions.

Exercise 1.9 *(a) Prove that* $R(3,4) = 9$

(b) Prove that $R(4,4) \leq 18$
(For a matching lower bound see Exercise ?? *in Section 5.1.)*

Exercise 1.10 *Recall that the Ramsey number* $R(k,\ell)$ *can also be formulated as the smallest* n *such that any* red/blue *coloring of the edges of* K_n *contains a monochromatic* K_k *in* red *or a monochromatic* K_ℓ *in* blue.

(a) Define the multicolor Ramsey-number $R_r(k_1,\ldots,k_r)$.

(b) Prove that $2^r \leq R_r(3,\ldots,3) \leq \lfloor er! \rfloor$

Ramsey's Theorem established the finiteness of $R(k,\ell)$ for every k,ℓ. Subsequenetly the quantitative determination of these numbers and their extensions grew into a central branch of the field of combinatoricsand the problem became the catalist behind many of the main developments of the field in general. The symmetric case, when $k=\ell$, in particular represents one of the great open problems of combinatorics. The upper and lower bounds

$$\sqrt{2} \leq \liminf_{k\to\infty} \sqrt[k]{R(k,k)} \leq \liminf_{k\to\infty} \sqrt[k]{R(k,k)} \leq 4,$$

were proved in 1947 and 1930, respectively, and progress on either of them was small and difficult.

$R(2,2)=2$ 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.

1.4 Basics of the probabilistic method

The above counting argument of Erdős is considered today as the introduction of a general proof technique in combinatorics, called the *probabilistic method*. The method is used to establish the *existence* of a certain type of object without actually constructing it. This approach could be particularly helpful when other, deterministic attempts to construct an object with desired properties prove to be unfruitful.

The simple, yet revolutionary idea is that one constructs an appropriate *probability space of objects*, rather than a particular instance of the desired object, such that in the probability space the desired objects occur with nonzero probability. The point in the choice of the probability space is that the probability of the desired objects should be *provably* nonzero. Miraculously, it is often the case that while concentrating on satisfying the desired property of one particular object proves to be hopelessly hard, proving that an overwhelming majority of objects have the desired property is relatively easy.

While other instances of this existence proof technique appeared earlier in several branches of mathematics, no one before used it as systematically as Erdős, who almost single-handedly developed it into a *method*.

For the moment we will be content with the simplest of conclusions.

- If the probability of objects having some property P is not 0, then *there exists* an object with property P.

Often it is more convenient to talk in terms of expected value.

- If the expected value of a random variable is C then *there exists* an object for which the variable's value is *at least* C and *there exists* an object for which the variable's value is *at most* C.

1.4.1 The union bound

The union bound: Let $A_1, \ldots, A_i, \ldots$ be any (finite or infinite) set of events. Then

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]\,.$$

In order to see the union bound at work let us reformulate the above Ramsey argument in the language of probability. This makes much sense conceptually, because thinking in terms of probability theory allows one to later apply the full machinery of the field. Even though every single fact of discrete probability can in principle be expressed just as counting, with some of the more tricky ones this could be extremely cumbersome, if not close to impossible to carry out. The more general point of view of probability has proven to be more and more fruitful ever since the groundbreaking counting result of Erdős.

First we define the "appropriate" probability space: the random graph $G(n, 1/2)$ with edge probability $1/2$. For each of the $\binom{n}{2}$ pairs of vertices we flip a fair coin, and we put the corresponding edge in the graph if the result of the coin flip is "head". Note that in this probability space all the $2^{\binom{n}{2}}$ labeled graphs on vertex set $[n] = \{1, \ldots, n\}$ occur with the same probability.

The calculations in the previous subsection can be used to estimate the probability that $G(n, 1/2)$ has clique number at least k.

Fix a subset $S \subseteq [n]$ of size k. Let A_S be the event that the vertices of S form a clique in $G(n, 1/2)$. For A_S to happen, all the $\binom{k}{2}$ coin flips corresponding to the pairs of S must turn out to be heads, while all other coin flips can be arbitrary. Hence

$$\mathbb{P}[A_S] = \frac{1}{2^{\binom{k}{2}}}.$$

The probability that the clique number is at least k is the union of A_S over all $S \subseteq [n], |S| = k$. We can then use the union bound,

$$\mathbb{P}[\omega(G(n,1/2)) \geq k] = \mathbb{P}\left[\bigcup_{S \in \binom{[n]}{k}} A_S\right] \leq \sum_{S \in \binom{[n]}{k}} \mathbb{P}[A_S] = \frac{\binom{n}{k}}{2^{\binom{k}{2}}} \leq \left(\frac{en\sqrt{2}}{k\sqrt{2}^k}\right)^k.$$

One can bound $\mathbb{P}[\alpha(G(n,1/2)) \geq k]$ similarly. Hence, say for $n = \sqrt{2}^k$, the union bound gives

$$\mathbb{P}[\omega(G(n,1/2)) \geq k \text{ or } \alpha(G(n,1/2)) \geq k] \;\leq\; 2 \cdot \left(\frac{e\sqrt{2}}{k}\right)^k.$$

For large enough k this probability is strictly less than 1 and thus the *existence* of a k-Ramsey graph on $\sqrt{2}^k$ vertices is proved.

But we get more. Let us play with the numbers a bit. By the above calculation, for every $k \geq 12$, at least 99.999 percent of all graphs on $\sqrt{2}^k$ vertices have both their clique number and their independence number strictly less than k. So we **know** that almost every graph is "good" for us in this sense. But should we just aim to hold one of these good Ramsey-graph in our hands ... we are going to see how elusive they are! Returning to the haystack parabole of the introduction: in the second part of these notes we will play the role of the hungry horse and try to finally pull out a hay instead of a needle. We will fail, but experience a lot of nice mathematics in the process. We will see that the situation is the exact opposite of the one we encountered with Turán-type problems. Optimal structures of Turán-type problem are very non-random; to find them one tries to distance oneself from random. Optimal structures of Ramsey-type problems are random-like in some sense and our explicit constructions will try to imitate some of that randomness.

1.4.2 Linearity of expectation

Linearity of Expectation Let $X_1, \ldots, X_n$ be random variables. Then

$$\mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i].$$

To demonstrate this method we prove the asymptotic version of Theorem 1.2 (Turán's Theorem). That is, for any graph G with clique number $\omega(G) < k$, we must estimate $e(G)$ from above. We turn this around and rather for any given edge number e we estimate the clique number from below. First we use the linearity of expectation to derive a *lower* bound on the clique number of *any* graph in terms of the degrees of its vertices. Since the number of edges is linked to the degrees via $e(G) = \frac{1}{2}\sum d(v)$, this will also imply a lower bound in terms of the number of edges using a standard convexity argument.

Theorem 1.7 *For any graph G, we have*

$$\omega(G) \geq \sum_{v \in V(G)} \frac{1}{n - d(v)} \geq \frac{n^2}{n^2 - 2e(G)}.$$

Proof. Let $V = \{v_1, \ldots, v_n\}$ be the vertex set of G. For each permutation π of $[n]$, let us design a clique $C(\pi)$ of G in the following fashion: a vertex $v_{\pi(i)} \in C(\pi)$ if and only if $v_{\pi(i)}$ is adjacent to $v_{\pi(j)}$ for every $j < i$. For example, $v_{\pi(1)}$ is always in $C(\pi)$, but $v_{\pi(2)}$ is in there only if it is adjacent to $v_{\pi(1)}$. By definition, $C(\pi)$ is a clique for every permutation π.

The idea is to calculate the expected value of the cardinality of $C(\pi)$ if we select a permutation π uniformly at random among all permutations of $[n]$. While calculating the expectation of $|C(\pi)|$ looks impossibly complicated at first sight, as is often the case, if we divide the problem into subproblems, then the linearity of expectation can come to the rescue.

Let X_i be the characteristic random variable of the event that $v_i \in C(\pi)$. (That is, $X_i = 1$ if $v_i \in C(\pi)$ and $X_i = 0$ if $v_i \notin C(\pi)$.) Then $|C(\pi)| = \sum_{i=1} X_i$.

We are interested in the probability that a vertex v_i is contained in $C(\pi)$. This event occurs if and only if all its non-neighbors come only *after* v_i in the permutation π. In other words, if we restrict π to v_i and its $n - 1 - d(v_i)$ non-neighbors, then v_i is the first element. This has probability $\frac{1}{n-d(v_i)}$, hence,

$$\mathbf{E}[X_i] = \mathbb{P}[v_i \in C(\pi)] = \frac{1}{n - d(v_i)}.$$

By the linearity of expectation

$$\mathbf{E}[|C(\pi)|] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{1}{n - d(v_i)}.$$

Since the expected value of this random variable is $\sum_{i=1}^{n} \frac{1}{n-d(v_i)}$, there must exist a *particular* permutation σ (maybe more than one) for which $|C(\sigma)|$ is at least $\sum_{i=1}^{n} \frac{1}{n-d(v_i)}$. As $C(\sigma)$ is a clique, the first inequality in the statement is proved.

For the second inequality note that the function $x \to \frac{1}{x}$ is convex on the positive reals, so the average value of the function over a set of positive real numbers is larger than the value of the function at the average of that set (by Jensen's inequality):

$$\sum_{i=1}^{n} \frac{1}{n - d(v_i)} \geq n \frac{1}{n - \bar{d}(G)} = \frac{n^2}{n^2 - 2e(G)},$$

where $\bar{d}(G) = \frac{2e(G)}{n}$ is the average degree of G. □

Now Turán's Theorem follows immediately.

Proof. (*of Theorem 1.2*) Let G be K_k-free graph, that is, $\omega(G) \leq k - 1$. Hence by the previous theorem $k - 1 \geq \frac{n^2}{n^2 - 2e(G)}$, which rearranges to give

$$e(G) \leq \left(1 - \frac{1}{k-1}\right) \frac{n^2}{2},$$

which proves Theorem 1.2. □

Remark: 1. It is worthwhile to ponder for a moment why this strange proof of Turán's Theorem can work. We bound the clique number as the expected number of the random variable $|C(\pi)|$. Since this estimate is eventually tight, an overwhelming majority of $C(\pi)$ have to be maximum cliques. This in fact is true in the extremal extremal case, i.e., when G is the Turán graph: every single permutation σ gives a maximum clique, as $C(\sigma)$ contains the first vertex (according to π) from each of the $k - 1$ vertex classes of the Turán graph.

2. The way $C(\pi)$ is created is in some kind of greedy way, but maybe not the greediest possible. Let us again imagine a process where the vertices come one by one according to the random permutation π and we decide online whether we put them in the set $C(\pi)$ (and not change our decision later). The greediest of approaches would put a vertex into $C(\pi)$ if it is adjacent to every vertex that is already put in $C(\pi)$. The $C(\pi)$ produced this way will be a clique alright, but the problem is with the analysis of its size. This process creates a complex system of dependencies between the decisions and the probability of a particular vertex v_i being in $C(\pi)$ might potentially depend on the structure of the whole graph (and not only the degree of v_i). For this reason we chose to strengthen the requirement for membership in $C(\pi)$: not only the current elements of $C(\pi)$, but *all* preceding vertices had to be adjacent to v_i. Observe that for the Turán graph the two algorithms produce the same $C(\pi)$.

Part I

Turán-type problems

Chapter 2

Complete Bipartite Graphs

Complete bipartite graphs are structurally very simple, but we will still see how the study of their Turán numbers leads us into uncharted geometric and algebraic territory and leaves us with many open questions.

2.1 Forbidding $K_{2,s}$

The following upper bound on the Turán number of $K_{2,2}$ appeared even earlier than Turán's Theorem, as an auxiliary lemma in a number theoretic paper of Paul Erdős. It then went into relative oblivion as a combinatorial statement. Paul Erdős later referred to it as the point when he "missed discovering Extremal Graph Theory".

Theorem 2.1 (Erdős; 1938) *We have*

$$ex(n, K_{2,2}) \leq \frac{1}{2}n^{3/2} + \frac{1}{4}n.$$

Proof. Take a graph $G \not\supseteq K_{2,2}$. Let us double-count its "cherries", i.e., its subgraphs isomorphic to $K_{2,1}$.

On the one hand, for a fixed vertex v there are $\binom{d(v)}{2}$ cherries centered at it. The total number of cherries is then equal to $\sum_v \binom{d(v)}{2}$.

On the other hand, since there is no $K_{2,2}$ in G, every pair of vertices u, w, is the set of endpoints of at most one cherry. We then have

$$\sum_v \binom{d(v)}{2} = \#\text{ of cherries } \leq 1 \cdot \binom{n}{2}.$$

By Jensen's inequality ($\binom{x}{2} = \frac{x(x-1)}{2} : \mathbb{R} \to \mathbb{R}$ is convex!) we obtain

$$\sum_v \binom{d(v)}{2} \geq n\binom{\bar{d}(G)}{2},$$

where $\bar{d}(G) = \sum_v d(v)/n = 2e(G)/n$ is the average degree of G. This implies $\bar{d}(G) \leq \frac{\sqrt{4n-3}+1}{2} \leq \sqrt{n} + \frac{1}{2}$, and hence

$$e(G) = \frac{n\bar{d}(G)}{2} \leq \frac{1}{2}n^{3/2} + \frac{1}{4}n.$$

□

The natural question arises: Is this upper bound on the Turán number of $K_{2,2}$ a good one? Is there a $K_{2,2}$-free graph with $\Omega\left(n^{3/2}\right)$ edges?

2.1.1 How about a randomized construction?

Still elated by the success of the random graph for the symmetric Ramsey problem, let us try some natural randomized approaches.

We look at the random graph $G(n,p)$ (i.e. the probability space of graphs, where each of the $\binom{n}{2}$ edges appears with probability p, independently from all other edges).

The first thought is to determine the largest function $p = p(n)$ for which the random graph does not contain a $K_{2,2}$ with positive probability, or rather with probability tending to a strictly positive number. For this, an always good indication is to calculate the expected number of (unlabeled) $K_{2,2}$s in $G(n,p)$. Let $\#K_{2,2}$ denote the random variable representing the number of unlabeled copies of $K_{2,2}$ in $G(n,p)$. Then

$$\mathbf{E}[\#K_{2,2}] = \frac{n(n-1)(n-2)(n-3)}{8}p^4.$$

For $p = \frac{1}{n}$ this expectation is less than 1, hence there is an instance of a graph in $G(n,p)$ which is $K_{2,2}$-free. "Of course there is" – one can say – "the empty graph for example!" The empty graph though would not help greatly our quest for a dense $K_{2,2}$-free graph. But random graph theory tells us more: once this expectation is $o(1)$, then $G(n,p)$ is $K_{2,2}$-free with probability tending to 1. Hence for large n with probability at least 0.99 $G(n,p)$ is $K_{2,2}$-free, provided $p = o(1/n)$. Random graph theory also tells us that for large n with probability at least 0.99 $G(n,p)$ has at least $\frac{1}{2}\binom{n}{2}p$ edges. Thus with probability at least 0.98 the random graph $G(n,p)$ is $K_{2,2}$-free *and* has an almost linear number of edges! In particular there is an instance of such a graph. Wow! Short meditation curbs the enthusiasm a bit, as one comes up with an explicit construction of a $K_{2,2}$-free graph with $n-1$ edges: the path P_n. Unfortunately we cannot really hope for more from the random graph in this simple manner: it turns out that for $p \gg \frac{1}{n}$, that is when $\mathbf{E}[\#K_{2,2}] \to \infty$, the probability that $G(n,p)$ is $K_{2,2}$-free tends to 0. These are standard facts from random graph theory, which we do not prove here, because they just show that *most* graphs with $\omega(n)$ edges are *not* $K_{2,2}$-free. That is, should we want hit upon a denser $K_{2,2}$-free graph, we indeed need to find a "needle in the haystack".

The Alteration Method

Despite this discouraging news on the random front one can still try to put a twist on the random approach. The idea is not to shoot for a $K_{2,2}$-free graph right away, but to first obtain with random methods a graph G', possibly with many many more than $\Theta(n)$ edges, which might contain some $K_{2,2}$s, but not so many. Indeed, if the number of $K_{2,2}$s is, say, at most half the number of edges, then deleting one edge from each $K_{2,2}$ makes

G' $K_{2,2}$-free, and the resulting graph G still has half of the edges of the original graph G'.

This "semi-random" technique is called the *deletion* or *alteration method*.

To this end we would like to find p *as large as possible*, such that

$$\mathbf{E}[e(G(n,p)) - \#K_{2,2}] \geq \frac{1}{2}\binom{n}{2}p.$$

For such a probability p there surely exists an instance of a graph G' for which we have $e(G') - \#\{K_{2,2} \text{ in } G'\} \geq \frac{1}{2}\binom{n}{2}p$. If we remove an edge from each copy of $K_{2,2}$ appearing in G', then we obtain a $K_{2,2}$-free graph in which the number of edges is at least $\frac{1}{2}\binom{n}{2}p$.

To evaluate the hopelessly complicated-looking random variable $e(G(n,p)) - \#K_{2,2}$ we of course use the linearity of expectation.

$$\begin{aligned}
\mathbf{E}[e(G(n,p)) - \#K_{2,2}] &= \mathbf{E}[e(G(n,p))] \;-\; \mathbf{E}[\#K_{2,2}] \\
&= \binom{n}{2}p \;-\; \frac{n(n-1)(n-2)(n-3)}{8} \cdot p^4 \\
&\geq \frac{1}{2}\binom{n}{2}p.
\end{aligned}$$

Solving we get

$$1 \geq \frac{(n-2)(n-3)}{2}p^3.$$

Hence for $p = cn^{-2/3}$ (with small enough c), there exists a graph G' for which we have $e(G') - \#\{K_{2,2} \text{ in } G'\} \geq \frac{1}{2}\binom{n}{2}p = c'n^{4/3}$. So after removing an edge from each $K_{2,2}$ of G we have a $K_{2,2}$-free graph with $c'n^{4/3}$ edges.

This is now at least a superlinear number of edges, but still far away from the order $n^{3/2}$ of the upper bound. We don't claim that this simple construction is the only possible random idea one can have, but we certainly don't have any better. We don't know how to use more sophisticated techniques to give a better randomized lower bound.

2.1.2 A heuristic geometric construction of a $K_{2,2}$-free graph with $n^{3/2}$ edges

Let us recall what should a $K_{2,2}$-free construction look like if it harbours any hopes for making the upper bound of Theorem 2.1 tight. During that proof we made two estimates: one was using Jensen's Inequality over the degree sequence, the other used the fact that in a $K_{2,2}$-free graph every pair of vertices have at most one common neighbour. Hence if our construction does not intend to lose more than just a constant factor of the edges of that upper bound then

- it should be close to being regular. (Jensen's Inequality is tight when the values over which it is applied are nearly equal.)
- at least a constant factor of all pairs of vertices should have a common neighbour.

Upon a first look, the first property seems pretty general and maybe the less problematic to achieve, while the second is the more substantial for the specific problem of constructing dense $K_{2,2}$-free graphs. The latter one might also inspire us to think of the following *analogous* scenario when any pair of something determines a unique other something: in classical Euclidean plane geometry any two points determine a unique line and any two lines contain (at most) one point in their intersection.

We will use this intuition to construct a $K_{2,2}$-free graph; however, it will be infinite. You might wonder what $n^{3/2}$ should mean when n is not finite... Please loosen up for the moment: The whole subsection is somewhat nonsensical, but nevertheless, it is a good introduction to what comes next.

We construct a bipartite graph G with one partite set A being the set of all points in the (Euclidean) plane $\mathbb{R}^2$, and the other partite set B being the set of all lines in the plane. We join vertices $p \in A$ and $l \in B$ with an edge if and only if the point p is on the line l.

If this graph contained a $K_{2,2}$, that would mean that there are two different lines that have two different points in common. This is of course not possible, and that is why G is $K_{2,2}$-free.

Let us now look at the number of edges. The set of points in the plane is just the set of ordered pairs with both coordinates from $\mathbb{R}$. So the cardinality of A is $|\mathbb{R}|^2$. Every line in the plane can be uniquely represented by its slope and its signed distance from the origin. Therefore $|B|$ is of cardinality $|\mathbb{R} \cup \{\infty\}| \cdot |\mathbb{R}|$. Now this is again roughly $|\mathbb{R}|^2$ (we ignored a lower order term of $|\mathbb{R}|$:-)). So the order of the graph is

$$|V(G)| \approx 2 \cdot |\mathbb{R}|^2.$$

On the other hand, on each line there are $|\mathbb{R}|$ points meaning that every vertex from B has degree $|\mathbb{R}|$. Since G is bipartite the total number of edges in G is

$$|E(G)| = \underbrace{|\mathbb{R} \cup \{\infty\}| \cdot |\mathbb{R}|}_{|B|} \cdot |\mathbb{R}| \approx |\mathbb{R}|^3 \approx \left(\frac{|V(G)|}{2}\right)^{3/2}.$$

2.1.3 Explicit constructions of dense $K_{2,2}$-free graphs

How should we make sense of the previous geometric idea, i.e. how to "finitize" it? First one has to realize that the required property, the "$K_{2,2}$-freeness", depends only on the fact that a system of two linear equations in two variables (which are not a constant multiple of each other) has at most one solution. Then the answer is quite obvious: instead of the field of real numbers $\mathbb{R}$, take a finite field $\mathbb{F}_q$, in which the same rules apply regarding solving an equation system. Recall that the q-element field $\mathbb{F}_q$ exists if and only if q is a prime power. For primes p, the set $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ of residue classes with the usual $\pmod p$ addition and multiplication is the p-element field.[1]

[1] See the Appendix for a quick review of the basic facts about finite fields.

Construction 0. We construct a bipartite graph G with one partite set A being all points in the *affine plane* $\mathbb{F}_q^2$, and the other partite set B being all lines in $\mathbb{F}_q^2$. Lines can be defined similarly to lines in $\mathbb{R}^2$: as the set of solutions of a linear equation $a_1x_1+a_2x_2 = a_3$, where not both of a_1 and a_2 are 0. This graph is $K_{2,2}$-free. Like in the previous section, two different lines do not have two common points, since the corresponding linear equation system (of two equations) has *at most* one solution.

We have q^2 points in A and $\frac{(q^2-1)q}{q-1} = q^2 + q$ lines in B. Indeed, there are $(q^2-1)q$ ways to select a_1, a_2, and a_3, and the corresponding equations determine the same line (i.e., have the same set of solutions) if and only if the coefficients are multiplied with the same non-zero field element. This makes up to the total of $n = 2q^2 + q$ vertices of G. The degree of a vertex in B is the cardinality of a line in $\mathbb{F}^2$. This number is q for every line, as the solution set is just a coset of the (1-dimensional) kernel of the linear map $(a_1, a_2) : \mathbb{R}^2 \to \mathbb{R}$ of rank 1. (or, more concretely, one can express the variable x_i corresponding to one of the non-zero coefficients a_i, $i = 1$ or 2, and then a substitution of an arbitrary field element to the other variable x_{3-i} gives a unique solution (x_1, x_2)). Then,

$$e(G) = (q^2 + q) \cdot q = \frac{1}{2\sqrt{2}} n^{3/2} + \frac{1}{8} n - O(\sqrt{n}) \approx \left(\frac{n}{2}\right)^{3/2}.$$

Now some might be happy, since the number of edges in this construction not only beats the random construction, but matches the order of magnitude of the upper bound. Ever discontent, however, we can still choose to be unsatisfied by the constant factor of the leading term being smaller than the $\frac{1}{2}$ from Theorem 2.1.

To fix that, we might have a couple of ideas.

Idea 1. We feel that there is some loss in terms of the number of edges when two lines do *not* intersect. In an affine plane this does happen when the two lines are parallel. Taking a projective plane $PG(q, 2)$ instead, where *every two* lines intersect, might be a good idea.
It turns out that this idea improves only the second order term.

Idea 2. The property of $K_{2,2}$-freeness comes from the structure of lines and points, which in turn depends on solving systems of linear equations. Since points and lines are coordinatized by elements of the same set ($\mathbb{F}_q^2 \setminus \{(0,0)\}$ or $PG(q, 2)$), we do not necessarily have to distinguish between them, and still encounter the same linear equation systems (and hence $K_{2,2}$-freeness). Pairwise identifying points and lines based on their coordinates cuts the number of vertices in half, but keeps the degrees essentially intact. This will lead to an improvement of the constant factor of the leading term.

Before we utilize these two observations to get a better construction, let us recall some basic notions about projective planes, in particular the projective plane $PG(q, 2)$ over the q-element field consisting of a set $\mathcal{L}$ of lines and a set $\mathcal{P}$ of points. Recall that $\mathbb{F}_q^*$ denotes the nonzero elements of $\mathbb{F}_q$.

Points: The set of points $\mathcal{P}$ are the equivalence classes of $\mathbb{F}_q^3 \setminus \{(0,0,0)\}$, where two triples are equivalent if they are non-zero constant multiples of each other,

$$[x_0, x_1, x_2] := \left\{ (cx_0, cx_1, cx_2) \in \mathbb{F}_q^3 \setminus \{(0,0,0)\} \ : \ c \in \mathbb{F}_q^* \right\} .$$

Hence each equivalence class contains $q-1$ triples and the number of points in the projective plane is $\frac{q^3-1}{q-1} = q^2 + q + 1$.

Lines: Given a triple $(a_0, a_1, a_2) \in \mathbb{F}_q^3 \setminus \{(0,0,0)\}$ we define the line $L(a_0, a_1, a_2)$ as follows:

$$L(a_0, a_1, a_2) := \left\{ [x_0, x_1, x_2] \in \mathcal{P} \ : \ a_0x_0 + a_1x_1 + a_2x_2 = 0 \right\} .$$

This definition makes sense, since containment in $L(a_0, a_1, a_2)$ obviously does not depend on which representative of $[x_0, x_1, x_2]$ we use in the equation. The set of lines $\mathcal{L}$ in $PG(q,2)$ consists of all the lines $L(a_0, a_1, a_2)$ defined above. Of course $L(a_0, a_1, a_2) = L(ca_0, ca_1, ca_2)$ for every $c \in \mathbb{F}_q^*$. So the number of lines is also $q^2 + q + 1$.

Remark: The points and lines of a projective plane can conveniently be pictured as the 1- and 2-dimensional linear subspaces, respectively, of the the affine 3-space (that is, the lines and planes going through the origin in $\mathbb{F}_q^3$). Any two planes going through the origin intersect in a line, providing the crucial property that any two lines of our projective plane intersect in a point.

The projective plane $PG(q,2)$ has the following properties:

- There are $q+1$ lines through a point.
- Every line has $q+1$ points on it.
- Every pair of lines intersect in exactly one point.
- For every pair of points there is exactly one line containing both.

Exercise 2.1 *Prove the above properties of* $PG(q,2)$.

Let us now return to improving Construction 0.

Construction 1. (Eszter Klein; 1938) Utilizing Idea 1., let us first consider the point/line incidence graph G of the projective plane $PG(q,2)$. Then G is a $K_{2,2}$-free, $(q+1)$-regular bipartite graph with partite sets of order $q^2 + q + 1$. That is $v(G) = 2(q^2 + q + 1)$ and

$$e(G) = \underbrace{(q^2 + q + 1)}_{n/2}(q+1) = \frac{n}{2}\left(\sqrt{\frac{n}{2} - \frac{3}{4}} + \frac{1}{2}\right) = \left(\frac{n}{2}\right)^{3/2} + \frac{n}{4} - O(\sqrt{n}).$$

One can see that there is a slight improvement in the second order term compared to our original construction.

Exercise 2.2 *Define $ex(n, m, H)$ as the largest number e, such that there is an H-free bipartite graph with partite sets of size n and m, respectively, containing e edges. Show that $ex(q^2+q+1, q^2+q+1, K_{2,2}) = (q^2+q+1)(q+1)$ for every prime power q, i.e. the above graph is* an *optimal construction.*

Construction 2. Utilizing Idea 2, we can make our original graph non-bipartite the following way. Every line defined in Construction 0 which has a non-zero constant term a_3 can be brought to the normal form $a_1x_1 + a_2x_2 = 1$, where $(a_1, a_2) \neq (0, 0)$. Contracting the pairs of vertices which represent the line and the point corresponding to the same pair $(a_1, a_2) \neq (0, 0)$, we obtain a graph where all the previous point/line incidence relations are transformed into an edge. After deleting multiple edges and loops, all the degrees become (essentially) the same as in Construction 0, but the number of vertices is cut in half. The $K_{2,2}$-freeness of Construction 0 is preserved because for any pair (a_1, a_2) and (a_1', a_2') of vertices, the number of common neighbours (x_1, x_2) is determined by the very same linear equation system as in Construction 0.

Formally, define a graph G with vertex set $\mathbb{F}_q^2 \setminus \{(0, 0)\}$, where the distinct vectors $(a_1, a_2), (b_1, b_2) \in V(G)$ are adjacent if

$$a_1b_1 + a_2b_2 = 1.$$

This is a $K_{2,2}$-free graph with $v(G) = q^2 - 1$ vertices. For the degree one can see that for any (a_1, a_2), $a_1x_1 + a_2x_2 = 1$ has exactly q solutions (x_1, x_2) (each of them non-zero). This would mean that the graph is q-regular, but this is not quite so. One of the solutions could be equal to (a_1, a_2), which would correspond to a *loop* in G. Since we are interested in simple graphs, these edges are not allowed. In any case, every vertex has degree at least $q - 1$, so the number of edges in our graph is

$$e(G) \geq \frac{1}{2}(q^2 - 1)(q - 1) = \frac{n^{3/2}}{2} - \frac{n}{2} + O(\sqrt{n}).$$

Exercise 2.3 *Prove that there are exactly $q - 1$, or $q + 1$ loops in G depending on whether or not -1 is a quadratic residue in $\mathbb{F}_q$. Conclude that the number of edges in our graph is in fact $n^{3/2}/2 - O(\sqrt{n})$.*

We should be extremely happy at this point, since this construction matches our upper bound even in the multiplicative constant of the leading term—we have an asymptotically correct answer!

Of course there are always those for whom *nothing* is good enough... And indeed, why not shoot for the moon and find *the* best construction? For that we need to combine Ideas 1 and 2.

Construction 3. (Brown, 1966; Erdős, Rényi, T. Sós, 1966) Similarly to Construction 2, we again contract pairs of vertices, but this time of of Construction 1. Formally, we define the so-called *polarity graph* G on the points of the projective plane $PG(q, 2) = (\mathcal{P}, \mathcal{L})$ in the following way.

Let $V(G) = \{[x_0, x_1, x_2] \in \mathcal{P}\}$ and $E(G) = \{\{x, y\} : x \neq y, x_0y_0 + x_1y_1 + x_2y_2 = 0\}$. By the properties of the projective plane, discussed above, the polarity graph is $K_{2,2}$-free, $n = q^2 + q + 1$ and the degree of any vertex in G is either $q + 1$ or q (because of possible loops). In the next exercise we will see that there are exactly $q + 1$ vertices of degree q. Thus the number of edges of G is

$$e(G) = \frac{1}{2}(q^2 + q)(q + 1) = \frac{1}{2}(n - 1)\left(\sqrt{n - \frac{3}{4}} + \frac{1}{2}\right) = \frac{1}{2}n^{3/2} + \frac{1}{4}n + O(\sqrt{n}).$$

Exercise 2.4 *Show that the polarity graph indeed contains exactly $q + 1$ vertices of degree q.*

Füredi [**?**]showed that

$$ex(q^2 + q + 1, K_{2,2}) = \frac{1}{2}q(q + 1)^2$$

for every prime power $q > 13$. Hence the polarity graph is *an* optimal construction. The understanding of the exact value of $ex(n, K_{2,2})$, when $n \neq q^2 + q + 1$ is quite sparse. For small n, up to 31, extensive computer searches have found the precise value of $ex(n, K_{2,2})$.

2.1.4 The upper bound on $ex(n, K_{t,s})$ and an application

A famous question of Erdős is the following: How many unit distances can n points in the plane determine? (This plane now is the usual Euclidean plane.) Erdős offered $500 for a solution [**?**].

Initial attempts at a construction might easily result in a linear number of pairs of points at unit distance. So at first sight even the existence of a point set with $nf(n)$ unit distances, where $f(n) \to \infty$, is surprising. The best known constructions have $f(n) = n^{\frac{c}{\log\log n}}$. However, it is conjectured by Erdős that there cannot be substantially more than linearly many pairs at unit distance: $n^{1+\varepsilon}$ should be an upper bound for any constant $\varepsilon > 0$.

For a point set P in the plane let us define its *unit-distance graph*, which has P as its vertex set and two points adjacent if their distance is 1. The problem of Erdős is then asking for the maximum number of edges a unit-distance graph with n vertices can have.

Remark: Another notorious open problem on the unit distance graph concerns its chromatic number when the vertex set is the whole Euclidean plane. Solving this problem is even worth 1000 dollars [**?**].

Exercise 2.5 *Prove that the chromatic number of the unit distance graph of the plane lies between 4 and 7.*
Remark. *These are the best known bounds.*

The neighbourhood of a vertex v in the unit distance graph lies on a unit circle around v. Two unit circles intersect in at most two points, thus for every two points in the plane, there are at most two common neighbours in the unit distance graph. Hence the unit distance graph of the whole plane (an infinite graph!) does not contain $K_{2,3}$ as a subgraph. This property will enable us to give an upper bound on the number of edges of unit distance graphs on finitely many vertices.

To this end we are going to give an upper bound on the number of edges of a graph that does not contain a $K_{t,s}$.

Theorem 2.2 (Kővári, T. Sós, Turán; 1954) *Let $s \geq t \geq 2$ be arbitrary integers. Then*

$$ex(n, K_{t,s}) \leq \frac{1}{2}(s-1)^{1/t} n^{2-1/t} + \frac{1}{2}(t-1)n \approx c_{s,t} n^{2-\frac{1}{t}}.$$

Proof. Let G be a graph containing no $K_{t,s}$. We are going to double-count the number of t-stars (i.e., $K_{1,t}$) in G. For every vertex v there are $\binom{d(v)}{t}$ t-stars centred at v. On the other hand, for every t-tuple T there are at most $(s-1)$ t-stars that have T as endpoints, since otherwise G would contain a $K_{t,s}$. Thus, we obtain

$$\binom{n}{t}(s-1) \geq \ \#\text{ of } K_{1,t} \text{ in } G = \sum_v \binom{d(v)}{t}.$$

We would like to apply Jensen's inequality, but there is a slight technical problem, since the function $x \to \binom{x}{t} = \frac{x(x-1)\cdots(x-t+1)}{t!}$ is not convex in the interval $[0,\infty)$, only for $x \geq t-1$. For that we assume that G is a maximal $K_{t,s}$-free graph, which implies $\delta(G) \geq t-1$. Then we have

$$(s-1)\frac{n^t}{t!} \geq \sum_v \binom{d(v)}{t} \geq n\binom{\bar{d}(G)}{t} \geq n\frac{(\bar{d}(G)-(t-1))^t}{t!},$$

From this we obtain $\bar{d}(G) \leq (s-1)^{1/t} n^{1-1/t} + t - 1$, which implies the theorem. □

In particular, we get $ex(n, K_{2,3}) \leq cn^{3/2}$, which means that we cannot expect to have a unit-distance graph with more than $cn^{3/2}$ edges. The best known upper bound [?, ?] in this problem of Erdős stands at $cn^{4/3}$. Any improvement would be extremely interesting.

Exercise 2.6 *(Unit distance problem in the euclidean 3-space.) Show that for any set P of n points in $\mathbb{R}^3$ there are at most $O(n^{5/3})$ pairs of points in $\binom{P}{2}$ that have Euclidean distance exactly one.*

Remark. *The best known upper bound on the number of unit distances in the eucledian 3-space is $O(n^{3/2})$ (due to Zahl [?], and Kaplan, Matoušek, Safernová, Sharir [?]). A point set with $n^{4/3}\log\log n$ pairs of unit distances was constructed by Erdős [?] in 1960.*

Exercise 2.7 *What happens in $\mathbb{R}^4$? How many unit-distances can there be among n points in $\mathbb{R}^4$?*

Obviously, it is "easier" to avoid $K_{2,3}$s than $K_{2,2}$s: any $K_{2,2}$-free graph is $K_{2,3}$-free as well. Even more, adding an edge to a $K_{2,2}$-free graph obviously produces a $K_{2,3}$-free graph, so $ex(n, K_{2,3})$ is strictly larger than $ex(n, K_{2,2})$. It is larger not by much though: according to Theorem 2.2 they are of the same order of magnitude.

The following conjecture predicts that a similar phenomenon is true for arbitrary pairs of integers $s \geq t \geq 2$: the upper bound of Kővári, T. Sós and Turán should be essentially tight and thus the order of magnitude of $ex(n, K_{t,s})$ depends only on the smaller of the parameters.

Conjecture 4 *For any two fixed integers $s \geq t \geq 2$*

$$ex(n, K_{t,s}) = \Theta\left(n^{2-\frac{1}{t}}\right).$$

In Sections 2.2 and 2.3 we will investigate further the status of this conjecture for $\min\{t, s\} \geq 3$, but first let us finish off the case when $\min\{t, s\} = 2$.

2.1.5 A lower bound for $ex(n, K_{2,s})$

In the last subsection we proved an upper bound on $ex(n, K_{t,s})$ in which the exponent of n depended only on the smaller of the two parameters t and s, namely for $t \leq s$ we had

$$ex(n, K_{t,s}) \leq \frac{1}{2}(s-1)^{1/t}n^{2-1/t} + \frac{1}{2}(t-1)n.$$

In particular if a $K_{t,s}$-free graph is found with $\Theta(n^{2-1/t})$ edges then $ex(n, K_{t,s'})$ is determined up to a constant factor for *every* $s' \geq s$. This is the case for $t = 2$, where we have that

$$\frac{1}{2}n^{3/2} \lessapprox ex(n, K_{2,2}) \leq ex(n, K_{2,3}) \leq \cdots \leq ex(n, K_{2,s}) \lessapprox \frac{1}{2}\sqrt{s-1}n^{3/2}.$$

Füredi generalized the affine construction (Construction 2) we presented in Subsection 2.1.3 to establish the asymptotics of $ex(n, K_{2,s})$ for every s. He introduced an equivalence relation on the vertex set, modified the definition of adjacency accordingly and thus reduced the number of vertices by a constant factor while keeping the degrees intact.

Theorem 2.3 (Füredi, 1996)

$$ex(n, K_{2,s}) \gtrapprox \frac{1}{2}\sqrt{s-1}n^{3/2}.$$

Proof. We choose a prime p such that $p-1$ is divisible by $s-1$ (see the Remark after the proof). Recall that the multiplicative group $\mathbb{F}_p^* = \{1, 2, \ldots, p-1\}$ of the p-element field is cyclic. Hence there is a unique subgroup $H < \mathbb{F}_p^*$ of order $|H| = s-1$. (If g is a generating element of $\mathbb{F}_p^*$, then $g^{\frac{p-1}{s-1}}$ generates H.)

On the set $\mathbb{F}_p^2 \setminus \{(0,0)\}$ we define the relation $\sim$ by

$$(a,b) \sim (a',b') \quad \text{if} \quad \exists h \in H \ : \ (a',b') = (ha, hb).$$

This is obviously an equivalence relation since H is a subgroup, and it partitions the set $\mathbb{F}_p^2 \setminus \{(0,0)\}$ into equivalence classes. These classes will be the vertices of our graph G. Note that each class has $s-1$ elements. We denote by $\langle a,b \rangle$ the equivalence class containing (a,b) and formally define

$$V(G) = \{\langle a,b \rangle : (a,b) \in \mathbb{F}_p^2 \setminus \{(0,0)\}\}.$$

We have $|V(G)| = n = \frac{p^2-1}{s-1}$.

Now consider two vertices $\langle a,b \rangle \neq \langle a',b' \rangle$. We define that $\langle a,b \rangle$ and $\langle a',b' \rangle$ are adjacent in G if $aa' + bb' \in H$.

Note that for $s = 2$ the definition of G agrees with Construction 2 of the previous section. Observe the change in the definition of an edge: we not only allow $aa' + bb'$ to be equal to 1 (as in Construction 2), but also to any other h from the subgroup H. This is not solely for the purpose of achieving a larger degree but also *necessary* for having the adjacency definition consistent with the equivalence classes. Indeed, if $\langle a,b \rangle$ and $\langle x,y \rangle$ are adjacent, and we also have $(a,b) \sim (a',b')$ and $(x,y) \sim (x',y')$, then $ax + by = h \in H, a'a^{-1} = b'b^{-1} = h' \in H, x'x^{-1} = y'y^{-1} = h'' \in H$ implying $a'x' + b'y' = hh'h'' \in H$, that is, the adjacency relation is well-defined.

How many neighbors does a vertex $\langle a,b \rangle$ have? Exactly the number of pairs (x,y) for which $ax + by \in H$, divided by $s-1$. One of the coordinates of (a,b) is nonzero, say $b \neq 0$. Then for any fixed $x \in \mathbb{F}_p$ and $h \in H$ there is a unique $y = \frac{h-ax}{b}$ satisfying $ax + by = h$. One can select x in p ways, h in $s-1$ ways, so the total number of solutions (x,y) of $ax + by \in H$ is $p(s-1)$. Since solutions come in equivalence classes of size $s-1$, we obtain that for $\langle a,b \rangle$ there are exactly p equivalence classes $\langle x,y \rangle$ with $ax + by \in H$. One of these might be equal to $\langle a,b \rangle$ itself, if $a^2 + b^2$ happens to be in H. This would constitute a loop in our graph; however, in any case every vertex has degree at least $p-1$.

It remains to show that G is $K_{2,s}$-free. To this end consider two distinct vertices $\langle a,b \rangle \neq \langle a',b' \rangle$. All common neighbors $\langle x,y \rangle$ of $\langle a,b \rangle$ and $\langle a',b' \rangle$ satisfy

$$\begin{aligned} ax + by &= h, \\ a'x + b'y &= h', \end{aligned}$$

for some $h, h' \in H$. Let us now fix h and h'. We distinguish two cases.

Case 1. The matrix of the above system of linear equations is nonsingular. Then there exists a unique solution (x,y).

Case 2. The matrix is singular. Then there exists λ such that $a = \lambda a'$ and $b = \lambda b'$. Note that since $\langle a,b \rangle \neq \langle a',b' \rangle$ the pairs (a,b) and (a',b') are not equivalent, hence λ cannot be an element of H. If we multiply the second equation by λ and subtract it from the first, we get $0 = h - \lambda h'$, or $\lambda = h(h')^{-1} \in H$, which is a contradiction. Hence if Case 2 applies then there are no solutions.

So for each fixed $h, h' \in H$ we have at most one solution (x, y). Solutions again come in equivalence classes of $s-1$ elements, giving that the total number of common neighbors of $\langle a, b\rangle$ and $\langle a', b'\rangle$ is at most $\frac{(s-1)^2}{s-1} = s-1$.

We thus constructed a $K_{2,s}$-free graph with

$$|E(G)| \geq \frac{1}{2}n(p-1) \geq \frac{1}{2}\sqrt{s-1}n^{3/2} - \frac{1}{2}n.$$

□

Remark:

1. Note that in the above construction, the number of vertices is of very special form, it is $\frac{p^2-1}{s-1}$ where p is a prime congruent to 1 modulo $s-1$. In order to provide a construction for *every* n, which then gives the correct asymptotics of $ex(n, K_{2,s})$, we need that the primes chosen in the proof of Theorem 2.3 are *frequent enough* for every s. Dirichlet's classic theorem states that in every infinite arithmetic progression, satisfying the obvious necessary condition, there are infinitely many primes. More precisely, for any pair (k, l) of relatively prime integers there are infinitely many primes of the form $ak + l$, $a \in \mathbb{Z}$. The density version of Dirichlet's Theorem also ensures that the number of these primes up to a large integer N is exactly what one expects according to the Prime Number Theorem, namely $\approx \frac{1}{\phi(k)}\frac{N}{\log N}$, where $\phi(k)$ is Euler's function, the number of positive integers up to k that are relatively prime to k. However, for us this is still not enough: we need that in *every small enough interval there is such a prime*. Such type of statements are the topic of intensive research in number theory. In particular by a result of Huxley and Iwaniec (1975) for every sufficiently large n there exists a prime $p \equiv 1 \pmod{s}$ such that $\sqrt{sn} - n^{1/3} < p < \sqrt{sn}$.

2. Selecting $s-1 = p-1$ would give us that *any* nonzero multiple of (a, b) is equivalent with it, that is $V(G)$ is the 1-dimensional projective space (line). The defined graph is the clique, which is then $K_{2,p}$-free indeed, since the projective line contains only $p+1$ points.

3. From the analysis of the two cases in the proof of Theorem 2.3 it could seem at first sight that, since Case 2 leads to a contradiction, every two vertices have exactly $s-1$ common neighbors. This is false however, and not only because of the presence of the loops, but also because Case 2 leads to a contradiction only if there was a common neighbor in the first place. It could happen, though rarely, that two vertices do not have a common neighbor at all. This occurs only if $(a, b) = (\lambda a, \lambda b)$ with some $\lambda \notin H$.

2.2 Forbidding $K_{3,s}$

The value of $ex(n, K_{t,s})$ for $t = 2$ is fully settled now up to lower order terms. Before going further to $t \geq 3$ we repeat our randomized approach for obtaining a dense H-free graph for arbitrary H. In particular, for $H = K_{t,s}$ we will check how the random attempt relates to the Kővári-Sós-Turán upper bound.

2.2.1 A quick detour to random graphs

The plan is again to find a probability p as large as possible such that in the random graph $G(n,p)$ we have

$$\mathbf{E}[e(G(n,p)) - \#\{\text{copies of } H \text{ in } G(n,p)\}] \geq \frac{1}{2}\binom{n}{2}p.$$

This definitely holds if

$$\frac{1}{2}\binom{n}{2}p \geq n^{n(H)}p^{e(H)} \geq \mathbf{E}[\#\{\text{copies of } H \text{ in } G(n,p)\}],$$

or if $p = O\left(n^{-\frac{n(H)-2}{e(H)-1}}\right)$.

Hence, we know that with some $p = cn^{-\frac{n(H)-2}{e(H)-1}}$ there exists a graph G such that after deleting one edge from every copy of H there are still $\frac{1}{2}\binom{n}{2}p = \Theta\left(n^{2-\frac{n(H)-2}{e(H)-1}}\right)$ edges left. We proved the following.

Proposition 2.4 *For any H,*

$$ex(n,H) = \Omega\left(n^{2-\frac{n(H)-2}{e(H)-1}}\right).$$

Applying this for $H = K_{t,s}$ we obtained a $K_{t,s}$-free graph with $\Theta\left(n^{2-\frac{t+s-2}{ts-1}}\right)$ edges. Observe, however, that $\frac{t+s-2}{ts-1}$ is *strictly* larger than $\frac{1}{t}$ for any $s \geq t \geq 2$. Hence the order of the random lower bound on $ex(n,K_{t,s})$ is *always smaller* than the order of the Kövári-Sós-Turán upper bound. So we are still out there looking for those needles in the haystacks...

2.2.2 An infinite construction

Similarly to our $K_{2,2}$-free construction we start our investigation of dense $K_{3,3}$-free graphs by an infinite geometric construction providing a good heuristic.

Motivated by our approach to the unit distance problem in the euclidean 3-space (see Exercise 2.6) we take the points of $\mathbb{R}^3$ as the vertices of our graph, and make two vertices connected with an edge if their distance is exactly one. That is, with the notation of the previous section, we consider the unit-distance graph of the 3-dimensional euclidean space. The set of neighbors of a vertex is the set of all points on the unit sphere centered in that vertex. If there were a $K_{3,3}$ in the graph, then three unit spheres would intersect in at least three points. The intersection of two unit spheres (unless it is empty) is a circle with radius strictly less than one (this radius might be 0, in case the spheres intersect in only one point). A third unit sphere can intersect this circle in at most two points, proving that G does not contain a $K_{3,3}$.

The unit sphere minus a point is homeomorphic to the plane, so the degree of each vertex is of order $|\mathbb{R}|^2$. Since $n(G) = |\mathbb{R}|^3$, the number of edges in the graph is roughly $\frac{1}{2}n^{5/3}$.

2.2.3 A finite construction

In this section we "finitize" the above heuristic, but the transformation will be more problematic than it was in the case of the $K_{2,2}$-free construction. There it was clear right away what caused that the graph is $K_{2,2}$-free: two lines intersect in at most one point or a linear equation system has at most one solution. These are all properties carrying over to any field. Here it is not clear at this point what property (or rather imperfectness) of the field $\mathbb{R}$ causes that three unit spheres intersect in at most two points.

Theorem 2.5 (Brown, 1966)

$$ex(n, K_{3,3}) \gtrapprox \frac{1}{2}n^{5/3}.$$

Proof. We set $V(G) = \mathbb{F}_p^3$. For every $a \in \mathbb{F}_p^3$ we define

$$S_\alpha(a) = \{x \in \mathbb{F}_p^3 \ : \ (x_1 - a_1)^2 + (x_2 - a_2)^2 + (x_3 - a_3)^2 = \alpha\},$$

where $\alpha \in \mathbb{F}_p$ is a constant to be determined later. We define two vertices a and b to be adjacent in G if $b \in S_\alpha(a)$. Obviously the adjacency relation is symmetric.

Is G a $K_{3,3}$-free graph? Suppose not. Assume for contradiction that there exist distinct vertices $a, b, c \in V(G)$ with $|S(a) \cap S(b) \cap S(c)| \geq 3$. For $x \in S(a) \cap S(b) \cap S(c)$ we have

$$\begin{aligned}
(x_1 - a_1)^2 + (x_2 - a_2)^2 + (x_3 - a_3)^2 &= \alpha, && (2.1)\\
(x_1 - b_1)^2 + (x_2 - b_2)^2 + (x_3 - b_3)^2 &= \alpha, && (2.2)\\
(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 &= \alpha. && (2.3)
\end{aligned}$$

One obtains that for at least three different x,

$$\begin{aligned}
\text{(2.2) - (2.1):} \quad & 2x_1(a_1 - b_1) + 2x_2(a_2 - b_2) + 2x_3(a_3 - b_3) + \\
& + b_1^2 - a_1^2 + b_2^2 - a_2^2 + b_3^2 - a_3^2 = 0,\\
\text{(2.3) - (2.2):} \quad & 2x_1(b_1 - c_1) + 2x_2(b_2 - c_2) + 2x_3(b_3 - c_3) + \\
& + c_1^2 - b_1^2 + c_2^2 - b_2^2 + c_3^2 - b_3^2 = 0,\\
\text{(2.1) - (2.3):} \quad & 2x_1(c_1 - a_1) + 2x_2(c_2 - a_2) + 2x_3(c_3 - a_3) + \\
& + a_1^2 - c_1^2 + a_2^2 - c_2^2 + a_3^2 - c_3^2 = 0
\end{aligned}$$

holds.

The matrix of the last system of equations is

$$A = \begin{pmatrix} a_1 - b_1 & a_2 - b_2 & a_3 - b_3 \\ b_1 - c_1 & b_2 - c_2 & b_3 - c_3 \\ c_1 - a_1 & c_2 - a_2 & c_3 - a_3 \end{pmatrix}.$$

One can easily see that the rank of A is either 1 or 2. If $\mathrm{rk}(A) = 2$, then all solutions of the system lie on a line (we already assumed that there exists a solution). On the other

hand, if $\mathrm{rk}(A) = 1$, then a, b, c lie on a line. Therefore, if there is a $K_{3,3}$ with partite classes $\{a, b, c\}$ and $\{d, e, f\}$ in G, then either a, b, c lie on a line, or d, e, f lie on a line.

Could it happen that $S_\alpha(a)$, for some α and some a, contains three points from a line? Or maybe even the full line? A sphere??? — By translation, we can assume that such a special line, containing three points of the sphere $S_\alpha(a)$, also passes through the origin. Let the line consist of the points τv where $v \in \mathbb{F}_p^3 \setminus \{(0,0,0)\}$ is fixed and $\tau \in \mathbb{F}_p^*$ is arbitrary. By our assumption

$$\tau^2 \sum_i v_i^2 - 2\tau \sum_i v_i a_i + \sum_i a_i^2 = \alpha$$

holds for at least three τ. This is only possible when

$$\sum_i v_i^2 = 0, \quad \sum_i v_i a_i = 0, \quad \sum_i a_i^2 = \alpha.$$

But then, assuming w.l.o.g. that $v_1 \neq 0$, we get

$$\begin{aligned} v_1^2\alpha &= v_1^2 \sum_i a_i^2 = v_1^2 a_1^2 + v_1^2(a_2^2 + a_3^2) = (-v_2a_2 - v_3a_3)^2 + v_1^2(a_2^2 + a_3^2) \\ &= (-v_2a_2 - v_3a_3)^2 + (-v_2^2 - v_3^2)(a_2^2 + a_3^2) \\ &= v_2^2a_2^2 + 2v_2a_2v_3a_3 + v_3^2a_3^2 - v_2^2a_2^2 - v_2^2a_3^2 - v_3^2a_2^2 - v_3^2a_3^2 = -(v_2a_3 - v_3a_2)^2, \end{aligned}$$

and

$$-\alpha = \left(\frac{v_2a_3 - v_3a_2}{v_1}\right)^2.$$

Hence if we select α such that $-\alpha$ is *not a square*, then we arrive at a contradiction and our graph does *not* contain a $K_{3,3}$. Recall that for an odd prime power p, the number of (non-zero) quadratic residues is $\frac{p-1}{2}$ which is the same as the number of quadratic non-residues.

Exercise 2.8 *Let k be an arbitrary field. Prove that if $-\alpha \in k$ is a square, then the corresponding sphere-graph (defined in the 3-dimensional space over k) not only contains a $K_{3,3}$, but also a $K_{n^{1/3},n^{2/3}}$. (In case k is an infinite field we mean a $K_{|k|,|k|}$.) In particular, our heuristics would fail badly in the complex 3-space.*

To finish the proof of Theorem 2.5 we still need to see that the Brown-graph contains enough edges, i.e., the sphere $S_\alpha(a)$ (the neighborhood of vertex a) contains enough vertices — for a well-chosen α. For the $K_{2,2}$-free graphs of Subsection 2.1.3 this was a piece of cake, since to count edges we just had to solve linear equations. Here we have some difficulty, since the equations are quadratic. Of course intuitively we feel that the cardinality of $S_\alpha(a)$ is around $|\mathbb{F}_p|^2$ for any α, as $S_\alpha(a)$ is a *surface* in the three-dimensional space over $\mathbb{F}_p$.

Even more, let us take the average over all possible "radii" α. For any $a \in \mathbb{F}_p^3$ we have

$$\sum_{\alpha \in \mathbb{F}_p} |S_\alpha(a)| = p^3,$$

so by averaging there is at least one "radius" α for which $|S_\alpha(a)| \geq p^2$. For Brown's construction to work we need this for a somewhat special α: for which $-\alpha$ is not a square. These αs are abundant, half of the nonzero elements are such. Still, theoretically, it is possible that for all "bad radii" the corresponding spheres contain $2p^2$ points, while for all "good radii" the spheres are empty. In the following we will see that this is not the case and even under much more general circumstances, the number of points on a surface does not deviate much from the expected value. We include a probabilistically motivated proof due to Wolfgang Schmidt. The proof imitates the technique of the second moment method, used frequently in probabilistic combinatorics.

Let us fix positive integers $d_1, \ldots, d_n \in \mathbb{N}$ and consider the following equation

$$a_1 x_1^{d_1} + a_2 x_2^{d_2} + \cdots + a_n x_n^{d_n} = a_0,$$
$$a_0, \ldots, a_n \in \mathbb{F}_p.$$

Let $N(a_0, \ldots, a_n)$ denote the number of solutions $(x_1, \ldots, x_n) \in \mathbb{F}_p^n$ of this equation. We have

$$\sum_{a_1, \ldots, a_n} \sum_{a_0 \in \mathbb{F}_p} N(a_0, \ldots, a_n) = \sum_{a_1, \ldots, a_n} p^n = p^{2n}.$$

Then, the average value of $N(a_0, \ldots, a_n)$ over $(a_0, \ldots, a_n)$ is p^{n-1}. The next theorem shows that the function $N(\cdot)$ *never* deviates from its average by much.

Theorem 2.6 *Let $c_0, c_1, \ldots, c_n \in \mathbb{F}_p \setminus \{0\}$. Then we have*

$$|N(c_0, \ldots, c_n) - p^{n-1}| \leq \left[\left(\frac{p}{p-1} \right)^{n/2} \prod_{i=1}^{n} (d_i, p-1) \right] p^{(n-1)/2}.$$

Remark: 1. Estimating the number of solutions of higher degree equations over finite fields is a classic and well-studied area of mathematics full of beautiful ideas and hard theorems. There is an even more general theorem about the number of solutions to high degree equations over finite fields (due to Weil), but its proof well exceeds the possibilities of our course.
2. A more precise and elementary proof of what is needed for the Brown-graph is sought for in Exercise 2.9.
3. Assuming $c_1, \ldots, c_n \neq 0$ is necessary otherwise the theorem is (easily) not valid. The assumption $c_0 \neq 0$ is not crucial, see Exercise 2.10.

Proof. (W. Schmidt) Before starting, let us note that we can assume w.l.o.g. that $d_i | p-1$ for every $i = 1, \ldots, n$, since for every $\gamma \in \mathbb{F}_p$,

$$\#\{x \in \mathbb{F}_p \ : \ x^{d_i} = \gamma\} = \#\{x \in \mathbb{F}_p \ : \ x^{(d_i, p-1)} = \gamma\}.$$

We will look at the following sum, which is appropriate to measure the *average* deviation from the average.

$$\sum_{a_0,\dots,a_n} (N(a_0,\dots,a_n) - p^{n-1})^2$$

For a moment, consider N as a random variable determined by $a_0,\dots,a_n$, which are chosen independently, uniformly at random from $\mathbb{F}_p$. Then the above sum (divided by p^{n+1}) is its variance $var(N)$, measuring how much the values of the random variable can deviate from its average; exactly what we are interested in.

Our plan is first to bound the variance from above, i.e., to show that the deviation of N from p^{n-1} is not too large *on the average*. This still does not say *anything* about the deviation of each individual term of the sum. But then we prove that N takes on only a few (in fact constantly many) values, moreover any such value occurs on a large (i.e., constant) fraction of the $(n+1)$-tuples $(c_0,\dots,c_n)$. Hence we will conclude that $N(c_0,\dots,c_n)$ cannot deviate from p^{n-1} too much, because otherwise N would deviate from p^{n-1} by much on a large fraction of its domain, which would imply that the variance would be too large.

$$\begin{aligned} p^{n+1}\cdot var(N) &= \sum_{a_0,\dots,a_n} (N^2(a_0,\dots,a_n) - 2N(a_0,\dots,a_n)p^{n-1} + p^{2(n-1)}) \\ &= \sum_{a_0,\dots,a_n} N^2(a_0,\dots,a_n) - 2p^{2n}\cdot p^{n-1} + p^{n+1}\cdot p^{2(n-1)} \\ &= \sum_{a_0,\dots,a_n} N^2(a_0,\dots,a_n) - p^{3n-1} \end{aligned}$$

We estimate the above sum of squares similarly to calculating $\sum N(a_0,\dots,a_n)$: by double-counting.

$$\begin{aligned} \sum_{a_0,\dots,a_n} N^2(a_0,\dots,a_n) &= \sum_{a_0,\dots,a_n} \Bigg(\sum_{\substack{x_1,\dots,x_n \\ a_1x_1^{d_1}+\dots+a_nx_n^{d_n}=a_0}} 1 \Bigg) \Bigg(\sum_{\substack{y_1,\dots,y_n \\ a_1y_1^{d_1}+\dots+a_ny_n^{d_n}=a_0}} 1 \Bigg) \\ &= \sum_{x,y\in\mathbb{F}_p^n} \sum_{\substack{(a_0,\dots,a_n) \\ a_1x_1^{d_1}+\dots+a_nx_n^{d_n}-a_0=0 \\ a_1y_1^{d_1}+\dots+a_ny_n^{d_n}-a_0=0}} 1 \end{aligned}$$

The point of exchanging the summation is obvious: now, instead of having equations of high degree, we have systems of *linear* equations (with the x_i and y_i being fixed and the a_i being the variables). We denote the matrix of this homogeneous system by

$$A := \begin{pmatrix} -1 & x_1^{d_1} & \dots & x_n^{d_n} \\ -1 & y_1^{d_1} & \dots & y_n^{d_n} \end{pmatrix}.$$

If $\mathrm{rk}(A) = 2$, then there are p^{n-1} solutions $a = (a_0, \ldots, a_n)$ to $Aa^T = (0,0)^T$.

If $\mathrm{rk}(A) = 1$, then there are p^n solutions $(a_0, \ldots, a_n)$.

Now we only have to count how many times we encounter the second case. How often do we have $\mathrm{rk}(A) = 1$? In other words, for how many vectors $x, y \in \mathbb{F}_p^n$ do we have $x_i^{d_i} = y_i^{d_i}$ for all $i = 1, \ldots, n$? For a fixed i, with $x_i \neq 0$ there are exactly d_i solutions y_i satisfying $x_i^{d_i} = y_i^{d_i}$. If $x_i = 0$, then of course $y_i = 0$ as well. In any case, for any fixed $(x_1, \ldots, x_n)$ we have *at most* $d_1 \cdots d_n$ appropriate $(y_1, \ldots, y_n)$ giving us a matrix A of rank 1.

The sum can then be estimated by

$$\begin{aligned} \sum_{x,y \in \mathbb{F}_p^n} \sum_{\substack{(a_0,\ldots,a_n) \\ a_1 x_1^{d_1} + \cdots + a_n x_n^{d_n} - a_0 = 0 \\ a_1 y_1^{d_1} + \cdots + a_n y_n^{d_n} - a_0 = 0}} 1 &\leq p^{2n} p^{n-1} + p^n d_1 \cdots d_n (p^n - p^{n-1}) \\ &= p^{3n-1} + p^{2n-1}(p-1) d_1 \cdots d_n. \end{aligned}$$

Finally, we can upper bound the variance,

$$p^{n+1} \cdot var(N) = \sum_{a_0,\ldots,a_n} (N(a_0, \ldots, a_n) - p^{n-1})^2 \leq p^{2n-1}(p-1) d_1 \cdots d_n. \tag{2.4}$$

At this point we can conclude that the *average deviation* of $N(a_0, \ldots, a_n)$ from p^{n-1} is at most $\sqrt{d_1 \cdots d_n} p^{(n-1)/2}$. This is roughly just the square-root of p^{n-1}, a promising sign.

Now consider the $(n+1)$-tuple $(c_0, \ldots, c_n)$ from the statement of the theorem. We claim that the variable N equals $N(c_0, \ldots, c_n)$ on a constant (roughly $\frac{1}{d_1 \cdots d_n}$) fraction of its domain $\mathbb{F}_p^{n+1}$. Indeed,

$$N(c_0, c_1, \ldots, c_n) = N(tc_0, tc_1 b_1^{d_1}, \ldots, tc_n b_n^{d_n})$$

for any $t \neq 0$ and $b_1, \ldots, b_n \neq 0$, since the two equations

$$\begin{aligned} c_1 x_1^{d_1} + \cdots + c_n x_n^{d_n} - c_0 &= 0 \\ tc_1 (b_1 x_1)^{d_1} + \cdots + tc_n (b_n x_n)^{d_n} - tc_0 &= 0 \end{aligned}$$

have the same number of solutions. More precisely, $(y_1, \ldots, y_n)$ is a solution of the first equation if and only if $(y_1/b_1, \ldots, y_n/b_n)$ is a solution of the second equation. It remains to determine the number of different $(n+1)$-tuples $(tc_0, tc_1 b_1^{d_1}, \ldots, tc_n b_n^{d_n})$. Each of $t, b_1, \ldots, b_n$ can be chosen in $p-1$ different ways, which gives $(p-1)^{n+1}$ choices altogether. However, some of the $(n+1)$-tuples they give rise to are identical. Such a thing happens if and only if $b_i^{d_i} = b_i'^{d_i}$ for every $i = 1, \ldots, n$ (Here we used that $c_i \neq 0$ for $i = 0, \ldots, n$). For fixed b_i there are exactly d_i such b_i', so the number of different $(n+1)$-tuples $(tc_0, tc_1 b_1^{d_1}, \ldots, tc_n b_n^{d_n})$ is $(p-1)^{n+1}/(d_1 \cdots d_n)$. The fact that all these

$(n+1)$-tuples have the same N-value $N(c_0, c_1, \ldots, c_n)$ combined with inequality (2.4) gives

$$\begin{aligned} \frac{(p-1)^{n+1}}{d_1 \cdots d_n} \left(N(c_0, \ldots, c_n) - p^{n-1}\right)^2 &\leq \sum_{a_0,\ldots,a_n} \left(N(a_0, \ldots, a_n) - p^{n-1}\right)^2 \\ &\leq p^{2n-1}(p-1)d_1 \cdots d_n, \end{aligned}$$

which proves the theorem. □

Corollary 2.7 $|S_\alpha(a)| \geq p^2 - 16p$

Proof. Substitute $n = 3$, $d_1 = d_2 = d_3 = 2$, $a_0 = \alpha$, $a_1 = a_2 = a_3 = 1$ in Theorem 2.6, and use $(\frac{p}{p-1})^{3/2} \leq 2$. □

To complete the proof of Theorem 2.5 note that indeed the Brown graph has the claimed number of edges, since by Corollary 2.7 the degree of each vertex is (roughly) p^2. □

In fact, for the special case of $S_\alpha(a)$ the deviation from the average p^2 can be calculated exactly. Let $QR(p)$ be the set of quadratic residues of $\mathbb{F}_p$.

Exercise 2.9 *Give an elementary proof that for any $a \in \mathbb{F}_p^3$ the sphere $S_\alpha(a)$ contains either $p^2 - p$ or $p^2 + p$ points depending on whether α and -1 are quadratic residues or not. Note that four cases need to be considered.*
For this purpose recall that the equation $x^2 + y^2 = \beta$, where $\beta \neq 0$ is fixed, has $p-1$ solutions $x, y \in \mathbb{F}_p$ if -1 is a quadratic residue in $\mathbb{F}_p$, and $p+1$ solutions if -1 is not a quadratic residue; furthermore, $x^2 + y^2 = 0$ has $2p-1$ solutions if $-1 \in QR(p)$, or 1 single solution if $-1 \in QNR(p)$.
Give a general exact formula for the number of solutions to $x_1^2 + \cdots + x_k^2 = \beta$, for any fixed $k \in \mathbb{N}, \beta \in \mathbb{F}_p$.

Exercise 2.10 *Prove that $N(c_0, c_1, \ldots, c_n)$ cannot deviate by "much" from the average even if $c_0 = 0$. The point is of course how much is "much"? (The assumption that all other $c_i \neq 0$ is still needed, otherwise the problem becomes smaller dimensional and the error term gets be worse.)*

The next exercise exhibits some of the difficulties one faces when stepping from dense $K_{3,3}$-free graphs to $K_{4,4}$-free graphs.

Exercise 2.11 *A natural thought to extend the idea of the Brown graph to $K_{4,4}-$ or $K_{4,1000}-$avoiding dense graphs is the following. Instead of three dimensions let us take four, i.e., our vertex set is $\mathbb{F}_p^4$. Let the neighborhood of a vertex x be determined by a four-dimensional sphere around it, in particular a vertex y is adjacent to x if and only if $\sum_i (y_i - x_i)^2 = 1$. According to Theorem 2.6, our graph has roughly $cn^{7/4}$ edges — the conjectured truth. Prove, however, that this graph contains a $K_{p-1,p-1}$. Prove that even taking a higher degree surface of the form $\sum_i (y_i - x_i)^{1000} = 1$ as the neighborhood of x instead of the sphere would not help us. (Note that Theorem 2.6 ensures that this graph also has roughly the correct number $cn^{7/4}$ of edges.)*

2.2.4 An upper bound

The above construction of Brown could also be called the "unit-distance graph" of $\mathbb{F}_p^3$ if we choose $\alpha = 1$, which we can do if $p \equiv 3 \pmod 4$ (since then -1 is a quadratic non-residue). After this, here is where we are standing in terms of the Turán number of $K_{3,3}$:

$$\frac{1}{2}n^{5/3} \lessapprox ex(n, K_{3,3}) \lessapprox \frac{1}{2}2^{1/3}n^{5/3}.$$

In 1996, Füredi proved that the upper bound can be improved to match the lower bound, so the construction of Brown (which is from 1966) is asymptotically optimal. This is the first (and so far lone) improvement on the classical $(s-1)^{1/t}n^{2-1/t}$ upper bound (which is from 1954). Remember, this simple bound was asymptotically tight for $t = 2$ and s arbitrary.

Before we start proving Füredi's upper bound we state a technical lemma which will be convenient later. Behind its artificial appearance it is quite easy: it only formalizes the same convexity calculation we made when proving the Kővári-Sós-Turán upper bound.

Lemma 2.7.1 *Let $v, k \geq 1$ be integers, and let $c, x_0, x_1, \ldots, x_k \geq 0$ be integers. Then*

$$\sum_{i=1}^{v} \binom{x_i}{k} \leq c\binom{x_0}{k}$$

implies

$$\sum_{i=1}^{v} x_i \leq x_0 c^{1/k} v^{1-1/k} + (k-1)v.$$

Proof. The small technical hurdle in proving this lemma is that the function $\binom{x}{k}$ is not convex on the full interval $[0, \infty)$. Hence we define $\widetilde{\binom{x}{k}}$ to be zero if $x \leq k-1$ and to be equal to $\binom{x}{k}$ otherwise. With this new definition $\widetilde{\binom{x}{k}}$ agrees with $\binom{x}{k}$ on integers, but it is a convex function not only on $\{x : x \geq k-1\}$, but also on $\{x : x \geq 0\}$.

So our assumption can be written as

$$\sum_{i=1}^{v} \binom{\tilde{x_i}}{k} \leq c\binom{x_0}{k}$$

Case 1. If $\sum x_i \leq (k-1)v$, then the statement is clear (the "error term" takes care).
Case 2. Otherwise, let $S = \frac{\sum x_i}{v} > k-1$. Then by convexity and Jensen's inequality

$$v\widetilde{\binom{S}{k}} \leq \sum_{i=1}^{v} \widetilde{\binom{x_i}{k}} \leq c\binom{x_0}{k}.$$

Our assumption for Case 2 ensures that $\widetilde{\binom{S}{k}} = \binom{S}{k}$, so we have

$$v\frac{S(S-1)\cdots(S-k+1)}{k!} \leq c\frac{x_0(x_0-1)\cdots(x_0-k+1)}{k!}.$$

Estimating both sides trivially, we have

$$v(S-k+1)^k \leq cx_0^k.$$

Solving this inequality for S we obtain

$$S \leq c^{1/k}v^{-1/k}x_0 + k - 1,$$

which is exactly what we wanted. □

Let us now turn to the proof of Füredi. Once we decide that Brown's construction should be asymptotically optimal and we try to prove that, we better first find the inaccuracy of the Kővári-Sós-Turán argument. That argument has two estimates: The one applying Jensen's inequality is tight for all graphs that are more or less regular, and the Brown graph is. So the only problem can occur with the crucial combinatorial idea, when we say that in a $K_{3,3}$-free graph each triple of vertices can be the endpoints of at most *two* $K_{1,3}$. This in fact fails badly in Brown's graph, where roughly half of the triples is not the endpoint of *any* $K_{1,3}$ at all! Such a phenomenon is not so surprising once we think back to our motivating infinite $K_{3,3}$-free unit-distance graph in $\mathbb{R}^3$. There the number of neighbors of a triple depends on the radius of its circumcircle: if the radius is less than 1, then there are two common neigbors, if the radius is larger than 1, then there are no common neighbors, and only those "degenerate" triples have one common neighbor whose circumcircle has radius exactly one.

Exercise 2.12 *Prove that in Brown's graph roughly half of the triples has two common neighbors and the other half has none. Even more: describe explicitly those triples of the Brown graph which do not have a common neighbor!*

Keeping this in mind, we will repeat the Kővári-Sós-Turán-argument *only for the triples, that do have a common neighbor*, i.e. have a chance to have two of them. This turns out to be the key observation we need to improve on the upper bound.

Theorem 2.8 (Füredi, 1996)

$$ex(n, K_{3,3}) \lessapprox \frac{1}{2}n^{5/3}.$$

Proof. Suppose G is a $K_{3,3}$-free graph. We fix a vertex $x \in V(G)$. There are $\binom{d(x)}{3}$ triples of vertices in the neighborhood of x. Since G is $K_{3,3}$-free, each of these triples can be fully contained in the neighborhood of at most *one other* vertex. This implies the following inequalities,

$$\sum_{y \neq x} \binom{|N(x) \cap N(y)|}{3} \leq \binom{d(x)}{3}, \quad \forall\, x \in V(G).$$

Applying Lemma 2.7.1 we have

$$\sum_{y \neq x} |N(x) \cap N(y)| \leq d(x)(n-1)^{2/3} + 2(n-1).$$

Summing up over all x we get

$$\sum_{x\in V(G)}\sum_{y\neq x}|N(x)\cap N(y)| \leq \sum_{x\in V(G)}\left[d(x)(n-1)^{2/3}+2(n-1)\right].$$

The left side counts twice the cherries $K_{1,2}$ in G by their endpoints. Counting at their midpoints and introducing $\bar{d}(G)=\sum_{x\in V(G)} d(x)/n$ for the average degree, we have

$$2\cdot\sum_{z\in V(G)}\binom{d(z)}{2} \leq n\bar{d}(G)(n-1)^{2/3}+2(n-1)n.$$

By Jensen's inequality we infer

$$2n\binom{\bar{d}(G)}{2} \leq n\bar{d}(G)(n-1)^{2/3}+2(n-1)n,$$

implying

$$\bar{d}(G)-1 \leq (n-1)^{2/3}+\frac{2(n-1)}{\bar{d}(G)}.$$

If $\bar{d}(G)\leq (n-1)^{2/3}$, then we are done. Otherwise by the above we have

$$\bar{d}(G) \leq (n-1)^{2/3}+\frac{2(n-1)}{\bar{d}(G)}+1 \leq (n-1)^{2/3}+2(n-1)^{1/3}+1.$$

Hence, with $e(G)=n\bar{d}(G)/2$ we concluded the proof of

$$ex(n,K_{3,3}) \leq \frac{1}{2}n^{5/3}+n^{4/3}+\frac{1}{2}n.$$

□

Exercise 2.13 *Improve the KST-upper-bound a bit. Show that for arbirary $s\geq 3$, we have $ex(n,K_{3,s}) \lessapprox \frac{\sqrt[3]{s-2}}{2}n^{5/3}$.*

Exercise 2.14 *Generalize the proof above and show that $ex(n,K_{4,4}) \lessapprox \frac{1}{2}n^{7/4}$. (Hint: Instead of lower bounding $\sum\binom{x_i}{3}$ in terms of $(\sum x_i)^3$ (which follows from the convexity of $\binom{x}{3}$) you might want to bound it from below in terms of the product of $\sum\binom{x_i}{2}$ and $\sum x_i$.)*

Open Problem. The asymptotics of $K_{3,s}$ is not known for any $s>3$. There are infinitely many values of s for which the upper and lower bounds are within a constant factor of $\sqrt[3]{2}$ of each other (we will discuss these results later), but there are also infinitely many values s where this constant factor separation is $\sqrt[3]{s-2}$.
Any improvement would be very interesting. The value of $ex(n,K_{3,4})$ is the first unknown.

2.3 Forbidding $K_{t,s}$

We have seen now many dense $K_{t,s}$-free constructions, but none so far where both parameters t and s are at least 4. While the Turán numbers of $K_{2,s}$ and $K_{3,s}$ are well-understood, the densest $K_{4,4}$-graph we know is essentially the Brown graph. This graph does not even have a $K_{3,3}$, consequently has only $\Theta(n^{5/3})$ edges—very far from the KST-upper bound of $n^{7/4}$ for $K_{4,4}$.

So what does become so hard when t and s are both at least 4? Phrasing it mysteriously, besides us being not creative enough, the problem is that $4 = 2 + 2$.

For $t \leq s$, it is natural to choose a $K_{t,s}$-free construction to live in some (affine or projective) t-dimensional space over a finite field $\mathbb{F}_q$, at least this what we have done so far. The vertex set is typically the space itself and the neighborhood of each vertex is chosen to be some hypersurface, so that the number of edges does work out well. One of course has to prove that these $(t-1)$-dimensional surfaces contain the appropriate number, $\approx q^{t-1} \approx n^{1-\frac{1}{t}}$, of points; this is sometimes easier, sometimes harder, but could always be done. Then one must also show that the intersection of any t of these $\approx q^t$ neighborhood hypersurfaces contains at most $s-1$ points. The exercises leading up to this section tried to demonstrate that this is the more problematic issue, and in fact the first and most critical point is whether these t-wise intersections contain at most *constantly many* points. Then it is usually only a bonus that by what one is able to bound this constant, let it be $t-1$ or something else. In Exercise ?? and Exercise 2.11 we investigated promising constructions, which broke down in a strong way: we found complete bipartite graphs whose order tended to infinity with the order of the graph. Part of the problem was degeneracies: a line or some low degree curve was part of several neighborhood surfaces.

Neighborhoods are $t-1$ dimensional hypersurfaces, common neighborhoods are the intersections of these. Two "average" hypersurfaces intersect in a $(t-2)$-dimensional surface, and taking one more average hypersurface to the intersection always reduces the dimension by one. Hence when one takes t average hypersurfaces in t-space one expects that their intersection is 0-dimensional, that is the union of constantly many points. Our experience shows that for $t \geq 4$ it is hard to select q^t hypersurfaces in the t-dimensional space such that *any* t of them has a 0-dimensional intersection.

In the first section we pursue the idea of choosing the neighborhood hypersurfaces randomly. We will succeed in formalizing the intuitive idea that the intersection of t "avergae" hypersurfaces has a 0-dimensional, and thus constant size, intersection.

In the two subsequent sections we introduce two related explicit constructions that give better estimates on the intersection size.

2.3.1 Random algebraic constructions

We start by revisiting the simplest random construction and take a closer look at why it fails.

Since the constant factor of the number of edges will not play any role in this section,

we prefer to work in the bipartite random graph model $G(n,n,p)$, that allows a technically cleaner treatment. Let L and R be two disjoint sets of vertices of size n each. For each pair $u \in L$ and $v \in R$ of vertices we put an edge with probability $p = n^{-\frac{1}{t}}$, with the choices being mutually independent.

We choose this probability so that the number of edges in expectation n^2p, will be of the order $n^{2-\frac{1}{t}}$—the order we expect to be the right one for the Turán number of the complete bipartite graph $K_{t,s}$ for any $t \leq s$.

We already convinced ourselves that at such high probability the random graph will be filled with copies of $K_{t,t}$ and even the alteration methods works only for much smaller edge probability.

Let us fix a set $U \subseteq L$ of size t and study the distribution of the size of the common neighborhood $N(U) = \cap_{u \in U} N(u)$ of U. For each vertex $v \in R$ let $I(v) = I_U(v)$ denote the indicator random variable of the event that $v \in N(U)$. Then $\mathbb{P}[I(v) = 1] = p^t = \frac{1}{n}$. The random variables $I(v)$ are mutually independent for $v \in V$, since they depend on pairwise disjoint sets of pairs of vertices.

This means that the random variable $|N(U)| = \sum_{v \in R} I(v)$ has binomial distribution with parameters n and $\frac{1}{n}$. Its expectation is $n \cdot \frac{1}{n} = 1$, so on the average U does *not* even form the larger side of a $K_{t,2}$, not to mention forming one of the sides of a $K_{t,t}$. So, at least on average, everything looks rosy! How come then, that still, with overwhelming probability, $G(n,n,p)$ contains not only $K_{t,t}$, but even $K_{t,g(n)}$ for some function g of n tending to infinity?

Having a binomial distribution with constant expectation, the distribution of $|N(U)|$ can be estimated by the Poisson distribution of the same expectation. In particular, for every constant $s \in \mathbb{N}$,

$$\mathbb{P}[|N(U)| = s] = \binom{n}{s} \frac{1}{n^s} \left(1 - \frac{1}{n}\right)^{n-s} \to \frac{1}{e} \cdot \frac{1}{s!},$$

This means that while on the average each of the $\binom{n}{t}$ t-subsets of L will have a single common neighbor in R, at the same time each has a constant (independent of n) chance of having a common neighborhood of size s, for any constant s (say, also for $s = t^{t^{t^t}}$).

For example when $t = 4$, the probabilty that a fixed four-element set $U \subseteq L$, has 100 common neighbor tends to a constant. This constant is tiny, something of the order 10^{-157}, but independent of n. Consequently the expected number of 4-subsets of L hosting a $K_{4,100}$ in $G(n,n,p)$ is of the same order n^4 as the number of four elements sets. Calculating the variance and applying Chebyshev's inequality to the random variable X counting such 4-sets, one can show that the value of X is likely to be very close to its expectation. Hence $\Theta(n^4)$ copies of $K_{4,100}$ do appear with overwhelming probability.

In a way the problem is the long smooth tail of the Poisson distribution: any constant value appears with constant probability. The cause of the long smooth tail is the full mutual independence of the n^2 random bits we use to create our random graph. This definitely has positive effects: it is very useful for example to have the number of edges what we want it to be with high probability, and it is certainly very pleasent that we

can understand and analyse the properties of the random graph relatively easily. On the other hand we have also seen the adverse effects of this excessive independence: the very subgraph of constant size we would like to avoid, say $K_{4,100}$, also appears in it almost surely. And not only that, but its copies are populating all parts of the random graph so densely, that even plastic surgery can not get rid of them all without making the graph essentially empty.

This is what we circumvent with the introduction of algebra into a probabilistic model. Instead of using n^2 independent random bits to create our random bipartite graph on two parts of size n each, we define a random algebraic model where we use only $\sim \log n$ bits to create a random graph on the same number of vertices. This limited amount of independence will be enough to maintain the expected number of edges and also to provide effective enough bounds on the size of common neighborhoods.

The key algebraic theorem we use to cut away a significant chunk of the Poisson-tail that was killing us above, provides a quantitative correlation between two different measures of "how large" an algebraic variety over $\bar{\mathbb{F}}_q$ is. One is the algebro-geometric notion of dimension, the other is the number theoretic concept of the number of its elements with coordinates in $\mathbb{F}_q$. To illustrate, consider t linear functions $f_1, \dots, f_t \in \mathbb{F}_q[y_1, \dots, y_m]$. The set of their common zeroes is the intersection of t hyperplanes. This set is an affine subspace of a certain dimension d and hence its size is q^d. The *Lang-Weil bound* is a far-reaching generalization of this counting fact for the situation when the f_i are polynomials of higher degree. It states that under certain assumptions the number of common zeroes of a set of polynomials $f_1, \dots, f_t \in \mathbb{F}_q[y_1, \dots, y_m]$ in $\mathbb{F}_q^m$ is of the order $q^{\dim V}$, where V denotes the set of common zeroes of the f_i over the algebraic closure $\bar{\mathbb{F}}_q$. Since the dimension is an integer, the possible number of common zeroes of the polynomials is quite limited, for example it is never $\log q$ or $q^{\frac{9}{10}}$ (for q large enough).

For our purposes the following spezialized consequence of the Lang-Weil bound, with no assumptions on the polynomials, will be convenient.

Theorem 2.9 *[?] For every $t, d \in \mathbb{N}$, there exists a constant $C = C(t,d)$ such that for arbitrary polynomials $f_1, \dots, f_t \in \mathbb{F}_q[y_1, \dots y_t]$ of degree at most d, the number of their common zeroes $y \in \mathbb{F}_q^t$ is either less than C or more than $q - C\sqrt{q}$.*

The proof of the Lang-Weil bound and in turn of the above theorem is beyond the scope of our lecture.

This theorem implies that in any graph where the neighborhoods are hypersurfaces defined by polynomials, there are no common neighborhoods of size between C and $q - C\sqrt{q}$. Therefore this part of the tail will be gone!

Construction. The vertex set will be two copies of $\mathbb{F}_q^t$, which we denote by L and R. For a polynomial $F(x,y) \in \mathbb{F}_q[x_1, \dots, x_t, y_1, \dots, y_t]$ we define a graph $G = G(f)$ on the vertex set $L \cup R$. Vertices $u \in L$ and $v \in R$ form an edge if $F(u,v) = 0$. Note that neighborhood of a vertex $u \in L$ is the hypersurface $\{y \in \mathbb{F}_q^t : F(u,y) = 0\}$.

We set $d = d(t) = t^2 - t + 2$. For the randomized construction, we choose the polynomial F uniformly at random from the set $\mathcal{P}_d \subset \mathbb{F}_q[x_1, \ldots, x_t, y_1, \ldots, y_t]$ of polynomials that have degree at most d both in the variables $x_1, \ldots, x_t$ and the variables $y_1, \ldots, y_t$. The probability space is the uniform product space $\mathbb{F}_q^{C_d}$, where C_d is the number of monomials satisfying the degree condition, where the coordinates correspond to the coefficients of the monomials.

As a first sanity check, let us see that for every pair $u \in L$ and $v \in R$ the probability that they form an edge in $G(F)$ is the right one, i.e. $n^{-\frac{1}{s}} = \frac{1}{q}$. This follows immediately by writing $F = P + Q$ as the sum of its non-constant terms P and its constant term Q and noting that for any polynomial $g \in \mathcal{P}_d$ with constant term 0, the probability of $F(u, v) = 0$ conditioned on $P = g$ is exactly $\frac{1}{q}$. This is because $F(u, v) = 0$ happens if and only if Q is equal to the constant $-g(u, v) \in F_q$.

This already implies that the number of edges has the right expectation. Writing $e(G) = \sum_{u \in L, v \in R} \mathbb{1}_{uv \in E(G)}$ as the sum of indicator random variables and using the linearity of expectations we have that

$$\mathbb{E}[e(G)] = \sum_{u \in L, v \in R} \mathbb{E}\left[\mathbb{1}_{uv \in E(G)}\right] = \sum_{u \in L, v \in R} \mathbb{P}[F(u, v) = 0] = n^2 \cdot \frac{1}{q} = n^{2 - \frac{1}{s}}.$$

We will show that the expected number of t-sets in L which host a copy of a $K_{t,C}$ in $G(f)$ for the constant $C = C(t, d)$ from Theorem 2.9 will be much smaller than q^{t-1} and hence deleting all incident edges to these t-sets produces a $K_{t,C}$-free graph with the right number of edges.

Let us fix a vertex set $U \subseteq L$ of size $|U| = t$. For every $u \in U$, consider the polynomials $F(u, y) \in \mathbb{F}_q[y_1, \ldots y_s]$. The number of their common zeroes $v \in \mathbb{F}_q^s$ is equal to $|N(U)|$. By Theorem ?? $\mathbb{P}[|N(U)| \geq C] = \mathbb{P}\left[|N(U)| \geq q - C\sqrt{q}\right]$. Here is where the tail-cutting happnes: to bound the number of $K_{t,C}$ in G it is then sufficient to bound the number of $K_{t,q-C\sqrt{q}}$. And we will be able to prove this to be very small, since q is not a constant.

To upper bound the upper bound of random variable eventually our only tool is Markov's Inequality, the effectiveness of which one can always try to boost using a higher moment, in particular if the variable in question is the sum of random variables with at least a limited amount of independence. This is our plan as well to bound $\mathbb{P}[|N(U)| \geq s]$. Recall that for a vector $v \in R$, $I(v)$ denotes the indicator random variable of the event that $v \in N(U)$, that is, $F(u, v) = 0$ for every $u \in U$. Then $|N(U)| = \sum_{v \in V} I(v)$. We will first establish the d-wise independence of the random variables $I(v)$ in order to bound $\mathbb{E}\left[|N(U)|^d\right]$ and use it in

$$\mathbb{P}[|N(U)| \geq s] = \mathbb{P}\left[|N(U)|^d \geq s^d\right] \leq \frac{\mathbb{E}\left[|N(U)|^d\right]}{s^d}.$$

The following lemma establishes the d-wise indpendence of random variables $I(v)$.

Lemma 2.9.1 *Let $t, r, d \in \mathbb{N}$ be such that $t, r \leq \min\{d, \sqrt{q}\}$. For every $U \subseteq L$ of size $|U| = t$ and $V \subseteq R$ of size $|V| = r$ we have that*

$$\mathbb{P}[F(u, v) = 0 \text{ for every } u \in U, v \in V] = \frac{1}{q^{tr}},$$

where F is polynomial chosen uniformly random from $\mathcal{P}_d$.

Proof. We call a set of vectors *simple* if the first coordinates are pairwise distinct. First we prove the statement in the case when both U and V are simple.

We write $F = P + Q$, where Q is the part containing the monomials of F the form $x_1^i y_1^j$ for $0 \leq i \leq t-1$ and $0 \leq j \leq r-1$ and P is the rest. First we generate P and then Q. Let $g \in \mathcal{P}$ be an arbitrary polynomial without the above monomials. We claim that the probability of $F(u,v) = 0$ happening for every $u \in U$ and $v \in V$, conditioned on $P = g$, is $\frac{1}{q^{tr}}$. Indeed, $F(u,v) = 0$ happens if and only if for the tr uniformly random coefficients $\beta_{i,j} \in \mathbb{F}_q$ we have that

$$\sum_{j=0}^{r-1} \sum_{i=0}^{t-1} \beta_{i,j} u_1^i v_1^j = -g(u,v),$$

which gives a linear equation system with tr equations in tr variables. Its matrix $(u_1^i v_1^j)_{u,v,i,j}$ is the tensor product of the two Vandermonde matrices $(u_1^i)_{u,i}$ and $(v_1^j)_{v,j}$. The determinant is conseuqently the product of $(\det(u_1)_i)^t$ and $(\det(v_1)_j)^r$. Both of these determinants are non-zero, since the u_1 are all distinct and the v_1 are all distinct. Therefore there is a unique solution $(\beta_{i,j}) \in \mathbb{F}_q^{tr}$ satisfying the equation system. Each vector in $\mathbb{F}_q^{tr}$ has the same probability to appear as the coefficients of Q, so the conditional probability is indeed $\frac{1}{q^{tr}}$. Since this holds for every g, we have that the unconditioned probability of $F(u,v) = 0$ is also $\frac{1}{q^{tr}}$.

Let now $U \subseteq L$ and $V \subseteq R$ be arbitrary. We will find invertible linear transformations T and S of $\mathbb{F}_q^t$, such that $T(U)$ and $S(V)$ are both simple. Let $\tilde{F}(x,y) := F(T(x), S(y))$. Note that $F \to \tilde{F}$ is one-to-one map from $\mathcal{P}_d$ to $\mathcal{P}_d$, hence the distribution of $\tilde{F}$ is also uniform on $\mathcal{P}_d$. Since $T(U)$ and $S(V)$ are simple and have sizes t and r, respectively, the probability that $\tilde{F}(u,v) = F(T(u), S(v)) = 0$ for every $u \in U$ and $v \in V$ is $\frac{1}{q^{tr}}$. This is then equal to the probability of $F(u,v) = 0$ for every $u \in U$ and $v \in V$, because the distribution of $\tilde{F}$ and F are the same.

In order to create the invertible linear function T, we define its first coordinate $T_1 : \mathbb{F}_q^t \to \mathbb{F}_q$ and extend it so the function is invertible. A linear function is determined uniquely by its values on any basis, so the number of potential T_1 is q^t. Among these, for any $u \neq u'$ there are exactly q^{t-1} for which $T_1(u) = T_1(u')$. One can see this for example by extending $u - u'$ to a basis, so the values of T_1 are freely chosen on all but the first element of this basis. Consequently there are at least $q^t - \binom{t}{2} q^{t-1} > 0$ linear functions T_1, such that $T_1(u) \neq T_1(u')$ for any $u, u' \in U$. Here we used that $t < \sqrt{q}$.

The construction of S is analogous, but we use that $r < \sqrt{q}$. □

Now using the linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[|N(U)|^d\right] &= \mathbb{E}\left[\left(\sum_{v\in V} I(v)\right)^d\right] = \sum_{v_1\in V}\cdots\sum_{v_d\in V}\mathbb{E}\left[I(v_1)\cdots I(v_d)\right]\\ &= \sum_{v_1\in V}\cdots\sum_{v_d\in V}\mathbb{P}\left[\{v_1,\ldots,v_d\}\subseteq N(U)\right]\end{aligned}$$

By Lemma 2.9.1 the probability in the summand is $\frac{1}{q^{tr}}$ where r is the cardinality of the set $\{v_1,\ldots,v_d\}$. We classify the sum according to r and obtain that

$$\mathbb{E}\left[|N(U)|^d\right] = \sum_{r=1}^{d}\binom{n}{r}M_r\frac{1}{q^{tr}} < \sum_{r=1}^{d}n^r M_r\frac{1}{q^{tr}} = \sum_{r=1}^{d}M_r =: m_d$$

where M_r denotes the number of surjective functions from a d-element set to an r-element set.

We now proved that similar to the expectation, the dth moment of $|N(U)|$ is also bounded by a constant.

This is good because then we have that

$$\sum_{U\subseteq L,|U|=t}\mathbb{P}\left[|N(U)|\geq s\right] \leq \frac{\mathbb{E}\left[|N(U)|^d\right]}{s^d} \leq \frac{m_d}{s^d}.$$

So for the expected number X of t-sets that form a side of a $K_{t,C(t,C(t,d(t)))}$ we have

$$\mathbb{E}[X] = 2\binom{n}{t}\frac{m_d}{(q/2)^d} \leq c'_d n^t q^d = c'_d q^{t^2}q^{-t^2+t-2} = c'q^{t-2}$$

for large q.

Let us now choose a ploynomial $f\in\mathcal{P}_d$ for which the value of the random variable $e(G)-q^tX$ is at least its expectation $q^{2t-1}-c'q^{2t-2}$. From the graph $G(f)$ delete all edges incidedent to one vertex in every bad t-set. In the resulting graph H there are obviously no $K_{t,C(t,C(t,d(t)))}$, yet the number of edges is at least $q^{2t-1}-c'q^{2t-2} = \Omega\left(n^{2-\frac{1}{t}}\right)$.

2.3.2 The norm-graphs

The next exercise demonstrates some more what can occur when we step into the fourth dimension. It is a prelude for what is coming in the following section.

Exercise 2.15 *Let the vertex set of a graph G be $\mathbb{F}_p^4$. Let (a,b,c,d) be adjacent to (a',b',c',d') if and only if $(a+a')(b+b')(c+c')(d+d')=1$. Prove that G contains a $K_{n^{1/4},n^{1/4}}$.*

The problem here arises because the variables could be "separated" from each other; there are four of them and $4 = 2 + 2$, each pair is responsible for a curve, and these create the large complete bipartite graph. This difficulty is overcome in the next section by introducing much-much higher degree in the equation, but still keeping the simple structure of the equation in the exercise. The key to this is the existence of the *Frobenius automorphism*. If the characteristic of a field $\mathbb{F}$ is q, then the mapping $X \to X^q$, $X \in \mathbb{F}$, is an automorphism of $\mathbb{F}$, the Frobenius automorphism. Indeed, the function is of course a bijection since q and $|\mathbb{F}^*|$ are relatively prime. The multiplication is clearly interchangeable with the mapping, while the addition is interchangeable since for any $Z \in \mathbb{F}$, $qZ = 0$, and hence

$$(X+Y)^q = X^q + \binom{q}{1}X^{q-1}Y + \cdots + \binom{q}{q-1}XY^{q-1} + Y^q = X^q + Y^q.$$

In the following we will define a sequence of dense $K_{t,s}$-free graphs for *arbitrary* t. Although we are not able to say anything new about $ex(n, K_{t,s})$ when $s = t$, we will at least determine its order of magnitude when s is *some* function of t. Namely, the *norm-graphs*, defined in the next section, are $K_{t,t!+1}$-free and have $cn^{2-1/t}$ edges thus matching the Kővári-Sós-Turán upper bound. Later $t!+1$ will be improved to $(t-1)!+1$ by a modified construction called the *projective norm-graph*.

Let q be a prime power and let t be a positive integer. The *norm-graph* $G_{q,t} = G$ is defined as follows. Let $V(G) = \mathbb{F}_{q^t}$, $E(G) = \{\{A, B\} : N(A+B) = 1\}$, where $N(\cdot) : \mathbb{F}_{q^t} \to \mathbb{F}_q$ is the norm function[2]

$$N(A) = A \cdot A^q \cdot \cdots \cdot A^{q^{t-1}} = A^{(q^t-1)/(q-1)}.$$

We have $|V(G)| = q^t =: n$. For a fixed $A \in V(G)$, the number of solutions X to the equation $N(X + A) = 1$ is exactly $\frac{q^t-1}{q-1}$. (Remember, for any $\alpha \in \mathbb{F}$ in a finite field $\mathbb{F}$, $|\{x : x^l = \alpha\}|$ is either 0 or $|\mathbb{F}^*|/(|\mathbb{F}^*|, l)$.) Hence, excluding the possible loops at those A where $N(2A) = 1$,

$$|E(G)| \geq \frac{1}{2}q^t\left(\frac{q^t-1}{q-1} - 1\right) \geq \frac{1}{2}n^{2-1/t}.$$

In the remaining of the section we study why G is $K_{t,t!+1}$-free. What does the presence of a $K_{t,s}$ in G mean? If a $K_{t,s}$ is present in G, then there exist $D_1, \ldots, D_t \in \mathbb{F}_{q^t}$ such that

[2]The *norm* of the field extension $\mathbb{F}_{q^l}$ over $\mathbb{F}_q$ is the map N_l defined on $\mathbb{F}_{q^l}$ by $N_l(A) = A \cdot A^q \cdots A^{q^{l-1}}$. We drop the subscript l throughout, as it will be apparent from the context. Clearly N is a multiplicative function: if $A, B \in \mathbb{F}_{q^l}$ then $N(AB) = N(A)N(B)$. From $N(A)^q = N(A)$ we infer that $N(A) \in \mathbb{F}_q$ for every $A \in \mathbb{F}_{q^l}$. Indeed, the roots of the polynomial $x^q - x$ are precisely the elements of $\mathbb{F}_q$, and it vanishes at $N(A)$.

the system of equations

$$\begin{array}{rcl} (X+D_1)(X^q+D_1^q)\cdots(X^{q^{t-1}}+D_1^{q^{t-1}}) &=& 1 \\ (X+D_2)(X^q+D_2^q)\cdots(X^{q^{t-1}}+D_2^{q^{t-1}}) &=& 1 \\ \vdots \qquad\qquad\qquad\qquad \vdots &&\vdots \\ \vdots \qquad\qquad\qquad\qquad \vdots &&\vdots \\ (X+D_t)(X^q+D_t^q)\cdots(X^{q^{t-1}}+D_t^{q^{t-1}}) &=& 1 \end{array} \tag{2.5}$$

has s solutions $X \in \mathbb{F}_{q^t}$. Note that here we used the comfortable fact shown above, that the mapping $A \mapsto A^q$ is a field automorphism (the *Frobenius automorphism*), and thus, in particular, $(A+B)^q = A^q + B^q$.

To obtain a bound on the number of solutions we consider a much more general setup.

Lemma 2.9.2 (Key Lemma) *Let $\mathbb{F}$ be a field, and $a_{ij}, b_i \in \mathbb{F}$ for $1 \le i,j \le t$, such that $a_{i_1 j} \neq a_{i_2 j}$ if $i_1 \neq i_2$. Then*

$$\begin{array}{rcl} (x_1-a_{11})(x_2-a_{12}) \cdots (x_t-a_{1t}) &=& b_1 \\ (x_1-a_{21})(x_2-a_{22}) \cdots (x_t-a_{2t}) &=& b_2 \\ \vdots \qquad\qquad\qquad\qquad \vdots &&\vdots \\ \vdots \qquad\qquad\qquad\qquad \vdots &&\vdots \\ (x_1-a_{t1})(x_2-a_{t2}) \cdots (x_t-a_{tt}) &=& b_t \end{array} \tag{2.6}$$

has at most $t!$ solutions in $\mathbb{F}^t$.

Remark: 1. The Key Lemma is easily proved when $b_1 = \cdots = b_t = 0$. To satisfy the first equation one must select a factor $(x_{\pi(1)} - a_{1\pi(1)})$ to be 0. This can be done t ways. Once the variable $x_{\pi(1)} = a_{1\pi(1)}$ is fixed, one can simplify the remaining $t-1$ equations with the factor $(x_{\pi(1)} - a_{i\pi(1)})$, since $a_{1\pi(1)} \neq a_{i\pi(1)}$ by the condition of the lemma. We end up with $t-1$ equations of the same kind as in the lemma and can proceed by induction.
2. The statement of the Key Lemma is best possible, which is witnessed by the case when all $b_i = 0$. Indeed, $(a_{1\pi(1)}, \ldots, a_{t\pi(t)})$ is a (distinct) solution for every $\pi \in S_t$.
3. Observe that the scenario is similar to the one we faced during the solution of Exercise 2.15. There the four equations similar to (2.6) occasionally had $\sim |\mathbb{F}|$ solutions. Nevertheless, according to the Key Lemma, this could only occur if some coordinate of at least two of the four points agreed. In the setup of the norm-graph the Frobenius authomorphism helps us to avoid this degeneracy.

Corollary 2.10 *For any prime power q and integer $t \ge 1$ the norm-graph $G_{q,t}$ does not contain a $K_{t,t!+1}$. In particular*

$$ex(n, K_{t,t!+1}) = \Theta\left(n^{2-1/t}\right).$$

Proof. Applying the Key Lemma with $\mathbb{F} = \mathbb{F}_{q^t}$, $a_{ij} = -D_i^{q^{j-1}}$, $x_i = X^{q^{i-1}}$, and $b_i = 1$ we have the $t!$ bound on the number of solutions of the system (2.5). To check the condition of the Key Lemma we recall that the Frobenius automorphism is in particular a bijection, so for a fixed j the $-D_i^{q^{j-1}}$ are all distinct. □

To get the bound on the number of solutions of system (2.6) of the Key Lemma we make use of the following standard tool.

Claim 1 *Let K be an algebraically closed field, $A = K[x_1, \ldots, x_t]$, $f_1, \ldots, f_r \in A$, $B = K[f_1, \ldots, f_r]$, and*

$$F : K^t \to K^r, \quad F(x) := (f_1(x), \ldots, f_r(x)).$$

If A is integral over B and B is integrally closed in $QF(B)$, then for any $b \in K^r$

$$|F^{-1}(b)| \leq d := \dim_{QF(B)} QF(A).$$

Proof. Let $F^{-1}(b) = \{P_1, \ldots, P_s\} \subseteq K^t$ and choose a polynomial $g \in A$ such that $g(P_i)$ are all distinct (this is easy since K is infinite). Let $h(y) \in QF(B)[y]$ be the monic minimal polynomial of g over $QF(B)$. Then $q := deg(h) \leq d$, otherwise $\{1, g, g^2, \ldots g^d\} \subseteq QF(A)$ would be a linearly independent set over $QF(B)$, a contradiction.

The coefficients of h are integral over B, so they are from B (since B is integrally closed in $QF(B)$). Why are the coefficients of h integral over B? Since one of its roots, g is integral, integrality carries over to all its roots and the coefficients are polynomials of the roots of h. Indeed, if L is the splitting field of h, then any zero $g' \in L$ of h can be mapped to g by an automorphism of L which fixes $QF(B)$, so the integrality of g over B implies the integrality of g' over B.

So there exists coefficients $c_1, \ldots c_q \in B$ such that

$$g^q + c_1 g^{q-1} + \ldots c_q = 0.$$

If we substitute $P_1, \ldots P_s$ into this polynomial equation, we obtain s equations over K.

Note that $c_j(P_i)$ does **not** depend on i, since $F(P_i)$ is constant and B is generated by the coordinate functions of F. Let $e_j = c_j(P_i)$.

Since $g(P_1), \ldots, g(P_s)$ are distinct solutions to the equation

$$g^q + e_1 g^{q-1} + \ldots e_q = 0,$$

we have $s \leq q \leq d$. □

Proof. (of the Key Lemma) We naturally set $f_i(x) = (x_1 - a_{i1})(x_2 - a_{i2}) \cdots (x_t - a_{it})$ and $r = t$. By Claim 1 we need to prove

- $A = K[x_1, \ldots, x_t]$ is integral over $B = K[f_1, \ldots, f_r]$ (this implies that A is a finitely generated B-module, since it is a finitely generated B-algebra),
- B is integrally closed in $QF(B)$, and
- $\dim_{QF(B)} QF(A) = t!$.

Integrality of A over B: We use induction on t.
The case $t = 1$ is trivial: $A = K[x_1] = K[x_1 - a_{11}] = B$.

Let $t > 1$, and let $R \geq B$ be an arbitrary valuation ring. Since the integral closure of B is the intersection of all valuation rings containing it, we are done once we prove $R \geq A$.

Assume $x_t \notin R$. This implies $x_t - a_{it} \notin R$ for all $i = 1, \ldots, t$. Then

$$\frac{1}{x_t - a_{it}} \in \mathfrak{m} \lhd R \quad \stackrel{f_i \in R}{\Longrightarrow} \quad g_i := \frac{f_i}{x_t - a_{it}} \in \mathfrak{m},$$

for all $i = 1, \ldots, t$, where $\mathfrak{m}$ is the unique maximal ideal of the valuation ring R.

By induction $K[x_1, \ldots, x_{t-1}]$ is integral over $K[g_1, \ldots, g_{t-1}]$, we have

$$g_1, \ldots, g_{t-1} \in R \quad \Rightarrow \quad K[x_1, \ldots, x_{t-1}] \leq R.$$

The polynomials $g_1, \ldots, g_{t-1}, g_t$ have no common zero, thus by the Nullstellensatz there are $h_1, \ldots, h_t \in K[x_1, \ldots, x_{t-1}]$ such that

$$\sum_{i=1}^{t} h_i g_i = 1.$$

This is a contradiction, since the left side is in $\mathfrak{m}$ ($h_i \in R$!!!).

B is integrally closed in $QF(B)$: Since by the above A is integral over B and $x_1, \ldots, x_t$ are algebraically independent over K, we have that $f_1, \ldots, f_t$ are algebraically independent over K. Hence $B \cong A$ and thus is a UFD. The claim now follows since *every* UFD is integrally closed in its field of fractions.

Computing the rank: Let $\mathfrak{m} = (f_1, \ldots, f_t) \lhd B = K[f_1, \ldots, f_t]$, note that $\mathfrak{m}$ is maximal. Since A is a finitely generated B-module, $A_{\mathfrak{m}}$ is a finitely generated $B_{\mathfrak{m}}$-module, and therefore $A/\mathfrak{m}A \cong A_{\mathfrak{m}}/\mathfrak{m}A_{\mathfrak{m}}$ is a finite dimensional $B/\mathfrak{m} \cong B_{\mathfrak{m}}/\mathfrak{m}B_{\mathfrak{m}} \cong K$-vector space.

If $\bar{x}_1, \ldots, \bar{x}_s \in A_{\mathfrak{m}}/\mathfrak{m}A_{\mathfrak{m}}$ is a basis over K, then by Nakayama's Lemma we get that $x_1, \ldots, x_s$ generate $A_{\mathfrak{m}}$ over $B_{\mathfrak{m}}$, and hence $x_1, \ldots, x_s$ generate $QF(A)$ over $QF(B)$.

Thus $\dim_{QF(B)} QF(A) \leq \dim_K A/\mathfrak{m}A$. Then we have what we want since

$$A/\mathfrak{m}A \stackrel{\text{Claim 2}}{=} A/\cap \mathfrak{m}_\sigma \stackrel{\substack{\text{Chinese} \\ \text{Remainder}}}{\cong} \bigoplus_{\sigma \in S_t} A/\mathfrak{m}_\sigma \cong \bigoplus_{\sigma \in S_t} K,$$

where $\mathfrak{m}_\sigma := (x_1 - a_{1\sigma(1)}, \ldots, x_t - a_{t\sigma(t)}) \lhd A$ is maximal.

So to conclude the proof of the Key Lemma we need to show the following.

Claim 2 *Let $\mathfrak{m}_\sigma = (x_1 - a_{1\sigma(1)}, \ldots, x_t - a_{t\sigma(t)})$. Then*

$$\mathfrak{m}A = \bigcap_{\sigma \in S_t} \mathfrak{m}_\sigma.$$

Proof. $\boxed{\subseteq}$ Obvious.
$\boxed{\supseteq}$ Let

$$g_\sigma = \frac{f_1 \cdots f_t}{\prod_{i=1}^{t}(x_i - a_{i\sigma(i)})}.$$

The polynomials f_i, $i = 1, \dots t$ and g_σ, $\sigma \in S_t$ have no common zero, so there exists $h_i, h_\sigma \in K[x_1, \dots, x_t]$ such that $\sum f_i h_i + \sum g_\sigma h_\sigma = 1$.

Let $g \in \cap \mathfrak{m}_\sigma = \prod \mathfrak{m}_\sigma$, and let $g = \sum h_i f_i g + \sum h_\sigma g_\sigma g$. We know that $\sum h_i f_i g \in \mathfrak{m}A$, so $g_\sigma g \in \mathfrak{m}A$ (for $\forall \sigma \in S_t$) would imply $g \in \mathfrak{m}A$.

Fix a permutation $\sigma \in S_t$. We have $g = \sum g^*$, where a typical term $g^* = g' \cdot \prod_{\tau \in S_t} m_\tau$ with $g' \in A$ and $m_\tau \in \{x_1 - a_{1\tau(1)}, \dots, x_t - a_{t\tau(t)}\}$ for all $\tau \in S_t$. In particular let j be the index for which $m_\sigma = x_j - a_{j\sigma(j)}$. Then

$$f_j | g^* g_\sigma \quad \Rightarrow \quad g^* g_\sigma \in \mathfrak{m}A \quad \Rightarrow \quad g g_\sigma \in \mathfrak{m}A.$$

□

Hence we concluded the proof of the Key Lemma. □

2.3.3 The projective norm-graphs

Even denser $K_{3,3}$-free graphs — an elementary construction

The graph $H = H_{q,3}$ is defined as follows. The vertex set $V(H)$ is $\mathbb{F}_{q^2} \times \mathbb{F}_q^*$. Two distinct vertices (A, a) and $(B, b) \in V(H)$ are connected if and only if $N(A + B) = ab$, where $N(X) = X^{1+q}$ is the norm of $X \in \mathbb{F}_{q^2}$ over $\mathbb{F}_q$. Of course $N(X) \in \mathbb{F}_q$ and it is clear that $|V(H)| = q^3 - q^2$. If $N(A+X) = ax$, then (A, a) and $X \neq -A$ determine x. Thus for any fixed $(A, a) \in V(H)$, there are exactly $q^2 - 1$ solutions (X, x) to $N(A + X) = ax$. This implies that, excluding possible loops, the degree of each vertex is at least $q^2 - 2 \geq n^{2/3}$.

We prove that H is $K_{3,3}$-free and hence provides an improvement (in the second order term) over Brown's construction for a dense $K_{3,3}$-free graph. (The Brown-graph has $\frac{1}{2}n^{5/3} - \frac{1}{2}n^{4/3}$ edges for infinitely many values of n.)

Theorem 2.11 *The graph $H = H_{q,3}$ contains no subgraph isomorphic to $K_{3,3}$. Thus there exists a constant C such that for every $n = q^3 - q^2$ where q is a prime power*

$$ex(n, K_{3,3}) \geq \frac{1}{2}n^{\frac{5}{3}} + \frac{1}{3}n^{\frac{4}{3}} + C.$$

Remark: At this point it is worthwhile to recall that the upper bound of Füredi (Theorem 2.8) is

$$ex(n, K_{3,3}) \leq \frac{1}{2}n^{\frac{5}{3}} + n^{\frac{4}{3}} + \frac{1}{2}n.$$

Proof. The statement of Theorem 2.11 is a direct consequence of the following.
If $(D_1, d_1), (D_2, d_2), (D_3, d_3)$ are distinct elements of $V(H)$, then the system of equations

$$\begin{array}{rcl} N(X + D_1) & = & xd_1 \\ N(X + D_2) & = & xd_2 \\ N(X + D_3) & = & xd_3 \end{array} \tag{2.7}$$

has at most two solutions $(X, x) \in \mathbb{F}_{q^2} \times \mathbb{F}_q^*$.

Observe that if the system has at least one common solution (X, x), then
(i) $X \neq -D_i$ for any $i = 1, 2, 3$ and
(ii) $D_i \neq D_j$ if $i \neq j$.

The latter is true, because if $D_i = D_j$, then the presence of a common neighbor implies $d_i = d_j$.

Because of *(i)* we can divide the first two equations by the last one and get rid of x. The norm is a multiplicative function, so we obtain

$$N\left(\frac{X + D_i}{X + D_3}\right) = \frac{d_i}{d_3},$$

for $i = 1, 2$.

We can divide each equation by $N(D_i - D_3)$, since these are nonzero by *(ii)*. Then we can substitute $Y = 1/(X + D_3)$, $A_i = 1/(D_i - D_3)$ and $b_i = d_i/(d_3 N(D_i - D_3))$ and obtain the following two equations:

$$\begin{array}{rcl} N(Y + A_1) & = & (Y + A_1)(Y^q + A_1^q) = b_1 \\ N(Y + A_2) & = & (Y + A_2)(Y^q + A_2^q) = b_2. \end{array} \tag{2.8}$$

Here we used the fact that $(A + B)^q = A^q + B^q$ for all A, B in $\mathbb{F}_{q^2}$.

We need the following simple special case of our Key Lemma from the previous section.

Lemma 2.11.1 *Let K be a field and $a_{ij}, b_i \in K$ for $1 \leq i, j \leq 2$ such that $a_{1j} \neq a_{2j}$. Then the system of equations*

$$\begin{array}{rcl} (x_1 - a_{11})(x_2 - a_{12}) & = & b_1, \\ (x_1 - a_{21})(x_2 - a_{22}) & = & b_2 \end{array} \tag{2.9}$$

has at most two solutions $(x_1, x_2) \in K^2$.

Although this is a special case of the Key Lemma (for $t = 2$) we include an elementary proof, which does not use any commutative algebra. We note that even for $t = 3$, we do not know of any proof of the Key Lemma which is simpler than the one in the previous section.

Proof. Subtracting the first equation from the second we get

$$(a_{11} - a_{21})x_2 + (a_{12} - a_{22})x_1 + a_{21}a_{22} - a_{11}a_{12} = b_2 - b_1.$$

Here we can express x_1 in terms of a linear function of x_2, since $a_{12} \neq a_{22}$. Substituting this back into one of the two equations of (2.9) we obtain a quadratic equation in x_2 with a non-zero leading coefficient (since $a_{11} \neq a_{21}$). This equation has at most two solutions in x_2 and each one determines x_1 uniquely. □

In order to finish the proof of Theorem 2.11 we apply Lemma 2.11.1 with $x_1 = Y, x_2 = Y^q$, $a_{11} = -A_1, a_{12} = -A_1^q$, and $a_{21} = -A_2, a_{22} = -A_2^q$. The conditions of the lemma hold since $-A_1^q = a_{12} = a_{22} = -A_2^q$ would mean $A_1 = A_2$, which is impossible by *(ii)*. Hence the system of equations (2.8) has at most two solutions Y. These solutions are in one-to-one correspondence with the solutions (X, x) of the equations (2.7), so Theorem 2.11 is proved. □

Exercise 2.16 *The* k-color Ramsey number $R_k(G)$ *is the smallest integer* m*, such that no matter how the edges of* K_m *are colored with* k *colors, there exists a monochromatic copy of* G*.*

Show that $R_k(K_{3,3}) = (1 + o(1))k^3$*.*

*(*Hint: *For the lower bound use the projective norm-graphs and the Key Lemma.)*

The general projective norm-graphs

The proof of Theorem 2.11 in the previous subsection is completely elementary. In order to prove the properties of the projective norm-graphs for $t > 3$ we need the Key Lemma for arbitrary t.

Let us define the projective norm-graph $H = H_{q,t}$ for any $t > 2$. The vertex set of H is $\mathbb{F}_{q^{t-1}} \times \mathbb{F}_q^*$. Two distinct vertices (A, a) and $(B, b) \in V(G)$ are adjacent if and only if $N(A + B) = ab$, where the norm is understood over $\mathbb{F}_q$, that is, $N(x) = x^{1+q+\cdots+q^{t-2}}$. Note that $|V(H)| = q^t - q^{t-1}$. If (A, a) and (B, b) are adjacent, then (A, a) and $B \neq -A$ determine b. Thus H is almost regular with possible degrees $q^{t-1} - 1$ and $q^{t-1} - 2$.

Theorem 2.12 *The graph* $H = H_{q,t}$ *contains no subgraph isomorphic to* $K_{t,(t-1)!+1}$*.*

Proof. The proof is a straightforward generalization of the proof of Theorem 2.11 with the remark that we need to use Lemma 2.9.2 (for $t - 1$ equations) instead of Lemma 2.11.1. □

Therefore, the following improvement over the previous section holds.

Corollary 2.13 *For every fixed* $t \geq 2$ *and* $s \geq (t-1)! + 1$ *we have*

$$ex(n, K_{t,s}) \geq \frac{1}{2} n^{2-\frac{1}{t}} - O(n^{2-\frac{1}{t}-c}),$$

where $c > 0$ *is an absolute constant.*

The improvement is most visible for small values of t, for example:

Corollary 2.14

$$ex(n, K_{4,7}) = \Theta(n^{7/4}).$$

Open Problem. The value of $ex(n, K_{4,4})$ is wide open. The conjecture is of course $\Theta\left(n^{7/4}\right)$, but we can't even separate it from $ex(n, K_{3,3})$, that is, we don't know whether

$$\lim_{n\to\infty} \frac{ex(n, K_{4,4})}{n^{5/3}} \to \infty.$$

Settling this would already be a major advance, even though most experts would be willing to bet a significant amount of money on that it is true.

Füredi's construction of dense $K_{2,s}$-free graphs (Theorem 2.3) showed that

$$\lim_{s\to\infty}(\liminf_{n\to\infty} ex(n, K_{2,s})n^{-3/2}) = \lim_{s\to\infty} \frac{1}{2}\sqrt{s-1} = \infty.$$

Exercise 2.17 *Use the projective norm-graphs together with Füredi's idea (which improves the $K_{2,2}$-free Construction 2 to a $K_{2,s}$-free construction) to give a $K_{3,s}$-free construction whose number of edges comes within a factor $\sqrt[3]{2} + o(1)$ of the KST upper bound of $ex(n, K_{3,s})$ for every $s \geq 3$ of the form $s = 2r^2 + 1$, $r \in \mathbb{Z}$. (The $o(1)$ above is understood as $s \to \infty$.)*
More generally, prove that for any fixed t

$$\lim_{s\to\infty}(\liminf_{n\to\infty} ex(n, K_{t,s})n^{-(2-1/t)}) = \infty.$$

Chapter 3

Even Cycles

In this chapter we go on studying graphs free of some fixed even cycle. This choice seems natural as, besides complete bipartite graphs, even cycles are the other front-runner for the title of "arguably the simplest family of bipartite graphs". Note that $K_{2,2} = C_4$ is the member of both families. You will find other similarities between complete bipartite graphs and even cycles: our knowledge of their Turán numbers are equally scarce.

Graphs without short cycles were studied much earlier in a different context, mostly in the context of finite geometry and group theory. Later they popped up in computer science in connection with *expander graphs*, i.e., graphs in which every "small" subset of the vertices has a "large" neighborhood (the set "expands"). While having no short cycle is not required of an expander, it just so happens that some of the best expanders tend to have large girth[1] and thus are relevant in our context as well. This is not a complete coincidence though: the ultimate d-regular expander, the infinite $d-1$-ary tree, has "infinite girth".

3.1 The Moore-bound

As any C_{2k}-free graph is also $K_{k,k}$-free, the Kövári-Sós-Turán upper bound on $ex(n, K_{k,k})$ of course also bounds $ex(n, C_{2k})$. This is of the order $n^{2-\frac{1}{k}}$. But C_{2k} has much fewer edges than $K_{k,k}$, so we sense that there must be further reasons that limit the number of edges in $2k$-cycle-free graphs more significantly. The following heuristic "proof" shows that $ex(n, C_{2k}) \lessapprox \frac{1}{2}n^{1+\frac{1}{k}}$, using that any two disjoint paths of length k, starting at the same vertex, must end at distinct vertices (otherwise their union would form a $2k$-cycle).

Let G be a C_{2k}-free graph with n vertices and average degree $\bar{d}$. We fix a vertex v_0 and count paths of length k starting at v_0. If v_0 is chosen properly, then it has roughly $\bar{d}$ neighbors, all its neighbors have roughly $\bar{d}-1$ further neighbors, etc ... Hence there are roughly $\bar{d}(\bar{d}-1)^{k-1}$ paths starting at v_0, all of which must have a different endpoint. Thus $n \geq \bar{d}(\bar{d}-1)^{k-1}$, which implies $\bar{d} \leq n^{1/k} + 1$. So we are done, right?

Well, except that some details would still need to be hammered out. There are two problems with the above argument, each of them is associated with the appearance of the word "roughly". First of all we need to find a vertex v_0, that has the property that most

[1] The *girth* of a graph is the length of its shortest cycle.

of the vertices which have distance at most $k-1$ to it have degree roughly $\bar{d}$. For the moment, let us overcome this problem by assuming that G is d-regular, i.e. $\bar{d} = d$. The second problem in the proof arises when we claim that the $\bar{d}(\bar{d}-1)^{k-1}$ paths of length k starting at v_0 all have distinct endpoints. Some of these paths may intersect each other *earlier*, which means that even if their endpoint were the same, their union would not form a $2k$-cycle. Hence we cannot infer the crucial lower bound on the number of vertices. This difficulty will disappear the moment we assume not only that G contains no cycle of length $2k$, but also that it doesn't contain shorter cycles, i.e. that its girth is *at least* $2k+1$.

To conclude, the above heuristic argument does prove the following classical combinatorial result.

Proposition 3.1 *(Moore bound) Let G be a d-regular graph of girth at least $2k+1$. Then*

$$v(G) \geq 1 + \sum_{i=0}^{k-1} d(d-1)^i.$$

Here, in the sum, we account for *all* the vertices that are on paths of length k starting at v_0 (and not only the endpoints of these paths).

Exercise 3.1 *Derive the* even-Moore bound. *That is, show that if G is a d-regular graph with girth $2k$, then*

$$v(G) \geq 2\sum_{i=0}^{k-1} (d-1)^i.$$

The Moore-bound in particular would imply that $ex(n, C_3, C_4, \ldots, C_{2k}) \leq \frac{1}{2}n^{1+1/k} + \frac{1}{2}n$ had we not had the convenient assumption of the regularity of G. Alon, Hoory, and Linial (2002) overcame this difficulty and showed that a natural extention of the Moore-bound is valid for arbitrary graphs. We will discuss their result in the next section, but for now we will be content with a simpler version, which makes the above heuristic precise at the price of a constant factor two.

For this purpose we recall a simple trick to pass from average degree to minimum degree. This claim is very useful in a wide variety of settings, whenever we only care about the order of magnitude of the final answer.

Proposition 3.2 *Every graph G has a subgraph $H \subseteq G$ with minimum degree*

$$\delta(H) > \frac{\bar{d}(G)}{2}.$$

Proof. Let $d = \bar{d}(G)$ be the average degree of G. If we find a vertex v in G with degree $d(v) \leq d/2$, we kick it out of G. We repeat this process until there is no such vertex, and we denote the obtained graph by H. This H is certainly the kind of subgraph we

want, unless it is empty. Could it be empty? How many edges did we delete in the process? In each step at most $d/2$ edges were deleted, altogether at most $(n-1)d/2$. Since $e(G) = \frac{d}{2}n$, not all edges were deleted. Thus H is not empty and has no vertex of degree at most $d/2$. □

Proposition 3.3 $ex(n, C_3, \ldots, C_{2k}) \leq n^{1+1/k} + n$.

Proof. Let G have n vertices, average degree d and girth at least $2k+1$. By Proposition 3.2, G has a subgraph $H \subseteq G$ with minimum degree $\delta(H) > d/2$. Obviously, the girth of H is also at least $2k+1$. Then, by the above heuristic argument (which is now precise if we replace $\bar{d}$ with $\delta = \delta(H)$), we have

$$n \;\geq\; v(H) \;\geq\; 1 + \delta(\delta-1)^{k-1}.$$

For the number of edges of G this implies

$$e(G) = n\frac{d}{2} < n\delta \leq n\left(n^{1/k}+1\right).$$

□

Exercise 3.2 *Show that the Turán number of any tree is linear in n. More precisely, prove that for any tree T with t edges,*

$$\frac{(t-1)n}{2} + O(1) \leq \mathrm{ex}(n, T) \leq (t-1)n.$$

For every t exhibit a tree with t edges, for which the lower bound is tight.

Remark. Erdős and Sós (1962) conjectured that the lower bound $\frac{(t-1)n}{2} + O(1)$ is the Turán number of *any* tree with t edges. Ajtai, Komlós, Simonovits, and Szemerédi announced that they proved the conjecture.

3.2 The Moore-bound for arbitrary graphs

The goal of this section is to generalize the Moore bound for arbitrary graphs. When a graph G does not contain cycles of length up to $2k$, then the k-neighborhood of *any* vertex is a tree. Knowing the degrees of vertices and their adjacencies one can count exactly the number of vertices in it. The exact summation is

$$\sum_{v_1 \in N(v_0)} \sum_{v_2 \in N(v_1)\setminus\{v_0\}} \cdots \sum_{v_k \in N(v_{k-1})\setminus\{v_{k-2}\}} 1. \tag{3.1}$$

This sum is certainly a lower bound on the number n of vertices of G, no matter how we choose the starting vertex v_0. The value however can vary greatly depending on the

starting vertex. If we choose v_0 in a very sparse part of G, with many low degree vertices, we will reach much fewer vertices in k steps and our bound on n will be weaker. In the Moore bound we had the extra assumption of d-regularity, hence it did not matter where we started. In Proposition 3.3 we chose a vertex from a subgraph of G, where vertices on each level of the tree had a relatively large degree. But to be able to identify a part of G where *every degree* is relatively large, we had to give up on a significant part of G and thus be content with half of the average degree.

In this section we aim at regaining this constant factor two. We hope that by choosing a vertex that is "average" in some sense, then, even if the size of many of the neighborhoods in the above sum are less than the average degree, there will also be sufficiently many other neighborhoods whose size is more than the average degree, so that the overall sum is at least as large as it would be when the graph was regular.

This is the content of the next theorem of Alon, Hoory and Linial (2002). They showed that the Moore bound is valid for *any* graph, with d replaced with the average degree of G. The proof we include here is based on the more conceptual argument of Babu and Radhakrishnan (2014), using the information theoretic notion of entropy.

Theorem 3.4 (Alon,Hoory, Linial (2002)) *Let $G = (V, E)$ be an n-vertex graph with average degree $\bar{d} \geq 2$ and girth $g(G) \geq 2k + 1$. Then*

$$n \geq 1 + \bar{d} \sum_{i=0}^{k-1} (\bar{d} - 1)^i.$$

This clearly implies the constant factor two improvement over Proposition 3.3. Indeed, if G is graph on n vertices with no cycle of length at most $2k$, then by the above $n > (\bar{d} - 1)^k$, implying $\bar{d} < n^{1/k} + 1$.

Corollary 3.5 *For every integer $k \geq 2$, $ex(n, \{C_3, C_4, ..., C_{2k}\}) < \frac{1}{2} n^{1+\frac{1}{k}} + \frac{n}{2}$.*

Proof. [*of Theorem 3.4*] As we planned, we choose a root vertex v_0 randomly and hope that on average its k-neighbourhood will be large. But how? Intuitively it is sort of clear that for a general graph it is *not* a good idea to pick v_0 uniformly at random. A starting vertex v_0 in (3.1) with a tiny degree is already greatly disadvantaged from the start to reach an average number of vertices overall, such vertices should have tiny probability to become the root.

Another consideration is that once we are convinced that a particular distribution π of the vertices is optimal for choosing our starting vertex, why wouldn't we want to have the same distribution for the neighbors of the starting vertex, or the neighbors' neighbors, so that they also have large neighborhoods? In other words we might want to try out the so-called stationary distribution of the uniform random walk on G, which ensures that the distribution of, say, the fifty-third vertex on the walk is the same as the one of the starting vertex. This is the following natural distribution

$$\pi(v) := \frac{d(v)}{\sum_u d(u)};$$

Observe that the probability of becoming the root vertex is proportional to the degree, that is, a vertex with twice as large degree is twice as likely to be the root.

Our main lemma says that this simple dependence of the random selection on the vertices, involving nothing more than their degrees, is already sufficient to guarantee that the expectation of the size of *any* of the ith neighborhoods is at least as much as it would be in a regular graph. For a vertex $v \in V(G)$ and integer $i \in \mathbb{N}_0$, let $n_i(v)$ denote the number of vertices in G whose distance to v is exactly i.

Lemma 3.5.1 *For an arbitrary integer $i \geq 2$, and graph $G = (V, E)$ with girth $g(G) > 2i$, minimum degree $\delta(G) \geq 2$, and average degree $\bar{d}$, we have*

$$\mathbb{E}_{v \sim \pi}\left[n_i(v)\right] \geq \bar{d}(\bar{d}-1)^{i-1}$$

Before proving the lemma, let us see how it implies our theorem. First note that we can assume without loss of generality that the minimum degree $\delta(G) \geq 2$. Indeed, if there was a vertex of degree at most 1, then deleting it produces a graph G' of average degree $\bar{d}' \geq \bar{d} \geq 2$. By induction the theorem is valid for G', so using that the function $x \mapsto x(x-1)^{i-1}$ is monotone increasing for any $i \in \mathbb{N}_0$ and real $x \geq 1$, we have

$$n = v(G') + 1 \geq 1 + \bar{d}' \sum_{i=0}^{k-1} (\bar{d}'-1)^i + 1 \geq 2 + \bar{d} \sum_{i=0}^{k-1} (\bar{d}-1)^i.$$

Therefore we can assume $\delta(G) \geq 2$ and apply the lemma to our G. By definition, $n \geq \sum_{i=0}^{k} n_i(v)$ for *any* vertex $v \in V$. Now if we choose v randomly according to the distribution π, use linearity of expectation and Lemma 3.5.1, we can conclude that *there exists* a vertex v_0 with

$$n \geq \sum_{i=0}^{k} n_i(v_0) \geq \mathbb{E}_{v \sim \pi}\left[\sum_{i=0}^{k} n_i(v)\right] = \sum_{i=0}^{k} \mathbb{E}_{v \sim \pi}\left[n_i(v)\right] \geq 1 + \bar{d} \sum_{i=0}^{k-1} (\bar{d}-1)^i.$$

□

For the proof of the lemma we define a (directed) random walk of length on G. This facilitates that we do give a non-zero random consideration to *all* paths of length i, which is necessary to not lose any of the vertices at distance i from the starting vertex. In order to make sure that the random walk is in fact a *path*, we disallow that the walk repeats the same edge in the other direction immediately after its use. There will be no restriction in using edges more than once, only the immediate repetition is forbidden. This does not cause any problems for us due to the lack of short cycles: no vertex repetition will occur in the first k steps.

This walk will still have the beautiful property that the distribution of the vertices at *any* fixed stage is the same as the starting distribution π.

Formally, we define a sequence $v_0, \overrightarrow{e_1}, v_1, \overrightarrow{e_2}, v_2, \ldots$ of random variables, such that v_j is a V-valued random variable representing the jth vertex of the walk, and $\overrightarrow{e_j}$ is a $\overrightarrow{E}$-valued random variable such that $\overrightarrow{e_j} = (v_{j-1}, v_j)$, representing the j edge of the walk.

Here $\overrightarrow{E} = \{(u,v),(v,u) : uv \in E\}$ denotes the set of directed edges involving edges of G.

1) Choose the initial vertex v_0 at random with probability $\pi(v)$;

2) Choose an edge $\overrightarrow{e_1}$ uniformly at random from edges incident to v_0 and let v_1 be the other endpoint of $\overrightarrow{e_1}$;

3) For $j \geq 2$, choose an edge $\overrightarrow{e_j}$ uniformly at random from all edges incident to v_{j-1}, except (v_{j-1}, v_{j-2}), and let v_j be the other endpoint of $\overrightarrow{e_j}$;

While this random walk in principle could return to an already visited vertex later, this will not affect us, because the graph we apply it to will have girth larger than $2k$ and we stop the walk before a cycle could appear.

In the next lemma we show that the above random walk has the property that its ith vertex has the same distribution for every i. This will be crucial in establishing that not only the first neighborhood of the starting vertex will be average, but all the ith neighborhoods.

Lemma 3.5.2 *For every $j \in \mathbb{N}$, v_j is distributed according to π and $\overrightarrow{e_j}$ is distributed uniformly over $\overrightarrow{E}$.*

Proof. We calculate the distributions of v_j and $\overrightarrow{e_{j+1}}$ by induction on $j \geq 0$. For $j = 0$, v_0 has distribution π by definition. For $\overrightarrow{e_1}$ we need to show that $\mathbb{P}\left[\overrightarrow{e_1} = (u,w)\right] = \frac{1}{n\bar{d}}$ for every $(u,w) \in \overrightarrow{E}$. Since $\overrightarrow{e_1} = (u,w)$ implies $v_0 = u$, we have

$$\begin{aligned}\mathbb{P}\left[\overrightarrow{e_1} = (u,w)\right] &= \mathbb{P}\left[v_0 = u \wedge \overrightarrow{e_1} = (u,w)\right] = \mathbb{P}\left[v_0 = u\right] \cdot \mathbb{P}\left[\overrightarrow{e_1} = (u,w) \mid v_0 = u\right] \\ &= \frac{d(u)}{n\bar{d}} \cdot \frac{1}{d(u)} = \frac{1}{n\bar{d}}.\end{aligned}$$

Let now $j \geq 1$. For an arbitrary $(u,w) \in \overrightarrow{E}$ we have

$$\begin{aligned}\mathbb{P}\left[\overrightarrow{e_{j+1}} = (u,w)\right] &= \sum_{\overrightarrow{e} \in \overrightarrow{E}} \mathbb{P}\left[\overrightarrow{e_{j+1}} = (u,w) \mid \overrightarrow{e_j} = \overrightarrow{e}\right] \cdot \mathbb{P}\left[\overrightarrow{e_j} = \overrightarrow{e}\right] \\ &= \sum_{x \in N(u) \setminus \{w\}} \mathbb{P}\left[\overrightarrow{e_{j+1}} = (u,w) \mid \overrightarrow{e_j} = (x,u)\right] \cdot \mathbb{P}\left[\overrightarrow{e_j} = (x,u)\right] \\ &= \sum_{x \in N(u) \setminus \{w\}} \frac{1}{d(u)-1} \cdot \frac{1}{n\bar{d}} = (d(u)-1) \cdot \frac{1}{d(u)-1} \cdot \frac{1}{n\bar{d}} = \frac{1}{n\bar{d}}.\end{aligned}$$

The distribution of v_j follows from that of $\overrightarrow{e_j}$ For an arbitrary $v \in V$ we have $\mathbb{P}\left[v_j = v\right] = \sum_{u \in N(v)} \mathbb{P}\left[\overrightarrow{e_j} = (u,v)\right] = \frac{d(v)}{n\bar{d}}$.

□

In order to prove Lemma 3.5.1, we will introduce the concept of *entropy* and see some of its basic properties.

Entropy

Definition: Let $p : \Omega \to (0, 1]$ be a discrete probability distribution, that is $\sum p(x) = 1$. The *entropy of* p is

$$H(p) = \mathbb{E}_{x \sim p}[-\log_2(p(x))] = -\sum_x p(x) \log_2(p(x)).$$

For a discrete random variable X its entropy $H(X)$ is just the entropy of its probability distribution.

Remark:

1. The entropy of a random variable depends only on its distribution, and not on its values.

2. The entropy can be thought of as a measure of uncertainty of the random variable. Or, that how much new information is obtained, on average, by learning the outcome of one experiment by the random variable. In other words, how many bits one would need on average in order to communicate an outcome of X.

3. To make sense, in the definition we assumed that $p(x) > 0$ for every $x \in \Omega$. Often it is convenient to talk about the entropy of a distribution p over some set Ω without prior knowledge about whether indeed all values $x \in \Omega$ can occur with non-zero probability. We will do this by adopting the convention that $p(x) \log_2 p(x) := \lim_{z \to 0} z \log_2 z = 0$ when $p(x) = 0$.

4. For a random variable, it will be convenient to introduce the notion of its *range* $R(X) = \{x \in \Omega : p(x) > 0\}$. In other words the $R(X)$ is the support of the induced probability distribution.

5. The base of the logarithms will be supressed, they are all of base 2 in this section.

Examples.

a) If X is Bernoulli random variable with success probability $p \in [0, 1]$, then $H(X) = -p \log p - (1 - p) \log(1 - p)$ is the so-called *binary entropy function*.

b) If X is the uniform distribution with an n-element range, then $H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n$.

Next we list the most basic properties of entropy. All these can be well-justified using the intuitive meaning of entropy concept. In your homework you are asked to rigorously prove them.

Intuitively it seems clear that any deviation from the uniform distribution should decrease the amount of uncertanity. The next proposition states that uncertanity is indeed the largest, when all atomic events have the same probability.

Proposition 3.6 *If X is a random variable taking on n values, then $H(X) \leq \log n$.*

Proof. Homework. □

We often deal with not only one, but a whole sequence of random variables. Two random variables X and Y on the same probability space $(\Omega, \mathbb{P})$, with values in Ω_1 and Ω_2, respectively, naturally define their joint distribution with values in $\Omega_1 \times \Omega_2$ by $p_{X\times Y}(x, y) = \mathbb{P}[X = x \wedge Y = y]$.

Definition: Given two random variables X, Y, the *joint entropy* of X and Y is

$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y).$$

Note that in case the variables X and Y are not independent then the range of their product might be much smaller than the product of their respective ranges. This will be highly applicable in our situation.

In many situations we would like to obtain information about the result of a random process consisting of several steps. Each of the steps is simple to describe, but the cumulative effect is unclear. In order to be able to decipher the random information hidden in the sequence step by step, we introduce the notion of conditional entropy.

Definition: The *conditional entropy of X given Y* is

$$H(X|Y) := \mathbb{E}_y [H(X|\{Y = y\})] = \mathbb{E}_y \left[-\sum_x \mathbb{P}[X = x|Y = y] \log \mathbb{P}[X = x|Y = y] \right].$$

The intuitive meaning of the conditional entropy is that given knowledge of the value of Y, how much extra information, on average, is involved in learning the value of X as well. Learning the value of the joint distribution means learning the values of both coordinates. We can do this in two steps, coordinate by coordinate. Then the obtained information can be broken down into two terms: the information gained by learning the second coordinate, and then with this extra knowledge, the information obtained by learning the first. This is formulated in the next proposition.

Proposition 3.7 *If X and Y are two random variables, then*

$$H(X, Y) = H(Y) + H(X|Y).$$

More generally, for any sequence of random variables $X_1, \ldots, X_i$ we have

$$H(X_1, \ldots, X_i) = H(X_1) + \sum_{j=2}^{i} H(X_j|X_1, \ldots X_{j-1}).$$

Proof. Homework. □

We are now ready to prove Lemma 3.5.1

Proof.[*of Lemma 3.5.1*] In a graph with no cycle of length at most $2i$ the size $n_i(x)$ of the ith neighborhood of an arbitrary vertex x is the numer of paths of length i starting at x. This is exactly the size of the support of the random variable $(\overrightarrow{e_{1,x}}, v_{1,x}, \ldots, \overrightarrow{e_{i,x}}, v_{i,x})$, where $v_{j,x}$ is the jth vertex, and $\overrightarrow{e_{j,x}}$ is the jth (directed) edge in the non-returning random walk defined above, starting at x.

Hence by the concavity of the log-function and Proposition 3.6, we have

$$\log(\mathbb{E}_{x\sim\pi}[n_i(x)]) \geq \mathbb{E}_{x\sim\pi}[\log n_i(x)] \geq \mathbb{E}_{x\sim\pi}\left[H(\overrightarrow{e_{1,x}}, v_{1,x}, \ldots, \overrightarrow{e_{i,x}}, v_{i,x})\right].$$

You might have wondered why overload the expression with all these random variables depending on each other, when already the random variable $\overrightarrow{e_{i,x}}$ alone has a range of size $n_i(x)$. Now the sequence will come in handy as we can use Proposition 3.7 for the step by step entropy calculation. The following basic observations for the conditional entropies follow from the definition of the random walk.

(i) $H(v_{j,x}|\overrightarrow{e_{1,x}}, v_{1,x}, \ldots v_{j-1,x}, \overrightarrow{e_{j,x}}) = 0$ for every $j = 1, \ldots, i$;

(ii) $H(\overrightarrow{e_{1,x}}) = \log d(x)$;

(iii) $H(\overrightarrow{e_{j,x}}|\overrightarrow{e_{1,x}}, v_{1,x}, \ldots, \overrightarrow{e_{j-1,x}}, v_{j-1,x}) = \mathbb{E}_{x_{j-1}\sim v_{j-1,x}}\left[\log\left(d(x_{j-1}) - 1\right)\right]$ for every $j = 2, \ldots, i$.

For (i) note that given the value of $\overrightarrow{e_{j,x}}$, learning the value of $v_{j,x}$ carries no extra information, since the value is just the endpoint of $\overrightarrow{e_{j,x}}$, determined. Formally, the random variable $v_{j,x}$ conditioned on any fixed value of $(\overrightarrow{e_{1,x}}, v_{1,x}, \ldots v_{j-1,x}, \overrightarrow{e_{j,x}})$ has a single value in its support: the endpoint of $\overrightarrow{e_{j,x}}$. And the entropy of a single-valued random variable is $-\log 1 = 0$.

For (ii) note that $\overrightarrow{e_{1,x}}$ is a uniform random variable with a range of size $d(x)$.

Finally for (iii) note that for any fixed value $(f_1, x_1, \ldots, f_{j-1}, x_{j-1})$ in the range of $(\overrightarrow{e_{1,x}}, v_{1,x}, \ldots, \overrightarrow{e_{j-1,x}}, v_{j-1,x})$, the variable $\overrightarrow{e_{j,x}}$ conditioned on it is uniform over a range of size $d(x_{j-1}) - 1$, so the entropy of this particular conditioned random variable is $\log(d(x_{j-1}) - 1)$. These are the values that has to be averaged out in the definition of the conditional entropy, over all choices $(f_1, x_1, \ldots, f_{j-1}, x_{j-1})$ from the range. The values are a function of only the last coordinate x_{j-1}, and that is why we average over the distribution $v_{j-1,x}$ of the last vertex.

By Proposition 3.7, the linearity of expectation, and $(i) - (iii)$ above, we have that

$$\mathbb{E}_{x\sim\pi}\left[H(\overrightarrow{e_{1,x}}, v_{1,x}, \ldots, \overrightarrow{e_{i,x}}, v_{i,x})\right] = \mathbb{E}_{x\sim\pi}\left[\log d(x)\right] + \sum_{j=2}^{i} \mathbb{E}_{x\sim\pi}\left[\mathbb{E}_{x_{j-1}\sim v_{j-1,x}}\left[\log(d(x_{j-1}) - 1)\right]\right]$$

To evaluate this expression recall that first choosing the start vertex x according to π and then the vertex x_{j-1} according to $v_{j-1,x}$ means that x_{j-1} is distributed like the $(j-1)$th vertex v_{j-1} of the random walk. Hence by Lemma 3.5.2 x_{j-1} is distributed

according to π for every $j \geq 1$. This simplifies our expression significantly.

$$\begin{aligned}\mathbb{E}_{x\sim\pi}\left[H(\overrightarrow{e_{1,x}}, v_{1,x}, \ldots, \overrightarrow{e_{i,x}}, v_{i,x})\right] &= \mathbb{E}_{x\sim\pi}\left[\log d(x)\right] + \sum_{j=2}^{i} \mathbb{E}_{x_{j-1}\sim\pi}\left[\log(d(x_{j-1})-1)\right] \\ &= \mathbb{E}_{x\sim\pi}[\log(d(x)(d(x)-1)^{i-1})].\end{aligned}$$

By the definition of π this is then further equal to

$$\begin{aligned}&= \frac{1}{n\bar{d}} \sum_x d(x) \log(d(x)(d(x)-1)^{i-1}) \\ &\geq \frac{1}{n\bar{d}} n\bar{d} \log(\bar{d}(\bar{d}-1)^{i-1}) = \log(\bar{d}(\bar{d}-1)^{i-1}).\end{aligned}$$

where we used that the function $x \mapsto x \log\left(x(x-1)^{i-1}\right)$ is convex on $x \geq 2$ and that the minimum degree $\delta(G) \geq 2$.

Hence $\mathbb{E}_{x\sim\pi}[n_i(x)] \geq \bar{d}(\bar{d}-1)^{i-1}$ as the logarithm function is increasing. □

Exercise 3.3 *Prove the analogous statement for graphs with even girth $g = 2k$.*

3.3 Upper bound for the size of C_{2k}-free graphs

The goal of this section is to rectify the second simplifying assumption we made at the beginning of Section 3.1 for the heuristic argument to imply Proposition 3.1. Namely, that we assumed that G does not have any cycle of length *at most* $2k$ instead of forbidding just $2k$-cycles.

Clearly, $ex(n, \{C_3, \ldots, C_{2k}\}) \leq ex(n, C_{2k})$. Erdős [?], already in a paper in 1963, stated without proof that the order of magnitude of the simple upper bound of Proposition 3.3 for $ex(n, C_3, \ldots, C_{2k})$ also holds for $ex(n, C_{2k})$. Bondy and Simonovits (1974) were the first to publish a proof, as part of the solution of a more general problem of Erdős about finding not only $2k$-cycles, but cycles of many different length in dense enough graphs. Their constant factor of the leading term $n^{1+\frac{1}{k}}$, unlike the one in Proposition 3.3 and Theorem 3.4, did depend on k. Since then the proof and the constant factor was polished several times [?, ?, ?]. The current record of $80\sqrt{k}\log k$ was established by Bukh and Jiang [?]. Here we include the simpler argument of Verstraëte [?] for a slightly weaker constant factor.

Theorem 3.8 $ex(n, C_{2k}) \leq 8(k-1)n^{1+1/k}$.

Proof. For our proof let us recall that the *radius* $rad(H)$ of a graph H is the distance from a "central vertex" of H to a farthest vertex. Formally,

$$rad(H) = \min_{v\in V(H)} \max_{x\in V(H)} dist_H(v, x),$$

where $dist_H(v,x)$ is the length of the shortest v,x-path in H.

The following two lemmas will imply our theorem. The first one provides us with cycles of every even length from a long interval, starting with a length at most $2rad(H)$. And all we need to assume for this is that the average degree at least 8 and the graph is bipartite. Our second statement accommodates the use of the first one by allowing us to pass to a subgraph with radius at most k, while still keeping the average degree an appropriately large constant.

Lemma 3.8.1 *Let H be a bipartite graph with average degree $\bar{d}(H) \geq 8$ Then we have cycles of $2\left\lfloor \frac{\bar{d}(H)}{4} \right\rfloor$ consecutive even lengths in H such that the shortest of them has length at most $2\,\mathrm{rad}(H)$.*

Lemma 3.8.2 *For every graph G with $e(G) \geq cn^{1+1/k}$, there exists a subgraph $H \subseteq G$ with average degree $\bar{d}(H) \geq c$ and radius $\mathrm{rad}(H) \leq k$.*

Assuming these two statements, Theorem 3.8 follows easily. By Theorem 2.1 we can assume that $k \geq 3$. Let G' be a graph with n vertices and at least $8(k-1)n^{1+1/k}$ edges. We find a $2k$-cycle in G'. First we take a bipartite subgraph $G \subseteq G'$ with $4(k-1)n^{1+1/k}$ edges. Then by Lemma 3.8.2 we have a subgraph $H \subseteq G$ with $\bar{d}(H) \geq 4(k-1)$ and $\mathrm{rad}(H) \leq k$. Since $4(k-1) \geq 8$, by Lemma 3.8.1 we can find cycles in H of $2(k-1)$ consecutive even lengths, such that the shortest is of length at most $2\,\mathrm{rad}(H) \leq 2k$. Thus one of these will exactly be $2k$ and Theorem 3.8 is proved. □

Proof. (of Lemma 3.8.2) By Proposition 3.2 we may assume that G has minimum degree at least $cn^{1/k}$. (Otherwise we pass to a subgraph.) We will build H starting from an arbitrary vertex v, taking all vertices at distance *at most* i. We denote the graph induced by these vertices by H_i. By definition $H_0 \subseteq H_1 \subseteq \cdots \subseteq H_k \subseteq \cdots$. We will verify that the average degree of H_r is at least c for some $r \leq k$. Then we are done, since no vertex in H_r is farther than k away from v, which implies that $\mathrm{rad}(H_r) \leq k$.

The idea is very simple. Since the minimum degree of G is large, there are a lot of edges emanating from H_{i-1}. Either a lot of these edges *share* their other endpoint, in other words $V(H_i)$ is small, implying H_i is dense enough and we are done. Otherwise most of the edges emanating from $V(H_{i-1})$ have *distinct* other endpoints, so $V(H_i)$ is large, the levels expand. And they do so each time by a factor $\Theta(n^{1/k})$, the minimum degree. There is no room, however, for such a big level expansion to happen more than k times and thus the radius is bounded.

Let us formalize the above by setting

$$r = \min\left\{i : \bar{d}(H_i) \geq c\right\}.$$

This minimum surely exists since $\bar{d}(G) \geq cn^{1/k} \geq c$.

Our goal is to show that $r \leq k$. Because all edges having an endpoint in H_{i-1} are fully contained in H_i and $\delta(G) \geq cn^{1/k}$, we know that $2e(H_i) \geq cn^{1/k}v(H_{i-1})$. On the

other hand, for $i < r$ we have $2e(H_i) = \bar{d}(H_i)v(H_i) < cv(H_i)$ by the definition of r. We conclude

$$n \geq v(H_{r-1}) > n^{1/k}v(H_{r-2}) > \cdots > n^{\frac{r-1}{k}},$$

which implies $k > r-1$. Hence the average degree of H_r is at least c for some $r \leq k$. Since the radius of H_r is at most $r \leq k$, the lemma is proved. □

Proof. (of Lemma 3.8.1) Let $s = \left\lfloor \frac{\bar{d}(G)}{4} \right\rfloor$. Take a central vertex $v_0 \in V(H)$, one that realizes the radius r of H. Let V_1 be the set of its neighbors, let V_2 be the neighbors' neighbors and so on; let V_i be the set of vertices whose distance from v_0 is exactly i. By our assumption on the radius, $V(G) = \cup_{i=1}^{k} V_i$.

Our plan is to find a family of paths of any even length up to $2s$, such that each of these paths alternate between two levels V_ℓ and $V_{\ell+1}$, starting and ending at the former. Then we hope to use these paths, extended by some shortest paths from their endpoints towards v_0, to create cycles of $2s$ consecutive even length. Finding paths of lengths up to $2s$ will not be difficult: a long enough cycle contains all such paths and H is dense enough to contain a long cycle between some two levels. But to conclude that the extensions of these paths towards v_0 will all create even cycles of *different* lengths, we need to ensure that all these extensions meet at the very same level V_q.

To ensure this we need to choose our paths with a bit more foresight. It turns out that a long enough cycle with an extra edge already hosts enough paths to provide us with the extra flexibility necessary. The following technical lemma conveniently formulates what we will need.

Lemma 3.8.3 *Let K be a cycle with one extra edge, and let $\chi : V(K) \to \{red, blue\}$ be a two-coloring of the vertices of K which does use both colors and is* not *proper. Then for every $\ell < n(K)$ there exists a path of length ℓ whose endpoints are colored* differently.

Proof. Homework. □

We claim that there must exist an index $\ell < r$ such that $e(H[V_\ell \cup V_{\ell+1}]) \geq s|V_\ell \cup V_{\ell+1}|$. Indeed, otherwise we have

$$\begin{aligned} e(H) &= e(H[V_0 \cup V_1]) + \cdots + e(H[V_{r-1} \cup V_r]) \\ &< s|V_0 \cup V_1| + \cdots + s|V_{r-1} \cup V_r| < s \cdot 2|V(H)|. \end{aligned}$$

implying $\bar{d}(H) < 4s$, a contradiction. Here we used that by construction of the levels, all edges of H go between neighboring levels.

By Proposition 3.2 we can find a subgraph $H_0 \subseteq H[V_\ell \cup V_{\ell+1}]$ with $\delta(H_0) > s \geq 2$.

We take a path of maximum length in H_0. Maximality of the path implies that all neighbors of an end-vertex w of the path must be on the path. Also, each two neighbors of w must be at path-distance at least two since H_0 contains no triangle. Therefore, the furthermost neighbor of w on the path is at least $1 + 2s$ away (along the path) from it. This means that there exists a cycle of length at least $2s + 2$ in H_0. Moreover, since

$\delta(H_0) \geq 3$, the endpoint w has at least three neighbors on the cycle. Let $K \subseteq H[V_\ell \cup V_{\ell+1}]$ be the graph consisting of this cycle with an extra edge.

Let T be a spanning tree of H with root v_0 built by Breadth First Search. We look at a minimal subtree T' of T such that $V(T') \supseteq V_\ell \cap V(K)$. By q we denote the number of levels of T'; note that $q \leq \ell < r = \mathrm{rad}(H)$. Let u be the root of T'.

Because of its minimality, T' has at least two branches at its root. Choose one of the branches and denote it by T''. Let $R = T'' \cap V_\ell$ and $B = V(K) \setminus R$. Observe that another branch of T' must also hit V_ℓ, so $B \cap V_\ell \neq \emptyset$. Apply Lemma 3.8.3 with the two-coloring of $V(K)$ defined by R and B: there exists an R,B-path of each length $< 2 + 2s$. For this, note that, since H is bipartite and $B \cap V_\ell \neq \emptyset$, *there is* a B,B-edge in K.

Any R,B-path of even length starts and ends in V_ℓ. Let $2j$ be an even number less than $2s + 2$. Expanding an R,B-path of length $2j$ with the unique paths of length q from its endpoints to the root of T' gives us a cycle of length $2j + 2q$. Hence we obtained cycles of $2s/2$ consecutive even lengths in H. The shortest of them is of length $2 + 2q \leq 2\,\mathrm{rad}(H)$. □

3.4 Dense C_{2k}-free graphs

After establishing upper bounds and wondering about how good they might be, it is always instructive to check out what the simple random construction of Subsection 2.2.1 offers. By Proposition 2.4 there exists a graph G with $e(G) = \Omega(n^{2-\frac{2k-2}{2k-1}})$ that does not contain a C_{2k} and therefore $ex(n, C_{2k}) = \Omega\left(n^{1+\frac{1}{2k-1}}\right)$. In fact, the same idea (taking a random graph of appropriate edge probability, deleting an edge from each forbidden subgraph, and using linearity of expectation to lower bound the number of edges), also gives the same order lower bound for the girth problem: $ex(n, C_3, C_4, \ldots, C_{2k}) = \Omega\left(n^{1+\frac{1}{2k-1}}\right)$. That is, the average degree of the random construction is of order $n^{\frac{1}{2k-1}}$, roughly the square-root of what we have in the upper bound.

3.4.1 Benson's construction

Early on, research about even-cycle-free graphs was motivated by problems of group theory. In the early sixties, following the work of the great group theorist Jacques Tits, interest arose in regular graphs with even girth that are extremal in a very strong sense, namely their number of vertices achieve the even-Moore bound $n = 2\sum_{i=0}^{g/2-1}(d-1)^i$ *precisely*. Feit and Higman, and independently Singleton proved that, besides the trivial case $d = 2$, this can be achieved only when the girth g is $6, 8$ or 12.

Simultanously, extremal graphs realizing these special cases were sought. The incidence graphs of the projective planes (from Subsection 2.1.3) is an infinite series of extremal examples realizing the Moore-bound for $g = 6$. Feit and Higman stated without proof that such graphs "certainly exist" for $g = 8$ and 12 as well. Implicitely they

were constructed first by Tits in form of the finite geometric concept of *generalized quadrangle* and *generalized hexagon*. Singleton gave a construction for $g = 8$, which was subsequently simplified by Benson (1966), who also provided explicit construction of a sequence of extremal graphs for $g = 12$. His graphs contain exactly $n = 2\sum_{i=0}^{g/2-1}(d-1)^i$ vertices, they have girth g, and are d-regular. Translating the essence to our language: Benson constructed C_6-free and C_{10}-free graphs $\Theta(n^{4/3})$ and $\Theta(n^{6/5})$ edges, respectively. These values match the order of the upper bound in Theorem 3.8.

Benson's approach generalizes the idea of the incidence graph of the projective plane: there C_4-freeness followed simply from the fact that two lines don't intersect in more than one point. Benson's girth-8 graph will also be a point-line incidence graph of a projective space of higher dimension: this will ensure right away that there are no 4-cycles and one can concentrate only on avoiding 6-cycles. A 6-cycle in the language of point-line incidence graphs corresponds to three lines determining a non-degenerate "triangle" (here we mean the geometric concept, not K_3). To avoid such "triangles of lines" we restrict ourselves to the points of such a surface in the space, which does not contain a triangle. (Note that, intuitively, "properly curving" surfaces in the euclidean space do not contain a triangle of lines.) We move to dimension 4 where a surface will contain roughly q^3 points, while the degree is of course still $q+1$, since each line contains this many points. This indicates that we might expect the "right" number of edges $\sim \frac{1}{2}q^4 \sim \frac{1}{2}n^{4/3}$.

We consider the projective space $PG(4,q)$ over $\mathbb{F}_q$, where the points have five coordinates, conveniently indexed by $-2,-1,0,1$ and 2 (for symmetry reasons). Let $Q_4 \subseteq PG(4,q)$ be a quadratic surface defined by the quadratic form $Q(x) = x_0^2 + x_1x_{-1} + x_2x_{-2}$, that is, $Q_4 = \{x : Q(x) = 0\}$. Let G_8 be the point-line incidence graph of Q_4.

Theorem 3.9 (i) $|Q_4| = q^3 + q^2 + q + 1$

(ii) *Q_4 contains $q^3 + q^2 + q + 1$ lines*

(iii) *every line contains $q + 1$ points*

(iv) *every point is contained in $q + 1$ lines of Q_4*

(v) *G_8 is C_6-free*

Remark: We could have chosen any quadratic surface instead of Q as they are all *equivalent*, that is, for any quadratic form there exists a nonsingular linear transformation which transforms it to Q.

Exercise 3.4 *Prove parts $(i) - (iv)$.*

Proof. (of (v)) Let $a, b \in Q_4$ be two distinct points on the surface Q_4. If the line $\ell_{ab} = \{\lambda a + \mu b : (\lambda,\mu) \in \mathbb{F}_q^2 \setminus \{(0,0)\}\}$ spanned by a and b is fully contained in Q_4, then $B_Q(a,b) = 0$ for the associated bilinear form defined by

$$B_Q(x,y) := \frac{1}{2}(Q(x+y) - Q(x,x) - Q(y,y)) = y_0x_0 + \frac{1}{2}(x_1y_{-1} + x_{-1}y_1 + x_2y_{-2} + x_{-2}y_2).$$

Indeed, $Q(\lambda x + \mu y) = \lambda^2 Q(x) + 2\lambda\mu B_Q(x,y) + \mu^2 Q(y)$.

Let us assume that G_8 contains a 6-cycle. This means that there are three distinct points $a, b, c \in Q_4$, such that the respective connecting lines $\ell_{ab}, \ell_{bc}, \ell_{ca}$ are all distinct and fully contained in Q_4. These lines are distinct if and only if the five-dimensional vectors a, b, and c are linearly independent. Consider now the 3×5-dimensional non-singular matrix M with rows a, b, and c. The kernel of this matrix is then 2-dimensional. However, we find three linearly independent vectors $\hat{a} = (a_2, a_1, 2a_0, a_{-1}, a_{-2})$, $\hat{b} = (b_2, b_1, 2b_0, b_{-1}, b_{-2})$ and $\hat{c} = (c_2, c_1, 2c_0, c_{-1}, c_{-2})$ in the kernel, since $B_Q(x,y) = 0$ for all nine choices of $x = a, b, c$ and $y = \hat{a}, \hat{b}, \hat{c}$. The vectors $\hat{a}, \hat{b}, \hat{c}$ are indeed linearly independent since the original vectors a, b, c are linearly independent, a contradiction. □

Remark: "Geometrically" the bilinear form $B_Q(x,y)$ is the equation of the "tangent hyperplane" of Q_4 at $x \in Q_4$. This is of course the same as the "tangent hyperplane" at $y \in Q_4$, since the whole line ℓ_{xy} is on the surface.
Note that if $a, b, c \in Q_4$ and $\ell_{ab}, \ell_{bc}, \ell_{ca} \subseteq Q_4$ then $a + b + c$, and in fact the whole plane spanned by a, b, and c is contained in Q_4 as well.

Corollary 3.10

$$ex(n, C_6) = \Theta(n^{4/3}).$$

To construct a C_{10}-free graph, Benson uses the same basic idea of point-line incidence graphs, but the analysis gets more complicated. The space is now $PG(n,6)$ and the surface on which the construction lives is $Q_6 = \{x : x_0^2 + x_1x_{-1} + x_2x_{-2} + x_3x_{-3} = 0\}$. We have $|Q_6| = 1 + q + q^2 + \cdots + q^5$ and degree of each line is $q + 1$, no matter what. We cannot, however, take all the lines on Q_6, since on the one hand there are too many of them, and on the other hand there are not only triangles but full planes contained in Q_{10}, so C_6 would abound. Benson selects $1 + q + q^2 + \cdots + q^5$ lines of Q_6 and proves that the corresponding point-line incidence graph is $(q+1)$-regular and has girth 12.

Exercise 3.5 *Show that no 3-dimensional space is fully contained in Q_6*

These constructions of Benson show that for $k = 2, 3$ and 5,

$$\frac{1}{2 \cdot \sqrt[k]{2}} n^{1+\frac{1}{k}} + O(n) \leq ex(n, C_3, \ldots, C_{2k}, C_{2k+1}) \leq ex(n, C_3, \ldots, C_{2k}) \leq ex(n, C_{2k})$$

Erdős and Simonovits (1982) showed that the first inequality is essentially tight by asymptotically extending the even-Moore bound to non-regular graphs. They showed that for any fixed integer $k \geq 2$, we have

$$ex(n, C_3, \ldots, C_{2k}, C_{2k+1}) \leq \frac{1}{2 \cdot \sqrt[k]{2}} n^{1+\frac{1}{k}} + O(n^{1-\frac{1}{k}}).$$

Recall that in Exercise 2.2 we saw that $ex(n, C_4, C_3, C_5, C_7, \ldots) \approx \frac{1}{2\sqrt{2}} n^{3/2}$, which is of course much weaker than the theorem of Erdős and Simonovits. For $k = 2$, they

strengthened their result and obtained that even $ex(n, C_4, C_5) \approx \frac{1}{2\sqrt{2}} n^{3/2}$. It is not known, however, what happens to $ex(n, C_3, C_4)$.

Open Problem Is it true that $ex(n, C_3, C_4) \approx \frac{1}{2\sqrt{2}} n^{3/2}$?

By Theorem 2.1 we know that $ex(n, C_3, C_4) \leq ex(n, C_4) \approx \frac{1}{2} n^{3/2}$, while the point-line incidence graph of the projective plane shows that $ex(n, C_3, C_4) \geq \frac{1}{2\sqrt{2}} n^{3/2}$.

3.4.2 Wenger's construction

Twenty-five years after Benson, Wenger found the simplest construction of dense C_6- and C_{10}-free graphs which I can imagine to be available. For a field $\mathbb{F}$, and integer $k \in \mathbb{N}$ let us define a curve in $\mathbb{F}^k$ given by $M_k = \{(1, \alpha, \alpha^2, \ldots, \alpha^{k-1}) : \alpha \in \mathbb{F}\}$. This curve is called the *moment curve*. One of its important properties is that every k different vectors on it are linearly independent. To see this, it is enough to check the determinant of the matrix containing the coordinates of k distinct points of the curve. It is a *Vandermonde-determinant*, and as such nonzero.

$$|A| = \begin{vmatrix} 1 & \alpha_1 & \alpha_1^2 & \ldots & \alpha_1^{k-1} \\ 1 & \alpha_2 & \alpha_2^2 & \ldots & \alpha_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_k & \alpha_k^2 & \ldots & \alpha_k^{k-1} \end{vmatrix} = \prod_{1 \leq i < j \leq k} (\alpha_j - \alpha_i) \neq 0.$$

The graph $W_{k,q}$ we define is going to be a bipartite point-line incidence graph living in the k-dimensional affine space over $\mathbb{F}_q$. We prefer to think of a line as the set $\{\beta \vec{s} + \vec{v} : \beta \in \mathbb{F}_q\}$ of q points, where $\vec{s} \in \mathbb{F}_q^k$, $\vec{s} \neq \vec{0}$ is the "slope" of the line.

There are $\frac{(q^k-1)q^k}{(q-1)q} \approx q^{2k-2}$ lines in $\mathbb{F}_q^k$; way too much more than the number of points, for $k > 2$. In order to cut down on the number of lines we only consider those whose "slope" is on the moment curve and call them *distinguished*. The number of distinguished lines is $\frac{q \cdot q^k}{q} = q^k$. Let $W = W_{k,q}$ be the point-line incidence graph with vertex set

$$V(W) = \mathbb{F}_q^k \cup \{\ell : \ell \text{ is a distinguished line}\}.$$

Each line contains q points, and through each point q lines pass with a slope from the moment curve. Hence W is q-regular, where $q = \left(\frac{n}{2}\right)^{1/k}$ with $n = |V(W)|$.

Obviously, there is no C_4 in W since it is a point-line incidence graph, and this holds for every $k \geq 2$.

Another general fact about these graphs is that if two lines ℓ_1, ℓ_2 intersect in $\mathbb{F}_q^k$ (i.e., they have a common neighbor in G), then their slope is different.

Now let $k \geq 3$. We claim that there is no C_6 in W. If there was a C_6, say $p_1, \ell_1, p_2, \ell_2, p_3, \ell_3$ in W, then the slopes $\vec{s}_1, \vec{s}_2, \vec{s}_3 \in M_k$ of the three distinguished lines ℓ_1, ℓ_2, ℓ_3 would be all distinct. Since p_1 and p_2 are distinct points on line ℓ_1 we have $\vec{p}_2 - \vec{p}_1 = \alpha_1 \vec{s}_1$, where $\alpha_1 \in \mathbb{F}_q^*$. Similarly, $\vec{p}_3 - \vec{p}_2 = \alpha_2 \vec{s}_2$ and $\vec{p}_1 - \vec{p}_3 = \alpha_3 \vec{s}_3$, where $\alpha_i \in \mathbb{F}_q^*$. Summing up these equations gives a nontrivial linear combination of three vectors from the moment curve M_k, a contradiction for $k \geq 3$.

One cannot extend the previous argument for C_8 and $k \geq 4$. The problem is that the slope vectors $\vec{s}_1, \vec{s}_2, \vec{s}_3, \vec{s}_4 \in M_k$ of the four distinguished lines participating in an 8-cycle would not necessarily be all distinct. The ones having a common neighbor will of course be different, but it could happen (and will) that the opposite ones are parallel, i.e., $\vec{s}_1 = \vec{s}_3$, $\vec{s}_2 = \vec{s}_4$. Then we are dealing with a parallelogram, for the coefficients we have $\alpha_1 = -\alpha_3$, $\alpha_2 = -\alpha_4$, so the resulting linear combination of the four vectors $\alpha_1\vec{s}_1 + \alpha_2\vec{s}_2 + \alpha_3\vec{s}_3 + \alpha_4\vec{s}_4 = 0$ is trivial. Hence W contains a lot of C_8s for every $k \geq 2$.

Is this then the end of the story? What happens for $2k$-cycles?

Let $k \geq 5$. In case there was a 10-cycle, the slope vectors of the participating lines would again provide a linear combination $\alpha_1\vec{s}_1 + \alpha_2\vec{s}_2 + \alpha_3\vec{s}_3 + \alpha_4\vec{s}_4 + \alpha_5\vec{s}_5 = 0$, where each $\alpha_i \neq 0$. A minute of meditation convinces us that, since $\vec{s}_i \neq \vec{s}_{i+1}$, $i = 1, \ldots, 5$, there is an index j such that $\vec{s}_j \neq \vec{s}_i$ for all $i \neq j$. Hence the linear combination is nontrivial, a contradiction for $k \geq 5$. The graph W is C_{10}-free!

We have proved the following.

Theorem 3.11 (Wenger, 1991) *$W_{3,q}$ is a q-regular bipartite graph on $2q^3$ vertices with girth 8.*
$W_{5,q}$ is a q-regular bipartite graph on $2q^5$ vertices which is C_{10}-free.

For a similar reason as in the case of $k = 4$, the construction does not work for any $k > 5$.

In conclusion, Wenger's graphs have roughly the same number of edges as the corresponding graphs of Benson. The important difference is that while Benson's C_{10}-free graphs have girth 12, Wenger's C_{10}-free graphs do contain 8-cycles.

Exercise 3.6 *Let $W_{4,q}$ be the four dimensional Wenger graph. Let $\theta(3,4)$ be the graph consisting of the union of three internally disjoint paths of length four between two vertices.*

(a) Characterize the copies of C_8 and $\theta(3,4)$ in $W_{4,q}$.

(b) Establish the existence of $\theta(q,4)$ in $W_{4,q}$.

(c) Prove that $W_{4,q}$ is not vertex transitive.

3.4.3 Constructions for arbitrary k

We studied extensively the cases of C_4, C_6 and C_{10}, where the order of magnitude of $ex(n, C_{2k})$ is known. For arbitrary k it is worthwhile to compare the probabilistic lower bound of $\Omega(n^{1+1/(2k-1)})$ for $ex(n, C_{2k})$ with the best known constructions, which we do not get to discuss in detail. There are two different sequences of graphs both improving the constant factor $\frac{1}{2}$ of $\frac{1}{k}$ in the *exponent* to $\frac{2}{3}$. Note that this constant factor is 1 in the upper bound. The first such sequence to achieve this consists of the famous *Ramanujan graphs* of Margulis (1988) and Lubotzky, Phillips and Sarnak (1988). These graphs were primarily constructed to provide explicit constant degree expanders, but they work for

our purpose as well, since with appropriate parameters they are dense graphs of high girth. Quantitatively, they show that for arbitrary $k \geq 2$,

$$\Omega\left(n^{1+\frac{2}{3k+3}}\right) \leq ex(n, C_3, C_4, \ldots, C_{2k}, C_{2k+1}) \leq ex(n, C_{2k}).$$

Later Lazebnik, Ustimenko and Woldar (1995) gave a completely different and more elementary(?) construction of high girth, dense graphs (but not constant degree expanders!) which polished these bounds a bit:

$$\Omega\left(n^{1+\frac{2}{3k-3+\epsilon}}\right) = ex(n, C_3, C_4, \ldots, C_{2k}, C_{2k+1}) \leq ex(n, C_{2k}),$$

where $\epsilon = 0$ if k is odd, and $\epsilon = 1$ when k is even. Their construction could be considered a generalization of Wenger's graph. They provide the best known lower bound for every $k \geq 4$, $k \neq 5$.

3.5 Polishing the constant

The general question we address in this section is the following:
Meta Problem Is the large girth problem significantly different from the $2k$-cycle-free graph problem?!

In Sections 3.3 and 3.4 we saw that they certainly don't differ in the order of magnitude for $k = 2, 3$ and 5. But for arbitrary k, the answer is not known. To formulate this precisely let k be a fixed integer.

Conjecture 5 *There exists a constant c_k such that*

$$\limsup_{n\to\infty} \frac{ex(n, C_{2k})}{ex(n, C_3, C_4, \ldots, C_{2k})} = c_k.$$

For $k = 2, 3$ or 5, when the orders of magnitude are known to be equal one wonders about the value of c_k. As we saw in the previous section,

$$ex(n, C_3, \ldots, C_{2k}) \lessapprox \frac{1}{2} n^{1+1/k}.$$

Erdős and Simonovits conjectured that the same upper bound holds for $ex(n, C_{2k})$.

Currently, Benson's graphs give us

$$\begin{aligned} \frac{1}{2} n^{3/2} &\approx ex(n, C_4), \\ 0.396 \cdot n^{4/3} \leq \frac{1}{2^{4/3}} \cdot n^{4/3} &\lessapprox ex(n, C_6), \text{ and} \\ 0.435 \cdot n^{6/5} \leq \frac{1}{2^{6/5}} \cdot n^{6/5} &\lessapprox ex(n, C_{10}). \end{aligned}$$

In the next subsections we give improvements on $ex(n, C_{10})$ and $ex(n, C_6)$, respectively, which show that the conjecture of Erdős and Simonovits is false. While the first one is based on a simple trick, the latter further develops an idea we already used in Subsection 2.1.3 for the case of C_4-free graphs.

3.5.1 Denser C_{10}-free graphs

For the 10-cycle we can resolve the Meta Problem and show that the values of $ex(n, C_{10})$ and $ex(n, C_3, C_4, \ldots, C_{10})$ are *not equal* asymptotically. On the one hand, we saw in the previous section that

$$ex(n, C_3, C_4, \ldots, C_{10}) \lessapprox \frac{1}{2} n^{6/5}.$$

On the other hand, a denser C_{10}-free graph can be obtained by the following simple idea of Lazebnik, Ustimenko, and Woldar [**?**].

We start from Benson's bipartite graph G_{12} of girth 12, with bipartition (A, B). Then we split each vertex of B into four. Formally we create four disjoint set B_1, B_2, B_3, B_4 with $|B| = |B_i|$, $i = 1, 2, 3, 4$, and label their elements by $b^{(i)} \in B_i$, where $b \in B$. The neighborhood of vertex $b^{(i)}$ will be the same as the neighborhood of b. Assume that there is a C_{10} in G with vertices $a_1, b_1^{(i_1)}, a_2, b_2^{(i_2)}, a_3, b_3^{(i_3)}, a_4, b_4^{(i_4)}, a_5, b_5^{(i_5)}$. This 10-cycle corresponds to a subgraph C of G_{12}, which admits a closed walk $a_1, b_1, a_2, b_2, a_3, b_3, a_4, b_4, a_5, b_5$ of length 10, where *all* a_i are distinct and *not all* b_i are the same. Each a_i is passed through exactly once by the walk. Let us delete from C the edges used twice. Each such deletion will make one a_i isolated. The leftover is a graph with Eulerian components. Since not all b_i are the same, we did not delete all the edges. This nontrivial Eulerian component must contain a cycle, but that cannot happen here, since the Benson graph has girth 12.

We have $5(q^5 + q^4 + \cdots + q + 1) \approx 5q^5$ vertices, and $4(q+1)(q^5 + q^4 + \cdots + q + 1) \approx 4q^6$ edges. Therefore,

Proposition 3.12

$$ex(n, C_{10}) \gtrapprox \frac{4}{5^{6/5}} n^{6/5} \geq 0.579 n^{6/5}.$$

The analogous trick does work for C_6-free graphs as well, but the resulting graph does not raise the value of $ex(n, C_6)$ up to $\frac{1}{2} n^{4/3}$.

Exercise 3.7 *Construct a C_6-free graph with $\frac{2}{3^{4/3}} n^{4/3} \approx .0462 n^{4/3}$ edges.*

Hence there *is* a denser C_6-free graph than the Benson-graph, if we don't insits on the girth being larger than 6. To reach $\frac{1}{2} n^{4/3}$ we need yet another geometrically motivated idea.

3.5.2 Denser C_6-free graphs — polarities

We start this section by generalizing the idea of Construction 3. of Subsection 2.1.3.

Let $(\mathcal{P}, \mathcal{L})$ be a hypergraph[2] with vertex set $\mathcal{P}$ and edge set $\mathcal{L} \subseteq 2^{\mathcal{P}}$. The *incidence graph* $\Gamma = \Gamma(\mathcal{P}, \mathcal{L})$ of hypergraph $(\mathcal{P}, \mathcal{L})$ is the bipartite graph with vertex set $V(\Gamma) =$

[2] In case of a geometric motivation one calls the elements of $\mathcal{P}$ *points*, and the elements of $\mathcal{L}$ *lines*. In this case the set $I := \{(p, \ell) : p \in \ell\} \subseteq \mathcal{P} \times \mathcal{L}$ is called an *incidence relation* on $(\mathcal{P}, \mathcal{L})$, and one says that the triple $(\mathcal{P}, \mathcal{L}, I)$ is a *(rank two) geometry*.

$\mathcal{P} \cup \mathcal{L}$, and edge set $E(\Gamma) = \{p\ell : p \in \mathcal{P}, \ell \in \mathcal{L}, p \in \ell\}$. An automorphism of Γ of order two, which exchanges $\mathcal{P}$ and $\mathcal{L}$ is called a *polarity* of $(\mathcal{P}, \mathcal{L})$. That is, function $\pi : \mathcal{P} \cup \mathcal{L} \to \mathcal{P} \cup \mathcal{L}$ is a polarity if

(i) $\pi(\mathcal{P}) = \mathcal{L}$ and $\pi(\mathcal{L}) = \mathcal{P}$,

(ii) for every $p \in \mathcal{P}$ and every $\ell \in \mathcal{L}$, we have that $p \in \ell$ if and only if $\pi(\ell) \in \pi(p)$,

(iii) $\pi^2 = \mathrm{id}$.

Example. Let $\mathcal{P}$ be the set of points of the projective plane over $\mathbb{F}_q$ and let $\mathcal{L}$ be the set of lines of the same plane. Recall that a point p was described by three coordinates (x_0, x_1, x_2) and a line l described by the triple $[y_0, y_1, y_2]$ contains all those points (x_0, x_1, x_2) which satisfy $x_0y_0 + x_1y_1 + x_2y_2 = 0$. Here the function $(x_0, x_1, x_2) \to [x_0, x_1, x_2]$ defines a polarity.

In case a polarity exists, we can define the polarity graph of the hypergraph which, compared to the incidence graph, cuts the number of vertices in half while keeps almost all degrees intact. Formally, the *polarity graph* Γ^π is the graph with $V(\Gamma^\pi) = \mathcal{P}$, and $E(\Gamma^\pi) = \{p_1p_2 : p_1 \neq p_2, p_1 \in \pi(p_2)\}$.

Example. Construction 3. of Subsection 2.1.3 is the polarity graph of the hypergraph defined (in the previous example) by the points and lines of the projective plane over $\mathbb{F}_q$.

We say that $p \in \mathcal{P}$ is an *absolute point* of π if $p \in \pi(p)$. Absolute points are the ones that would define a loop in the polarity graph, had we not excluded that possibility. The set of all absolute points is denoted by $N_\pi = \{p : \ p \text{ is absolute point of } \pi\}$.

Lemma 3.12.1 *The following properties hold.*

(i) *If $p \in N_\pi$, then we have $d_{\Gamma^\pi}(p) = d_\Gamma(p) - 1$. Otherwise, $d_{\Gamma^\pi}(p) = d_\Gamma(p)$.*

(ii) *$|V(\Gamma^\pi)| = \frac{1}{2}|V(\Gamma)|$ and $|E(\Gamma^\pi)| = \frac{1}{2}\left(|E(\Gamma)| - |N_\pi|\right)$,*

(iii) *If $C_{2k+1} \subseteq \Gamma^\pi$, then $C_{4k+2} \subseteq \Gamma$,*

(iv) *If $C_{2k} \subseteq \Gamma^\pi$, then there are two copies of C_{2k} in Γ such that one is the polar dual of the other,*

(v) *$g(\Gamma^\pi) \geq \frac{1}{2}g(\Gamma)$.*

Exercise 3.8 *Prove the lemma.*

With this definition in hand we try to generalize the idea of Construction 3. of Subsection 2.1.3 and define an appropriate geometry which admits a polarity. As it turns out, Wenger's graph (and Benson's as well), at least for certain prime powers, is the incidence graph of one. Let $\mathcal{P}_* = \{(a, b, c) : a, b, c \in \mathbb{F}_q\}$, $\mathcal{L}_* = \{[d, e, f] : d, e, f \in \mathbb{F}_q\}$, and $(a, b, c) \in [d, e, f]$ if and only if $e - b = da$, $f - c = ea$.

Exercise 3.9 *Prove that the incidence graph of the above geometry is nothing else but Wenger's C_6-free construction.*

Exercise 3.10 *[Lazebnik, Ustimenko, Woldar [?]] Let $q = 2^{2\alpha+1}$ and define π_* by*

$$\begin{aligned} \pi_* : (a,b,c) &\mapsto \left[a^{2^{\alpha+1}}, (ab)^{2^\alpha} + c^{2^\alpha}, b^{2^{\alpha+1}}\right], \\ \pi_* : [d,e,f] &\mapsto \left(d^{2^\alpha}, f^{2^\alpha}, (df)^{2^\alpha} + e^{2^\alpha+1}\right). \end{aligned}$$

(a) *Prove that π_* is a polarity.*

(b) *Prove that set of absolute points is*

$$N_{\pi_*} := \{(a, b, a^{2^{\alpha+1}+2} + ab + b^{2^{\alpha+1}}) : a, b \in \mathbb{F}_q\}.$$

(c) *Show that Γ^{π_*} is C_6-free and conclude that $ex(n, C_6) \geq \frac{1}{2}n^{4/3} - \frac{1}{2}n^{2/3}$ for every $n = 2^{3(2\alpha+1)}$ with $\alpha \in \mathbb{N}$.*

Hence there is a C_6-free graph which is denser than the one obtained by the doubling trick of Exercise 3.7. However, the result of the previous exercise compared with the result of the previous section does not exclude the possibility that $ex(n, C_6)$ and $ex(n, C_3, C_4, C_5, C_6)$ are asymptotically the same, $\frac{1}{2}n^{4/3}$. This was done in a recent paper by Füredi, Naor and Verstraëte (2005), but only for an infinite sequence of n, not for all n. They gave a random twist to the polarity graph of the Wenger graph to improve the constant factor above $\frac{1}{2}$—but again only for an infinite sequence of n. It is still only a conjecture that $\lim_{n\to\infty} ex(n, C_6)n^{-4/3}$ exists and we do not have an educated guess what its value should be.

Theorem 3.13 *For infinitely many values of n we have*

$$0.534n^{4/3} \lessapprox ex(n, C_6).$$

Proof. Let G be the polarity graph defined in Exercise 3.10. We define a doubling trick similar to the one in the previous subsection. Since G is not bipartite to begin with, we have to be more careful. Let $A \subset V(G)$ be an arbitrary subset of the vertices, and let $B = V(G) \setminus A$. Let us fix an arbitrary orientation τ of the edges of $G[A]$. Let us define a graph $H = H(G, A, \tau)$ on the vertex set $V(H) = V(G) \cup A'$, where the set A' is independent, has the same size as A, and contains a "twin" v' of each vertex $v \in A$. All edges inside $V(G)$ remain, moreover we include exactly one new twin edge for each edge incident to A as follows. Vertices in A' have the same neighbors from B, as their respective twins in A. Furthermore, for each edge $wz \in E(G)$ inside A, exactly one of the two corresponding pairs between A and A' is put in $E(H)$. That which one, depends on the τ-orientation of wz. Formally,

$$\begin{aligned} E(H) \;=\; & E(G) \cup \{w'v : w' \in A', v \in B, wv \in E(G)\} \cup \\ & \cup \{w'z : w' \in A', z \in A, wz \in E(G), \tau \text{ orients } wz \text{ from } w \text{ towards } z\} \end{aligned}$$

Claim 3 *If G contains neither C_3, C_4 nor C_6, then $H = H(G, A, \tau)$ is C_6-free for any A and τ.*

Proof. Suppose to the contrary that there is a 6-cycle C in H. The cycle C, through the twin edges, determines a closed walk W of length 6 in G. The edges of W must form a tree, since G does not contain C_3, C_4 and C_6, and containing a C_5 does not allow a walk of six edges to be closed. Hence each edge of W is traversed at least twice by W. Since G is C_6-free, there must be at least one vertex a' of C in A'. Since A' is independent, both neighbors x and y of a' in C, must be in $V(G)$. Then W tranverses both ax and ay, and traverses both (at least) twice. So the edges ax and ay must also be in C. That is a contradiction, because then the 6-cycle C would contain the edges of a four-cycle a, x, a', y, a contradiction. □

To finish the proof of Theorem 3.13 we just need to select the appropriate A — which will be a random one. The number of vertices increases by $|A|$ while all edges having an endpoint in A will have a twin.

Exercise 3.11 *Finish the proof of the theorem by calculating the expected number of edges of H if A is selected uniformly at random among all subsets of size k for an appropriate value of k.*

□

Füredi, Naor and Verstraëte also improved the best known upper bound which currently stands at $0.627n^{4/3} \gtrapprox ex(n, C_6)$.

The definition of the polarity in Exercise 3.10 could seem a bit ad hoc and definitely mysterious in terms of how one can find it. The remark below, following [**?**], tries to shed some light (without proof).

Remark: A *generalized m-gon* is a geometry $(\mathcal{P}, \mathcal{L}, I)$ whose incidence graph Γ is d-regular, with girth $2r$ and diameter r. (This is a special case of the definition given by Tits but it will do for the purposes of this paragraph.)

For $d = 2$, generalized r-gons exist for for all $r \geq 3$: the incidence structure of the vertices and sides of a usual r-gon. For $d \geq 3$ generalized r-gons are called *thick* and their occurrence is very rare. Feit, Higman (1964) and Singleton (1963, 1966) proved that thick generalized r-gons can exist only for $r = 3, 4$ or 6.

The incidence graph of the projective plane is a thick generalized triangle existing for every $d = q + 1$, where q is a prime-power. Benson's C_6-free and C_{10}-free graphs are examples of thick generalized quadrangles and hexagons, respectively.

The concept of generalized r-gons was introduced by Tits in an influential paper in group theory. Here is how group theorists constructed thick regular r-gons (i.e., dense C_4-, C_6-, C_{10}-free graphs.) Let G be a group, and let $P_1, P_2 \leq G$ be two subgroups of G. We denote the sets of left cosets by $\mathcal{P} := (G : P_1)$, $\mathcal{L} := (G : P_2)$, and define the incidence relation by

$$I := \{(S, T) : S \in \mathcal{P}, T \in \mathcal{L}, S \cap T \neq \emptyset\}.$$

For appropriately chosen groups G, P_1 and P_2 the corresponding geometry turns out to be a thick regular generalized r-gon. Let $q = p^\alpha$ be a prime power and let $A_2(q), B_2(q)$, and $G_2(q)$ denote the respective rank two Chevalley groups. All three of these are sequences of well-studied, classic groups of Lie-type over finite fields, providing infinite sequences of finite simple groups. $A_2(q)$ is a projective special linear group and as such, also known under the name $PSL_3(q)$. If $G \cong A_2(q), B_2(q)$ or $G_2(q)$, and both $P_1, P_2 \leq G$ are maximal subgroups containing the normalizer B of a fixed p-Sylow subgroup of G, then $(\mathcal{P}, \mathcal{L}, I)$ is a $(q+1)$-regular r-gon for $r = 3, 4$, and 6, respectively, and we have $|\mathcal{P}| = |\mathcal{L}| = 1 + q + q^2 + \ldots q^{r-1}$.

It is known that if q is appropriate, then these geometry have a polarity. In particular, for generalized triangles (of type $A_2(q)$), it exists for every prime power q. For generalized quadrangles (of type $B_2(q)$), it exists if $q = 2^{2\alpha+1}$. For generalized hexagons (of type $G_2(q)$), it exists if $q = 3^{2\alpha+1}$.

The affine part. In order to check the existence of these polarities we pass to the so-called affine part of these geometries. The normalizer B acts on $\mathcal{P}$ and $\mathcal{L}$ by left multiplication, and the orbits of B have length $1, q, \ldots, q^{r-1}$, respectively. The restriction of I to the *largest* orbits $\mathcal{P}_*$ and $\mathcal{L}_*$ (of size q^{r-1} each) defines the *affine subgeometry* $(\mathcal{P}_*, \mathcal{L}_*, I_*)$ of $(\mathcal{P}, \mathcal{L}, I)$. By Γ_* we denote its incidence graph.

It is known that if there is a polarity of the geometry, then there is one that maps $\mathcal{P}_* \cup \mathcal{L}_*$ into itself. Thus for the above cases, there exists a polarity π_* of the affine part $(\mathcal{P}_*, \mathcal{L}_*, I_*)$.

The nice thing about the incidence graphs and polarities of the affine part is that they admit a simple coordinatized description. In order to check their properties one does not need to have any idea where they came from. Of course one won't be able to come up with one without intuition. In Exercise 3.10 we looked at a coordinatization of a polarity of the affine part of the generalized quadrangle (of type $B_2(q)$), which eventually improved the leading constant of $ex(n, C_6)$ from $1/2\sqrt[3]{2}$ to $1/2$. Similar simple coordinatized description exists for the affine part of the generalized hexagon, but that would not lead us to a better construction, since for $ex(n, C_{10})$ the leading constant was already proved to be larger than $1/2$ in the last subsection.

3.6 Dense regular graphs with large girth

The starting point of our discussion in this chapter was the Moore-bound (see Proposition 3.1 and Exercise 3.1), providing a lower bound on the minimum number of vertices in d-regular graphs with girth g. A *Moore-graph* is a d-regular graph with girth g, having exactly as many vertices as the corresponding (odd or even) Moore-bound. The existence of Moore-graphs is decided for many values of the two parameters. For $g = 3$ and arbitrary $d \geq 3$ the complete graph K_{d+1} provides the unique example, while for $g = 4$ and arbitrary $d \geq 2$ the complete bipartite graph $K_{d,d}$ is the unique Moore-graph. For $g = 5$, it was proven by Hoffmann and Singleton using spectral methods, that Moore-

graphs can only exist when $d = 2, 3, 7$, or 57. In the first three cases, there are unique examples: C_5, the Petersen graph, and the Hoffmann-Singleton graph. The existence of a 57-regular graph of girth 5 on $1 + 57 + 57 \cdot 56 = 3250$ vertices is still one of the tantalizing mysteries of algebraic graph theory. Then of course there is always the cycle C_g of length g, providing the unique example for $d = 2$ and arbitrary $g > 3$, but otherwise the existence of Moore-graph for $g > 5$ is very limited. Bannai and Ito [?] and Damerell [?] have shown that no Moore graph with odd girth $g \geq 7$ and $d > 2$ can exist. The even girth case was settled by a theorem of Feit and Higman [?], which implies that Moore-graphs with even girth g and $d > 2$ cannot exist unless $g = 4, 6, 8$, or 12. In the latter three cases the Benson-graphs provide Moore-graphs whenever d is of the form $q + 1$ with q being a prime power. The question of existence for $g = 6, 8$, and 12 is open for other values of d.

The limited range of parameters when the Moore-bound can be tight motivates the definition of the *cage number* $c(d, g)$, representing the smallest number n of vertices on which there is a d-regular graph with girth g. We will see later that this quantity is well-defined.

Above we have overviewed the cases when the cage number can be equal to the Moore-bound and found that it can never happen if the girth g is larger than 12. For our investigation of the Turán number $ex(n, C_{2k})$ these construction were relevant, because we were interested in C_{2k}-free constructions for some *fixed value of* k and tried to create as dense graph as possible, which is equivalent to achieving a given degree d of regularity with as few vertices as possible.

Now we will concentrate the other end of the spectrum: keep the degree d of regularity a fixed constant and let g be large/tend to infinity.

In that case the order of magnitude of both the even and the odd Moore bound is $(d-1)^{\lfloor (g-1)/2 \rfloor} = \Omega\left((d-1)^{\frac{1}{2}g}\right)$, that is exponential in g with the base $\sqrt{d-1}$.

In the next exercise we describe an upper bound due to Erdős and Sachs.

Exercise 3.12 *Let $f(d, g)$ be the smallest n such that there exists a d-regular graph with girth at least g on n vertices.*

- *Let G be a graph on $2m \geq 4\sum_{i=0}^{g-2}(d-1)^i$ vertices such that (i) G has girth at least g and (ii) every vertex $v \in V(G)$ has degree $d-1$ or d, and G has the largest possible number of edges among graphs with these properties. Prove that G is d-regular*
- *Conclude that $f(d, g) \leq 4\sum_{i=0}^{g-2}(d-1)^i$*

Remark: The above bound is due to Erdős and Sachs and it is roughly the square of the Moore bound. They also derive that every d-regular graph with girth at least g *does have* a cycle of length g, so the cage number $c(d, g)$ exists and is equal to $f(d, g)$. That is, we obtain

$$c(d, g) \leq (d-1)^{(1+o(1))g},$$

that is roughly the square of the lower bound.

In the exercise we proved the upper bound using an implicit inductive argument. This bound turned out be very difficult to topple. Actually the explicit construction of a fixed d-regular graph with girth at least g on just exponentially many $c(d)^g$ vertices turned out to be not an easy task for any constant $c(d)$.

In the 1980's Margulis, and independently Lubotzky, Phillips and Sarnak, obtained construction using groups and sophisticated algebraic number theory to construct graphs for every fixed $d = q + 1$ and arbitrary large g and number of vertices $c(d, g) \leq (d - 1)^{\left(\frac{3}{4}+o(1)\right)g}$. The main goal of these constructions were to give explicit constant degree expander graphs, with second eigenvalue as small as possible, the lack of short cycles was only side product. Later Lazebnik, Ustimenko, and Woldar gave alternative constructions that give similar, but slightly better bound.

The description and proof of correctness of these constructions are beyond the scope of our lecture notes. Our goal in this section is to introduce a much simpler construction of Margulis, with a somewhat larger, but still exponential, number of vertices. This construction will provide at least partial glimpse into how the more complicated ones came around. We largely follow the treatment of [?].

The simple canonical way to construct graphs that are regular is by Cayley graphs. They are also natural candidates to examine for girth problems as cycles in Cayley graphs have a simple description based on the generators.

For a group $\langle G, \cdot \rangle$ and a subset $S \subseteq G$ of generators with the property $S = S^{-1}$, and $1 \notin S$, we define the Cayley graph $C = C(G, S)$ as follows. The vertex set is the group G and two group elements g and $h \in G$ are adjacent if $gh^{-1} \in S$. Note that $gh^{-1} \in S$ if and only if $hg^{-1} = (gh^{-1})^{-1} \in S$, because $S = S^{-1}$. Moreover there are no loops since $1 \notin S$.

If $S = \{s_1, \ldots s_d\}$, then the neighbors of any group element g are $gs_1, \ldots gs_d$. In particular C is $|S|$-regular. Furthermore, there exists a cycle of length ℓ in C if and only if there is a relation $s_{i_1} \cdot \cdots \cdot s_{i_\ell} = 1$ of minimal size that expresses the unit element as a product of ℓ elements $s_{i_1}, \ldots, s_{i_\ell}$ of S. (That is, there is no proper subsequence $s_{i_j}, \ldots, s_{i_k}$, $1 \leq j \leq k \leq \ell$ whose product is the identity element.)

To demonstrate the idea of Cayley graphs and establish explicitely that for *any* degree d, there exists a d-regular graph with arbitrary large girth, we describe first a simple construction with much worse parameters. Let r be arbitrary and let $T = T_{d,r}$ be the full d-ary tree of depth r, with root vertex w. Our Cayley graph will be over the symmetric group S_V of permutations of the vertex set $V = V(T)$. To define the generators, we fix an arbitrary proper d-coloring $\chi : E(T) \to [d]$. This is easy to find by first coloring the d edges incident to the root with distinct colors and then proceeding down the tree level by level, always coloring properly the $d - 1$ uncolored edges at each new vertex with the remaining colors. For each color $i = 1, \ldots, d$ we define a permutation $\pi_i \in S_V$ as follows. For a vertex $u \in V(C)$ which has a neighbor z such that $\chi(uz) = i$, we set $\pi_i(u) = z$ (since χ is proper, there is no more than one such neighbor z). Otherwise, $\pi_i(u) = u$. Observe that this happens only for a leaf vertex u, if the color i is not exactly the one that appears on the sole edge incident to u.

We claim that the girth of this graph is at least $2r+1$, which shows our promised statement as r was chosen arbitrarily. If there is a cycle of length g in C, then there exists a product $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_g} = id_V$. Let us follow the position of the root vertex in this product. $\pi_{i_1}(w)$ is definitely on the first level as w has all colors, hence also i_1. Then $\pi_{i_2}\pi_{i_1}(w)$ is definitely on the second level of the tree T, since $i_2 \neq i_1$ and the vertex $\pi_{i_1}(w)$ has all colors but i_1 on its incident edges towards the second level. Similarly, $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_\ell}(w)$ is on the ℓth level of T for every $\ell \in [r]$. Finally, $\pi_{i_1}\pi_{i_2}\cdots\pi_{i_r}(w)$ is a leaf vertex. The next permutation, $\pi_{i_{r+1}}$ should leave this leaf fixed, since $i_r \neq i_{r+1}$, and then we need r more step back to the root. So the cycle has length at least $2r+1$.

Exercise 3.13 *Show that the girth of the above graph is in fact at least* $4r+2$.

The number of vertices in the above example is at least doubly exponential in r and hence also in the girth $\geq 2r+1$:

$$\left(\sum_{i=1}^{r}(d-1)^i\right)! \geq \left(\frac{(d-1)^r}{e}\right)^{(d-1)^r} \geq (d-1)^{(d-1)^r}.$$

In the main construction of this section we will define a 4-regular graph whose order is single exponential in its girth g. It will not be as good as the above existence proof which gives $O(3^g)$ vertices, but the constant in the base will also not be outrageous. This graph, due to Margulis, could be considered the prequel to the other, more complicated constructions which have only $O(3^{\frac{3}{4}g})$ vertices.

The idea is to start with the ultimate 4-regular graph of large girth: the infinite 4-regular tree, and construct it as a Cayley graph. Then we factor this group appropriately to make it finite. Factoring of the group creates a canonical Cayley graph, which is the homomorphic image of the original Cayley graph. What happens to the properties of the original Cayley graph, in particular what are the degrees and how can cycles occur? It turns out that the new Cayley graph is also 4-regular and cycles can occur only under very controled conditions and we will be able to track their size. The proof we include here is due to Gábor Tardos.

The group will be $SL_2(\mathbb{Z})$, i.e. the multiplicative group of (2×2) integer matrices with determinant 1. Our generator set S will contain four members, the following matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad A^{-1} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \quad B^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$$

First we show that $C(SL_2(\mathbb{Z}), S)$ is the infinite 4-regular tree.

Claim 4 *For any vector $x \in \mathbb{R}^2$ with $x_1, x_2 \neq 0$ and $x_1 \neq x_2$, exactly three of $\|Ax\|$, $\|A^{-1}x\|$, $\|Bx\|$, $\|B^{-1}x\|$ are larger than $\|x\|$, where the norm $\|y\|$ of vector $y \in \mathbb{R}^2$ denotes the infinity norm* $\max\{|y_1|, |y_2|\}$.

Proof. Suppose wlog that $x_1 > x_2$. Then $|2x_1 + x_2|$ and $|-2x_1 + x_2|$ are both at least $2|x_1| - |x_2| > |x_1|$, so $\|Bx\|$ and $\|B^{-1}x\|$ are larger than $\|x\| = x_1$. Now among $|x_1 + 2x_2|$

and $|x_1 - 2x_2|$ exactly one is larger and one is smaller than $x_1 = \|x\|$ (that which one is which depends on whether x_1 and x_2 have the same sign or not). That means that exactly one of $\|Ax\|$ and $\|A^{-1}x\|$ is larger than $\|x\|$ and the other is smaller. □

Now let us assume that there is a product $M_g \cdots M_1 = I$, with factors from S, that give the identity matrix. We will show that g is large. Let us follow the movement of the particular vector $y = (1, \sqrt{2})$ when we start applying the sequentially the M_is. (But any vector with algebraically independent coordinates would do.) Let j be the index for which the infinity norm of the image $M_j M_{j-1} \cdots M_1 y =: x$ is the largest. That means that two of the neighbors of x, that is $M_{j+1}x = M_{j+1}M_j \cdots M_1 y$ and $M_j^{-1}x = M_{j-1} \cdots M_1 y$ have norm that is not larger than $\|x\|$. This is a contradiction to the previous Claim, since the vector x satisfies its conditions. Indeed, $M_j M_{j-1} \cdots M_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is an integer matrix with determinant one. So applying it to y the resulting vector $x = (a + b\sqrt{2}, c + \sqrt{2})$ (i) cannot have a 0-coordinate, because that would mean that the matrix must have a 0 row, and (ii) cannot have equal coordinates because then the matrix had two equal rows. (All this because 1 and $\sqrt{2}$ are algebraically independent.) So there are no cycles and hence the Cayley graph $C(SL_2(\mathbb{Z}), S)$ is indeed the infinite 4-regular tree.

Now let us take the same four matrices as generators, but in the group $SL_2(\mathbb{Z}_p)$. This is a finite group of size $\frac{(p^2-1)(p^2-p)}{p-1} = p^2(p-1) \sim p^3$.

Let us take a shortest cycle and let $M_1, \ldots, M_g$ be the corresponding generators. We know that modulo p the product $M_g \cdots M_1 = I$ is the idenity matrix. But we also know that over $\mathbb{Z}$ the matrix is NOT the identity matrix I. So at least one of the entries must be at least p in absolute value, that is $\|M_1, \ldots, M_g\| \geq p$. But multiplying an arbitrary matrix B with any one of the four generator matrices $M \in S$, the infinity norm of the new matrix MB is at most 3 times the infinity norm of M. The infinity norm of any of the generators is 2, so $2 3^{g-1} \geq p$. This means that the number of vertices is at most

$$|V(C(SL_2(\mathbb{Z}_p)))| \sim p^3 = O(27^g).$$

The bound is admittedly weaker than the bound $O\left((2.28)^g\right)$ of [**?**, **?**], but the proof is sweet and self-contained. It is also weaker than the implicit bound $O\left(3^g\right)$ of Erdős and Sachs, but the construction is explicit. So let's just not worry (and be happy).

Chapter 4

Other Bipartite Graphs

4.1 Forbidding degenerate graphs

4.1.1 A conjecture of Erdős

In this section we discuss a conjecture of Erdős which tries to grasp the essence of the Kővári-Sós-Turán upper bound on the number of edges in $K_{t,s}$-free graphs.

Definition: A graph G is *r-degenerate* if there is an ordering $v_{i_1}, \ldots, v_{i_n}$ of the vertices of G such that $|N(v_{i_j}) \cap \{v_{i_1}, \ldots, v_{i_{j-1}}\}| \leq r$ for every $j = 1, \ldots, n$.

One can "build" any r-degenerate graph G by adding vertex after vertex such that the number of edges never increases by more than r at any addition. An equivalent definition of the r-degeneracy of G is that every subgraph has minimum degree at most r.

Obviously every graph of maximum degree r is also r-degenerate, but there is much more. The complete bipartite graph $K_{4,999}$ has maximum degree 999, but is 4-degenerate: one can build it by adding first one by one the four vertices of the smaller class (which do not add any edges) and then the vertices of the larger class (which all have degree 4).

Erdős conjectured a far-reaching generalization of the Kővári-Sós-Turán bound. In that bound the exponent depended only on the smaller of the two parameters of the complete bipartite graph. Erdős' conjecture states that the point there was that t is the *degeneracy* of $K_{t,s}$.

Conjecture 6 (Erdős) *For every r-degenerate bipartite graph H, there is a constant c_H such that*

$$ex(n, H) \leq c_H n^{2-\frac{1}{r}}.$$

On the intuitive level one might expect this conjecture to be true because of the following. Suppose the average degree of a graph G on n vertices is $Cn^{1-1/r}$ where C is some large constant. We plan to *embed* our r-degenerate graph H into G vertex by vertex following the special ordering provided by the r-degeneracy. By embedding we mean a mapping $f : V(H) \to V(G)$ such that $uv \in E(H)$ implies $f(u)f(v) \in E(G)$. Let us say $j-1$ vertices of H are already embedded in G. The j^{th} vertex has at most r neighbors among those vertices of H which are embedded already. We will be successful in finding an image for v_{i_j} if and only if the common neighborhood of the (at most r)

images of the neighbors of v_{i_j} among the first $j-1$ vertices of H is nonempty (in G) and, even more, has an element which is not an image already. This last requirement can be overcome if the common neighborhood is not only non-empty but has at least $|V(H)|$ vertices. Anyway, the size of the common neighborhood of the possibly r images should not be empty (or should be at least a large constant size). If G were random with expected average degree d, then this would easily hold since the expected cardinality of the common neighborhood of two vertices is $d \cdot \frac{d}{n}$ and, in general, of r vertices it is $d \cdot \left(\frac{d}{n}\right)^{r-1} \approx C \geq |V(H)|$ if $C = C_H$ is large enough. In particular one easily embeds any r-degenerate graph of fixed order into $G(n, n^{-1/r})$.

Of course in the conjecture G is not random, but arbitrary. On the other hand we do not need that *all* subsets of size r of G have a nontrivial common neighborhood (which is true for the random graph), just that *some* well-chosen r-subsets have a common neighborhood.

Anyway, enough of the talk, let's walk the walk.

4.1.2 Towards the conjecture

In this section we prove two different weakenings of the conjecture. First we prove the full strength of the conjecture for bipartite graphs where every degree on *one side* is at most r. This is clearly a generalization of the Kővári-Sós-Turán upper bound, which discusses when the graph is $K_{r,r}$. Second we prove a weakening of the conjecture in a different direction. Namely, for *every* r-degenerate bipartite graph H, the conjecture is true with slightly larger average degree.

Bounded maximum degree on one side

Theorem 4.1 (Alon, Krivelevich, Sudakov; Füredi) *If $H = (A \cup B, E)$ is a bipartite graph such that in B every vertex has degree at most r, then there exists a constant $c(H)$ such that*

$$ex(n, H) < c_H n^{2-1/r}.$$

Proof. Let G be a graph on n vertices with $c_H n^{2-1/r}$ edges where c_H is to be determined later. To prove the theorem we need to find a copy of H in G. In other words, we have to define an embedding of H into G, that is an injective function $f : V(H) \to V(G)$, which maps an edge of H into an edge of G.

To this end we aim to find a subset $A_0 \subseteq V(G)$ of cardinality $a = |A|$ such that

$$\text{any } r\text{-subset of } A_0 \text{ has at least } a + b \text{ common neighbors in } G,$$

where $b = |B|$. Our plan is to embed A arbitrarily onto A_0 and then, one by one, the vertices of B into $V(G) \setminus A_0$.

More precisely, suppose we fixed an arbitrary bijection $f : A \to A_0$ and we are in the process of embedding the vertices of $B = \{v_1, \ldots, v_b\}$ into G. Assume we defined

properly the images $f(v_j)$ for $j < i$. Let $N_i = N_H(v_i) \subseteq A$ be the neighborhood of the next vertex v_i. By our assumption, $|N_i| \leq r$. A proper image $f(v_i)$ of v_i should satisfy two properties:

- $f(v_i)$ should be adjacent to every vertex in $f(N_i) = \{f(v_j) : v_j \in N_i\}$, and
- $f(v_i)$ should be different from every vertex in $A_0 \cup \{f(v_j) : j < i\}$.

Both properties can easily be satisfied if indeed A_0 has the magic property that any r-subset of it has at least $a + b$ common neighbors in G.

Hence to finish the proof it is enough to find such an A_0. We will find A_0 with random methods. Let d be the average degree of G. How are we going to find such a magic A_0? Well, we make sure that the *whole* A_0 has a lot (i.e. roughly r) of common neighbors, so the r subsets of A_0 have somewhere to start collecting their common neighbors.

Let T be a random subset of $V(G)$ obtained by picking r vertices one by one independently and uniformly at random, with repetition. Let $A' = A'(T) = \{v : T \subseteq N(v)\}$ be the random subset consisting of those vertices whose neighborhood completely contains T. In other words, A' is the largest subset whose common neighborhood contains the random set T. Note that for any subset $R \subseteq V(G)$ the probability that the whole R is part of A' is the same as the probability that each of the r random vertices of T is a vertex in the common neighborhood $N(R)$ of R. Since these r choices are independent, $\mathbb{P}[R \subseteq A'] = \left(\frac{|N(R)|}{n}\right)^r$. In particular for any vertex v, $\mathbb{P}[v \in A'] = \left(\frac{d(v)}{n}\right)^r$.

We estimate the cardinality of A'. We find that by selecting c_H a large enough constant, the expected size of A' can be made an arbitrary large constant. By the linearity of expectation,

$$\mathbf{E}[|A'|] = \sum_{v \in V} \left(\frac{d(v)}{n}\right)^r \geq \frac{1}{n^r} n \left(\frac{\sum_{v \in V} d(v)}{n}\right)^r = \frac{1}{n^{r-1}} \left(\frac{2|E(G)|}{n}\right)^r = (2c_H)^r,$$

where the inequality follows from the convexity of the function $g(x) = x^r$.

We would like to keep only such a portion of A' for which all r-subsets have a lot of neighbors. For this reason we will delete one vertex from each "bad" r-subset, namely from those which have at most $a+b-1$ common neighbors. Let $Y = Y(A')$ be the random variable counting these bad r-subsets of A'. Then by the linearity of expectation,

$$\begin{aligned} \mathbf{E}[Y] &= \sum_{R \text{ bad}} \mathbb{P}[R \subseteq A'] = \sum_{R \text{ bad}} \left(\frac{|N(R)|}{n}\right)^r \leq \sum_{R \text{ bad}} \left(\frac{a+b-1}{n}\right)^r \\ &\leq \binom{n}{r} \left(\frac{a+b-1}{n}\right)^r \leq \frac{(a+b-1)^r}{r!}. \end{aligned}$$

Finally, if c_H is chosen large enough,

$$\mathbf{E}[|A'| - Y] \geq (2c(H))^r - \frac{(a+b-1)^r}{r!} > a - 1.$$

Hence there exist a set T and a corresponding $A' = A'(T)$ such that $|A'| - Y(A') \geq a$. Let A' be such a set. For every bad r-tuple of A' delete one vertex of it from A'. The remaining set A_0 has at least a elements and every r-subset of it has at least $a + b$ common neighbors, exactly what we wanted. □

Arbitrary r-degenerate bipartite graphs

The trick featured in the previous proof (which by the way earned the epithet "simple yet surprisingly powerful" by the authors) is also applicable in proving a weaker version of the conjecture for *arbitrary* r-degenerate graphs.

Theorem 4.2 *Let H be an r-degenerate bipartite graph. Then*

$$ex(n, H) = O(n^{2-\frac{1}{10r}}).$$

Remark: The constant 10 in the exponent can be (and was) improved by the same method to 4, but to reduce it to the conjectured 1 new ideas seem to be needed.

Proof. Let h denote the number of vertices of H and let A and B be its two partite sets. Assume that we are given a graph G on n vertices with at least $n^{2-1/10r}$ edges, where $n > h^{10}$. Once we prove that such a G contains a copy of H, we proved our theorem.

So how to find our, possibly very delicate, H in the sparse wastelands of G? The core of our embedding procedure is the following lemma.

Lemma 4.2.1 *The graph G contains disjoint sets $A_1, B_1 \subseteq V(G)$ such that*

- *every r-subset of A_1 has at least $n^{0.1}$ common neighbors in B_1, and*
- *every r-subset of B_1 has at least $n^{0.1}$ common neighbors in A_1.*

Once we have the sets A_1, B_1 guaranteed by Lemma 4.2.1, the embedding of H into G is quite straightforward. Let $v_1, \ldots, v_h$ be an r-degenerate ordering of the vertices of H. We will embed H vertex-by-vertex following this ordering such that vertices of A are embedded into A_1 and vertices of B are embedded into B_1. Suppose we already defined an appropriate embedding $f : \{v_1, \ldots, v_{i-1}\} \to V(G)$ and now we are looking for an image for v_i. Assume w.l.o.g. that $v_i \in A$ and let us identify the set of its neighbors $N_i := N(v_i) \cap \{v_1, \ldots, v_{i-1}\}$ which have already been embedded. In order to extend the embedding f to v_i we must find a vertex in B_1 which is a common neighbor of the set $f(N_i)$ of images of all the embedded neighbors of v_i *and* is not an image already itself. By our assumption $|f(N_i)| \leq r$, so by Lemma 4.2.1 there exist at least $n^{0.1}$ common neighbors of $f(N_i)$ in B_1. We must avoid at most $h - 1 < n^{0.1}$ already existing images, hence we can surely select a proper image for v_i from $N_i \setminus f(\{v_1, \ldots, v_{i-1}\})$.

This process at termination created an embedding of H into G and the proof of Theorem 4.2 is complete. □

To complete the argument we need to prove Lemma 4.2.1.

Proof. (of Lemma 4.2.1) First we define a bipartite subgraph G_1 of G with parts $\tilde{A}$ and $\tilde{B}$, such that $|\tilde{A}| = \lceil \frac{n}{2} \rceil$, $|\tilde{B}| = \lfloor \frac{n}{2} \rfloor$, and $|E(G_1)| \geq \frac{1}{2}|E(G)| \geq \frac{1}{2}n^{2-1/10r}$.

Exercise 4.1 *Prove that* any *graph G has such a bipartite subgraph G_1.*

To find the appropriate A_1 and B_1 we use twice the simple yet amazingly powerful sampling trick which was helpful already in the proof of Theorem 4.1.

First we create a subset $A_1 \subseteq \tilde{A}$ of cardinality $|A_1| > n^{0.6}$ such that

$$\text{every } 3r\text{-subset of } A_1 \text{ has at least } n^{0.1} \text{ neighbors in } \tilde{B}.$$

Note that in order to have further leverage to ensure the required property for the subsets of B_1 as well, we require a somewhat stronger property than it is needed at the end: all $3r$-subsets of A_1 have a lot of common neighbors (instead of just r-subsets).

Let $T_1 \subseteq \tilde{B}$ be a set of $4r$ vertices, selected independently and uniformly at random from $\tilde{B}$ with replacement. Let

$$A_0 = A_0(T_1) = \{v \in \tilde{A} : T_1 \subseteq N_{G_1}(v)\}.$$

We can bound the expectation of the cardinality of A_0 similarly to the previous subsection.

$$\begin{aligned} \mathbf{E}[|A_0|] &= \sum_{v\in\tilde{A}} \mathbb{P}[v \in A_0] = \sum_{v\in\tilde{A}} \left(\frac{|N_{G_1}(v)|}{|\tilde{B}|}\right)^{4r} \geq |\tilde{A}| \sum_{v\in\tilde{A}} \left(\frac{|E(G_1)|}{|\tilde{A}||\tilde{B}|}\right)^{4r} \geq 2^{4r-1} n^{1-\frac{4r}{10r}} \\ &\geq 2n^{0.6}. \end{aligned}$$

Similarly to the previous subsection, we can also ensure that there are no "bad" $3r$-subsets, where a $3r$-subset R of $\tilde{A}$ is *bad* if it has less than $n^{0.1}$ neighbors in $\tilde{B}$. Let $Y = Y(A_0)$ be the random variable counting the bad $3r$-subsets in A_0. Then we have

$$\begin{aligned} \mathbf{E}[Y] &= \sum_{R \text{ bad}} \mathbb{P}[R \subseteq A_0] = \sum_{R \text{ bad}} \mathbb{P}[N_{G_1}(R) \supseteq T_1] = \sum_{R \text{ bad}} \left(\frac{|N_{G_1}(R)|}{|\tilde{B}|}\right)^{4r} \\ &\leq \sum_{R \text{ bad}} \left(\frac{n^{0.1}}{|\tilde{B}|}\right)^{4r} \leq \binom{|\tilde{A}|}{3r} \left(\frac{n^{0.1}}{|\tilde{B}|}\right)^{4r} = \left(\frac{e|\tilde{A}|}{3r|\tilde{B}|}\right)^{3r} \left(\frac{n^{0.4}}{|\tilde{B}|}\right)^{r} < 1. \end{aligned}$$

Hence $\mathbf{E}[|A_0| - Y] > n^{0.6}$, and in particular there is a choice of T_1 such that that for the corresponding $A_0 = A_0(T_1)$ we have that the value $|A_0| - Y(A_0) > n^{0.6}$. Let A_1 be the set we obtain from A_0 after deleting one vertex from every bad $3r$-subset of A_0. Then $A_1 \subseteq \tilde{A}$ is of cardinality $|A_1| > n^{0.6}$ and every $3r$-subset of A_1 has at least $n^{0.1}$ neighbors in $\tilde{B}$.

Now we turn to constructing B_1. Let $T_2 \subseteq A_1$ be a set of $2r$ vertices, selected independently and uniformly at random from A_1 with replacement. Let

$$B_1 = B_1(T_2) = \{v \in \tilde{B} : T_2 \subseteq N_{G_1}(v)\}.$$

This time we do not need to bound the expectation of the cardinality of this random set, we are only interested in bounding the number of "bad" r-subsets and making sure there

is none. Here an r-subset of $\tilde{B}$ is called *bad* if it has less than $n^{0.1}$ common neighbors in A_1. Let $Z = Z(B_1)$ be the random variable counting the bad r-subsets contained in B_1. Then we have

$$\begin{aligned}\mathbf{E}[Z] &= \sum_{R \text{ bad}} \mathbb{P}[R \subseteq B_1] = \sum_{R \text{ bad}} \mathbb{P}[N_{G_1}(R) \supseteq T_2] = \sum_{R \text{ bad}} \left(\frac{|N_{G_1}(R) \cap A_1|}{|A_1|}\right)^{2r} \\ &\leq \sum_{R \text{ bad}} \left(\frac{n^{0.1}}{|A_1|}\right)^{2r} \leq \binom{|\tilde{B}|}{r} \left(\frac{n^{0.1}}{|A_1|}\right)^{2r} \leq \frac{|\tilde{B}|^r}{r!} \left(n^{-0.5}\right)^{2r} \leq \frac{n^r}{2^r r!} n^{-r} < 1\end{aligned}$$

Hence there is a choice of T_2 such that that in the corresponding B_1 contains no bad r-subset, i.e.,

every r-subset of B_1 has at least $n^{0.1}$ neighbors in A_1.

Great, this fulfills half of what is required from the pair (A_1, B_1). But why is the other half true? At this point it is not even clear that B_1 is not empty, not to mention that every r-subset R of A_1 should have at least $n^{0.1}$ neighbors in B_1! To see this we need to utilize what we know about the $3r$-subsets of A_1.

For an r-subset R of A_1 the set $S = R \cup T_2$ is a subset of A_1 containing at most $3r$ elements. Hence it has at least $n^{0.1}$ common neighbors in $\tilde{B}$. That's good: we have then a lot of common neighbors of R as well, but are they in B_1? Yes!!! By definition, all common neighbors of T_2 *are* in B_1. So $N_{G_1}(S) \subseteq N_{G_1}(T_2) \subseteq B_1$. In particular the set $N_{G_1}(S)$ represents at least $n^{0.1}$ common neighbors of the set R in B_1! This concludes the proof of Lemma 4.2.1. □

4.2 The Turán number of the cube

Now we return to the original question Turán posed in the paper which started the field of Extremal Graph Theory. In today's terminology Turán asked about the Turá number of the graphs of the Platonic solids. As it was mentioned in the first lecture, this question was resolved for the tetrahedron by Turán in the same paper and asymptotically by Erdős, Stone and Simonovits for the octahedron, the dodecahedron and the icosahedron. The lone case still unresolved is the one of the cube Q_3. Our best lower bound is of order $n^{3/2}$, the same as for C_4. Since the cube is a bipartite graph which has maximum degree 3 on one of its sides (in fact on both), Theorem ??? implies an upper bound $O(n^{5/3})$. We take now a closer look at an improvement, which is the best bound known today.

We start the proof with a technical concept and lemma. When dealing with an upper bound for some Turán number and caring only about the order of magnitude we could easily assume that the minimum degree is at least the half of the average degree (Claim ??). We simply deleted small degree vertices and when the process has stopped we still have a nonempty graph. One moment thought convinces us that a similar assumption is not at all so trivial to see for the maximum degree. I.e., can we assume

that the maximum degree is within a constant factor of the average degree? The next technical lemma says "yes".

A graph H is called *γ-relaxed regular* if $\Delta(H) \leq \gamma \cdot \delta(H)$.

Lemma 4.2.2 *Let $\alpha > 0$ be a constant. If we have a graph G with $e(G) \geq cn^{1+\alpha}$, then there exists $H \subseteq G$ such that H is γ-relaxed regular, where $\gamma = 10 \cdot 2^{1+1/\alpha^2}$,*

$$n(H) \geq n(G)^{\frac{1-\alpha}{1+\alpha}} \text{ and } e(H) \geq \frac{2}{5}cn(H)^{1+\alpha}.$$

Let $ex_\gamma(n, G)$ denote the maximum number of edges number a G-free γ-relaxed regular graph on n vertices can have. Obviously, $ex(n, G) \geq ex_\gamma(n, G)$.

Corollary 4.3 *Let $0 < \alpha < 1$ and $\gamma = 10 \cdot 2^{1+1/\alpha^2}$. If $ex_\gamma(n, G) = O(n^{1+\alpha})$, then $ex(n, G) = O(n^{1+\alpha})$.*

Theorem 4.4 $ex(n, Q_3) \leq cn^{8/5}$.

Proof. Let G be a graph, which does not contain a cube Q_3. We will show that the number of its edges is $O(n^{8/5})$. W.l.o.g. G is bipartite. Let $\gamma = 10 \cdot 2^{34/9}$. By Corollary 4.3 we can also assume that G is γ-relaxed regular.

We are going to double-count copies of $K_{2,2}$ in G. First, we count them by their edges. For a fixed edge $f = uv$, the number of $K_{2,2}$s containing f is exactly the number of edges connecting a vertex in $N(u)$ and a vertex in $N(v)$. Since G contains no Q_3, we know that the graph induced on $N(f) := N(u) \cup N(v)$ contains no C_6 (note that here it is important that G is bipartite). Thus we can apply our upper bound on the number of edges in C_6-free graphs for the subgraph induced by $N(f)$. To apply this upper bound in a meaningful way, it is crucial that we do control the size of $N(e)$, i.e., it is very important that G is γ-relaxed regular. Hence for the number N of $K_{2,2}$'s we have

$$\begin{aligned} 4N &= \sum_{f \in E(G)} e(G[N(f)]) \leq \sum_{f \in E(G)} c_1 \cdot |N(f)|^{4/3} \\ &\leq \sum_{f \in E(G)} c_1 (2\gamma \bar{d})^{4/3} = c_1 e(G) \left(\frac{4\gamma e(G)}{n}\right)^{4/3} \\ &= K \frac{e(G)^{7/3}}{n^{4/3}}, \end{aligned} \tag{4.1}$$

where $\bar{d} = 2e(G)/n$ is the average degree of G.
On the other hand, we can count $K_{2,2}$s by their non-adjacent vertex pairs. Let $codeg(u, v)$ denote the number of common neighbors of vertices u and v. Then we have

$$\begin{aligned} 2N &= \sum_{\{u,v\} \subseteq V} \binom{codeg(u,v)}{2} \geq \binom{n}{2} \binom{\sum_{u,v} \frac{codeg(u,v)}{\binom{n}{2}}}{2} \geq Kn^2 \left(\frac{\sum_{u,v} codeg(u,v)}{n^2}\right)^2 \\ &= K \frac{\left(\sum_{u,v} codeg(u,v)\right)^2}{n^2}, \end{aligned}$$

where we used the convexity of the function $\binom{x}{2}$. We can count the sum of the co-degrees another way:

$$\begin{aligned}\sum_{\{u,v\}\subseteq V} codeg(u,v) &= \#\text{ of } K_{1,2}\text{'s in } G = \sum_{x\in V}\binom{d(x)}{2} \geq n\binom{\overline{d}(G)}{2} \\ &\geq K'\frac{e^2}{n},\end{aligned}$$

and hence

$$2N \geq K''\frac{e^4}{n^4}. \tag{4.2}$$

Combining (4.1) and (4.2) we obtain

$$\begin{aligned}K\frac{e(G)^{7/3}}{n^{4/3}} &\geq 2K''\frac{e(G)^4}{n^4}, \\ Kn^{8/3} &\geq 2K''e(G)^{5/3}, \\ K'''n^{8/5} &\geq e(G).\end{aligned}$$

□

Part II

Ramsey-type problems

Chapter 5

The symmetric Ramsey-problem

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

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

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

Recall the exponential upper and lower bounds

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

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

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

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

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

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

5.1 Initial Constructions

5.1.1 Paley graphs

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

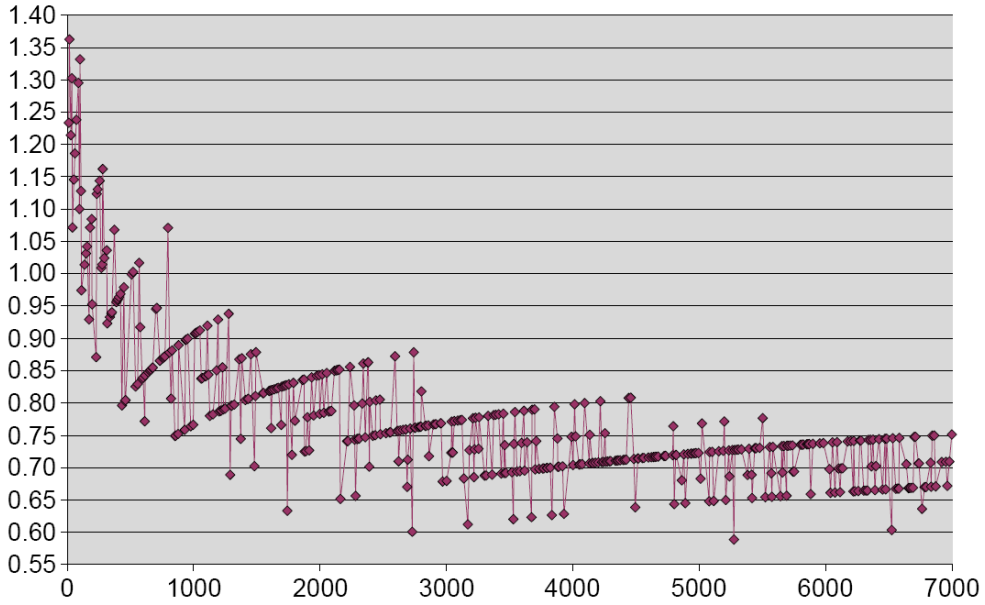

Figure 5.1: The quotient $\frac{\log n(P_p)}{\omega(P_p)}$ in the Payley-graph P_p for the primes $p \leq 7000$

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

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

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

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

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

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

5.1.2 Beating the Turán Construction

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

The Abbott product

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

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

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

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

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

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

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

Set families with intersection restrictions

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

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

5.2 What sort of explicit?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

As expected, all our constructions are strongly explicit.

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

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

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

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

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

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

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

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

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

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

A morally explicit superpolynomial Ramsey graph.

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

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

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

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

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

Then by the union bound

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.3 Limiting the randomness

5.3.1 d-wise independent sample spaces

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.4 Approximating randomness

5.4.1 Almost independent sample spaces

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

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

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

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

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

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

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

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

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

which is ϵ-close to d-wise independent.

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

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

Linear tests

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and the lemma is proved. □

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

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

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

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

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

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

Almost independent sample spaces via the quadratic character

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

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

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

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

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

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

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

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

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

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

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

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

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

Better Ramsey-graphs

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

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

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

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

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

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

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

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

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

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

which is polynomial in n.

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

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

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

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

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

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

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

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

5.5 Ramsey graphs via intersection theorems

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

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

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

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

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

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

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

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

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

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

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

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

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

a contradiction. □

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

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

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

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

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

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

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

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

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

□

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 6

The asymmetric Ramsey-problem

6.1 Constructive $R(K_3, K_k)$

We continue our study of explicit constructions for the Ramsey problem with the asymmetric case. We want lower bound $R(3,k)$ constructively. The general bound of Erdős and Szekeres gives

$$R(3,k) < \binom{k+1}{2} = O(k^2).$$

This was improved by Ajtai, Komlós and Szemerédi (1981) to $k^2/\log k$. Their argument is considered the first application of the "nibble-method", which then went on to become one of the more succesful technique of probabilistic combinatorics.

From the other side, an ugly, but routine application of the Local Lemma gives a lower bound of $\left(\frac{k}{\log k}\right)^2$. In 1995 Kim, in remarkable tour de force of probabilistic combinatorics, managed to remove one log-factor and established the correct order of magnitude $\frac{k^2}{\log k}$. He received the Fulkerson prize for his paper. Curiously, Kim is using the very same nibble method that was invented while proving the matching upper bound.

The first explicit construction of triangle-free graph where the vertex set is of super-linear size in terms of the independence number is due to Erdős who showed

$$R(3,k) = \Omega\left(k^{2\log 2/3(\log 3 - \log 2)}\right) = \Omega\left(k^{1.13}\right)$$

constructively. Later this has been improved by Chung, Cleve and Dagum to

$$\Omega\left(k^{\log 6/\log 4}\right) = \Omega\left(k^{1.29}\right),$$

and even further improved by Alon to $\Omega\left(k^{4/3}\right)$. Later Alon polished his approach, giving a construction with $\Omega(k^{3/2})$ vertices. His proof bounds the independence number through the second eigenvalue of the adjacency matrix of the graph. The second part of this section is devoted to this construction, but first we look at a more recent, simpler one, found by Codenotti, Pudlak and Resta [**?**].

6.1.1 A simple one

We start by describing a weaker construction that demonstrates the basic trick of [**?**]; this will later be generalized in Section 6.2. The final twist to this idea will be added in

a subsequent exercise. Let B be either the Benson-graph or the Wenger-graph of girth 8 described in Section 3.4. From B we construct a new graph G defined on the edges of B. Let $V(G) := E(B)$ and two vertices x_1y_1, x_2y_2 of G with $x_1 \neq x_2$, $y_1 \neq y_2$ are adjacent if x_1y_2 or x_2y_1 is an edge of B.

First, we check that there is no triangle in G. Suppose to the contrary that there are vertices $x_1y_1, x_2y_2, x_3y_3 \in V(G)$ which form a K_3. Then the six vertices $x_1, x_2, x_3, y_1, y_2, y_3$ of B span three other edges, six altogether. Consider the subgraph of B induced by these six vertices. Were two of the neighborhoods of x_1, x_2 and x_3 intersecting $Y = \{y_1y_2, y_3\}$ in more than one vertex, we would immediately arrive at a contradiction to the C_4-freeness of B. Hence the neighborhood of any x_i in Y must contain exactly two vertices, and any pair of them should intersect in exactly one vertex. That is, the six edges must form a six-cycle in B, yet another contradiction. So G is indeed triangle-free.

Let $k = |V(B)|$. Then $|V(G)| = |E(B)| = \Theta\left(k^{4/3}\right)$. We set to prove that every independent set of G is of size $O(k)$ giving an explicit lower bound of order $k^{4/3}$ to $R(3,k)$. Let $I \subseteq V(G)$ be an independent set of G. The members of I are edges in B and we claim they form a star-forest in B. Indeed, should there be a path $e_1e_2e_3$ of three edges in I, the two terminal edges e_1 and e_3 would be adjacent in G, contradicting the independence of I. Hence $|I| \leq k-1$ and we are done.

Exercise 6.1 *Improve further the above construction by using as B, instead of the Benson graph, the point/line incidence graph of the projective plane (introduced in Subsection 2.1.3 as Construction 1.) and modifying slightly the definition of an edge in G. Let $V(G) := E(B)$ and let P be one of the partite sets of B, say the points of the projective plane. Fix an arbitrary ordering $\prec$ on P. Two vertices x_1y_1, x_2y_2 of G are adjacent if $x_1 \prec x_2$, $y_1 \neq y_2$ and $x_1y_2 \in E(B)$. (So roughly "half" of the edges are kept compared to the previous construction.) Show that this graph is an explicit lower bound of order $k^{3/2}$ for $R(3,k)$.*

6.1.2 Alon's Construction

As in all our Ramsey-type constructions, we try to imitate randomness. We aim to bound the independence number of the constructed quasi-random graph via its second eigenvalue. Cayley graphs are very symmetric and their eigenvalues are strongly connected to the charaters of the underlying group. Let us be given a group H where the operation $+$ is written additively. We define a graph G on the vertex set $H = V(G)$. The adjacency relation will be given by a subset $S \subseteq H$ the in the following fashion: $g, h \in H$ are adjacent if $g-h \in S$. To make sure that the adjacency relation is symmetric we also require that $S = -S$. In order to avoid loops we assume $0 \notin S$. Note that the neighborhood of each vertex $g \in V(G)$ is the set $g + S$.

Cayley graphs are often good quasi-random graph provided the subset S is chosen "random enough" inside the underlying group. For example, if one chooses S to be a proper subgroup of H, then $G(H,S)$ is very not quasi-random, it is not even connected:

it contains $|H|/|S|$ disjoint cliques, each being a coset of the subgroup $S < H$. In other words, a subgroup of H is not a random-like subset of H.

Passing to Cayley graphs from general graphs represents a simplification of our task, in some sense we reduce the "dimension of our problem": instead of trying to construct all the edges of a quasi-random graph, we are content with describing a quasi-random neighborhood of one vertex and then make sure that all vertices look the same locally. Definition and analysis gets simpler, but of course, we reduce the playing field significantly.

Let us start by looking at what makes a Cayley graph triangle-free.

Claim 5 *A Cayley graph $G(S, H)$ is triangle-free if and only if the equation $s_1 + s_2 + s_3 = 0$ has no solution s_1, s_2, s_3 in S.*

Proof. Suppose that $u, v, w \in V(G)$ form a triangle. Then $v = u + s_1$ for some $s_1 \in S$, $w = v + s_2$ for some $s_2 \in S$, $u = w + s_3$ for some $s_3 \in S$. Hence $u = w + s_3 = v + s_2 + s_3 = u + s_1 + s_2 + s_3$ implying $s_1 + s_2 + s_3 = 0$. For the other direction, let us assume that $s_1 + s_2 + s_3 = 0$ for some $s_1, s_2, s_3 \in S$. Then for any $u \in V(G)$, the vertices $u, u + s_1$ and $u + s_1 + s_2$ form a triangle. It is clear that there is an edge between any pair of these three vertices, and then this also implies that they do represent three different vertices, since G is loop-free. □

In a vector space over the two-element field $\mathbb{F}_2$ the linear equation $s_1 + s_2 + s_3 = 0$ is equivalent to the vectors s_1, s_2 and s_3 being linearly dependent. Hence triangle-freeness of a Cayley graph $G(\mathbb{F}_2^n, S)$ is equivalent to the set S being three-wise linearly independent (cf. the three-wise independence of sample spaces as defined in Subsection 5.3.1). Consequently, the idea of using the parity check matrices of BCH-codes arises quite naturally.

The first try

First we obtain a weaker bound on $R(3, k)$ by introducing one of the basic ideas of Alon's construction. Let G be the Cayley graph on the group $\mathbb{Z}_2^{2k}$ with a "neighborhood set" $S \subseteq \mathbb{Z}_2^{2k}$, that consists of the column vectors of the parity check matrix of the binary BCH-code of designed distance 5. To recall, this construction uses the 2^k-element field $\mathbb{F}_{2^k}$. On the one hand $\mathbb{F}_{2^k}$ is a k-dimensional vector space over $\mathbb{F}_2$, that is, its elements can be written as $0-1$-vectors of length k and the addition in $\mathbb{F}_{2^k}$ is just the usual vector addition in $\mathbb{Z}_2^k$. On the other hand, the multiplication of $\mathbb{F}_{2^k}$ completely "messes up" the additive structure. We will in fact use the vector notation to denote the elements of the field but will switch between interpreting the 0-1-vector as an element of $\mathbb{F}_{2^k}$ or $\mathbb{Z}_2^k$. For two vectors $u, v \in \mathbb{Z}_2^k$ let $[u, v]$ denote their concatenation in $\mathbb{Z}_2^{2k}$.

Let $S = S^{(2)}_{simple} = \{[z, z^3] \in Z_2^{2k} : z \in \mathbb{F}_{2^k} \setminus \{0\}\}$. Here, of course, z^3 denotes the binary vector of length k corresponding to the field element z^3.

The number of vertices in this Cayley graph is $n = 2^{2k}$, while the degree d of regularity is equal to $|S| = 2^k - 1 \approx \sqrt{n}$.

This graph certainly does not contain any triangle, as that would mean that there are three different $z_1, z_2, z_3 \in \mathbb{F}_{2^k} \setminus \{0\}$ such that

$$\begin{aligned} z_1 + z_2 + z_3 &= 0 \text{ and} \\ z_1^3 + z_2^3 + z_3^3 &= 0. \end{aligned}$$

This is of course impossible since substituting $z_1 + z_2$ for z_3 in the second equation (we used that $x = -x$ over characteristic 2) we get $0 = z_1^3 + z_2^3 + (z_1 + z_2)^3 = z_1^3 + z_2^3 + z_1^3 + 3z_1^2 z_2 + 3z_1 z_2^2 + z_2^3 = z_1 z_2 (z_1 + z_2) = z_1 z_2 z_3$. Hence one of the z_i must be 0 which is not allowed. Thus there is no triangle in G.

Exercise 6.2 *Argue why using squares instead of cubes in the definition of S would not be a good idea. That is, what is the problem with the neighborhood set $\{[z, z^2] : z \in \mathbb{F}_{2^k}\}$?*

How about the largest independent set? We intend to use Corollary A.37 and bound the largest non-principal Fourier coefficients of the characteristic vector of the neighborhood set S.

By Corollary A.37 we know that $\alpha(G(S)) \leq n\frac{\Phi(S)}{|S|}$. What is now $\Phi(S)$? For that, what is $\widehat{\mathbb{1}}_S = \sum_{s \in S} \chi(s)$? Recall that every character of Z_2^{2k} is determined by an element of Z_2^{2k} the following way: $\chi_\beta(\alpha) = (-1)^{\langle \alpha, \beta \rangle}$. Then

$$n\widehat{\mathbb{1}}_S(\chi_\beta) = \sum_{s \in S} \chi_\beta(s) = \sum_{s \in S} (-1)^{\langle \beta, s \rangle}.$$

Imagine that the elements of S written as colums of the $2k \times (2^k - 1)$ -dimensional matrix M. We are interested in the vector $\beta^T M$, whose components are the bits $\langle \beta, s \rangle$. In fact we are rather interested in the Hamming weight $w = w(\beta)$ of $\beta^T M$: for each 1-coordinate of $\beta^T M$ we have a -1 term in the corresponding Fourier coefficient, while for each 0-coordinate we have a 1. That is the corresponding Fourier coefficient is $d - 2w$. Obviously, we want this to be as small as possible, that is, we want roughly the same number of ones and zeros in the vector $\beta^T M$, for every $\beta \in \mathbb{Z}_2^{2k}, \beta \neq 0$. This is certainly easy to check for a β of the form $[\beta_1, 0]$ where $\beta_1 \in \mathbb{Z}_2^k, \beta_1 \neq$: there are exactly one more ones than zeros in $\beta^T M$. Similar is true for β of the form $[0, \beta_2]$ where $\beta_2 \in \mathbb{Z}_2^k, \beta_1 \neq$ provided $z \to z^3$ is a bijection, that is, if $3 \not| 2^k - 1$ or k is odd.

Now comes the big cannon, a theorem of Carlitz and Uchiyama (which is itself a nontrivial consequence of the theorem of Weil about the Rieman hyothesis over finite fields) tells us that the Hamming weight of $\beta^T M$ is roughly half of the length for *any* $\beta \in \mathbb{Z}_2^{2k}, \beta \neq 0$. More precisely, $|w - 2^{k-1}| \leq 2^{k/2}$. This means that for every $\beta \neq 0$, $|n\widehat{\mathbb{1}}_S(\chi_\beta)| \leq 2^{k/2}$ and thus $\Phi(S) \leq 2^{k/2} \approx \sqrt{d}$.

Combining this with our bound on the independence number we have $\alpha(G(S)) \leq n/\sqrt{d} = n^{3/4}$. Hence we have shown $ExpR(3, k) \gtrapprox k^{4/3}$.

Remark. The matrix M is the parity check matrix of a binary BCH code (with designed distance 5), and as such, the linear combinations of its rows form the dual of the code.

That is, the vector $\beta^T A$, which is just linear combination of the rows with coefficients β_i, is an element of this dual code. The theorem of Carlitz and Uchiyama then tells us that the weight of any code word in the dual of a BCH code is roughly half of the length.

Increasing the degree

Are we at the end of the road for the BCH-idea? One certainly senses the presence of some leftover in the triangle-freeness proof: the full power of the BCH-matrix is not exploited. We only use that any *three* columns of M are linearly independent, while it is also true that any *four* of them are.

Observe that our bound on the independence number depends on $\Phi(S)/|S|$. If we assume that S is perfectly quasi-random in the sense that $\Phi(S) \approx \sqrt{|S|}$, then the upper bound on $\alpha(G)$ depends solely on how big the degree $d = |S|$ is. In the previous subsection we had $d = \Theta(n^{1/2})$, so there could be some room for improvement.

How can we increase the degree? Considering larger BCH-matrices and taking not only z and z^3, but z^5, z^7, etc..., will increase the length of our vectors, together with the parameter of their linear independence property, but the degree of the graph goes down (see Exercise 6.3). The Carlitz-Uchiyama bound remains valid, however, thus the Hamming weight of $\beta^T M$ remains roughly half of the length, just the error term worsens by a constant factor.

Theorem 6.1 (Carlitz-Uchiyama) *Let $M = M_h$ be the $hk \times (2^k - 1)$ matrix whose columns are the vectors of the form $[z, z^3, z^5, \ldots, z^{2h-1}]$, with $z \in \mathbb{F}_{2^k}^*$. Then for every $\beta \in \mathbb{Z}_2^k, \beta \neq 0$, the Haming weight $w = w(\beta)$ of $\beta^T M$ satisfies*

$$\left|w - 2^{k-1}\right| \leq (h-1)2^{k/2}.$$

Remark. The matrix M_h is the parity-check matrix of the binary BCH-code of designed distance $2h+1$. As it was proved in Subsection 5.3.1, any $2h$ columns of M_h are linearly independent.

Exercise 6.3 *Describe an explicit construction showing $ExpR(C_5, K_k) = \Omega(k^{6/5})$, using the parity check matrix of the BCH-code of designed distance 7. Explain why the constructed graph is also K_3-free, but not C_4-free or C_6-free.*

In the previous exercise we worked over a larger group $\mathbb{Z}_2^{3k}$ and our neighbor set S_{simple} consisted of all vectors of the form $[z, z^3, z^5]$, $z \in \mathbb{F}_{2^k}^*$. This resulted in degree only of order $n^{1/3}$, so-so for C_5-freeness, but too little for a K_3-free construction. To increase the degree, let us partition S_{simple} into two, for the moment arbitrary, subsets W_0 and W_1 and let us define our new neighbor set as $S = \{w_0 + w_1 : w_0 \in W_0, w_1 \in W_1\}$. Note that the degree d is $|S| = |W_0||W_1|$; if $w_0 + w_1$ was equal to $w_0' + w_1'$, then, since any four elements of S_{simple} are linearly independent, we had $w_0 = w_0'$ and $w_1 = w_1'$. To maximize the degree, we will select W_0 and W_1 with almost equal size and then $d \approx n^{2/3}$.

Alon's graph is the Cayley graph G on $\mathbb{Z}_2^{3k}$ with the neighborhood set S, where the partition $W_0 \cup W_1$ of S_{simple} will be chosen later appropriately.

Why is there no triangle in G? Were there a triangle then by Claim 5 there would be three different elements $w_0 + w_1, w_0' + w_1', w_0'' + w_1''$ of S which sum up to 0. On the other hand we know that any *six* members of S_{simple} are linearly independent so $w_0 + w_0' + w_0'' + w_1 + w_1' + w_1''$ can only be 0 if every vector occurs an even number of times. Since the sets W_0 and W_1 are disjoint and the sum contains an odd number of elements from each, this is clearly impossible.

For estimating the independence number we again use the upper bound involving $\Phi(S)$.

$$\sum_{s\in S}\chi(s) = \sum_{s_0\in W_0}\sum_{s_1\in W_1}\chi(s_0+s_1) = \sum_{s_0\in W_0}\sum_{s_1\in W_1}\chi(s_0)\chi(s_1) = \left(\sum_{s_0\in W_0}\chi(s_0)\right)\left(\sum_{s_1\in W_1}\chi(s_1)\right).$$

For a character $\chi = \chi_\beta$, the sum $\sum_{s_0\in W_0}\chi_\beta(s_0) = \sum_{s_0\in W_0}(-1)^{\langle\beta,s_0\rangle}$ depends on the W_0-entries of the vector $\beta^T M$. By the same argument as in the previous subsection, $\sum_{s_0\in W_0}\chi_\beta(s_0)$ is equal to $|W_0| - 2x$ where x is the Hamming weight of the vector $\beta^T M$ restricted to the coordinates at W_0. Similarly, $\sum_{s_1\in W_1}\chi_\beta(s_1)$ is equal to $|W_1| - 2y$ where y is the Hamming weight of the vector $\beta^T M$ restricted to the coordinates at W_1. In conclusion, our main concern is to minimize the product $(|W_0| - 2x)(|W_1| - 2y)$. Of course the Carlitz-Uchiyama bound does tell us that the sum $x+y$, the Hamming weight of the vector $\beta^T A$, is roughly half of 2^k, but in principle it could happen that x and y are very non-equal resulting in $(|W_0| - 2x)(|W_1| - 2y)$ being very large.

In order to avoid this we must now specify the selection of W_0 and W_1 such that $x \approx y$. A random partition would certainly fit the bill, but considering the business we are in, we need to be, well, more explicit. We need to find a 0-1 vector of length $2^k - 1$, that is more or less independent of the previous row vectors of M. The message of Subsection 5.3.1 was that the linear independence of the columns corresponds to probabilistic independence of the rows and in particular the rows of the matrix of the BCH-code proovide a good approximation of independence. Hence the natural choice for the required quasi-random partition should be given by the first new row of the next BCH-matrix M_4. That involves the function z^7, more precisely its first digit in its expression as a bit vector of length k. For $i = 0$ and 1, let

$$W_i := \{[z, z^3, z^5] \in S_{simple} : \text{first digit of } z^7 \text{ is } i\}.$$

The easiest way to ensure that W_0 and W_1 are roughly half of 2^k is if we assume that $z \to z^7$ is a bijection of $\mathbb{F}_{2^k}$, which certainly happens if $7 \not| (2^k - 1)$ or $3 \not| k$.

After all the heuristic, how does one actually prove that this partition is a good choice? The set W_1 was chosen such that its characteristic vector of W_1 is a row of the one larger BCH-matrix (whose columns are the $4k$-vectors $[z, z^3, z^5, z^7]$). We know the Carlitz-Uchiyama bound holds for this matrix as well, maybe with a worse constant factor in the error term, but who cares: any linear combination of the rows of M_4

have Hamming-weight roughly 2^{k-1}. In particular, $[\beta, (1, 0 \ldots, 0)]M_4 = \mathbb{1}_{W_1} + \beta^T M_3$ is a linear combination of the rows of M_4. Compared to $\beta^T M_3$, the W_0-entries do not change, but all the W_1-entries flip, hence the Hamming-weight of this vector is $x + |W_1| - y = x + 2^{k-1} - 1 - y$. This is roughly 2^{k-1} by the Carlitz-Uchiyama bound, so $x \approx y$. Or more precisely, $|x - y| \leq 3 \cdot 2^{k/2}$.

We already know that $x + y \approx 2^{k-1} \pm 2 \cdot 2^{k/2}$, so we have that $x \approx y \approx 2^{k-2} \pm 4 \cdot 2^{k/2}$. Hence $||W_0| - 2x|$ and similarly $||W_1| - 2y|$ is $O(2^{k/2})$. meaning that $\Phi(S) = O(2^k) = \Theta(\sqrt{d})$, quasi-randomness at its best.

In conclusion we have shown that G is a triangle free graph with $\alpha(G) \leq n\Theta(\sqrt{d})/d = \Theta(n^{2/3})$. This establishes $ExpR(3, k) \geq ck^{3/2}$.

Can one improve further by making the graph denser? Maybe even up to the vicinity of the truth, $\Theta(n^2/\log^2 n)$? The next subsection shows that to be impossible in very strong sense.

Exercise 6.4 *Prove that the number of C_4 in Alon's (n, d, λ)-graph is asymptotically the same as the expected number of C_4 in the random graph $G(n, n^{-1/3})$.*

6.1.3 Turán property of quasi-random graphs

Exercise 6.5 *Prove that the independence of both of the previous constructions do have independence number $\Theta(n^{2/3})$. Even more, any K_3-free d-regular graph with $d = \Theta(n^{2/3})$ has an independent set of order $\Theta(n^{2/3})$.*

In the next section we will see that Alon's construction is best possible in a much stronger sense. There is not only one triangle in an $(n, d, \Theta(\sqrt{d}))$-graph with $d = \Omega(n^{2/3})$ but also there are so many triangles that one needs to delete half of the edges of the graph to kill of them.

The quasi-randomizable proof of Turán's theorem

There are many proofs of Turán's Theorem available (see the paper of Aigner [**?**] for six of them). The main difficulty in generalizing these classical arguments that they are all very much tailored to the complete graph K_n. Here we need an approach that uses only the quasi-random property of K_n, that is, that the edges are "distributed sufficiently evenly", a property shared by all (n, d, λ)-graphs with the appropriate parameters.

Our strategy will be the following.

1. We give a new(?) proof of Turán's theorem and identify the quasi-random properties that make it work.

2. Prove that our quasi-random graphs have these properties.

3. Prove that any graph having these properties are t-Turán.

To proceed with point 1. we prove the following slightly weaker version of Turán's theorem.

Theorem. (Turán, 1941) To make K_n K_t-free one must delete at least $\frac{n^2}{2(t-1)} - \frac{n}{2}$ edges.

Proof. We proceed by induction on t. The base case $t = 2$ is trivial.

Keeping in mind that we would like to generalize this proof to appropriate (n, d, λ)-graphs, we denote the base graph K_n by G and the degree of a vertex by $d = n - 1$.

Let R be a minimum set of edges, such that $G - R$ is K_{t+1}-free. The elements of R called *red* edges, the others are called *blue* edges. We set to prove $|R| \geq \frac{n^2}{2t} - \frac{n}{2}$.

First we introduce some notation. Let R_v be the *red neighborhood* of v, that is, the set of those vertices that are adjacent to v via a red edge. Let B_v the analogous notation for the *blue neighborhood*. The *red degree* of v is denoted by $r_v = |R_v|$ and the *blue degree* is denoted by $b_v = |B_v|$. We define $r = \frac{1}{n}\sum_{v \in V} r_v$ to be the average red degree and $b = \frac{1}{n}\sum_{v \in V} b_v$ to be the average blue degree. Note that for every $v \in V$,

$$b_v + r_v = d = b + r. \tag{6.1}$$

We call triangles with exactly one red edge *interesting*. The core combinatorial ingredient of the proof is a doublecounting argument estimating the number of interesting triangles. Let N_i be the number of triangles with i red edge.

First we count interesting triangles by their "apex": their vertex with two incident blue edges. For each v there are exactly as many interesting triangles with v as their apex, as many red edges there are in the blue neighborhood of v.

$$N_1 \quad = \quad \sum_{v \in V} |R \cap E\left(G[B_v]\right)| \tag{6.2}$$

To derive an upper bound we count the interesting triangles by their vertices with one red and one blue edges. For each vertex $v \in V$ the number $e(B_v, R_v) = |R_v||B_v|$ clearly counts these intersting tringles and some more: those with two red edges will also be counted. By summing up for all vertices we count all such triangles twice.

$$N_1 \quad \leq \quad N_1 + N_2 = \frac{1}{2}\sum_{v \in V} e(R_v, B_v) \tag{6.3}$$

To proceed with a lower bound we conclude that, since $E(G) \setminus R$ is blue K_{t+1}-free, the blue neighborhood of v must be blue K_t-free. Then by induction we have that there are at least $\frac{b_v^2}{2(t-1)} - \frac{b_v}{2}$ red edges inside B_v. The induction can be applied because $G[B_v]$ is itself a complete graph. Summing up for all v and using Jensen's inequality we obtain a lower bound.

$$N_1 \quad \geq \quad \sum_{v \in V} \frac{b_v^2}{2(t-1)} - \frac{b_v}{2} \geq n\left(\frac{b^2}{2(t-1)} - \frac{b}{2}\right) \tag{6.4}$$

To have an upper bound we use that in the complete graph there are $b_v r_v$ edges between B_v and R_v. Then we bound again using Jensen's inequality and (6.1).

$$N_1 \leq \frac{1}{2}\sum_{v\in V} b_v r_v = \frac{1}{2}\sum_{v\in V} b_v d - b_v^2 \leq \frac{1}{2}\left(nbd - nb^2\right) = \frac{n}{2}br \tag{6.5}$$

Then we do what we gotta do: divide, and multiply, add and subtract.

$$\begin{aligned} n\left(\frac{b^2}{2(t-1)} - \frac{b}{2}\right) &\leq \frac{n}{2}br \\ \frac{b}{t-1} - 1 &\leq r \\ b - (t-1) &\leq r(t-1) \\ n - 1 - r - (t-1) &\leq r(t-1) \\ n - t &\leq rt \\ \frac{n}{t} - 1 &\leq r \end{aligned}$$

$$\Longrightarrow |R| = \frac{rn}{2} \geq \frac{n^2}{2t} - \frac{n}{2}$$

□

3-Turán property of quasi-random graphs

Let us see first how easily (6.2) and (6.3) imply our theorem for triangles.

In fact let us state this as an independent combinatorial lemma, valid for an arbitrary red/blue-coloring of an arbitrary graph G.

Lemma 6.1.1 *Let G be an arbitrary graph whose edges are two-colored with red and blue. Let R be the set of red edges, and let B_v and R_v be the blue and red neighborhoods of v, respectively. Then*

$$\sum_{v\in V} |R \cap E(G[B_v])| \leq \frac{1}{2}\sum_{v\in V} e_G(R_v, B_v). \tag{6.6}$$

Theorem 6.2 *Let G be an (n, d, λ)-graph. Then G is 3-Turán provided*

$$\lambda \ll \frac{d^2}{n}.$$

Remark Just a brief remark explaining the bound on λ We will use the above combinatorial lemma and we are interested in the number of edges insides B_v being roughly what it is expected to be. For this, B_v has to be large enough. Recall the bound $\lambda n/d$ on the independence number of (n, d, λ)-graphs. This comes from the edge-distribution inequlity of (n, d, λ)-graphs; when the set under consideration is so small that the error term in the inequality takes over, then the set can be independent, or a clique, or anything else, we cannot control it. But when the set size is larger order than $\lambda n/d$ then we have a pretty good idea about the number of edges induced by it. The neighborhood of v consists of d vertices, so for the blue neighborhood B_v (which is mostly expected to be o the same order) to be large enough, d should be of larger order compared to the threshold $\lambda n/d$.

Proof. We use the notation of the previous subsection Let R be a minimum set of edges, such that $G - R$ is K_3-free. The elements of R called *red* edges, the others are called *blue* edges.

We use the above combinatorial lemma for our situation. To estimate the LHS of (6.6) we note that all edges of G inside B_v must be red. Hence we can use Corollary A.24 to estimate them. Summing up for all v and using Jensen's inequality we obtain a lower bound.

$$\sum_{v\in V} |R \cap E(G[B_v]| \;\geq\; \sum_{v\in V} e_G(B_v) \geq \sum_{v\in V} \left(\frac{b_v^2}{2}\frac{d}{n} - \lambda b_v\right) \geq n\frac{b^2}{2} - \lambda n b \tag{6.7}$$

To have an upper bound of the RHS for the complete graph we used that there are $b_v r_v$ edges between B_v and R_v, and then we bounded using Jensen's inequality and (6.1).

Here to estimate $e(R_v, B_v)$ we use again Corollary A.24.

$$\begin{aligned}
\frac{1}{2}\sum_{v\in V} e(B_v, R_v) \;&\leq\; \frac{1}{2}\sum_{v\in V}\left(\frac{d}{n}b_v r_v + \lambda\sqrt{b_v r_v}\right) && (6.8)\\
&\leq\; \frac{1}{2}\sum_{v\in V}\left(b_v d - b_v^2\right) + \sum_{v\in V}\lambda\sqrt{b_v r_v} && (6.9)\\
&\leq\; \frac{1}{2}\left(nbd - nb^2 + n\sqrt{br}\right)) = \frac{n}{2}br + \lambda\sqrt{br} && (6.10)
\end{aligned}$$

After rearranging we obtain

$$b - \frac{2\lambda n}{d}\left(1 + \sqrt{\frac{r}{b}}\right) \leq r.$$

If $r > b$, then we are done: more than half of the edges of G is red.

Otherwise $r/b \leq 1$ and $b > d/2$, so

$$\frac{d}{2} - \frac{4\lambda n}{d} \leq r.$$

This implies

$$\frac{d}{2} - \frac{4o\left(\frac{d^2}{n}\right)n}{d} = \frac{d}{2} - o(d) \leq r,$$

so $|R| = rn/2 \geq n\left(\frac{d}{2} - o(d)\right) = \left(\frac{1}{2} - o(1)\right)|E(G)|$. □

Exercise 6.6 *Prove that the* 3*-Turán condition for quasirandom graph is tight for any reasonable degree above the critical value* $n^{2/3}$*. More precisely: For any* d*,* $\Omega(n^{2/3}) \leq d \leq n$ *construct a triangle-free* $d_1 = \Theta(d)$*-regular graph* G_1 *with* $n_1 = O(n)$ *vertices and second eigenvalue* $\lambda(G_1) = O(d_1^2/n_1)$*. (Hint: Consider the* r-blow-up *of Alon's construction: replace all vertices by an independent set of* r *vertices.)*

Properties of quazi-random graphs

Proving the t-Turán property

6.2 Constant clique versus large clique

What is known about the asymmetric Ramsey problem if the fixed small graph is a K_4, or more generally, a clique K_l of constnt order? The gap between the known upper and lower bounds for $R(K_l, K_k)$ is growing, the lower bound being roughly the square-root of the upper bound. Both bounds are obtained by random methods. The upper bound of k^{l-1} follows from the Erdős-Szekeres Theorem. A log-power improvement is obtained by the nibble method of Ajtai-Komlós and Szemerédi. The best known lower bound of $(k/\log k)^{(l-1)/2}$ can be obtained by various methods, one of them being the Local Lemma.

For $R(l, k)$, where l is fixed, even the probabilistic bounds are not tight. Namely, we have

$$\left(\frac{k}{\log k}\right)^{(l+1)/2} \leq R(l, k) \leq \frac{k^{l-1}}{\log k}.$$

Let us see what kind of lower bound we can hope for on $ExpR(K_l, K_k)$, when $l > 3$ is a constant. Alon's graph in the previous section was triangle-free and quasi-random with degree $\Theta\left(n^{2/3}\right)$ It is quite possible that there exists a much denser, perfectly quasi-random K_4-free graphs, that give a good explicit lower bound on the asymmetric Ramsey-function $R(K_4, K_k)$ Or there might be perfectly quasi-random K_l-free graphs whose nears 1 when the constant l tends to infinity. The upper bound on the independence number of (n, d, λ)-graphs is $\lambda n/d$ and this bound is often tight. Even if our quasi-random graph is as good as it gets and $\lambda = \Theta(\sqrt{d})$, the upper bound on the independence number transforms into $\Theta(n/\sqrt{d}) \geq \Theta(\sqrt{n})$. In other words, this way no quasi-random Ramsey-graph will provide a lower bound better than $\Theta(k^2)$ on $ExpR(K_l, K_k)$, no matter how large the constant l is! This is of course in striking constrast with the known probabilistic lower bound, whose order grows with a factor $\sqrt{k}$ with each increase of l by one.

Two authors of the two explicit triangle-free Ramsey graphs of the previous section teamed up to overcome this difficulty. Their construction further develops the idea of

the "simple" triangle free graph of Subsection ?? and not-known to be quasi-random. A quasi-random explicit construction plays an essential role and the analysis makes use of several cornerstone results of combinatorics.

Theorem 6.3 (Alon-Pudlak) *There exists a constant $\epsilon > 0$ such that*

$$k^{\epsilon\sqrt{\frac{\log l}{\log\log l}}} \leq R(l,k)$$

constructively.

Proof. For an arbitrary graph G we define its *clique graph* $CQ_s(G)$ Let $V(CQ_s(G))$ to be set of all s-cliques of G. Two s-cliques K and L are adjacent in $CQ_s(G)$ iff there exists an edge in $E(G)$ connecting a vertex from $K \setminus L$ and a vertex from $L \setminus K$.
Remark: For $s = 1$, the clique-graph is equal to its base graph: $G = CQ_1(G)$. When $s = 2$ and G is the C_6-free Benson- or Wenger- graph we get back the construction of the previous section.

Our construction will the clique-graph of an appropriate base graph G with an appropriate clique-order s. The next proposition bounds the independence number of clique-graphs in terms of the independence number of its base graph, under very general circumstances.

Proposition 6.4 *Let G be a graph. Then, for every $s \geq 1$, we have*

$$\alpha(CQ_s(G)) \leq \alpha(G).$$

Proof. Let $L_1, \ldots, L_\alpha$ be a maximum independent set of $CQ_s(G)$. We will find vertices $x_i \in L_i \setminus \cup_{j\neq i} L_j$ for every $i = 1, \ldots, \alpha$. The set $\{x_1, \ldots, x_\alpha\}$ should then be an independent set of G because $L_1, \ldots, L_\alpha$ is an independent set of $CQ_s(G)$, and our claim is proved.
Let us find an appropriate x_1, the others are found similarly. For every $i \neq j$, either $L_i \cap L_1 \subseteq L_j \cap L_1$, or the other way around, otherwise L_i and L_j would be adjacent in $CQ_s(G)$. Therefore, the sets $L_i \cap L_1$, $i = 2, 3, \ldots, \alpha$, are ordered by inclusion, and there is a maximal one, say $L_j \cap L_1$. Since $|L_1| = |L_j|$, there exists $x_1 \in L_1 \setminus (L_j \cap L_1) = L_1 \setminus \cup_{i=2}^{\alpha} L_i$. □

We take the projective norm graph $G = G_{q,t}$ as our base graph, and our construction will be its clique graph $CQ_s(G_{q,t})$ with an appropriate s, in fact $s = \lceil \frac{t}{2} \rceil$.

We will need the following properties of the norm graph.

1. $|V(G)| = n = q^{t-1}(q-1)$.

2. $G_{q,t}$ is $(q^{t-1} - 1)$-regular (with loops counted).

3. $G_{q,t}$ does not contain $K_{t,(t-1)!+1}$.

4. Let the eigenvalues of the adjacency matrix of $G_{q,t}$ be $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$. Then $\lambda_1 = q^{t-1} - 1$, and $\lambda = \max_{i=2}^{n} |\lambda_i| = q^{(t-1)/2} \leq \sqrt{n}$.

5. For $s = \lceil \frac{t}{2} \rceil$, the # of K_s in $G_{q,t}$ is

$$(1 + o(1)) \binom{n}{s} n^{-\binom{s}{2}/t}.$$

First we argue that Properties 1-5. imply that we have the appropriate construction in hand to prove Theorem 6.3. Proposition 6.4 tells us that $\alpha(CQ_k(G)) \leq \alpha(G) \leq n$. (In fact, through Property 4. one can derive a stronger bound of $n^{\frac{1}{2}+\frac{1}{t}}$, but this would only effect our result by a constant factor in the exponent. See the relevant exercise later.) By 5. we have

$$|V(CQ_s(G))| \sim c_s n^{s - \frac{\binom{s}{2}}{t}} \sim c_s n^{\frac{t}{2} - \frac{t^2}{8t}} = c_s n^{3t/8} \geq c_s \alpha^{3t/8}. \tag{6.11}$$

In order to prove the theorem we want to see that the clique number of $CQ_s(G)$ does not depend on n rather depends only on t. That is, we would like to show that $\omega(CQ_s(G)) \leq f(t)$ for an appropriate function $f(t)$.

For this we need the following classic and notorious fact/problem of Extremal Combinatorics

Exercise 6.7 *(Sunflower Lemma of Erdős and Rado)*
(a) Let $\mathcal{F}$ be a family of s-sets such that $|\mathcal{F}| > s!(x-1)^s$. Prove that $\mathcal{F}$ contains a sunflower with x petals, *i.e. sets $A_1, \ldots, A_x$, such that $A_{i_1} \cap A_{i_2} = \cap_{j=1}^{x} A_j$ for every $i_1 \neq i_2$.*
(b) Construct an s-uniform family with $(x-1)^s$ members containing NO sunflower with x petals.
(The open questions arising from this Lemma are among the favorite problems of Erdős. For example, already the first special case is open. What is the largest number of sets in an s-uniform family containing NO sunflower with 3 petals? The above say that the answer is between 2^s and $s!2^s$. There are small improvements to these bounds, but the big question whether the answer is exponential (i.e. C^s with some constant C) or much larger is not known.)

Let x be such an integer that $s!(x-1)^s < \omega = \omega(CQ_s(G)) \leq s!x^s$. Then the Sunflower Lemma says that for a maximum clique $B_1, \ldots, B_\omega$ there exists a sunflower with x petals $B_{i_1}, B_{i_2}, \ldots, B_{i_x}$ among them. Hence the subset $\cup_{j=1}^{x} B_{i_j}$ of $V(G)$ is of order $\leq sx$, and induces at least $\binom{x}{2}$ edges in G (since $B_{i_1}, B_{i_2}, \ldots, B_{i_x}$ is a sunflower and forms a clique in the clique-graph). If x could be arbitrarily big compared to $s = \lceil t/2 \rceil$, then this edge-number would get arbitrarily close to the square of the number of vertices sx. But never forget that by property 3., G does not contain a bipartite graph $K_{t,(t-1)!+1}$, so **no**

part of it can have close to quadratic number of edges. Quantitatively, by the Kővári-T. Sós-Turán upper bound,

$$\begin{aligned}\frac{1}{2}(t-1)!^{1/t}(sx)^{2-1/t}+\frac{1}{2}(t-1)sx &\geq ex(sx,K_{t,t!+1})\geq \binom{x}{2},\\ 2(t-1)!^{1/t}s^{2-1/t} &\geq x^{1/t},\\ c(t-1)!s^{2t-1} &\geq x,\end{aligned}$$

Hence

$$\omega \leq s!x^s \leq cs!(t-1)!^s(s^{2t-1})^s \leq ct^{t/2}t^{t/2}t^{(2t-1)\frac{t}{2}} = ct^{t^2+t/2},$$

and thus

$$t > \Omega\left(\sqrt{\frac{\log\omega}{\log\log\omega}}\right).$$

Combining this with (6.11) we obtain

$$|V(CQ_s(G)| \geq \alpha^{\Omega\left(\sqrt{\frac{\log\omega}{\log\log\omega}}\right)},$$

completing the proof of the Theorem 6.3. □

Let us now turn to the proof of Properties 1-5. of the projective norm graph $G_{q,t}$. The properties 1-3. we already discussed when we introduced the norm-graphs.

Eigenvalues of the projective normgraph

Property 4 is a nice elementary application of character sums. We notice that all eigenvalues are (possibly signed) absolute values of Gaussian sums, hence can be calculated precisely. We find that the second eigenvalue is asymptotically the square root of the degree, thus the projective norm-graphs are as quasi-random as it gets. Estimation of the eigenvalues of algebraically defined graphs is often hard or applies deep theorems of algebra or algebraic geometry. Besides, these techniques rarely give the precise answer. Our treatment is elementary, uses only basic facts about groups and fields while providing the exact values.

Let us recall the definition of the projective normgraphs $G_{q,t}$ Let q be the power of an odd prime, $t > 1$ be an arbitrary integer The projective norm-graph $G = G_{q,t}$ is defined as follows. The vertex set $V(G)$ is the direct product $\mathbb{F}_{q^{t-1}} \times \mathbb{F}_q^*$. A vertex (A,a) is adjacent to (B,b) if and only if $N(A+B) = ab$, where $N : \mathbb{F}_{q^{t-1}} \longrightarrow \mathbb{F}_q$ is the usual norm function, i.e. $N(X) = X^{(q^{t-1}-1)/(q-1)}$. It is easy to see (as we saw in Section 2.3) that $G_{q,t}$ has $n := q^{t-1}(q-1)$ vertices. For any fixed $(A,a) \in V(G_{q,t})$ and any $B \in \mathbb{F}_{q^{t-1}}$, $B \neq -A$, there is a unique $b \in \mathbb{F}_q^*$ such that (A,a) is adjacent to (B,b) in $G_{q,t}$. Hence, counting each loop into account once, $G_{q,t}$ is $(q^{t-1}-1)$-regular.

Theorem 6.5 *The second eigenvalue of $G_{q,t}$ is $\lambda = q^{(t-1)/2}$.*

Proof. As it is often the case the road to the eigenvalues leads through anticipating the eigenvectors.

Let M be the adjacency matrix of $G_{q,t}$. Let χ be an arbitrary additive character of $\mathbb{F}_{q^{t-1}}$ and ϕ be an arbitrary multiplicative character of $\mathbb{F}_q^*$. Let (χ,ϕ) denote the column vector whose coordinates are labeled by the elements of $V(G)$, and whose entry at the coordinate (A,a) is $\chi(A)\phi(a)$.

It turns out that (χ,ϕ) is "almost" an eigenvector of M. Indeed, the entry of the vector $M(\chi,\phi)$ at the coordinate (A,a) is

$$\sum_{\substack{B\in\mathbb{F}_{q^{t-1}}\\ b\in\mathbb{F}_q^*\\ N(A+B)=ab}}\chi(B)\phi(b)=\sum_{\substack{B\in\mathbb{F}_{q^{t-1}}\\ B\neq -A}}\chi(B)\phi\left(\frac{N(A+B)}{a}\right)=\sum_{\substack{C\in\mathbb{F}_{q^{t-1}}\\ C\neq 0}}\chi(C-A)\phi\left(\frac{N(C)}{a}\right)=$$

$$=\sum_{\substack{C\in\mathbb{F}_{q^{t-1}}\\ C\neq 0}}\chi(C)\phi(N(C))\overline{\chi(A)\phi(a)}$$

So M takes the vector (χ,ϕ) into a *constant multiple* of its *conjugate*:

$$M(\chi,\phi)=\left[\sum_{\substack{C\in\mathbb{F}_{q^{t-1}}\\ C\neq 0}}\chi(C)\phi(N(C))\right](\overline{\chi},\overline{\phi}). \tag{6.12}$$

Applying M once more we obtain that (χ,ϕ) is an eigenvector of M^2. Moreover *all* eigenvectors of M^2 are of this form as there are $q^{t-1}(q-1)$ such vectors — exactly the dimension of the ambient space. These vectors are pairwise orthogonal hence they are linearly independent.

Exercise 6.8 *Check that indeed, the set of vectors* $\{(\chi,\phi):\chi\in\widehat{\mathbb{F}}_{q^{t-1}},\psi\in\widehat{\mathbb{F}_q^*}\}$ *forms an orthogonal basis of* $\mathbb{C}^{q^{t-1}(q-1)}$.

So all eigenvalues of M^2 are of the form

$$\left|\sum_{\substack{C\in\mathbb{F}_{q^{t-1}}\\ C\neq 0}}\chi(C)\phi(N(C))\right|^2.$$

The eigenvalues of M^2 are the squares of the eigenvalues of M. These are reals, as M is symmetric. So all eigenvalues of M are of the form

$$\pm\left|\sum_{\substack{C\in\mathbb{F}_{q^{t-1}}\\ C\neq 0}}\chi(C)\phi(N(C))\right|.$$

This expression looks very similar to that of a Gaussian sum, as discussed in Theorem A.40. We only have to sort out what field we are exactly talking about. The additive character χ is of the field $\mathbb{F}_{q^{t-1}}$, but ϕ is a multiplicative character of a different field $\mathbb{F}_q$. Fortunately we also have the norm function $N : \mathbb{F}_{q^{t-1}} \to \mathbb{F}_q^*$ in the formula. Since both ϕ and N are multiplicative functions, their composition ϕN is a multiplicative character of the field $\mathbb{F}_{q^{t-1}}$.

We can now apply Theorem A.40 that estimates the absolute value of Gaussian sums and have that

$$\left| \sum_{\substack{C \in \mathbb{F}_{q^{t-1}} \\ C \neq 0}} \chi(C)\phi(N(C)) \right| \leq q^{(t-1)/2},$$

unless $\chi = \chi_0$ and $\phi = \phi_0$ are the principal characters of their respective groups when the corresponding eigenvalue is $q^{t-1} - 1$. (No surprise, as this is the degree of G.) □

Exercise 6.9 *Determine all eigenvalues, together with their multiplicities. (Hint: Try to guess the eigenvectors corresponding to eigenvalues* 1 *and* -1*.)*

Exercise 6.10 *Derive that the projective norm-graphs provide a constructive lower bound of* $\Omega(k^{4/3}(1 + o(1)))$ *for* $r(C_4, K_k)$*. (For which the best known lower bound, using the Local Lemma, is of order* $(k/\log k)^{3/2}$*.)*

The number of small cliques in a quasi-random graph

To prove Property 5. we need Property 4. and the quasi-random edge-distribution of (n, d, λ)-*graphs*, discussed in .... These random-like properties (n, d, λ)-graphs with small λ imply yet another series of random-like properties: Property 5. states that the # of K_s in G is roughly equal to the expected number of K_s in a truly random graph $G(n, p)$ with edge probability $p = n^{-1/t}$.

In fact we derive a much stronger statement by induction. We will see that there is nothing special about the number of *cliques* in Property 5, every small graph appears as many times in $G_{q,t}$ as it approximately would in a random graph $G(n, p)$ with edge probability $p = n^{-\frac{1}{t}}$.

Lemma 6.5.1 *Let* G *be an* (n, d, λ)*-graph. Let* H *be an arbitrary graph on* s *vertices, such that*

$$d = o\left(n^{1-\frac{1}{2s-3}}\right), \qquad \lambda = o\left(\frac{d^{s-2}}{n^{s-3}}\right).$$

Then the number of labeled copies of H *in* G *is*

$$(1 + o(1))n^s \left(\frac{d}{n}\right)^r,$$

where r *denotes the number of edges of* H*.*

Corollary 6.6 *Let $s < \frac{t}{2} + 1$ be an arbitrary positive integer. Then the number of labeled copies of K_s in the projective norm-graph $G_{q,t}$ is*

$$(1 + o(1))n^{s-\binom{s}{2}/t},$$

Remark: The lemma is true for arbitrary constant s provided the edge density of G is linear in n; this is the version proved by Chung, Graham and Wilson [?] (c.f. ....). For sparser graphs some upper bound on s, that is on how "big" "small subgraphs" the lemma is valid for, is absolutely essential. Such an upper bound is necessitated by the fact that we are only dealing with quasi-random graphs as opposed to truly random ones. In truly random graphs of edge density $n^{-1/t}$ the copies of $K_{t,(t-1)!+1}$ would be abound, while in our, only quasi-random, projective norm graph there are none.

Of course this is exactly the power of quasi-randomness: we turn its weak randomess to our advantage. In some properties that are not needed for our purposes we are content with something weaker than truly random, like here we do not have the lemma for arbitrary constant s. At the same time this relaxation of randomness lets us to achieve another essential component of our proof, not possessed by random graphs, the $K_{t,(t-1)!+1}$-freeness of the projective norm-graphs.

Proof. Induction on r. The statement is trivial for $r = 0$.

Assume now that $r > 0$ and let us denote d/n by p. Let $N(H)$ be the number of labeled copies of H in G, that is, more formally, the number of injections $f : V(H) \to V(G)$ which are homomorphisms. We set to prove $N(H) = (1 + o(1))n^s p^r$.

Let $uv = e$ be an arbitrary edge of H, there certainly exists one. We denote by H' the graph $H - \{u, v\}$ we obtain from H by deleting u and v. Throughout the proof let f' denote a general injective homomorphism of H' to G. Our approach is to fix an injective homomorphism f' on H' and try to count its extensions to u and v. Let $V(u, f')$ denote the set of possible images of u that would make an injective extension of f' to u which is a homomorphism. That is,

$$V(u, f') := \bigcap_{w \in N_H(u) \setminus \{v\}} N_G(f'(w)) \setminus f'(V(H')),$$

and $V(v, f')$ is defined analogously.

An extension of f' to $\{u, v\}$ is an injective homomorphism of H into G if and only if the images of u and v are in $V(u, f')$ and $V(v, f')$, respectively, they are different, and *adjacent* to each other. Hence $N(H)$ is exactly $e(V(u, f'), V(v, f'))$ minus the number of loops in $V(u, f') \cap V(v, f')$. We expect the number of loops to be of smaller order than the main term. By quasi-randomness, we expect to be able to estimate the main term $e(V(u, f'), V(v, f'))$ by $p|V(u, f')||V(v, f')|$ up to an error term of $\lambda\sqrt{|V(u, f')||V(v, f')|}$. We can think of $|V(u, f')||V(v, f')|$ more or less to be the number of extensions of f' to $H - e$, since here we do not have any restriction on whether the images of u and v are adjacent. Here the only error we have to take care of is when the images of u and v coinside, but there are only $|V(u, f') \cap V(v, f')|$ ways this could happen. Finally,

the number of embeddings of $H - e$ to G is equal, by the induction hypothesis, to $p^{r-1}n^s(1+o(1))$.

Now the calculations follow.

$$
\begin{aligned}
|N(H) - p^r n^s| \quad \leq \quad & \left| N(H) - \sum_{f'} e(V(u,f'),V(v,f')) \right| + && (6.13)\\
& \left| \sum_{f'} e(V(u,f'),V(v,f')) - \sum_{f'} p|V(u,f')||V(v,f')| \right| + && (6.14)\\
& \left| p \sum_{f'} |V(u,f')||V(v,f')| - pN(H-e) \right| + && (6.15)\\
& p\left|N(H-e) - p^{r-1}n^s\right| && (6.16)
\end{aligned}
$$

We will see that all four terms are $o(p^r n^s)$.

The fourth term is clearly $p \cdot o(p^{r-1}n^r)$ by induction.

We bound the first and the third term similarly. The first term is exactly $\sum_{f'} \mu(f')$, where $\mu(f')$ denotes the number of loops in $V(u,f') \cap V(v,f')$, the third term is exactly $\sum_{f'} |V(u,f') \cap V(v,f')|$. Since $\mu(f') \leq |V(u,f') \cap V(v,f')| \leq n$, both terms are upper bounded by

$$\sum_{f'} n = nN(H') = n \cdot (1+o(1))n^{s-2}p^{r'},$$

where r' denotes the number of edges in H' and we use the induction hypothesis for H'. Since $r - r' \leq 2s-3$ the condition $d \ll n^{1-\frac{1}{2s-3}}$ implies $n^{s-1}p^{r'} = o(n^s p^r)$.

To bound the order of the error term (6.14) we use Corollary A.24 and obtain

$$
\begin{aligned}
\sum_{f'} \lambda \sqrt{|V(u,f')||V(v,f')|} \quad &\leq \quad \lambda \sqrt{N(H')} \sqrt{\sum_{f'} |V(u,f')||V(v,f')|} \\
&= \quad \lambda(1+o(1))\sqrt{n^{s-2}p^{r'} \cdot n^s p^{r-1}} \\
&= \quad o(n^{s-\frac{r}{t}})
\end{aligned}
$$

The first inequality follows from the Cauchy-Schwarz Inequality, the second equality from induction and the above estimates on (??) and (6.16), while the third one from $\lambda = o(\frac{d^{s-2}}{n^{s-3}})$. □

6.3 Multicolored Ramsey-numbers

We saw in the previous sections how hard the asymmetric Ramsey problem is: often probabilistic constructions are far from the upper bounds and explicit ones are even

weaker. And this is for two colors only ... Asymmetric Ramsey problems for three and more colors seemed even more hopeless. To quote just one of the more desparete questions [**?**]: it was not even known that

$$\frac{R(K_3, K_3, K_m)}{R(K_3, K_m)} \to \infty.$$

Alon and Rödl [**?**] discovered a method that makes it possible to determine the values of certain asymmetric multicolor Ramsey numbers up to a logarithmic factor and occasionally even up to a constant factor! They construct nearly optimal Ramsey graphs when one of the colors forbids the clique K_m and all other color classes forbid some fixed small subgraph, like K_3 or C_4. Somewhat surprisingly, the larger the number of colors is the more effective the method becomes. Although the construction is eventually random, explicit quasi-random graphs play an essential role.

The basic aproach of how, say, $R(C_4, C_4, K_m)$ is determined up to a polylogarithmic factor is pretty simple: Take two random copies of an appropriate C_4-free graph G on the same vertex set and color their edges with color one and two, respectively. The remaining edges of K_n are colored with color three. Now the first two color classes are C_4-free by definition, so the success of the construction hangs on the independence number of the union of the two random copies of G. An obvious choice for the role of the C_4-free graph G is a graph with the smallest independence number. It seems plausible to beleive that it cannot hurt to start already with a graph whose independence number is small to begin with. The existence of such a graph with independence number $\widetilde{\Theta}\left(n^{2/3}\right)$ is proved by the Local Lemma in the random graph space $G(n,p)$ with edge-probability $p = n^{-2/3}$. The problem is: taking two copies of a random graph is practically equivalent to increasing the edge-probability to twice as large, and a constant factor does not make a visible impact on the independence number when one applies the Local Lemma.

The polarity graph G of Section Section **??**, which is the best known explicit construction for $ExpR(C_4, K_m)$, does have a larger independent set to begin with, the value of α is $\Theta\left(n^{3/4}\right)$. However, the explicit graph is much denser than the randomized C_4-free construction and this makes it possible to successfully apply the main realization of Alon and Rödl: The independence number of quasi-random graphs might be as large as $\frac{n}{d}\lambda$, but they have only a "small number" of "big" independent sets, of size $m = \frac{n}{d}\log^2 n$. Hence, by taking already two random copies of G there is a decent chance that none of the $\binom{n}{m}$ subsets of the vertex set will be one of these "big" independent sets.

The actual quantitative bound, to be proved in the next subsection, is the following.

Corollary 6.7 *Let G be an (n, d, λ)-graph. Then the number of independent sets of size $m = \log^2 n \frac{n}{d}$ in G is at most*

$$\left(\frac{\lambda}{\log^{1.9} n}\right)^m.$$

Let us draw the reader's attention to two surprising comparison, regarding the previous corollary. Recall that $\frac{n}{d+1}$ is a simple general *lower bound* on the independence

number of *any* d-regular graph and there are many independent sets of that size (see the next subsection). According to the corollary the number of independent sets of only $\log^2 n$ times this simple lower bound is already significantly bounded. The standard uooer bound on the independence number of a quasi-random graph is $\lambda\frac{n}{d}$ and this bound is often tight. In our applications the λ factor is usually a power of n and hence $\lambda \gg \log^2 n$.

Remark: It is worthwhile to note why the denser C_4-free quasi-random graph is essential for this approach to work. The size of the independent sets whose number we control is the smaller when the degree is larger. The degree in the C_4-free random construction with independence number $\widetilde{\Theta}\left(n^{2/3}\right)$ is around $\widetilde{\Theta}\left(\sqrt[3]{n}\right)$ so, even if we would know that this graph is quasi-random, the lemma starts working for independent sets of size only $\widetilde{\Theta}\left(n^{2/3}\right)$. The degree in the polarity graph is $\sqrt{n}$, so the lemma estimates the number of independent sets of size as small as $\widetilde{\Theta}\left(\sqrt{n}\right)$. In some sense, this approach uses the "fault" of the quasi-random polarity graph compared to the "truly" random graphs: truly random graphs have no chance to be C_4-free and be so dense at the same time; the existence of the polarity graph is a fortunate accident.

Once one knows a bound on the number of independent sets of size m, the following straightforward, but somewhat unconventional application of the probabilitic method provides a lower bound on the Ramsey number $R_k(H, K_m) := R(\underbrace{H, \ldots, H}_{k}, K_m)$.

Lemma 6.7.1 *Let G be an H-free graph on n vertices and let M denote the number of independent sets of size m in G.*
If, for a positive integer $k \geq 2$, $M^k < \left(\binom{n}{m}\right)^{k-1}$, then $r_k(H; K_m) > n$.

Proof. We prove the existence of an appropriate $(k+1)$-coloring of the edges of the complete graph on V, $|V| = n$. For $1 \leq i \leq k$, let G_i be a random copy of G on V, i.e., the graph obtained from G by mapping the veretices of G to V according to a random one to one mapping. Color an edge e of the complete graph on V by the smallest index i for which e belongs to G_i, if such an index exists, otherwise color it by $k+1$. Obviously there are no monochromatic H in any of the first k colors as for any $i \leq k$ the color class i forms a subgraph of the H-free graph G_i.

Fix a set I of m vertices. What is the probability that I forms a clique in the last color class? Well, that is equivalent to saying that I is independent in each of the k graphs G_i. Now comes the unconventional counting: I has to be one of the M independent sets of size m contained in G_i. There are $\binom{n}{m}$ sets of size m in $V(G)$, each has $1/\binom{n}{m}$ chance to be mapped to I according to a random mapping. Hence altogether the probability that an independent set is mapped onto I is $M/\binom{n}{m}$. The choices of the G_i are independent, so the probabilities multiply:

$$Pr[I \text{ is a clique in the } (k+1)\text{st color class}] = \left(\frac{M}{\binom{n}{m}}\right)^k.$$

This is true for any set $I \subseteq V$ of size m. By the union bound the probability that *some* set of size m is independent in the $(k+1)$st color class is at most

$$\binom{n}{m}\left(\frac{M}{\binom{n}{m}}\right)^k,$$

which is strictly less than 1 by our assumption.

Hence *there exists* a $(k+1)$-coloring where there are no monochromatic cliques of order m in color $(k+1)$.

□

Remark: What would be the "conventional" idea for probabilistic analysis? Fix an independent set I of order m and calculate the probability that two random copies of a d-regular graph do not hit any of its pairs. ..... In this approach only use the pure existence of a dense C_4-graph, but use nothing from their quasi-random properties.

6.3.1 Bipartite graphs

A general corollary of the previous two lemmas for multicolored Ramsey numbers can now be formulated.when we are in possession of a "perfectly quasi-random" H-free graph, that is an H-free graph with $\lambda = \Theta\left(\sqrt{d}\right)$.

Corollary 6.8 *Suppose there exists an H-free (n,d,λ)-graph with $d = n^\alpha$ and second eigenvalue $\lambda = \Theta(\sqrt{d})$. Then for for every integer $k \geq 2$.we have that*

$$R(\underbrace{H,\ldots,H}_{k},K_m) > \left(\frac{m}{\log^2 m}\right)^{\frac{1}{1-\alpha}}.$$

The C_4-free polarity graph, the polarity graph of Benson's C_6-free and C_{10}-free graph (pointed out by Lazebnik and Ustimenko), and the $K_{t,(t-1)!+1}$-free projective norm graphs are all (n,d,λ)-graphs with $\lambda = \Theta(\sqrt{d})$. Their existence implies the following.

Corollary 6.9 *We have*

$$\begin{aligned} r_k(C_4,K_m) &= \tilde{\Theta}\left(m^2\right). \\ r(C_6,C_6,K_m) &= \tilde{\Theta}\left(m^{3/2}\right). \\ r(C_{10},C_{10},K_m) &= \tilde{\Theta}\left(m^{5/4}\right). \\ r(K_{t,s},K_m) &= \tilde{\Theta}\left(m^t\right) \textit{ for every } s > (t-1)!. \end{aligned}$$

Exercise 6.11 *Prove the upper bounds in the previous corollary.*

Exercise 6.12 *Verify that Corollary 6.10 is applicable for $H = C_6$ and $H = C_{10}$*

6.3.2 Triangles

We start out with an exercise, which provides some upper bound on $R_k(K_3, K_m)$. In this section we find out that this simple upper bound is in fact tight up to a logarithmic factor

Exercise 6.13 *Prove that $R_k(K_3, K_k) = O_k(m^{k+1})$, for every integer $k \geq 1$.*

Corollary 6.10 and Alon's K_3-free (n, d, λ)-graph with parameters $d = \Theta(n^{2/3})$ and $\lambda = \Theta(n^{1/3})$ clearly imply a tight lower bound of $\widetilde{\Theta}(m^3)$ on $R(K_3, K_3, K_m)$.

However, the order of the upper bound of Exercise 6.13 on $R_k(K_3, K_m)$ grows with k. To obtain a matching lower bound using Corollary 6.10 one needs that $\frac{n}{d}\log^2 n = m = \widetilde{\Theta}\left(n^{1/(k+1)}\right)$. That means a triangle-free quasi-random graph with $d = \widetilde{\Theta}(n^{k/(k+1)})$; much too dense to have $\lambda = \Theta\left(\sqrt{d}\right)$ for $k \geq 3$. Indeed, we saw in in Subsection ??. that Alon's graph is the densest triangle-free (n, d, λ)-graph with $\lambda = \Theta\left(\sqrt{d}\right)$, so we need to relax our assumption on perfect quasi-randomness. Fortunately, for $k \geq 3$, this is possible.

Corollary 6.10 *Suppose there exists an H-free (n, d, λ)-graph with $d = n^\alpha$ and second eigenvalue*

$$\lambda < d^{\frac{k-1}{k}}(\log n)^{2-0.1k},$$

for some positive integer k. Then for the multicolor Ramsey number we have that

$$R(\underbrace{H, \ldots, H}_{k}, K_m) > \left(\frac{m}{\log^2 m}\right)^{\frac{1}{1-\alpha}}.$$

Theorem 6.11 *For every $k \geq 1$, we have*

$$R_k(K_3, K_m) = \widetilde{\Theta}\left(m^{k+1}\right).$$

Proof. Use an appropriate blow-up of Alon's graph as discussed in Exercise ?? □

Exercise 6.14 *Suppose we are given a K_3-free graph G with independence number $\sqrt{n \log n}$ (by Kim such a thing does exists!) Prove the statement of Theorem 6.11 using random shifts of blowups of G.*

Exercise 6.15 *Prove that*

$$r_k(K_{t+2}; K_m) = \tilde{\Omega}(m^{k(t+1)/2+1})$$

using the fact that we know, by the Local lemma, that $r(K_{t+2}, K_m) = \tilde{\Omega}(m^{(t+3)/2})$.

Corollary 6.12 *Suppose there exists an (n, d, λ)-graph with $d = \tilde{\Theta}(n^\alpha)$ and $\lambda = \Theta(\sqrt{d})$, containing no homomorphic image of H. Then for any $k \geq 2$, $r_k(H, K_m) = \tilde{\Omega}(n^{\frac{\alpha k}{2(1-\alpha)}+1})$.*

Proof. Apply Corollary ?? to the r-blow-up of H with $r = n^{\alpha(\frac{k}{2}-1)}$ and Lemma 6.7.1.

□

6.3.3 The number of big independent sets in pseudorandom graphs

Let us finally wrap up the proof of Theorem ??.

Yet another: One way to solve the exercise problem of $r_k(K_{3,3}$, which we had in Section... is to drop random copies of $K_{3,3}$-free graphs on the board of $n = (k/\log k)^3$-vertices. The probability that a particular edge won't be covered by any of the k graphs is $(1 - cn{-}1/3)^k < e^{-cn^{-1/3}k} = e^{-c\log k} = k^{-10} = n^{-2}$. So with nonzero probability every edge will be covered.

Putting down one copy the independence number is $n\lambda/d = n^{1/2} = (k/\log k)^{3/2}$. Putting down two copies we already showed that the independence number is $k\log^? k$ and for three copies it is k where it stays for any finite number of copies, but for k copies it goes down to 0. What is the speed it goes down?

The simplest gready algorithm attempts to construct an independent set of cardinality m by setting $I = \emptyset$, then taking an arbitrary vertex $v \in V(G)$, adding it to I, and then deleting v and all its neighbors from G, and repeating this procedure m times or until G becomes empty.

$$\frac{\sum_{v_1\in V}\sum_{v_2\in B(v_1)}\sum_{v_3\in B(v_1,v_2)}\cdots\sum_{v_m\in B(v_1,\ldots,v_{m-1})}1}{m!}.$$

If our graph is d-regular, then the above algorithm is surely succesful for $m \le \frac{n}{d+1}$, since in one iteration no more than $d+1$ vertices get deleted. There are $m!$ ways the above algorithm finds a particular independent set. So the number of independent sets of size m, for $m \le n/(d+1)$, is

$$\begin{aligned}
&\frac{\sum_{v_1\in V}\sum_{v_2\in B(v_1)}\sum_{v_3\in B(v_1,v_2)}\cdots\sum_{v_m\in B(v_1,\ldots,v_{m-1})}1}{m!}\\
&\ge \frac{b_0\cdot b_1\cdot\cdots\cdot b_{m-1}}{m!}\\
&\ge \frac{n(n-(d+1))\cdots(n-(m-1)(d+1))}{m!}\\
&\ge \left(\frac{n}{m}\right)^m \ge (d+1)^m
\end{aligned}$$

where $B(v_1,\ldots v_i) = V(G)\backslash(\cup_{j=1}^{i}\{v_j\}\cup N(v_j))$, and b_i is the minimum value of $|B(v_1,\ldots v_i)|$ where $v_1,\ldots v_i$ runs through all i-tuples of the vertices of G.

Lemma 6.12.1 *Let G be an (n,d,λ)-graph. Then for any integer $m \ge s := \frac{2n\log n}{d}$, the number of independent sets of order m in G is at most*

$$e^{3m}\left(\frac{\lambda}{d}\right)^{m-s}\left(\frac{n}{m}\right)^m.$$

Proof. Let us follow the simple iterative procedure which picks vertices of an independent set one after the other and always deletes the picked vertex together with its neighbors

from further consideration. We will count *ordered* independent sets consisting of m vertices, hence counting each of them $m!$ times.

For an independent set $\{v_1, \dots, v_i\}$ let us denote by by $B(v_1, \dots v_i)$ those vertices that are still eligible to be picked next into our independent set. That is,

$$B(v_1, \dots, v_i) = V(G) \setminus (\cup_{j=1}^{i}(N(v_j) \cup \{v_j\})).$$

If $\{v_1, \dots, v_m\}$ is an independent set, then we know that the sets $B_0 = V(G) \supseteq B(v_1) \supseteq B(v_1, v_2) \supseteq \cdots \supseteq B(v_1, \dots, v_{m-1}) \neq \emptyset$ form a decreasing chain ending in a nonempty set.

Heuristically, for G to have lots of independent sets, the sets $B(v_1, \dots, v_i)$ should be "large" for many indices i. This is equivalent to saying that v_i should have only a "few" neighbors in $B(v_1, \dots, v_{i-1})$. In a quasi-random graph, however, there cannot be too many such vertices: in *any* (relatively large) set B most vertices actually must have as many neighbors as the edge density suggests.

We expect that vertices have roughly $\frac{d}{n}|B|$ neighbors in B Let $C = C(B)$ be the set of those vertices, which have not more than half of this. Formally,

$$C = \left\{ u \in B : |N(u) \cap B| \leq \frac{d}{2n}|B| \right\}.$$

The set $C(B(v_1, \dots, v_{i-1}))$ consists of those vertices whose selection into the independent set does not reduce the size of the potential continuations by much. Quasi-randomness says that there are not so many of these, only at most $\frac{2\lambda}{d}n$.

Lemma 6.12.2 *Let G be a (n, d, λ)-graph, and $B \subseteq V$ an arbitrary subset of its vertices. For*

$$C = C(B) = \left\{ u \in B : |N(u) \cap B| \leq \frac{d}{2n}|B| \right\},$$

we have $|C| \leq \frac{2\lambda}{d}n$.

Proof. We apply Lemma A.23.1 and obtain

$$\begin{aligned} \lambda^2|B| &\geq \sum_{v \in V} \left(N_B(v) - |B|\frac{d}{n} \right)^2 \geq \sum_{v \in C} \left(N_B(v) - |B|\frac{d}{n} \right)^2 \\ &\geq \sum_{v \in C} \left(|B|\frac{d}{2n} \right)^2 = |C||B|^2\frac{d^2}{4n^2}, \end{aligned}$$

implying $\frac{4\lambda^2}{d^2}n^2 \geq |B||C| \geq |C|^2$. □

Let us count now ordered independent sets of size m in G. Suppose the sequence $(v_1, \dots, v_m)$ provides an independent set. We use the abbreviation $B_i = B(v_1, \dots v_i)$ and $C_i = C(B_i)$. For how many indices could it happen that $v_{i+1} \notin C_i$? Well, if that is the case then $|N(v_{i+1}) \cap B_i| \geq \frac{d}{2n}|B_i|$, i.e., $|B_{i+1}| < \left(1 - \frac{d}{2n}\right)|B_i|$. This cannot happen

for more than $\frac{2n}{d}\log n$ indices otherwise $|B_{m-1}| < \left(1-\frac{d}{2n}\right)^{\frac{2n}{d}\log n}|B_0| < e^{-\log n}n = 1$, meaning B_{m-1} would be empty contradicting the fact that v_m is in it. In conclusion, for most of the indices $v_{i+1} \in C_i$.

To count sequences $(v_1,\dots,v_m)$ that provide an independent set, first we select $s = \frac{2n}{d}\log n$ indicies i for which we would allow $v_i \notin C_{i-1}$. This can be done $\binom{m}{s}$ ways. Once these are fixed, there are at most n ways to select a vertex v_i for each such index. Once these vertices are selected, for each remaining index j we select, in an increasing order, a vertex v_j from the set C_{j-1} (that is defined by the first $j-1$ choices). Each of these sets is of order at most $\frac{2\lambda}{d}n$ by Lemma 6.12.2. Hence the number of sequences $(v_1,\dots,v_m)$ providing an independent set is

$$\binom{m}{s}n^s\left(\frac{2\lambda}{d}n\right)^{m-s}.$$

Substituting and dividing by $m!$ we obtain that the number of independent sets is

$$\frac{1}{m!}\binom{m}{s}n^s\left(\frac{2\lambda}{d}n\right)^{m-s}.$$

□

Using the previous lemma and a more careful analysis one is able to improve on Corollary 6.9 and occasionally determine the correct order of magnitude of $R_k(H,K_m)$, when $k \geq 3$.

Theorem 6.13 *Let H be a bipartite graph with Turán-number $ex(n,H) = O(n^{1+\alpha})$. Assume furthermore that there exists an H-free (n,d,λ)-graph with $d = \Theta(n^\alpha)$ and $\lambda = \Theta\left(\sqrt{d}\right)$. Then for every $k \geq 3$ we have*

$$R_k(H,K_m) = \Theta\left(\frac{m}{\log m}\right)^{\frac{1}{1-\alpha}}.$$

Exercise 6.16 *Prove Theorem 6.13. For the upper bound, apply the following theorem of Ajtai, Komlós and Szemerédi.*

Theorem 6.14 (Ajtai, Komlós, Szemerédi) *Let T be a graph on n vertices with average degree d. If T contains less than $nd^{2-\gamma}$ triangles for some positive constant $\gamma > 0$, then*

$$\alpha(T) \geq \Omega\left(\frac{n}{d}\log d\right).$$

Theorem 6.13 clearly determines $R_k(H,K_m)$ up to a constant factor for $k \geq 3$, when H is C_4, C_6, C_{10}, or $K_{t,s}$ with $s > (t-1)!$. The question of Erdős about $r(C_4,K_m)$ has baffled researchers for decades. Mankind is simply not able to decide wheather there exists a C_4-free graph on n vertices, whose independence number is as small as $n^{\frac{1}{2}+\epsilon}$.

We are in fact very far: the best randomized construction gives independence number $n^{2/3}$. It is then particularly surprising that the union of two C_4-free graphs can already have such a small independence number and for the union of three C_4-free graphs one can even determine the power of the log-factor. It would be interesting to decide the exact log-power for $k = 2$. This could mean a bit more clever (algebraic?) placement of the two C_4-free graphs on top of each other, or reducing the known upper bound of $m^2/\log^2 m$ on $r(C_4, K_m)$ by a couple of log-factors, the latter surely being a hard task.

Appendix A

Appendix

A.1 Useful Facts about Finite Fields and Projective Planes

We start by recalling some basic definitions and facts from algebra about (abelian) groups, rings and finite fields. We state some standard results without proof (though the reader is heartily invited to verify them).

A.1.1 Groups

Definition: A *group* $(G, \circ)$ is a set G together with a binary operation $\circ$ on G (meaning that $\circ : G \times G \to G$) such that the following three properties hold:

(i) the operation $\circ$ is *associative*, that is

$$\forall\, a, b, c \in G \;:\; a \circ (b \circ c) = (a \circ b) \circ c \;,$$

(ii) there is an *identity element* (or *unit*) $e \in G$ such that

$$\forall\, a \in G \;:\; a \circ e = e \circ a = a \;,$$

(iii) for each $a \in G$ there is an *inverse element* $a^{-1} \in G$ such that

$$a \circ a^{-1} = a^{-1} \circ a = e \;.$$

If the group also satisfies

(iv) $\forall\, a, b \in G \;:\; a \circ b = b \circ a$,

the group is called *abelian* (or *commutative*).

For the sake of simplicity we will (as usual) often abbreviate $(G, \circ)$ by writing G and also write ab for the operation in the group instead of $a \circ b$. When referring to abelian groups it will be convenient to use $a + b$ instead of $a \circ b$.

Definition: A group G is called *cyclic* if there is an element $a \in G$ such that for any $b \in G$ there is some integer i with $b = a^i$. Such an element a is called a *generator* of G.

Definition: A group G is called *finite* if it contains only finitely many elements. The cardinality $|G|$ of the underlying set is the *order* of the group.

Definition: A subset H of the group G is a *subgroup* of G if H is a group itself with respect to the operation of G. The (cyclic) subgroup of all powers of an element $a \in G$ is called the subgroup *generated by* a, which will be denoted by $\langle a \rangle$. If $\langle a \rangle$ is a finite group, then its order is called *order* of the element a.

Theorem A.1 *If H is a subgroup of G then the relation R_H on G defined by $(a, b) \in R_H$ if and only if $a = bh$ for some $h \in H$, is an equivalence relation.*

Remark: Recall that an equivalence relation R on a set S induces a partition of S: for a fixed element $s \in S$ we denote the *equivalence class* of s (i.e. all the elements of S equivalent to s with respect to R) by

$$[s] = \left\{ t \in S \ : \ (s, t) \in R \right\} .$$

The collection of all distinct equivalence classes then forms a partition of S.

The relation R_H is called *left congruence modulo H*. Its induced equivalence classes are called *left cosets of G modulo H*, and denoted by

$$aH = \{ah \ : \ h \in H\} \ , \text{ for } a \in G.$$

Similarly one can define the right congruence modulo H, and obtain the right cosets of G modulo H, denoted $Ha = \{ha \ : \ h \in H\}$ for some $a \in G$.

Theorem A.2 *If H is a finite subgroup of G then every (left or right) coset of G modulo H has the same number of elements as H.*

Definition: If the subgroup H of G only yields finitely many distinct left cosets of G modulo H then the number of such cosets is called the *index* of H in G. We denote the index by $|G : H|$.

Theorem A.3 *Let G be a finite group and H be a subgroup of G. Then*

$$|G| = |G : H| \cdot |H| \ .$$

In particular, the order of H divides the order of G and the order of any element $a \in G$ also divides the order of G.

For cyclic groups we can say even more about its subgroups.

Theorem A.4 *Let G be a cyclic group. Then every subgroup of G is also cyclic. If G is a finite cyclic group and d is a divisor of the group's order $|G|$, then there is exactly one subgroup of G that has order d, and exactly one subgroup of index d.*

In the following we want to introduce a tool for comparing structures of two groups. For this purpose we investigate properties of mappings between the groups preserving the group operations.

Definition: Let $(G, *)$ and $(H, \cdot)$ be groups. A mapping $f : G \to H$ is called a *homomorphism* of G into H if

$$\forall\ a, b \in G\ :\ f(a * b) = f(a) \cdot f(b)\ .$$

If, additionally, f is surjective we call f an *epimorphism*. If f is bijective, then f is called an *isomorphism* and we say the groups G and H are *isomorphic*.
A homomorphism $f : G \to G$ is called *endomorphism*, and an isomorphism $f : G \to G$ is called *automorphism*.

Definition: A subgroup H of a group G is called a *normal subgroup of* G if for all $a \in G$ and all $h \in H$ we have $aha^{-1} \in H$.

Theorem A.5 *A subgroup H of a group G is normal if and only if the left coset aH and the right coset Ha are equal for all $a \in G$.*

Normal subgroups play an important role because one can impose a group structure on the set of cosets. (Recall that left and right cosets are the same in that case.)

Theorem A.6 *Let H be a normal subgroup of G. Then the set of cosets of G modulo H forms a group with respect to the operation defined by*

$$(aH)(bH) := (ab)H\ .$$

Definition: We call the group defined in Theorem A.6 *factor group* (or *quotient group*) of G modulo H and denote it by G/H.

If G/H is a finite group its order is equal to the index of H in G. If G is also finite, by Theorem A.3 we have

$$|G/H| = \frac{|G|}{|H|}\ .$$

A.1.2 Rings

Definition: A *ring* $(R, +, \cdot)$ is a set R together with two binary operations $+$ and $\cdot$, such that

(i) R is an abelian group with respect to $+$,

(ii) $\cdot$ is an associative operation,

(iii) the *distributive law* holds, that is

$$\forall\ a, b, c \in R\ :\ a \cdot (b + c) = a \cdot b + a \cdot c \text{ and } (b + c) \cdot a = b \cdot a + c \cdot a\ .$$

We will only consider rings that are *commutative*:

(iv) $\forall\ a, b \in R\ :\ a \cdot b = b \cdot a$,

and have an *identity element*

(v) $\exists\ e \in R\ \ \forall\ a \in R\ :\ a \cdot e = e \cdot a = a$.

Throughout this course when talking about a "ring" we shall always mean a commutative ring with an identity element, that is a ring satisfying the axioms (i)-(v) from the definition above.

Some usual notations for abbreviating ring operations are: generally we write R when talking about a ring instead of writing $(R, +, \cdot)$, we use 0 for the unity with respect to $+$, the additive inverse of a is denoted $-a$, we write $a - b$ instead of $a + (-b)$, and ab stands for $a \cdot b$.

In a ring we have $0a = a0 = 0$ for all $a \in R$ which implies that $(-a)b = a(-b) = -ab$ for all $a, b \in R$.

Definition: An *integral domain* is a ring with identity $e \neq 0$ in which $ab = 0$ implies $a = 0$ or $b = 0$ (there are no *zero divisors*).

Definition: A subset S of a ring R is called a *subring* of R if S is closed under $+$ and $\cdot$, and forms a ring with respect to these operations.

Definition: A subset I of a ring R is called an *ideal* if I is an additive subgroup of R and for all $a \in I, r \in R\ :\ ra \in I$.

For a given element $a \in R$ we denote by (a) the smallest ideal containing a. Note that since we assumed that R contains an identity element $(a) = \{ra : r \in R\}$.

Definition: An ideal I of a ring R is called *principal* if there is an $a \in R$ such that $I = (a)$. We say that I is the principal ideal *generated* by a.

Note that ideals are normal subgroups of the additive group of a ring. Hence, it follows that an ideal I in a ring R partitions R into non-empty mutually disjoint subsets, called *residue classes* modulo I. We denote the residue class of $a \in R$ modulo I by $[a] = a + I$. Elements $a \in R$ and $b \in R$ are called *congruent* modulo I ($a \equiv b \bmod I$) if they are in the same residue class modulo I, that is $a - b \in I$. Note that the set of residue classes of a ring R modulo an ideal I forms a ring with respect to the following operations

$$\begin{aligned} [a] + [b] &= (a + I) + (b + I) = (a + b) + I = [a + b] \\ [a]\,[b] &= (a + I)(b + I) = (ab) + I = [ab] \ . \end{aligned}$$

This ring is called *residue class ring* of R modulo I and denoted by R/I.

An important example is the residue class ring $\mathbb{Z}/(n)$, where (n) is the principal ideal generated by n. We denote the residue class of an integer a modulo the natural number n by $[a]$. Then the elements of $\mathbb{Z}/(n)$ are

$$[0], [1], \ldots, [n-1].$$

A.1.3 Finite fields

Definition: A *field* is a ring in which the non-zero elements of R form a group with respect to $\cdot$ (multiplication). Fields that only contain finitely many elements are called *finite fields*.

Note that every field is an integral domain. The converse is not true in general, consider for example the integers with the usual addition and multiplication. However, the following holds:

Theorem A.7 *Every finite integral domain is a field.*

Theorem A.8 *$\mathbb{Z}/(p)$, the ring of residue classes of the integers modulo the principal ideal generated by a prime p, is a field.*

In the following we will give a more convenient representation for the finite fields $\mathbb{Z}/(p)$. We do this by introducing a more general framework. Basically, it will just formally explain why we can "add" and "multiply" in $\mathbb{Z}/(p)$ by simply evaluating sums and products modulo p.

Consider a bijective map ϕ from a ring R into a set S. Now by means of ϕ one can impose a ring structure onto S that converts ϕ into an isomorphism. More precisely, let s_1 and s_2 be elements in S and let r_1 and r_2 be its pre-images under ϕ in R, that is $\phi(r_1) = s_1$ and $\phi(r_2) = s_2$. Then define

$$\begin{aligned} s_1 + s_2 &:= \phi(r_1 + r_2) \\ s_1 s_2 &:= \phi(r_1 r_2)\ . \end{aligned}$$

In case R has additional properties (for example being an integral domain or a field), then these also hold for S.

Definition: For a prime number p, let $\mathbb{F}_p$ be the set $\{0, 1, \ldots, p-1\}$ of integers and let $\phi : \mathbb{Z}/(p) \to \mathbb{F}_p$ be the mapping defined by $\phi([a]) = a$, for $a = 0, 1, \ldots, p-1$. Then $\mathbb{F}_p$, together with the field structure induced by ϕ, is a finite field, called the *Galois field of order p*.

We will now present some properties of finite fields some of which will be used in the lecture.

Definition: Let R be a ring. If there is a positive integer n with $nr := \underbrace{r + r + \ldots + r}_{n} = 0$ for all $r \in R$, then we call the least such integer the *characteristic* of the ring R. If no such integer exists, then R is said to have characteristic 0.

Theorem A.9 *A finite field has prime characteristic.*

Theorem A.10 *Let F be a finite field. Then F has p^n elements for some $n \in \mathbb{N}$, where the prime p is the characteristic of F.*

Theorem A.11 *Let F be a finite field with q elements. Then for every $a \in F$ we have $a^q = a$.*

For the sake of completeness, we state the general theorem about existence and uniqueness of finite fields:

Theorem A.12 *For every prime p and every positive integer n there exists a unique (up to isomorphism) finite field with p^n elements.*

This uniqueness justifies speaking of *the* finite field with q elements which we will denote by $\mathbb{F}_q$, where it is understood that q is a power of a prime.

Theorem A.13 *Let $\mathbb{F}_q$ be a finite field with $q = p^n$ elements. Then every subfield of $\mathbb{F}_q$ has order p^m, where m is a positive divisor of n. Conversely, if m is a positive divisor of n then there is exactly one subfield of $\mathbb{F}_q$ with p^m elements.*

For the finite field $\mathbb{F}_q$ we denote by $\mathbb{F}_q^*$ the multiplicative group of nonzero elements of $\mathbb{F}_q$.

Theorem A.14 *Let $\mathbb{F}_q$ be a finite field, then the multiplicative group $\mathbb{F}_q^*$ is cyclic.*

A.1.4 Projective Planes, Spaces

Loosely speaking, a projective plane consists of a set of points and a set of lines together with an incidence relation determining whether a given point is on a given line or not.

Definition: A *projective plane* Π is a triple $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ where $\mathcal{P}$ is a set of elements called *points*, $\mathcal{L}$ is a subset of $2^{\mathcal{P}}$ called *lines*, and $\mathcal{I}$ is an *incidence relation* between points and lines such that the following holds:

(i) every pair of distinct lines is incident to a unique point,

(ii) every pair of distinct points is incident to a unique line,

(iii) there are four points such that no three of them are incident to a single line.

A projective plane is called finite, if $\mathcal{P}$ is a finite set.

One can show that there is a *duality* between points and lines (with respect to interchanging them in Axiom (iii)), from which the next result follows:

Theorem A.15 *Let Π be a finite projective plane. Then there is an integer $m \geq 2$ such that every point (line, resp.) of Π is incident to exactly $m+1$ lines (points, resp.), and Π contains exactly $m^2 + m + 1$ points (lines, resp.).*

The integer m from the previous Theorem A.15 is called *order* of the finite projective plane Π. We will be mainly interested in projective planes constructed algebraically via finite fields.

Remark: The prime power conjecture for projective planes (PPC) is a very famous conjecture saying that a projective plane of order m exists if and only if m is a prime power. It is known that there is no projective plane of order 6 and 10, where the latter has only recently be proved. However, the status of projective plane of order 12 remains open. In the following we show the easy direction of PPC, that for every prime power there is a corresponding projective plane.

For this purpose consider the finite q-element field $\mathbb{F}_q$. The projective plane $PG(q,2)$ over $\mathbb{F}_q$ is defined as follows:

- Points: The set of points $\mathcal{P}$ in $PG(q,2)$ are the equivalence classes of $\mathbb{F}_q^3 \setminus \{(0,0,0)\}$, where two triples are in relation if they are nonzero constant multiples of each other:
$$[x_0,x_1,x_2] = \left\{(cx_0,cx_1,cx_2) \in \mathbb{F}_q^3 \setminus \{(0,0,0)\} \ : \ c \in \mathbb{F}_q^*\right\} .$$
- Lines: Given a triple $(a_0,a_1,a_2) \in \mathbb{F}_q^3 \setminus \{(0,0,0)\}$ we define the line $L(a_0,a_1,a_2)$ as follows:
$$L(a_0,a_1,a_2) := \left\{[x_0,x_1,x_2] \in \mathcal{P} \ : \ a_0x_0 + a_1x_1 + a_2x_2 = 0\right\} .$$
This also serves as the definition of the incidence relation. The set of lines $\mathcal{L}$ in $PG(q,2)$ consists of all the lines $L(a_0,a_1,a_2)$ defined above.

As an exercise verify the axioms of a projective plane and prove the according statements of Theorem A.15 for $PG(q,2)$.

Analogs of the concept of projective planes can also be defined for dimensions higher than 2. However, we refrain from giving the abstract definition in favor of generalizing the example of projective planes over finite fields.

Consider the finite q-element field $\mathbb{F}_q$. Then the projective d-space over $\mathbb{F}_q$, denoted by $PG(q,d)$, is defined in the following way:

- Points: The points $\mathcal{P}$ in $PG(q,d)$ are the equivalence classes of $\mathbb{F}_q^{d+1} \setminus \{(0,\ldots,0)\}$, where two $(d+1)$-tuples are in relation if they are nonzero constant multiples of each other:
$$[x_0,\ldots,x_d] = \left\{(cx_0,\ldots,cx_d) \in \mathbb{F}_q^{d+1} \setminus \{(0,\ldots,0)\} \ : \ c \in \mathbb{F}_q^*\right\} .$$
- A k-dimensional subspace in $PG(q,d)$ is the set of all points whose coordinates satisfy $d-k$ linearly independent homogeneous linear equations
$$\begin{array}{ccccccc} a_{10}x_0 & + & \cdots & + & a_{1d}x_d & = & 0 \\ \vdots & & \ddots & & \vdots & = & \vdots \\ a_{d-k,0}x_0 & + & \cdots & + & a_{d-k,d}x_d & = & 0 \ , \end{array}$$
with coefficients $a_{ij} \in \mathbb{F}_q$.

A $(d-1)$-dimensional subspace of a projective d-space is called *hyperplane*, that is the set of points satisfying

$$a_0x_0 + \cdots + a_dx_d = 0 \ .$$

A k-dimensional subspace is therefore the intersection of $d-k$ hyperplanes. Furthermore, 1-dimensional subspaces are called *lines* and it is easy to check that the 0-dimensional subspaces are the points $\mathcal{P}$ in $PG(q,d)$.

As an exercise compute the number of points in a hyperplane and in a k-dimensional subspace of the projective d-space over $\mathbb{F}_q$.

A.2 The d-th power residues

For a prime power q let $P_{d,q} : \mathbb{F}_q^* \to \mathbb{F}_q^*$ denote the dth power function, that is $P_{d,q}(x) = x^d$.

The elements of the image

$$R_d(q) := Im\,(P_{d,q}) = \{x^d : x \in \mathbb{F}_q^*\}$$

of the map are called the *d-th power residue*s of $\mathbb{F}_q$. In the special case of $d = 2$ we use the notation $QR(q) := R_2(q)$ instead, and call the elements, i.e. those $a \in \mathbb{F}_q^*$ for which there exists an element $y \in \mathbb{F}_q^*$ with $y^2 = a$, *quadratic residue*s. If there is no such y then a is called a *quadratic non-residue*. Note that 0 is excluded from the list of quadratic residues and non-residues. The set of quadratic non-residues is denoted by $QNR(q)$.

As $P_{d,q}$ is group homomorphism, we the size of each non-empty inverse image is the same. Since $\mathbb{F}_q$ is a field, we also know the size of the inverse images.

Theorem A.16 *Let q be a prime power, $d|q-1$ a positive integer, and $\alpha \in \mathbb{F}_q^*$. Then*

- $\left|P_{d,q}^{-1}(\alpha)\right| = \begin{cases} d & \alpha \in R_d(q) \\ 0 & \alpha \notin R_d(q) \end{cases}$
- *$R_d(q) = P_{\frac{q-1}{d},q}^{-1}(1)$ and hence $|R_d(q)| = \frac{q-1}{d}$.*

Proof. We will use that $R_d(q) \subseteq P_{\frac{q-1}{d},q}^{-1}(1)$, since $(y^d)^{\frac{q-1}{d}} = 1$ for every $y \in \mathbb{F}_q^*$ by Lagrange's Theorem.

Note furthermore that the elements of $P_{d,q}^{-1}(\alpha)$ are the roots of the polynomial $x^d - \alpha$ in $\mathbb{F}_q$, which is of degree d, so

$$|P_{d,q}^{-1}(\alpha)| \leq d.$$

Using the partition $\mathbb{F}_q^* = \cup_{\alpha \in R_d(q)} P_{d,q}^{-1}(\alpha)$, we write

$$q-1 = |\mathbb{F}_q^*| = \sum_{\alpha \in R_d(q)} \left|P_{d,q}^{-1}(\alpha)\right| \leq |R_d(q)| \cdot d \leq \left|P_{\frac{q-1}{d},q}^{-1}(1)\right| \cdot d \leq \frac{q-1}{d} d = q-1.$$

Therefore all inequalities above must be equalities, implying the theorem. □

Theorem A.17 *Let q be an odd prime power, and $a \in \mathbb{F}_q^*$ then*

$$\begin{aligned} a \in QR(q) &\Leftrightarrow a^{\frac{q-1}{2}} = 1 \\ a \in QNR(q) &\Leftrightarrow a^{\frac{q-1}{2}} = -1. \end{aligned}$$

In particular

$$|QR(q)| = \frac{q-1}{2} = |QNR(q)|.$$

Proof. For the proof of Theorem A.17 recall two facts from algebra:

1: A polynomial of degree d, with coefficients from an integral domain R, can have at most d roots in R. (Proof by induction on the degree).

2: **Lagrange's Theorem:** In a finite group G, $x^{|G|} = 1$ for any $x \in G$.

On the one hand by Fact 1 the polynomial $x^{q-1} - 1 = 0$ cannot have more than $q-1$ roots in $\mathbb{F}_q$. On the other hand it does have $q-1$ roots by Lagrange's Theorem, because all the elements in $\mathbb{F}_q^*$ are roots.

Consequently, since $x^{q-1} - 1 = \left(x^{\frac{q-1}{2}} - 1\right)\left(x^{\frac{q-1}{2}} + 1\right)$ and the ring of polynomials over $\mathbb{F}_q$ is an integral domain, again Fact 1 implies that both factors $\left(x^{\frac{q-1}{2}} - 1\right)$ and $\left(x^{\frac{q-1}{2}} + 1\right)$ must have exactly $\frac{q-1}{2}$ roots.

If $a = y^2$ is a quadratic residue in $\mathbb{F}_q$ then $a^{\frac{q-1}{2}} = y^{q-1} = 1$ by Lagrange's Theorem. Hence, a is a root of $x^{\frac{q-1}{2}} - 1$ implying that $|QR(q)| \leq \frac{q-1}{2}$. On the other hand note that by Fact 1 the polynomial $x^2 - a$ has at most two roots for any quadratic residue a, hence

$$q - 1 = \left|\mathbb{F}_q^*\right| \leq \sum_{a \in QR(q)} |\{x : x^2 = a\}| \leq 2|QR(q)|.$$

Concluding, $|QR(q)| = \frac{q-1}{2}$ and thus $QR(q)$ must be equal to the set of roots of $\left(x^{\frac{q-1}{2}} - 1\right)$. Then it follows that also $|\mathrm{QNR}(q)| = \frac{q-1}{2}$, and $QNR(q)$ must be equal to the set of roots of $\left(x^{\frac{q-1}{2}} + 1\right)$. □

Corollary A.18 *The product of two quadratic residues or two non-residues is again a quadratic residue, whereas the product of a residue and a non-residue gives a non-residue.*

Corollary A.19

$$\begin{aligned} -1 \in QR(q) &\Leftrightarrow q \equiv 1 \pmod 4 \\ -1 \in QNR(q) &\Leftrightarrow q \equiv 3 \pmod 4. \end{aligned}$$

Observe that $x^2 = (-x)^2$, which, in the case of a prime-field $\mathbb{F}_p$, implies that

$$QR(p) = \left\{ y^2 : 0 < y \leq \frac{p-1}{2} \right\}.$$

A.3 Basic definitions and facts from commutative algebra

In this section we summarize the well-known definitions and theorems of commutative algebra which we need later on.

In the following every ring is commutative with an identity element (denoted by 1). An additive subgroup I of a ring A is an *ideal* if it is closed under multiplication by any element of A. An ideal $P \neq A$ is called *prime* ideal if for any $a, b \notin P$ $ab \notin P$. An ideal is *maximal* if it is not contained in any other proper ideal. Every maximal ideal is prime.

We call an element of A a *unit* if it has a multiplicative inverse in A. Every non-unit element is contained in a maximal ideal.

Two elements $a, b \in A$ are associates, if $a = ub$ for some unit $u \in A$. An element $0 \neq p \in A$ is a *prime* if the ideal pA, generated by p, is a prime ideal. An integral domain is called a *unique factorization domain* or *UFD* if every nonzero and non-unit element can be written uniquely (up to associates) as a product of primes.

An additively written commutative group M is called an $A-module$ if there is a multiplication on M by the elements of A: if $a \in A$ and $m \in M$ then $am \in M$ and all the meaningful associative and distributive laws hold. Besides, it is required that multiplication by $1 \in A$ is the identity on M. For example if I is an ideal of A then I is an A-module.

Integrality and integral closure are crucial concepts in our line of reasoning. Let B be a subring of A. (We assume that B contains the identity element of A.) An element $a \in A$ is *integral* over B, if a is a root of a monic polynomial (a polynomial with leading coefficient 1) with coefficients from B. We say that A is an *integral extension* of B, if every element of A is integral over B. The set of elements in A, which are integral over B form a ring. This ring is denoted by $\overline{B}$ and is called the *integral closure* of B in A. If $B = \overline{B}$, then we say that B is *integrally closed* in A.

A ring A is *finite* over a subring $B \subseteq A$ if A is a finitely generated B-module. (There is a finite subset X of A such that the elements of A can be obtained as B-linear combinations of elements from X.) The finiteness of A over B is equivalent to the following two conditions: (i) A is a finitely generated as an algebra over A; (ii) A is integral over B

Let $S \subseteq A$ be a multiplicatively closed subset of A, $1 \in S$. We can "make" each element in S a unit by defining the *ring of fractions* $S^{-1}A$. Let us consider an equivalence relation on the set $A \times S$: $(a, s) \sim (b, t)$ (or with other notation $a/s \sim b/t$) if and only if there exist an $r \in S$ such that $(at - bs)r = 0$. Let $S^{-1}A = (A \times S)/\sim$. $S^{-1}A$ has a natural ring structure. In the case of integral domains (rings with no zero-divisors) the definition corresponds to the usual rules of addition and multiplication of fractions. If A is an integral domain and we choose $S = A \setminus \{0\}$, then $S^{-1}A$ is called the quotient field of A and denoted by $QF(A)$.

If A is a UFD, then A is integrally closed in $QF(A)$. (The easy elementary argument which shows that $\mathbf{Z}$ is integrally closed in $\mathbf{Q}$ works in general.)

An integral domain A *has rank* r over a subring $B \subseteq A$ if $QF(A)$ is a degree-r extension of $QF(B)$. That is, $\dim_{QF(B)} QF(A) = r$.

If P is prime ideal of A, then $S = A \setminus P$ is a multiplicatively closed subset and we denote $S^{-1}A$ with A_P. We call A_P the *localization* of A at P. To explain this name we have to define the concept of *local rings*. We call a ring *local*, if it has only one maximal ideal. The localization A_P is a local ring. Its only maximal ideal is the image of P at the map $p \to p/1$, since all the other elements are units.

For any A-module M we can define $S^{-1}M$ similarly to $S^{-1}A$: considering the set $M \times S$ factorized by a similar equivalence relation. If $S = A \setminus P$, where P is a prime ideal, $S^{-1}M$ is denoted by M_P.

The *transcendence degree* of an integral domain A over a subring $B \subseteq A$ is the maximal integer m such that there exist $a_1, \ldots, a_m \in A$ without any algebraic relation over B among them. (That is, there is no nonzero polynomial $f(x_1, \ldots, x_m) \in B[x_1, \ldots, x_m]$, such that $f(a_1, \ldots, a_m) = 0$.) If there is no such m, then the transcendence degree is infinite.

We need to define valuation rings because they play an important (we could say *integral*) rôle in characterizing integral closures.

Definition: Let B be an integral domain, K its field of fractions. B is a *valuation ring* of K if, for each $x \neq 0$, either $x \in B$ or $x^{-1} \in B$ (or both).

Every valuation ring B is a local ring. The only maximal ideal consists of those elements $x \in K$ for which $x^{-1} \notin B$.

Theorem A.20 *(Corollary 5.22. of [?] or Theorem 10.4 of [?]) Let A be a subring of a field K. Then the integral closure $\overline{A}$ of A in K is the intersection of the valuation rings of K which contain A.*

Let A be a local ring and m its maximal ideal. Let M be a finitely generated A-module. M/mM is annihilated by m, hence it carries a natural A/m-module structure. Since m is a maximal ideal of A, A/m is a field. Thus, M/mM is a finite-dimensional vector space over A/m (A module over a field is a vector space). The following statement is an immediate consequence of Nakayama's Lemma.

Proposition A.21 *(Proposition 2.8. [?]) Let x_i $(1 \leq i \leq n)$ be elements of M whose images in M/mM form a basis of this vector space. Then the x_i generate M as a module over A.*

We will repeatedly use the following weak form of Hilbert's Nullstellensatz:

Theorem A.22 *Let K be algebraically closed field. Let $f_1, \ldots, f_t \in K[x_1, \ldots, x_n]$ polynomials having no common zero in K^n. Then $(f_1, \ldots, f_t) = (1)$, that is there exist $g_i \in K[x_1, \ldots, x_n]$ such that $\sum_{i=1}^{n} f_i g_i = 1$.*

For the basics of commutative algebra we refer to [?], [?], [?]; especially [?, Chap. 5].

A.4 Eigenvalues of graphs

In this section by a graph $G = (V, E)$ we understand a simple graph on $n = |V|$ vertices with at most one loop at each vertex. The adjacency matrix $A = A(G)$ of G is an $n \times n$ matrix, where the rows and columns are labeled with the vertices of G and the entry $a_{u,v} = 1$ if and only if $uv \in E$, otherwise $a_{uv} = 0$. The following special case of the Spectral Theorem is clearly relevant for adjacency matrices.

Theorem A.23 *Let M be a symmetric $n \times n$ matrix with real entries. Then all eigenvalues of M are real and there is an orthonormal basis consisting of eigenvectors M.*

Hence each graph has a multiset of n real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. If G is d-regular then d is clearly an eigenvalue with eigenvector $\mathbb{1}_{[n]} = (1, 1, \cdots, 1)$.

A.4.1 The second eigenvalue and quasirandomness

Ever since randomness was introduced in Theoretical Computer Science, great efforts have also been made for its elimination. Whenever a random graph is utilized to perform an algorithmic task efficiently, but random bits are expensive or a deterministic answer would simply be more desirable, the need for a replacement arises. This demand is one of the main motivations behind the interest in explicit constructions of families of *quasirandom graphs*. Quasirandom graphs possess certain random-like properties and can, in some cases, serve as substitutes of truly random graphs.

There are several different ways to understand and define the quasirandomness of a graph. Here we consider the one through the second eigenvalue, which is linked strongly to the graph's *edge distribution* and *expansion properties*; both crucial concepts for applications in combinatorics or computer science (see [?, Chapter 9] for more details). Given a graph G, let $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ be the eigenvalues of its adjacency matrix. The *second eigenvalue* of G is defined to be $\lambda = \lambda(G) = \max\{\lambda_2, |\lambda_n|\}$. Graphs whose second eigenvalue is smaller order than the largest one possess some properties of random graphs with appropriate edge probability. The larger this "spectral gap" is the more randomness the graph has.

Our central definition is the one of (n, d, λ)-graphs. A graph G is called (n, d, λ)-*graph* if its number of vertices is n, it is d-regular (with possibly one loop at some vertices, but no multiple edges) and its second eigenvalue is λ.

As the second eigenvalue of any graph (of maximum degree at most $n/2$) is at least the square root of the degree up to a constant factor, graphs with $\lambda(G) = \Theta(\sqrt{d})$ are of particular interest.

Exercise A.1 *Let G be an (n, d, λ)-graph with $d \leq n/2$. Show that $d \geq \lambda \geq \sqrt{d/2}$.*

Remark. Observe that here we restrict our discussion to regular graphs. There is a generalization of these statements to non-regular graphs via the Lagragian matrix of the

graph, but since that approach does not really add extra information for our purposes, we stick with the technically much less demanding regular case.

The next couple of lemmas show that (n, d, λ)-graphs with small λ behave like truly random graphs in some sense.

Lemma A.23.1 *Let $G = G(V, E)$ be an (n, d, λ)-graph and $B \subseteq V$ be an arbitrary subset of the vertices. Then*

$$\sum_{v \in V} \left(d_B(v) - |B|\frac{d}{n} \right)^2 \leq \lambda^2 |B| \left(1 - \frac{|B|}{n} \right)$$

Proof. Let A be the adjacency matrix of G (in case of a loop, there is a 1 at the diagonal.) with eigenvalues $d = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. By the above, there is an orthonormal basis of the form $\{u_1, u_2, \ldots, u_n\}$, where u_i is an eigenvector corresponding to eigenvalue λ_i and of course, $u_1 = \frac{1}{\sqrt{n}} \mathbb{1}_{[n]}$ is an eigenvector for the eigenvalue d.

The connection to combinatorics is provided by the simple facts that $d_B(v)$ is the v-entry of the vector $A\mathbb{1}_B$ and that no shift by a multiple of the leading eigenvector, $\mathbb{1}_{[n]}$, changes the "combinatorial structure" of $\mathbb{1}_B$.

The vector $\mathbb{1}_B$ can be expressed as a linear combination $\sum_{i=1}^n \mu_i u_i$ of the eigenvectors where the first coeeficient is $\mu_1 = (\mathbb{1}_B, u_1) = |B|/\sqrt{n}$. Hence the projection $u = \mathbb{1}_B - \mu_1 u_1 \in \mathbb{R}^V$ of $\mathbb{1}_B$ on the subspace generated by the lower eigenvectors is defined by

$$u_v = \begin{cases} 1 - b & \text{if } v \in B \\ -b & \text{if } v \notin B. \end{cases}$$

As $u = \sum_{i=2}^n \mu_i u_i$ we obtain that

$$(Au, Au) = \sum_{i=2}^n \lambda_i^2 \mu_i^2 \leq \lambda^2 \sum_{i=2}^n \mu_i^2 = \lambda^2 (u, u).$$

The inequality of the Lemma is just expressing this fact. Indeed, the entry of Au at vertex v is

$$\sum_{w \in V} a_{vw} u_w = (1 - b) d_B(v) - b(d - d_B(v)),$$

and the value of (u, u) is

$$\sum_{w \in V} u_w^2 = |B|(1 - b)^2 + (n - |B|)b^2.$$

□

Often we will use the following corollary of Lemma A.23.1. For two subsets $B, C \subseteq E(H)$ let

$$e(B, C) = \#\{(u, v) : u \in B, v \in C, uv \in E(H)\}.$$

Corollary A.24 *If G is an (n,d,λ)-graph, then for any two subsets $B,C \subseteq V$ of the vertices*

$$\left| e(B,C) - \frac{d}{n}|B||C| \right| \leq \lambda\sqrt{|B||C|}.$$

Proof. Since $e(B,C) - \frac{d}{n}|B||C| = \sum_{w\in C}\left(d_B(w) - \frac{d}{n}|B|\right)$, by the Cauchy-Schwarz inequality we have

$$\begin{aligned}
\left| e(B,C) - \frac{d}{n}|B||C| \right| &\leq \sum_{v\in C}\left| N_B(v) - |B|\frac{d}{n}\right| \\
&\leq \sqrt{|C|}\sqrt{\sum_{v\in C}\left(N_B(v) - |B|\frac{d}{n}\right)^2} \\
&\leq \sqrt{|C|}\sqrt{\sum_{v\in V}\left(N_B(v) - |B|\frac{d}{n}\right)^2} \\
&\leq \sqrt{|C|}\sqrt{\lambda^2|B|}
\end{aligned}$$

Here the last inequality follows from Lemma A.23.1. □

Now an upper bound on the independence number is immediate.

Corollary A.25 *Let G be an (n,d,λ)-graph. Then*

$$\alpha(G) \leq \frac{\lambda n}{d}.$$

Proof. Let $I \subseteq V$ be an independent set of maximum size. Then $|I| = \alpha(G)$ and $e(I,I) = 0$, so

$$\left| 0 - \frac{d}{n}|I|^2 \right| \leq \lambda|I|,$$

which implies the statement. □

Finally, we mention an important result concerning the case of linear degree, that is, when $d = pn$ for some constant p, $0 < p < 1$. Then many quasi-random properties turned out to be equivalent.

- Property P_1:For every$B,C \subseteq V, e(B,C) = p|B||C| + o(n^2)$
- Property P_2:For every$B \subseteq V, e(B) = p\binom{|B|}{2} + o(n^2)$
- Property P_3:$\lambda = o(n)$

  Let $N^*(H)$ denote the number of induced labeled copies of H in G.
- Property $P_4(s)$: For every graph H on s vertices $N^*(H) = n^s p^{e(H)}(1-p)^{\binom{s}{2}-e(H)}(1+o(1))$

  Let $N(H)$ denote the number of labeled copies of H in G.

- Property $P_5(s)$: For every graph H on s vertices $N(H) = n^s p^{e(H)}(1 + o(1))$
- Property P_6: $N(C_4) = n^4 p^4 (1 + o(1))$

Obviously all of these properties are satisfied by the random graph $G(n, p)$. In a seminal paper, Chung, Graham, and Wilson proved that they are equivalent for every graph (sequence)!

Theorem A.26 *Let G be a (sequence of) (n, d, λ)-graph(s) with $d = pn$, with $p, 0 < p < 1$, a constant. Then P_1, P_2, $P_3(s)$ for some $s \geq 4$, $P_4(s)$ for some $s \geq 4$, P_5, P_6.*

The most suprising that the weak-looking property P_6 about the number of C_4 implies the bound on the second eigenvalue and the number of arbitrary fixed subgraph

Observe that the theorem cannot be true in this form for $d \ll n$. The C_4-free polarity graph discussed in Subsection 2.1.3 has the best possible quasi-random second eigenvalue $\sqrt{d}$ and still contains NO C_4, while the corresponding random graph with edge-probability $p = n^{-\frac{1}{2}}$ contains $\Theta(n^4 p^4) = \Theta(n^2)$ C_4.

A.5 Cayley graphs and Characters

All what was said so far about eigenvalues applies for any d-regular graph. The graphs we construct are often defined algebraically, in which case they are often possible to cast as *Cayley graphs* and their eigenvalues are most conveniently expressed in terms of the group's *characters*.

A.5.1 Cayley graphs

Given a group H and a subset $S \subseteq H$ with the properties that $0 \notin S$ and $S = -S$ (that is, for every $a, b \in H$, $a - b \in S$ if and only if $b - a \in S$), we define the Cayley graph $G(H, S) = G$ as follows:

- $V(G) = H$
- $E(G) = \{ab : a - b \in S\}$.

Examples

1. The Cayley graph $G((\mathbb{Z}_n, +), \{1, -1\})$ is just the cycle C_n.
2. The Cayley graph $G(\mathbb{F}_q^*, QR(q))$ is the Payley graph we defined in ....

It turns out that eigenvalues of Cayley graphs are connected to the more general concept of group characters. Below we define the general notion, but soon will concentrate on abelian groups, which come up in our applications.

A.5.2 Basics of characters of Abelian groups

The following are based partly on the notes of Babai [?].

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

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

How many characters are there? Just a few? Or many? Maybe an infinite number? We show that there are exactly as many characters as group elements and their structure is really restricted: they themselves form a group isomorphic to H.

Examples. **1.** One immediate example of a character is the *principal character* χ_0, which is defined by

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

and exists for an arbitrary group H.

2. Another important example is the *quadratic residue character* ρ_q of the multiplicative group $(\mathbb{F}_q^*,\cdot)$ of a finite field:

$$\rho_q(x) = \begin{cases} 1 & \text{if } x \in QR(q) \\ -1 & \text{otherwise.} \end{cases}$$

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.

3. For the cyclic group $(\mathbb{Z}_n,+)$ an obvious choice transferring the (mod n) addition to complex multiplication is the character χ_1. For every $x \in \mathbb{Z}_n$ we define

$$\chi_1(x) = \omega^x,$$

where $\omega = e^{2\pi i \frac{x}{n}}$.

The fact that the quadratic residue character has only values 1 and -1 and the values of χ_1 are also roots of unity 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.*
- $\chi(-a) = \chi(a)^{-1} = \overline{\chi(a)}$

All the nth roots of unity, i.e., the values of the character χ_1, sum up to 0. This is again not a coincidence: the values of *any* non-principal character sum up to 0.

Proposition A.27 *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

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

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.28 *$\hat{H}$ 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.29 *$H \cong \hat{H}$.*

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

Proposition A.30 *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.31 *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.29 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.32 *(First orthogonality relation) For any $\chi, \psi \in \hat{H}$,*

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

Proof. Exercise. □

Corollary A.33 *$\hat{H}$ forms an orthonormal basis in $\mathbb{C}^H$.*

Corollary A.34 *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. □

Bounding the deviation from the average

For us, the main application of the discrete Fourier transform is the estimation of the deviation of a function from its average, via the non-principal Fourier coefficients.

Proposition A.35 *Let $f : H \to \mathbb{C}$ be an arbitrary function on an abelian group H. Then for every $a \in H$ we have*

$$\left| f(a) - \frac{1}{|H|} \sum_{x \in H} f(x) \right| \leq \Phi(f)|H|,$$

where $\Phi(f) = \max\{|\langle \chi, f \rangle| : \chi \in \widehat{H}, \chi \neq \chi_0\}$.

Proof. Conveniently, the Fourier coefficient of f corresponding to the principal character is equal to the average value of f. Indeed,

$$\langle \chi_0, f \rangle = \frac{1}{|H|} \sum_{x \in H} \overline{\chi}_0(x) f(x) = \frac{1}{|H|} \sum_{x \in H} f(x).$$

Then, writing f in the Fourier basis of characters, we have

$$\begin{aligned} \left| f(a) - \frac{1}{|H|} \sum_{x \in H} f(x) \right| &= \left| \sum_{\chi \in \widehat{H}} \langle \chi, f \rangle \chi(x) - \frac{1}{|H|} \sum_{x \in H} f(x) \right| = \left| \sum_{\substack{\chi \in \widehat{H} \\ \chi \neq \chi_0}} \langle \overline{\chi}, f \rangle \chi(a) \right| \\ &\leq \sum_{\substack{\chi \in \widehat{H} \\ \chi \neq \chi_0}} |\langle \overline{\chi}, f \rangle| \cdot |\chi(a)| \leq \Phi(f)|H|. \end{aligned}$$

□

Remark: The Fourier coefficients of a function $f : H \to \mathbb{C}$ naturally define a function $\hat{f}$ on the character group $\widehat{H}$. For every $\chi \in \widehat{H}$ we set

$$\hat{f}(\chi) := |H| \langle \overline{\chi}, f \rangle = \sum_{x \in H} \chi(x) f(x).$$

The function $\hat{f} : \hat{H} \to \mathbb{C}$ is called the (discrete) *Fourier transform* of f. The formula

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

obtained by writing f in the Fourier basis is usually called the *Inverse Fourier Transform*. Since our applications of the discrete Fourier transform do not really go beyond the basics, we prefer avoiding the us of the notation $\hat{f}$ in our proofs.

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 6

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

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.36 *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.37 *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.36 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.36 and Corollary A.37 in the previous section.

Proposition A.38 *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.39 (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*

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

Note that Proposition A.27 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.40 *(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.27 over and over again; the first three cases being quite straightforward. To appply Proposition A.27 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.27 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.27; this time employed for the multiplicative character ψ.

□