BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 43, Number 4, October 2006, Pages 439–561
S 0273-0979(06)01126-8
Article electronically published on August 7, 2006

EXPANDER GRAPHS AND THEIR APPLICATIONS

SHLOMO HOORY, NATHAN LINIAL, AND AVI WIGDERSON

An Overview

A major consideration we had in writing this survey was to make it accessible to mathematicians as well as to computer scientists, since **expander graphs**, the protagonists of our story, come up in numerous and often surprising contexts in both fields.

But, perhaps, we should start with a few words about graphs in general. They are, of course, one of the prime objects of study in Discrete Mathematics. However, graphs are among the most ubiquitous models of both natural and human-made structures. In the natural and social sciences they model relations among species, societies, companies, etc. In computer science, they represent networks of communication, data organization, computational devices as well as the flow of computation, and more. In mathematics, Cayley graphs are useful in Group Theory. Graphs carry a natural metric and are therefore useful in Geometry, and though they are "just" one-dimensional complexes, they are useful in certain parts of Topology, e.g. Knot Theory. In statistical physics, graphs can represent local connections between interacting parts of a system, as well as the dynamics of a physical process on such systems.

The study of these models calls, then, for the comprehension of the significant structural properties of the relevant graphs. But are there nontrivial structural properties which are **universally** important? Expansion of a graph requires that it is simultaneously sparse and highly connected. Expander graphs were first defined by Bassalygo and Pinsker, and their existence first proved by Pinsker in the early '70s. The property of being an **expander** seems significant in many of these mathematical, computational and physical contexts. It is not surprising that expanders are useful in the design and analysis of communication networks. What is less obvious is that expanders have surprising utility in other computational settings such as in the theory of error correcting codes and the theory of pseudorandomness. In mathematics, we will encounter e.g. their role in the study of metric embeddings, and in particular in work around the Baum-Connes Conjecture. Expansion is closely related to the convergence rates of Markov Chains, and so they play a key role in the study of Monte-Carlo algorithms in statistical mechanics and in a host of practical computational applications. The list of such interesting and fruitful connections goes on and on with so many applications we will not even

Received by the editors April 28, 2006, and, in revised form, May 10, 2006.
2000 *Mathematics Subject Classification.* Primary 01-01, 68-01, 05-01, 68Q01, 94-01; Secondary 68Q15, 68Q17, 94B05, 05C25, 05C35, 05C40, 05C50, 05C80, 05C90, 60J10, 35J99, 20F05, 20F69, 20C99.
Work supported in part by grants from the Israel Science Foundation and the Israel-U.S. Binational Fund.

be able to mention. This universality of expanders is becoming more evident as more connections are discovered. It transpires that expansion is a fundamental mathematical concept, well deserving to be thoroughly investigated on its own.

In hindsight, one reason that expanders are so ubiquitous is that their very definition can be given in at least three languages: combinatorial/geometric, probabilistic and algebraic. Combinatorially, expanders are graphs which are highly connected; to disconnect a large part of the graph, one has to sever many edges. Equivalently, using the geometric notion of isoperimetry, every set of vertices has a (relatively) very large boundary. From the probabilistic viewpoint, one considers the natural random walk on a graph, in which we have a token on a vertex, that moves at every step to a random neighboring vertex, chosen uniformly and independently. Expanders are graphs for which this process converges to its limiting distribution as rapidly as possible. Algebraically, one can consider the Laplace operator on the graph and its spectrum. From this perspective, expanders are graphs in which the first positive eigenvalue (of their Laplace operator) is bounded away from zero.

The study of expanders leads in different directions. There are structural problems: what are the best bounds on the various expansion parameters, and how do they relate to each other and to other graph invariants? There are problems concerning explicit constructions: how to efficiently generate expanders with given parameters. These are extremely important for applications. There are algorithmic problems - given a graph, test if it is an expander with given parameters. Finally, there is the problem of understanding the relation of expansion with other mathematical notions, and the application of expanders to practical and theoretical problems.

In the past four decades, a great amount of research has been done on these topics, resulting in a wide-ranging body of knowledge. In this survey, we could not hope to cover even a fraction of it. We have tried to make the presentation as broad as possible, touching on the various research directions mentioned above. Even what we do cover is of course incomplete, and we try to give the relevant references for more comprehensive coverage. We have also tried to mention in each section related research which we are not covering at all and to reference some of this as well.

The selection of material naturally reflects our interests and biases. It is rather diverse and can be read in different orders, according to the reader's taste and interests.

General background material on the computer science side includes the books on Computational Complexity (specifically, complexity classes) [Pap94, Sip97], on Algorithms [CLRS01] and on Randomized Algorithms [MR95], and the survey on the P versus NP problem [Wig06].

This article evolved from lecture notes for a course on expanders taught at the Hebrew University, Israel, in 2003 by Nati Linial and Avi Wigderson. We are greatly indebted to the scribes of the course notes: Ran Gilad-Bachrach, Danny Harnik, Boaz Barak, Udi Wieder, Eran Ofek, Erez Waisbard, Yael Vinner-Dekel, Yishai Beeri, David Statter, Eyal Bigman, Tamir Hazan, Elon Portugaly, Ariel Elbaz, Yuval Filmus, Michal Igell, Eyal Rozenman, Danny Gutfreund, and Yonatan Bilu. Also, we acknowledge that the proof that Margulis construction is an expander is taken (with slight changes) from course notes of Ravi Boppana, with Mukesh Dalal as scribe.

We are also grateful for the careful reading of this manuscript by Mark Goresky, Eyal Rozenman and Dave Xiao. Their many constructive comments significantly improved its presentation. Special thanks to Eyal Rozenman for his help with writing the section on Cayley graphs.

Contents

1. The magical mystery tour

We begin our discussion with three fundamental problems from three different domains. At first sight these problems seem to have very little to do with expander graphs, or even graph theory, but as we shall see, they can all be solved using expander graphs.

1.1. **Three motivating problems.**

1.1.1. *Hardness results for linear transformations.* The **P** *vs.* **NP** problem is arguably the most important open problem in theoretical computer science. Despite its great significance and despite intensive research efforts, very little progress has been made. But interesting aspects of computational complexity can be investigated in other, more restricted contexts. For example, we may consider evaluating polynomials over a field using only the field's arithmetic, or even evaluating linear transformations using only addition and multiplication by scalars from the field. Valiant [Val76] considered the following natural problem:

Problem 1.1. Let A be an $n \times n$ matrix over the field[1] $\mathcal{F}$. What is the least number of gates in a circuit that computes the linear transformation $x \mapsto Ax$? Each gate is specified by two field elements a and b. Such a gate receives two inputs x and y and outputs $ax + by$.

Aside from its profound theoretical importance, certain instances of this question have far-reaching technological significance. Consider the matrix $a_{r,s} = \omega^{rs}$ $(n-1 \geq r, s \geq 0)$, where ω is a primitive n-th root of unity. The transformation $x \mapsto Ax$ is the Discrete Fourier Transform, which is fundamental to many modern technologies involving signal processing, machine learning, etc. As observed by Cooley and Tukey [CT65], there is a circuit realizing this linear transformation (the so-called Fast Fourier Transform (FFT)) with only $O(n \log n)$ gates. Therefore the least number of gates in such a circuit is between $O(n \log n)$ and n (which are required just to input the vector x). This may seem like a small gap in our knowledge, but it is rather significant. The technological implications of a Very Fast Fourier Transform, i.e. an $O(n)$-sized circuit that computes the transform (should one exist), are hard to overestimate. On the other hand, it would be a great theoretical breakthrough to establish a matching lower bound of $\Omega(n \log n)$, or even rule out the existence of such a circuit with only $O(n)$ gates.

For every field $\mathcal{F}$, it is fairly easy to show that for most $n \times n$ matrices A, every circuit realizing A must have $\Omega(n^2/\log n)$ gates. This is based on a counting argument that compares the number of circuits with a given number of gates and the number of $n \times n$ matrices over the field. As is often the case in computational complexity, despite this abundance of computationally hard functions, we are unable to exhibit any **specific, explicit** linear transformation A that requires asymptotically more then $O(n)$ gates. In an attempt to solve this problem, Valiant [Val76] conjectured that *super regular* transformations are "hard" in this sense.

Definition 1.2 (Super Regular Matrix)**.** A matrix A is *super regular* if every square sub-matrix of A has full rank.

Valiant considered the graph layout of a circuit which realizes the linear transformation corresponding to a super regular matrix. His main observation was that this graph must be a *super concentrator*:

Definition 1.3 (Super Concentrator)**.** Let $G = (V, E)$ be a graph and let I and O be two subsets of V with n vertices, each called the input and output sets respectively. We say that G is a *super concentrator* if for every k and every $S \subseteq I$ and $T \subseteq O$ with $|S| = |T| = k$, there exist k vertex disjoint paths in G from S to T.

[1] This problem is interesting for finite as well as for infinite fields.

It is a simple exercise to show that indeed the underlying graph of any circuit for a super regular matrix is a super concentrator (with inputs and outputs retaining their meaning in both). Valiant conjectured that any super concentrator must have $\gg n$ edges. That would have implied that any circuit which computes a super regular matrix must have $\gg n$ gates. However, Valiant himself disproved the conjecture and presented super concentrators with $O(n)$ edges. As you may have guessed, this is where expanders come into the picture.

We note that this construction can actually be used to give a super regular matrix that has a linear sized circuit, which seems to put this approach to rest. This is not quite so, and Valiant's ideas were later realized, as follows: If we consider circuits with more than two inputs per gate but where the circuit's depth is restricted, then superlinear lower bounds for the number of edges in depth-limited super concentrators were proven [DDPW83]. Subsequently the desired superlinear lower bounds for computing the associated linear transformations in bounded-depth circuit model were derived [Lok01, RS03].

Even though this approach did not yield strong lower bounds on circuit sizes, these attempts have brought forward the importance of sparse super concentrators in network theory and other areas. Valiant's idea has eventually had a major impact on the field.

We now skip to a totally different problem.

1.1.2. *Construction of good error correcting codes.* One of the most fundamental problems in communication is noise. Suppose that Alice has a message of k bits which she would like to deliver to Bob over some (noisy) communication channel. The problem is that noise in the channel may corrupt the message so that Bob receives a message that differs from the one sent by Alice.

In his ground-breaking paper "A Mathematical Theory of Communication" [Sha48], Claude Shannon laid the foundations for Information Theory and the mathematical theory of communication. The problem of communicating over a noisy channel (which in the form below was suggested by Hamming [H50]) occupies a central part of this theory.

Problem 1.4 (communication over noisy channel)**.** Alice and Bob communicate over a noisy channel. A fraction p of the bits sent through the channel may be altered. What is the smallest number of bits that Alice can send, assuming she wants to communicate an arbitrary k-bit message, so that Bob should be able to unambiguously recover the original message?

To solve this problem, Shannon suggested creating a dictionary (or code) $C \subseteq \{0,1\}^n$ of size $|C| = 2^k$ and using a bijective mapping ("an encoding") $\varphi : \{0,1\}^k \to C$. To send a message $x \in \{0,1\}^k$, Alice transmits the n-bit encoded message $\varphi(x) \in C$. It is assumed that Bob receives a string $y \in \{0,1\}^n$ that is a corrupted version of the message actually sent $\varphi(x) \in C$. Bob finds the codeword $z \in C$ that is closest to y (the metric used is the Hamming distance: $d_H(u,v)$ is the number of coordinates i where $u_i \neq v_i$). He concludes that the message actually sent was $\varphi^{-1}(z)$. If the distance between every two words in C is greater than $2pn$, it is guaranteed that indeed $z = \varphi(x)$, and Bob correctly infers the message sent by Alice.

The problem of communicating over a noisy channel is thus reduced to the problem of finding a good dictionary: namely, a set C of n-bit strings of largest possible cardinality subject to the condition that every two strings in C are at a large Hamming distance.

Definition 1.5 (the rate and distance of a dictionary)**.** Let $C \subseteq \{0,1\}^n$ be a dictionary. Its **rate** and (normalized) **distance** are defined by:

$$\begin{aligned} R &= \frac{\log |C|}{n}, \\ \delta &= \frac{\min_{c_1 \neq c_2 \in C} d_H(c_1, c_2)}{n}. \end{aligned}$$

As we saw before, the distance of a dictionary controls its power to overcome noise. A code's rate measures its efficiency in channel utilization. At this point we can refine the problem and ask:

Problem 1.6 (refined communication problem)**.** Is it possible to design arbitrarily large dictionaries $\{C_k\}$ of size $|C_k| = 2^k$, with $R(C_k) \geq R_0$ and $\delta(C_k) \geq \delta_0$ for some absolute constants $R_0, \delta_0 > 0$? Moreover, can we make these codes explicit and efficiently encodable and decodable?

This problem and its relatives (optimizing the code's parameters, and the algorithms' efficiency, in this and other error models and communication settings) is the subject of Coding Theory, a rich and active field initiated by Shannon's work (see e.g. [MS77a, MS77b] and [vL99] for the general theory and Sudan's notes [Sud00] for complexity theoretic aspects of the field). It took over 20 years of research until even the basic Problem 1.6 was resolved, but below we present a simple solution to this problem using expander graphs. However, before we do that, let us present our third motivating problem.

1.1.3. *Deterministic error amplification for RP.* The field of probabilistic algorithms burst into existence within Theoretical Computer Science, with the fast primality tests of Rabin [Rab80] and of Solovay and Strassen [SS77]. Given a k-bit integer x and a string r of k random bits, these algorithms efficiently compute a boolean valued function $f(x,r)$ with the following property. If x is prime, then $f(x,r) = 1$ for all choices of r. Otherwise, if x is composite, $f(x,r) = 1$ with probability smaller than $1/2$ over a randomly chosen r. If $f = 1$, the algorithm declares x a prime, and otherwise declares it to be composite. It never fails on primes, and for *every* composite x its probability of failure is at most $1/2$.

The error bound $1/2$ may not be very satisfactory, and one would like to reduce it to some desired level. A very simple way to reduce our failure probability is to apply the same algorithm repeatedly with new randomly chosen r's. Repeating it (say) d times will reduce the probability of error to below 2^{-d}. On the other hand, the running time and the number of random bits used increase by a factor of d. Is there a way to reduce the error "deterministically" without using more random bits, or at least using less than the obvious procedure above? We will see several answers to this question in these notes, and this section contains an initial advance on the problem. The importance of minimizing the number of random bits may not be evident, but we can assure the reader that it is a basic theoretical problem and, moreover, that getting your hands on good random bits is a nontrivial practical problem.

The above-mentioned primality testing algorithms belong to the class **RP** of Randomized Polynomial-Time algorithms. It is in this general setting that we discuss our problem. Let $\{0,1\}^*$ denote the set of all finite binary strings. Then a language $\mathcal{L} \subseteq \{0,1\}^*$ is in the class **RP** if there exists a randomized algorithm A with a polynomial (in $|x|$) running time such that if $x \in \mathcal{L}$, then $A(x,r) = 1$ (with certainty), whereas if $x \notin \mathcal{L}$, the probability of $A(x,r) = 1$ is smaller than $1/16$. (The definition remains unchanged with any constant < 1 that we choose. The constant $1/16$ was chosen for notational convenience.) Note again that r is a uniformly chosen random string of k bits, with k polynomially dependent on the length $|x|$ of the input x. In this case we say that $\mathcal{L} \subseteq \{0,1\}^*$ has a (1-sided error) randomized polynomial time membership algorithm.

Problem 1.7 (Saving Random Bits)**.** Assume that $\mathcal{L} \subseteq \{0,1\}^*$ has a (1-sided error) randomized polynomial time membership algorithm. How many random bits are needed in order to reduce the probability of error to be $\leq \epsilon$? (Note that we seek a bound that should apply to **every** input.)

1.2. **Magical graphs.** In the previous section we presented three seemingly unrelated problems. We now introduce a new object: a "Magical Graph" that will enable us to solve all these problems. This object exhibits an "expansion" property (a "combinatorial isoperimetric inequality") to fit our three applications.

Definition 1.8 (Magical Graph)**.** Let $G = (L,R,E)$ be a bipartite graph. The vertex set consists of L and R, two disjoint subsets, henceforth the left and right vertex sets. We say that G is an $(n,m;d)$-**magical graph** if $|L| = n$, $|R| = m$, and every left vertex has d neighbors and the following two properties hold (where $\Gamma(S)$ denotes the set of neighbors of a set S in G):

(1) $|\Gamma(S)| \geq \frac{5d}{8} \cdot |S|$ for every $S \subseteq L$ with $|S| \leq \frac{n}{10d}$.
(2) $|\Gamma(S)| \geq |S|$ for every $S \subseteq L$ with $\frac{n}{10d} < |S| \leq \frac{n}{2}$.

As observed by Pinsker [Pin73] (for other but related expansion properties), such graphs exist. The proof is by a probabilistic argument and it implies that, in fact, most graphs are magical.

Lemma 1.9. *There exists a constant n_0 such that for every $d \geq 32$ and $n \geq n_0, m \geq 3n/4$ there exists an $(n,m;d)$-magical graph.*

Proof. Let G be a random bipartite graph with n vertices on the left and m vertices on the right, where each left vertex connects to a randomly chosen set of d vertices on the right. We claim that with high probability G is a magical graph. We start by proving that the first property holds with high probability.

Let $S \subseteq L$ have cardinality $s = |S| \leq \frac{n}{10d}$, and let $T \subseteq R$ have cardinality $t = |T| < \frac{5ds}{8}$. Let $X_{S,T}$ be an indicator random variable for the event that all the edges from S go to T. It is clear that if $\sum X_{S,T} = 0$, where the sum is over all choices of S and T as above, then the first property holds. The probability of the event $X_{S,T}$ is $(t/m)^{sd}$, and therefore using a union bound and the inequality

$\binom{n}{k} \le (ne/k)^k$, we get that:

$$\begin{aligned}\Pr[\sum_{S,T} X_{S,T} > 0] &\le \sum_{S,T} \Pr[X_{S,T} = 1] \;=\; \sum_{S,T} (t/m)^{sd} \\ &\le \sum_{s=1}^{n/10d} \binom{n}{s}\binom{m}{5ds/8}\left(\frac{5ds}{8m}\right)^{sd} \\ &\le \sum_{s=1}^{n/10d} \left(\frac{ne}{s}\right)^s \left(\frac{8me}{5ds}\right)^{5ds/8} \cdot \left(\frac{5ds}{8m}\right)^{sd} < 1/10.\end{aligned}$$

The last inequality follows since the s-th term is bounded by 20^{-s}.

Similarly, we bound the probability of violating the second property by an analogous expression, which is simpler to bound. For every $S \subset L$ with cardinality $\frac{n}{10d} < s = |S| \le \frac{n}{2}$, and $T \subset R$ with $t = |T| < |S|$, let $Y_{S,T}$ be an indicator random variable for the event that all the edges from S go to T. As before, we would like to prove that the probability of the event $\sum Y_{S,T} = 0$ is small:

$$\begin{aligned}\Pr[\sum_{S,T} Y_{S,T} > 0] &\le \sum_{S,T} \Pr[Y_{S,T} = 1] = \sum_{S,T} (t/n)^{sd} \le \sum_{s=n/10d}^{n/2} \binom{n}{s}\binom{m}{s}(s/m)^{sd} \\ &\le \sum_{s=1}^{n/2} \left[(ne/s)\cdot(me/s)\cdot(s/m)^d\right]^s < 1/10.\end{aligned}$$

As before, the last inequality follows by noting that for all s the quantity in square brackets is bounded by 10^{-4}. Therefore, most graphs are $(n, m; d)$-magical graphs. □

We now turn to the solution of the three problems presented above. Note that Lemma 1.9 is existential, whereas we need explicit constructions of magical graphs to resolve our three problems. The issue of explicit constructions is an important aspect of this field and of this article, but at present we show how to solve these problems using the existence of magic graphs as a "black box".

1.3. The three solutions.

1.3.1. *A super concentrator with $O(n)$ edges.* We will see how magical graphs allow us to construct super concentrators. These graphs exhibit incredibly high connectivity despite the fact that they have only $O(n)$ edges. There is a long and still ongoing search for super concentrators with n input and output vertices and Kn edges with K as small as possible. This "sport" has motivated quite a few important advances in this area. The current "world record" holders are Alon and Capalbo [AC04].

If G is an $(n, 3n/4; d)$-magical graph, then $|\Gamma(S)| \ge |S|$ for every $S \subset L$ with $|S| \le \frac{n}{2}$. By Hall's marriage theorem (e.g., [Die97, Theorem 2.1.2]), for every $S \subseteq L$ of size $|S| \le \frac{n}{2}$ there is a perfect matching from S to $\Gamma(S)$.

We use this fact to recursively construct a super concentrator C' with n vertices on each side. For n below n_0, simply observe that a complete bipartite graph is a super concentrator with n^2 edges.

For $n \ge n_0$ we construct a super concentrator C' with n inputs and outputs, using three building blocks: (i) Two copies $G_1 = (L_1, R_1, E_1)$ and $G_2 = (L_2, R_2, E_2)$

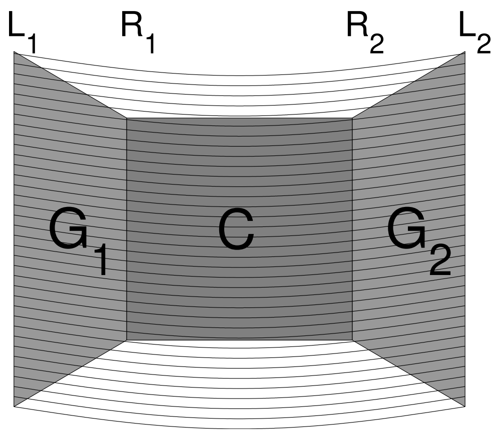


FIGURE 1. Constructing a super concentrator.

of our magical graph, where $|L_i| = n$, and $|R_i| = 3n/4$. (ii) A super concentrator C connecting the input set R_1 to the output set R_2. The input, output sets have size $3n/4$ and therefore C exists by induction. (iii) A perfect matching between L_1 and L_2. The input and output sets of our graph are L_1 and L_2 respectively. This is illustrated in Figure 1.

We need to prove that the graph we have constructed, C', is indeed a super concentrator and derive an upper bound on the number of its edges. Let S be a set of input vertices and T a set of output vertices such that $|S| = |T| = k$.

If $k \leq n/2$, then $|\Gamma_{G_1}(S)| \geq |S|$ and $|\Gamma_{G_2}(T)| \geq |T|$, since G_1, G_2 are magical graphs. Hence, by Hall's theorem there exists a perfect matching from S to $\Gamma_{G_1}(S)$ and from T to $\Gamma_{G_2}(T)$. Let $S' \subseteq \Gamma_{G_1}(S)$ be the set of vertices matched to vertices in S and likewise for T' and T. Since C is a super concentrator, the sets S' and T' can be connected by k disjoint paths. Consequently, S and T can be connected by disjoint paths in C'.

If the two sets S and T are large, i.e. $|S| = |T| = k > n/2$, then there must exist at least $k - n/2$ vertices in S that are matched to vertices in T by direct matching edges of (iii) above. Therefore we can delete the matched vertices from S and T and reduce the problem to the previous case of $k \leq n/2$. It follows that C' is a super concentrator.

We still need to provide an upper bound on the number of edges $e(n)$ in our n-inputs graph C'. We obtain the following recursion:

$$e(n) \leq \begin{cases} 2nd + n + e\,(3n/4) & \text{for } n > n_0 \\ n^2 & \text{for } n \leq n_0 \end{cases}.$$

Solving this recursion yields $e(n) \leq Kn$, where K is a constant that depends only on n_0 and d. Therefore we obtained a super concentrator with $O(n)$ edges as required.

A word about algorithms to construct such graphs: Suppose that we have an algorithm which constructs magical graphs of left size n in time $t(n)$. It should be clear that the above recursive construction yields an algorithm that constructs a super concentrator with input/output size n in time $O(t(n))$.

Finally, we note that super concentrators are but one example among a host of network construction problems in which expanders serve as a key building block.

These include the celebrated AKS sorting network [AKS83], and the variety of communication and computation networks which appear in [WZ93] and its extensive list of references.

1.3.2. *Construction of good error correcting codes.* We now turn to Shannon's problem concerning communicating over a noisy channel and present a solution due to Sipser and Spielman [SS96]. We observe a simple but useful property of magical graphs. Let G be such a graph with n left vertices and $3n/4$ right vertices. We show that for every nonempty $S \subset L$ with $s = |S| \leq \frac{n}{10d}$ there exists a vertex $u \in R$ with exactly one neighbor in S, namely, $|\Gamma(u) \cap S| = 1$. To see this, consider $e(S, \Gamma(S))$, the number of edges between S and $\Gamma(S)$. Clearly, $e(S, \Gamma(S)) = d \cdot |S| = ds$. On the other hand, since $\Gamma(S) \geq 5ds/8$, the average number of neighbors in S that a vertex in $\Gamma(S)$ has is at most $8/5 < 2$. But every vertex in $\Gamma(S)$ has at least one neighbor in S, so there must be some (indeed, many) vertices in $\Gamma(S)$ with exactly one neighbor in S.

We use the magical graph G to construct a code $C \subset \{0,1\}^n$ with rate at least $1/4$ and distance at least $1/10d$. To this end, represent the magical graph $G = (R, L, E)$ by a matrix A with row set R and column set L, where a_{ij} equals 1 or 0 depending on whether or not the i-th vertex in R is adjacent to the j-th vertex in L. The code is defined as the right kernel of A, viz. $C = \{x \in \{0,1\}^n \mid Ax = 0\}$. (Here calculations are done over the field with two elements.) Clearly C is a linear subspace of $\{0,1\}^n$ of dimension $\geq n/4$. Hence $|C| \geq 2^{n/4}$, yielding the claimed lower bound on the code's rate.

To prove a lower bound on the distance, first observe that since C is a linear code (i.e. a linear subspace of $\{0,1\}^n$) the smallest distance between two of its codewords equals the smallest weight of a nonzero codeword. Let $x \neq 0$ be an n-bit vector with support $S = \{j \in L : x_j = 1\}$. If $|S| < \frac{n}{10d}$, then, as we saw, there is some $i \in R$ with $|\Gamma(i) \cap S| = 1$. It follows that the i-th coordinate in Ax is 1, and so x is not in the right kernel of A and cannot be a codeword. It follows that the normalized distance of C is at least $1/10d$.

The above construction is a special case of a so-called LDPC (for Low Density Parity Check) code. This idea was first suggested by Gallager [Gal63] and has inspired (among many others) the works by Bassalygo, Pinsker and Margulis [Pin73, BP73, Mar73], the first to explicitly define expander graphs and construct them. After being nearly dormant for about 20 years, LDPC codes regained prominence in the 90's and are now believed to give simultaneously the best coding parameters as well as best algorithmic performance in various settings. For a survey of this fascinating field, see Richardson and Urbanke [RU].

Only fairly recently [CRVW02] did the art of explicit constructions of expanding graphs reach the level that makes the above simple argument feasible. It should also be mentioned that this construction not only yields codes with linear distance but also linear time iterative decoding. We will review these "lossless expanders" in Section 10.

As in the previous application, the time complexity of constructing the magical graph dominates the time to construct the (parity check matrix of the) appropriate code. This is yet another reason to seek efficient algorithms to construct these graphs. The next application calls for an even more concise and efficient description of these graphs.

1.3.3. *Deterministic error amplification for RP.* Our last problem revolves around deciding membership in a language $\mathcal{L} \in \mathbf{RP}$, with a given bound on the algorithm's error probability. The solution we present is due to Karp, Pippenger, and Sipser [KPS85]. It carries out **dependent** sampling of random strings using magical graphs.

As we explained above, we have to decide whether a given k-bit string x belongs to $\mathcal{L}$ or not. By assumption, there is a polytime algorithm that upon receiving x and a random k-bit string r calculates a function $f(x,r)$ such that $f(x,r) = 1$ whenever $x \in \mathcal{L}$, but $f(x,r) = 1$ with probability at most $1/16$ (the probability is over the choice of r) when $x \notin \mathcal{L}$.

To reduce the probability of error we will be considering several strings r. However, our goal is to reduce the failure probability below some set threshold while we utilize as few such strings r as possible. In other words, fix some $x \notin \mathcal{L}$ and let $B = \{r \in \{0,1\}^k \mid f(x,r) = 1\}$ be the set of strings r that are "bad" in that they fail on input x. We would like to make it as likely as possible that at least one of the r's we consider lies outside of B. The only information we have about the set $B \subseteq \{0,1\}^k$ is that it is not too big, $|B| \leq n/16$ where $n = 2^k$.

For any given integer d, we offer an algorithm for the membership problem that evaluates f only d times and fails with probability $\epsilon \leq \frac{1}{10d}$. The algorithm is rather simple. Fix an $(n,n;d)$-magical graph $G = (L,R,E)$ with $n = 2^k$, where each vertex in R and each vertex in L is associated with a unique k-bit string. To decide whether a given x is in $\mathcal{L}$, sample a k-bit string r which may be considered as a vertex in L. Let $r_1, \ldots, r_d \in R$ be the (strings associated with) the d neighbors of r. The algorithm outputs 1 iff $f(x,r_1) = f(x,r_2) = \ldots = f(x,r_d) = 1$.

Clearly this algorithm fails iff $x \notin \mathcal{L}$ and $r_1, \ldots, r_d \in B$, i.e. $\Gamma(r) \subseteq B$. Let $S \subset L$ be the set of left vertices that satisfy this condition (so we fail iff $r \in S$). Clearly $\Gamma(S) \subseteq B$. But we must have $|S| \leq \frac{n}{10d}$ or else we get a contradiction: $|B| \geq |\Gamma(S)| > (5d/8)(n/10d) \geq n/16$ (this is the moment of magic here...). This upper bound on $|S|$ means that we fail with probability at most $\frac{1}{10d}$ while using only the original k random bits. We can reduce the probability of error arbitrarily by increasing d appropriately. A key point is that we have reached this reduction in error probability *without using any additional random bits.*

Here are a few comments on this algorithm.

Unlike the previous two examples, the size n of the graph used is **exponential** in the natural size of the problem considered (the parameter k here). This means that for an efficient implementation of the new algorithm, our encoding of the magical graph must be much more efficient than in the previous applications. Specifically, given the name of a vertex (a k-bit string), we must be able to generate its d neighbors in time poly(k), which is far **smaller** than the size of the graph. We will later see that even this level of "explicitness" is achievable.

Next, with the d (dependent) samples used here, we can reduce the error to $O(1/d)$. This is much inferior to the exponential decay of the error as a function of d when we "waste" random bits and make d independent samples. We will later see that (other) dependent sampling via expanders (which uses only a few more random bits than the solution above) can achieve such an exponential decay as well.

Another comment concerns the 1-sided errors. Many probabilistic algorithms err both on inputs in the language $\mathcal{L}$ and those outside it, and the above amplification does not work as stated. However, we will later see that an appropriate modification of dependent sampling via expanders can achieve nearly optimal error reduction in such situations as well.

These problems and results have developed into a whole subfield of Theoretical Computer Science called *Randomness Extraction*. Two excellent surveys of these issues are [Gol97] and [Sha04].

2. Graph expansion and eigenvalues

2.1. Edge expansion and a combinatorial definition of expanders.

Let us introduce some conventions now. Unless we say otherwise, a graph $G = (V, E)$ is undirected and **d-regular** (all vertices have the same degree d; that is each vertex is incident to exactly d edges). Self loops and multiple edges are allowed. The number of vertices $|V|$ is denoted by n. Unlike the previous section, graphs need not be bipartite. For $S, T \subset V$, denote the set of edges from S to T by $E(S,T) = \{(u,v)|u \in S, v \in T, (u,v) \in E\}$. Here we think of every undirected edge as a pair of directed edges, so $E(S,T)$ is a set of **directed** edges. It will also be convenient to define $E(S)$ as the set of edges for which both vertices belong to S.

Definition 2.1.

(1) The **Edge Boundary** of a set S, denoted ∂S, is $\partial S = E(S, \overline{S})$. This is the set of edges emanating from the set S to its complement.

(2) The (edge) **Expansion Ratio** of G, denoted $h(G)$, is defined as:

$$h(G) = \min_{\{S \mid |S| \leq \frac{n}{2}\}} \frac{|\partial S|}{|S|}.$$

There are two important avenues for extending this definition. The first is in considering different notions of boundary. The most notable is vertex expansion, where we count the number of neighboring vertices of vertex sets S rather than the number of outgoing edges. See Sections 4 and 10 for more on this. The second avenue,proceeds to explore expansion as a function of the set size. See subsection 4.6.

Definition 2.2. A sequence of d-regular graphs $\{G_i\}_{i \in \mathbb{N}}$ of size increasing with i is a **Family of Expander Graphs** if there exists $\epsilon > 0$ such that $h(G_i) \geq \epsilon$ for all i.

Issues concerning the **explicit construction** of mathematical objects are fundamental to all of computer science, and expander graphs are no exception. There are two natural levels of efficiency to be considered in the construction of such graphs, which we have already seen in the examples of the previous section. In the first we require that an n-vertex graph should be generated "from scratch" in time polynomial in n. In the stronger version we demand that the neighborhood of any given vertex should be computable in time that is polynomial in the description length of the vertex (usually polynomial in $\log n$).

The technicalities of these definitions may seem odd to the uninitiated reader, but they reflect a very natural need. Expander graphs are to be used by various algorithms. The algorithms' performance will depend on efficiently obtaining the relevant information of the expanders being used.

Definition 2.3. Let $\{G_i\}_i$ be a family of expander graphs where G_i is a d-regular graph on n_i vertices and the integers $\{n_i\}$ are increasing, but not too fast (e.g. $n_{i+1} \leq n_i^2$ will do).

(1) The family is called **Mildly Explicit** if there is an algorithm that generates the j-th graph in the family G_j in time polynomial in j. (That is, G_j is computed in time $< Aj^B$ for some constants $A, B > 0$.)
(2) The family is called **Very Explicit** if there is an algorithm that on input of an integer i, a vertex $v \in V(G_i)$ and $k \in \{1, \cdots, d\}$ computes the k-th neighbor of the vertex v in the graph G_i. This algorithm's run time should be polynomial in its input length (the number of bits needed to express the triple (i, v, k)).

2.2. **Examples of expander graphs.**

(1) *A family of 8-regular graphs G_m for every integer m.* The vertex set is $V_m = \mathbb{Z}_m \times \mathbb{Z}_m$. The neighbors of the vertex (x, y) are $(x + y, y)$, $(x-y, y), (x, y+x), (x, y-x), (x+y+1, y), (x-y+1, y), (x, y+x+1), (x, y-x+1)$, (all operations are mod m).

   This family of graphs, due to Margulis [Mar73], is the first explicitly constructed family of expander graphs. Margulis' proof of expansion was based on representation theory and did not provide any specific bound on the expansion ratio h. Gabber and Galil [GG81] later derived such a bound using harmonic analysis. In Section 8 we show that Margulis' graphs are expanders. Note that this family is very explicit.
(2) *A family of 3-regular p-vertex graphs for every prime p.* Here $V_p = \mathbb{Z}_p$, and a vertex x is connected to $x + 1, x - 1$, and to its inverse x^{-1} (operations are mod p, and we define the inverse of 0 to be 0).

   Here, the proof of expansion depends on a deep result in Number Theory, the Selberg 3/16 theorem; see the discussion in subsection 11.1.2 for more details. This family is only mildly explicit, since we are at present unable to generate large primes deterministically. See [Gra05] for a survey of the Agrawal-Kayal-Saxenaan polytime primality testing algorithm.

2.3. **Graph spectrum and an algebraic definition of expansion.** The **Adjacency Matrix** of an n-vertex graph G, denoted $A = A(G)$, is an $n \times n$ matrix whose (u, v) entry is the number of edges in G between vertex u and vertex v. Being real and symmetric, the matrix A has n real eigenvalues which we denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. We can also associate with it an orthonormal system of eigenvectors $v_1, \ldots, v_n$ with $Av_i = \lambda_i v_i$. We often refer to the eigenvalues of $A(G)$ as the **Spectrum** of the graph G. The spectrum of a graph encodes a lot of information about it. Here are some simple illustrations of how certain properties of a d-regular graph are reflected in its spectrum:

- $\lambda_1 = d$, and the corresponding eigenvector is $v_1 = \mathbf{1}/\sqrt{n} = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$.
- The graph is connected iff $\lambda_1 > \lambda_2$.
- The graph is bipartite iff $\lambda_1 = -\lambda_n$.

As seen in the next theorem, the graph's second eigenvalue is closely related to its expansion parameter.

Theorem 2.4. *Let G be a d-regular graph with spectrum $\lambda_1 \geq \cdots \geq \lambda_n$. Then*

$$\frac{d-\lambda_2}{2} \leq h(G) \leq \sqrt{2d(d-\lambda_2)}.$$

This theorem is due to Cheeger [Che70] and Buser [Bus82] in the continuous case (see Section 4 for more on this). In the discrete case, it was proved by Dodziuk [Dod84] and independently by Alon-Milman [AM85] and by Alon [Alo86]. More concretely we see that $d-\lambda_2$, also known as the **Spectral Gap**, provides an estimate on the expansion of a graph. In particular, a d-regular graph has an expansion ratio $h(G)$ bounded away from zero iff its spectral gap $d-\lambda_2$ is bounded away from zero. The following lemma shows that a small second eigenvalue in a graph implies that its edges are "spread out", a hallmark of random graphs.

2.4. **The Expander Mixing Lemma.** Given a d-regular graph G with n vertices, we denote $\lambda = \lambda(G) = \max(|\lambda_2|, |\lambda_n|)$. In words, λ is the largest absolute value of an eigenvalue other than $\lambda_1 = d$. The following useful bound, observed by several researchers, probably appeared in print first in [AC89].

Lemma 2.5 (Expander Mixing Lemma)**.** *Let G be a d-regular graph with n vertices and set $\lambda = \lambda(G)$. Then for all $S, T \subseteq V$:*

$$\left| |E(S,T)| - \frac{d|S||T|}{n} \right| \leq \lambda\sqrt{|S||T|}.$$

A word of interpretation is in place here. The left-hand side measures the deviation between two quantities: one is $|E(S,T)|$, the number of edges between the two sets; the other is the expected number of edges between S and T in a random graph of edge density d/n, namely $d|S||T|/n$. A small λ (or large spectral gap) implies that this deviation (or **discrepancy** as it is sometimes called) is small, so the graph is nearly random in this sense.

When the spectral gap of G is much smaller than d, the upper and lower bounds in Theorem 2.4 differ substantially. This makes one wonder whether the spectral gap can be captured more tightly by some combinatorial invariant of the graph. A positive answer, and a converse to the Expander Mixing Lemma, was found recently by Bilu and Linial [BL]. We will not prove this result here.

Lemma 2.6 (Converse of the Expander Mixing Lemma [BL])**.** *Let G be a d-regular graph and suppose that*

$$\left| |E(S,T)| - \frac{d|S||T|}{n} \right| \leq \rho\sqrt{|S||T|},$$

holds for every two disjoint sets S, T and for some positive ρ. Then $\lambda \leq O(\rho \cdot (1 + \log(d/\rho)))$. The bound is tight.

Proof of the Expander Mixing Lemma. Let 1_S and 1_T be the characteristic vectors of S and T (i.e., the v-th coordinate of the vector 1_S is 1 if $v \in S$ and zero otherwise). Expand these vectors in the orthonormal basis of eigenvectors $v_1, \cdots, v_n$, viz., $1_S = \Sigma_i \alpha_i v_i$ and $1_T = \Sigma_j \beta_j v_j$. Recall that $v_1 = \mathbf{1}/\sqrt{n}$. Then

$$|E(S,T)| = 1_S A 1_T = (\Sigma_i \alpha_i v_i) A (\Sigma_j \beta_j v_j).$$

Since the v_i are orthonormal eigenvectors, this equals $\Sigma_i \lambda_i \alpha_i \beta_i$. Since $\alpha_1 = \langle 1_S, \frac{\mathbf{1}}{\sqrt{n}} \rangle = \frac{|S|}{\sqrt{n}}$, $\beta_1 = \frac{|T|}{\sqrt{n}}$, and $\lambda_1 = d$:

$$|E(S,T)| = d\frac{|S||T|}{n} + \Sigma_{i=2}^n \lambda_i \alpha_i \beta_i.$$

By the definition of λ:

$$\left| |E(S,T)| - d\frac{|S||T|}{n} \right| = |\Sigma_{i=2}^n \lambda_i \alpha_i \beta_i| \le \Sigma_{i=2}^n |\lambda_i \alpha_i \beta_i| \le \lambda \Sigma_{i=2}^n |\alpha_i \beta_i|.$$

Finally, by Cauchy-Schwartz:

$$\left| |E(S,T)| - d\frac{|S||T|}{n} \right| \le \lambda ||\alpha||_2 ||\beta||_2 = \lambda ||1_S||_2 ||1_T||_2 = \lambda \sqrt{|S||T|}. \qquad \square$$

In what follows, it is sometimes convenient to consider the normalized second eigenvalue $\lambda(G)/d$. A d-regular graph G on n vertices is called an (n,d)-graph. It is an (n,d,α)-graph if $\lambda(G) \le \alpha d$. Regular graphs with small α have a number of significant properties, some of which we collect below:

- An **independent set** in a graph is a set of vertices S, no two of which are adjacent, i.e. with $|E(S,S)| = 0$. It is an immediate consequence of the Expander Mixing Lemma that an independent set in an (n,d,α)-graph has cardinality at most αn.
- A **k-coloring** of a graph $G = (V,E)$ is a mapping $c : V \to \{1,\ldots,k\}$, such that $c(x) \neq c(y)$ for any two adjacent vertices x, y. The **chromatic number** of G, denoted $\chi(G)$, is the smallest k for which G has a k-coloring. The set $c^{-1}(j)$ is an independent set in G for every $k \ge j \ge 1$. Consequently, $\chi(G) \ge 1/\alpha$ for an (n,d,α)-graph G.
- The *distance* $d_G(x,y)$ between vertices x and y in a graph $G = (V,E)$ is the length of the shortest path between them. The **diameter** of G is defined as $\max_{x,y} d_G(x,y)$. Also $B(x,r) = \{y | d_G(x,y) \le r\}$ is *the ball of radius r around x.* We claim that an (n,d,α)-graph G has diameter $O(\log n)$. That certainly follows if we show that $|B(x,r)| > n/2$ for every vertex x and some $r \le O(\log n)$. This in turn follows from the expansion properties of G. That is, we show that $|B(x,r+1)| \ge (1+\epsilon)|B(x,r)|$ for some fixed $\epsilon > 0$ as long as $|B(x,r)| \le n/2$. The Expander Mixing Lemma implies that $|E(S,S)|/|S| \le d \cdot (|S|/n + \alpha)$ for every set S. Therefore, $|E(S,\overline{S})|/|S| \ge d \cdot ((1-\alpha) - |S|/n)$. But S has at least $|E(S,\overline{S})|/d$ neighbors outside itself, so the claim follows with $\epsilon = 1/2 - \alpha$.

2.5. **How big can the spectral gap be?** The question in the title has to be qualified, since the answer depends on the relationship between d and n. We are mostly interested in d fixed and large n. To illustrate how things may differ when d grows with n, consider the complete graph on n vertices K_n where every two vertices are adjacent and so $d = n-1$. Clearly the adjacency matrix of K_n is $J - I$ where J is the all-ones matrix and $I = I_n$ is the identity matrix. The spectrum of K_n is $[n-1, -1, -1, \cdots, -1]$. and $\lambda = 1$.

For the range we are interested in, $n \gg d$, the question was answered by N. Alon and R. Boppana (see A. Nilli [Nil91]):

Theorem 2.7 (Alon-Boppana). *For every (n,d)-graph,*

$$\lambda \geq 2\sqrt{d-1} - o_n(1).$$

The $o_n(1)$ term is a quantity that tends to zero for every fixed d as $n \to \infty$. More on this and a proof of this theorem appear in Section 5. Here is a very easy but somewhat weaker statement:

Claim 2.8. *For every (n,d)-graph G,*

$$\lambda \geq \sqrt{d} \cdot (1 - o_n(1)).$$

Proof. Let A be the adjacency matrix of G. It is not hard to see that trace(A^k) is the number of all walks of length k in G that start and end in the same vertex. In particular, all the diagonal entries in A^2 are $\geq d$. (Just move back and forth along any edge incident to the vertex in question.) Consequently, trace$(A^2) \geq nd$. On the other hand,

$$\text{trace}(A^2) = \sum_i \lambda_i^2 \leq d^2 + (n-1)\lambda^2.$$

It follows that $\lambda^2 \geq d \cdot \frac{n-d}{n-1}$, as claimed. □

2.6. **Four perspectives on expansion and how they compare.** We are now in a position to offer the reader a broader view of some of the main questions in the field. Expansion is defined in **combinatorial** terms and, as we shall see, this definition comes in several different flavors. This is closely related to the **spectral** theory of graphs. Finally, rapidly mixing random walks provide a **probabilistic** perspective.

In each of these three frameworks we consider mostly four types of questions:

- **Extremal:** How large/small can the pertinent expansion parameters be?
- **Typical:** How are these parameters distributed over random graphs?
- **Explicit construction:** Can one construct graphs for which these parameters (nearly) attain their optimum?
- **Algorithmic:** Given a graph, can you efficiently evaluate/estimate its expansion parameters?

It then becomes natural to consider some **comparative** problems: What can you conclude, say, about combinatorial-type expansion parameters from spectral information, etc.?

Here are some pointers to the present article where we either explain what is known about such question or provide some further references to the relevant literature.

2.6.1. *Extremal problems.* Here the most satisfactory answer comes from the spectral realm. The Alon-Boppana Theorem 5.3 tells us precisely how large the spectral gap can be in an (n,d)-graph.

The largest edge expansion $h(G)$ of an (n,d)-graph G is at most $d/2 - c\sqrt{d}$ for every $d \geq 3$ and sufficiently large n, where $c > 0$ is an absolute constant. This result is tight up to the value of c; see subsection 5.1.1. More interesting (and often more difficult) questions concern the expansion of smaller sets in the graph. Some discussion of this problem is to be found in Section 5 and subsection 4.6.

2.6.2. *Typical behavior.* Here the situation reverses. It is relatively not hard to analyze the (vertex/edge) expansion in random graphs by methods similar to those used in subsection 1.2. See subsection 4.6 for more details.

The typical behavior of the spectrum is harder to understand, and Section 7 is dedicated to an exposition of this fascinating story and the still lingering mysteries.

2.6.3. *Explicit constructions.* We have already mentioned the Margulis construction to which Section 8 is dedicated. The so-called Ramanujan Graphs due to Lubotzky-Phillips-Sarnak [LPS88] and Margulis [Mar88] are mentioned briefly in subsection 5.3, but are otherwise not discussed at depth here. We do survey some more combinatorial approaches to the problem, viz. subsection 6.4 and Section 11.

Direct estimates of the expansion, even for specific families of graphs, are even harder to come by and [LL06] is one of very few exceptions. In fact, the following question is quite nontrivial: Find explicit constructions of graphs in which small sets of vertices expand well. We will have quite a bit to say about this problem in Section 10.

2.6.4. *Algorithms.* The exact determination of $h(G)$, given G, is difficult (co-**NP**-hard) [BKV81]. This fact and the approximate version of the problem are briefly discussed in subsection 13.5. Likewise, we lack good estimates for the vertex isoperimetric parameter of a given graph or for the edge expansion of sets of a given size in a graph. These are among the most significant open questions in the theory. On the other hand, standard algorithms in linear algebra can be used to efficiently compute the spectrum of a given graph. For the analogous problem in the context of random walks see subsection 3.1.2.

2.6.5. *Comparisons.* As mentioned above, for random graphs, expansion is more accessible than spectral gap. On the other hand, eigenvalues are easily computable, while expansion is not. It is interesting to ask how well one theory reflects on the other when we seek (nearly) optimal graphs. Graphs with very large spectral gap are very good expanders: When $\lambda = o(d)$, the lower bound in Theorem 2.4 yields $h(G) \geq (\frac{1}{2} - o(1))d$. On the other hand, for d large, an (n,d)-graph G can have $h(G) \geq \Omega(d)$ while the spectral gap is small. Here is an illustration of how this can happen: Pick a small $\delta > 0$ and construct an $(n, (\delta{\cdot}d))$-graph G with $h(G) = \Omega(\delta{\cdot}d)$. Now add to it a collection of disjoint cliques of size $(1-\delta)d+1$ each. Clearly $h(G)$ does not decrease, but the spectral gap is at most δd.

Another interesting example can be obtained by considering the *line graph* H of an (n,d)-graph G that is a good expander. The vertex set of H is the edge set of G, and two vertices in H are adjacent iff the corresponding edges are incident in G. The graph H is an $(\frac{nd}{2}, 2d-2)$-graph. Its second eigenvalue is easily seen to be $\geq (1-o(1))d$, but if G has a large expansion ratio, then so does H.

Finally, we mention that Lemma 2.6 shows the near equivalence of discrepancy and spectral gap.

Connections with rapid mixing of random walks are discussed in subsection 3.1.

3. Random walks on expander graphs

A key property of the random walk on an expander graph is that it converges rapidly to its limit distribution. This fact has numerous important consequences at which we can only hint. In many theoretical and practical computational problems in science and engineering it is necessary to draw samples from some distribution $\mathcal{F}$ on a (usually finite but huge) set V. Such problems are often solved by so-called "Monte-Carlo" algorithms. One considers a graph G on vertex set V so that the limit distribution of the random walk on G is $\mathcal{F}$. A clever choice of G can guarantee that (i) it is feasible to efficiently simulate this random walk and (ii) the distribution induced on V by the walk **converges rapidly** to $\mathcal{F}$. Among the fields where this methodology plays an important role are Statistical Physics, Computational Group Theory and Combinatorial Optimization. We should mention approximation algorithms for the permanence of nonnegative matrices [JSV04] and for the volume of convex bodies in high dimension [Sim03] as prime examples of the latter. Excellent surveys on the subject are [JS96, Jer03]. As we briefly mention in subsection 4.5, some of this theory extends to the more general context of time-reversible Markov Chains [LW98, MT].

The main principle behind the topics we survey here is that the set of vertices visited by a length t random walk on an expander graph "looks like" (in some respects) a set of t vertices sampled uniformly and independently. The computational significance of this is that the number of random bits required in order to generate a length t walk on a (constant-degree) graph is significantly smaller than the number of random bits that are needed in order to independently sample t random vertices. We exhibit two applications of this idea: (i) a randomness-efficient error reduction procedure for randomized algorithms, and (ii) a strong hardness-of-approximation result for the maximum clique problem. Other computational applications of these ideas that we will not go into include derandomization of probabilistic space-bounded algorithms (see e.g. Nisan-Zuckerman [NZ96] and Impagliazzo-Nisan-Wigderson [INW94]).

3.1. **Rapid mixing of walks.** A **walk** on a graph $G = (V, E)$ is a sequence of vertices $v_1, v_2, \ldots \in V$ such that v_{i+1} is a neighbor of v_i for every index i. When v_{i+1} is selected uniformly at random from among v_i's neighbors, independently for every i, this is called a **random walk** on G. We usually initiate this random process by selecting the first vertex v_1 from some **initial probability distribution** π_1 on V. Clearly this induces a sequence of probability distributions π_i on V so that the probability that $v_i = x \in V$ equals $\pi_i(x)$ for every i and x. It is well known that for every finite connected nonbipartite graph G, the distributions π_i converge to a limit, or **stationary**, distribution. Moreover, it is easy to see that if G is regular, then this distribution is the uniform distribution on V.

This subsection deals with the speed of this convergence. There are several interesting ways to measure the distance between π_i and the limit distribution, and we will consider several norms and entropy measures. The main thrust is that in expanders the distance to the limit shrinks substantially with **every** step of the random walk and that this condition characterizes expander graphs. We now make this statement quantitative. We start with some definitions and notations.

Recall that an (n,d)-graph is a d-regular graph G on n vertices. It is called an (n,d,α)**-graph** if $|\lambda_2(G)|, |\lambda_n(G)| \leq \alpha d$, where $d = \lambda_1(G) \geq \ldots \geq \lambda_n(G)$ is the spectrum of G.

A vector $\mathbf{p} \in \mathbb{R}^n$ is called a **probability distribution vector** if its coordinates are nonnegative and $\sum_{i=1}^n p_i = 1$. The probability vector that corresponds to the uniform distribution on $\{1,\ldots,n\}$ is denoted by $\mathbf{u} = (1,\ldots,1)/n$. In this section we show that a random walk on the vertices of an expander converges rapidly to the stationary distribution.

Definition 3.1. A random walk on a finite graph $G = (V,E)$ is a discrete-time stochastic process $(X_0, X_1, \ldots)$ taking values in V. The vertex X_0 is sampled from some initial distribution on V, and X_{i+1} is chosen uniformly at random from the neighbors of X_i.

If G is a d-regular graph with adjacency matrix A, then its *normalized adjacency matrix* is defined as $\hat{A} = \frac{1}{d}A$. Here are some simple comments on this random walk.

- The random walk on $G = (V,E)$ is a Markov Chain with state set V and transition matrix $\hat{A}$.
- $\hat{A}$ is real, symmetric, and doubly stochastic; i.e. every column and every row sums up to 1.
- If $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$ are the eigenvalues of $\hat{A}$, then $\hat{\lambda}_1 = 1$ and $\max\{|\hat{\lambda}_2|, |\hat{\lambda}_n|\} \leq \alpha$.
- The corresponding eigenvectors are the same eigenvectors of A.
- Consider an experiment where we sample a vertex x from some probability distribution $\mathbf{p}$ on V and then move to a random neighbor of x. This is equivalent to sampling a vertex from the distribution $\hat{A}\mathbf{p}$.
- The matrix $\hat{A}^t$ is the transition matrix of the Markov Chain defined by random walks of length t. In other words $(\hat{A}^t)_{ij}$ is the probability a random walk starting at i is at j after t steps.
- The stationary distribution of the random walk on G is the uniform distribution, namely, $\mathbf{u}\hat{A} = \hat{A}\mathbf{u} = \mathbf{u}$. (This uses the symmetry of $\hat{A}$.)

3.1.1. *Convergence in the l_1 and l_2 norms.* The inner product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$. The l_1, l_2 and l_∞ norms are denoted as usual by

- $||\mathbf{x}||_1 = \sum_{i=1}^n |x_i|$,
- $||\mathbf{x}||_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^n x_i^2}$,
- $||\mathbf{x}||_\infty = \max_{1 \leq i \leq n} |x_i|$.

We now observe that if G is an (n,d,α)-graph and $\alpha < 1$, then regardless of the initial distribution $\mathbf{p}$, the random walk converges in l_1 exponentially fast to its limit (uniform) distribution. This will follow (via Cauchy-Schwartz) from a similar bound on l_2, which in turn follows from the fact that in l_2 the distance to the uniform distribution shrinks by a factor of α **at each step**.

Theorem 3.2. *Let G be an (n,d,α)-graph with normalized adjacency matrix $\hat{A}$. Then for any distribution vector $\mathbf{p}$ and any positive integer t:*

$$||\hat{A}^t \mathbf{p} - \mathbf{u}||_1 \leq \sqrt{n} \cdot \alpha^t.$$

Why use the l_1 norm to measure for the distance between two probability distributions p, q? A natural and commonly used metric is the **total variation distance** $\max_B |\Pr_p[B] - \Pr_q[B]|$, and it is not difficult to check that this equals $\frac{1}{2}||p-q||1$.

In other words, if the l_1 distance is small, then the two distributions p and q assign nearly equal probabilities to **every** event in the probability space. Theorem 3.2 follows immediately from the analogous l_2 bound below.

Theorem 3.3. *Let G be an (n, d, α)-graph with normalized adjacency matrix $\hat{A}$. Then for any distribution vector $\mathbf{p}$ and any positive integer t:*

$$||\hat{A}^t\mathbf{p} - \mathbf{u}||_2 \leq ||\mathbf{p} - \mathbf{u}||_2\alpha^t \leq \alpha^t.$$

Obviously it suffices to prove this bound for $t = 1$ (shrinkage per step) and use induction.

Lemma 3.4. *For every probability vector $\mathbf{p}$, $||\hat{A}\mathbf{p} - \mathbf{u}||_2 \leq \alpha||\mathbf{p} - \mathbf{u}||_2 \leq \alpha$.*

Proof. The uniform distribution $\mathbf{u}$ is invariant under the action of $\hat{A}$. Also, $\mathbf{p} - \mathbf{u}$ is orthogonal to $\mathbf{u}$ and thus shrinks in l_2-norm by a factor α under the action of $\hat{A}$. Consequently

$$||\hat{A}\mathbf{p} - \mathbf{u}||_2 = ||\hat{A}(\mathbf{p} - \mathbf{u})||_2 \leq \alpha||\mathbf{p} - \mathbf{u}||_2 \leq \alpha,$$

where the last inequality follows easily from the fact that $\mathbf{p}$ is a probability distribution. □

3.1.2. *Convergence in entropy.* Another important perspective of a random walk is offered by the **entropy** of the associated probability distributions. The entropy of probability distributions is a fundamental concept in the theory of communication, capturing the amount of "information", or alternatively "uncertainty", that it carries. When we take a random step, we "inject" more randomness into our distribution, indeed precisely the $\log d$ random bits that are needed to specify which of the d neighbors of the current vertex we move to next. One expects this injection to increase the amount of "randomness" in the distribution, namely its entropy. This is indeed always true in every regular graph, and expanders are those graphs for which the increase is significant.

This entropy viewpoint will be extremely important when we explain the zig-zag product and its use in combinatorial constructions of various expanders in Sections 9 and 10. In the same way that different norms capture different aspects of the probability distributions, there are several variations on the theme of entropy that do this. Let $[n]$ denote the set of integers $\{1, \ldots, n\}$. Then for a probability distribution $\mathbf{p}$ on $[n]$ we define:

- **Shannon entropy:** $H(\mathbf{p}) = -\sum_{i=1}^n p_i \log(p_i)$.
- **Rényi 2-entropy:** $H_2(\mathbf{p}) = -2\log(||\mathbf{p}||_2)$.
- **Min entropy:** $H_\infty(\mathbf{p}) = -\log(||\mathbf{p}||_\infty)$.

To see the connection between the last two quantities, note that if $\mathbf{p}$ is a probability distribution on $[n]$, then $\max p_i \geq \sum p_i^2 \geq \max p_i^2$. It follows that:

Proposition 3.5.

$$H_\infty(\mathbf{p}) \leq H_2(\mathbf{p}) \leq 2H_\infty(\mathbf{p}).$$

Here are some simple and useful properties that are common to all three, which the reader is invited to verify. As above, $\mathbf{p}$ is a probability distribution on an n-element set, and we denote a "generic" entropy by $\tilde{H}$.

- $\tilde{H}(\mathbf{p}) \geq 0$ with equality iff the distribution is concentrated on a single element.
- $\tilde{H}(\mathbf{p}) \leq \log n$ with equality iff the distribution is uniform.

- For any doubly stochastic matrix X (nonnegative matrix whose row and column sums are one), $\tilde{H}(X\mathbf{p}) \geq \tilde{H}(\mathbf{p})$. Equality holds iff $\mathbf{p}$ is uniform.

The last item shows that entropy increases with every step of the random walk on a regular graph. Making this quantitative depends on the choice of entropy measure. Below we do so for the Rényi 2-entropy in terms of the spectral bound α, which (not surprisingly) is just a restatement of the l_2 bound from the previous section. However, as noted above H_2 and H_∞ are very close to each other, and it is the latter we use in Section 10, so this interpretation will be important for us.

Before doing so, we remark that for Shannon entropy H, the precise relation between the increase in H and the spectral constant α is still unknown. However, one can define an analogous "entropy constant" which governs the increase "per step" in entropy. It is called the Log-Sobolev constant, and there are known quantitative relations between it and and the spectral constant (much like the relations between edge expansion and the spectral constant of the previous section). Using the Log-Sobolev constant to analyze the mixing time of random walks is a powerful method, but it is beyond the scope of this survey. For more on this, see e.g. [MT].

Let us write the distribution as $\mathbf{p} = \mathbf{u} + \mathbf{f}$, where $\mathbf{f} \perp \mathbf{u}$. We let μ capture how close $\mathbf{p}$ is to the uniform distribution, via $\mu = ||\mathbf{f}||/||\mathbf{p}|| \leq 1$ (e.g. $\mu = 0$ iff $\mathbf{p}$ is uniform). Then

$$||\hat{A}\mathbf{p}||^2 = ||\mathbf{u} + \hat{A}\mathbf{f}||^2 = ||\mathbf{u}||^2 + ||\hat{A}\mathbf{f}||^2 \leq ((1-\mu^2) + \alpha^2\mu^2)||\mathbf{p}||^2.$$

Hence

$$H_2(\hat{A}\mathbf{p}) \geq H_2(\mathbf{p}) - \log((1-\mu^2) + \alpha^2\mu^2) = H_2(\mathbf{p}) - \log(1 - (1-\alpha^2)\mu^2).$$

It follows that the 2-entropy never decreases and is, in fact, strictly increasing as long as the distribution p is not uniform. It is also clear that for better expanders (i.e., for smaller α) the 2-entropy grows faster.

3.2. **Random walks resemble independent sampling.** In the sequel, we imagine an abstract sampling problem in which an unknown set B in a universe of size n is "bad" in some sense, and we try to sample the universe so as to avoid the bad set as much as possible. Our task will be to do so, minimizing the number of random bits used. In a motivating example we saw already that set B includes all the bad random choices for a probabilistic algorithm, namely, those choices for which it gives the wrong answer. We now describe the advantages of imposing, out of the blue, an expander graph structure on the universe. Using it, we can choose a small sample using a random walk on the graph. Remarkably, the statistics of hitting B with such a (highly dependent) sample will be very close to that of a completely independent sample (provided we pick the degree and expansion of the graph appropriately).

Suppose that we are given an (n, d, α)-graph $G = (V, E)$ where the vertices in some subset $B \subseteq V$ are "bad". All we know about the set B is its cardinality $|B| = \beta n$. We wish to sample at least one vertex outside of B. We can certainly sample, uniformly at random, $t+1$ vertices $x_0, \ldots, x_t$ from V, and fail with probability $Pr[\forall i\ x_i \in B] \leq \beta^{t+1}$. This approach uses $(t+1)\log n$ random bits, and we will show that a similar performance can be achieved with substantially fewer random bits: namely, that if we choose a random starting vertex and carry out a random walk of length t from it, then our chance of failure, i.e., the probability that the whole random walk is confined to B, is exponentially small in t as well. To get started, let us reinterpret the Expander Mixing Lemma as the case $t = 1$ of this approach. Recall that the lemma says that:

$$\left| \frac{d|S||T|}{n} - |E(S,T)| \right| \leq \alpha d \sqrt{|S||T|} \leq \alpha d n,$$

for every two subsets $S, T \subseteq V(G)$ in an (n, d, α)-graph G. It is useful to rewrite this as:

$$\left| \frac{|S||T|}{n^2} - \frac{|E(S,T)|}{dn} \right| \leq \alpha.$$

Now consider two experiments in which we sample an ordered pair of vertices (i, j) in G, and consider it a success if $i \in S$ and $j \in T$. In the first version of this experiment both vertices are selected uniformly at random from $V = V(G)$, so our probability of success is $|S||T|/n^2$. In the second version we randomly pick $i \in V$, and then j is selected uniformly among the neighbors of i. In this version of the experiment, the probability of success is $|E(S,T)|/dn$. The Expander Mixing Lemma says that despite the different nature of the sample spaces in the two experiments, their success probabilities differ only by a small constant α. We now turn to the complete argument that pertains to longer random walks.

Let $G = (V, E)$ be an (n, d, α)-graph, and $B \subset V$ of cardinality $|B| = \beta n$. We carry out the following experiment: We pick $X_0 \in V$ uniformly at random and start from it a random walk $X_0, \ldots, X_t$ on G. Denote by (B, t) the event that this random walk is confined to B, i.e. that $\forall i\; X_i \in B$.

Theorem 3.6 (Ajtai-Komlós-Szemerédi [AKS87], Alon-Feige-Wigderson-Zuckerman [AFWZ95])**.** *Let G be an (n, d, α)-graph and $B \subset V$ with $|B| = \beta n$. Then the probability of the event (B, t) is bounded by*

$$\Pr[(B,t)] \leq (\beta + \alpha)^t.$$

Let $P = P_B$ be the orthogonal projection on the subspace of coordinates belonging to B. In matrix notation $P_{ij} = 1$ if $i = j \in B$ and 0 otherwise. We need the following simple observation:

Lemma 3.7. *The probability of event (B, t) is given by $Pr[(B,t)] = ||(P\hat{A})^t P\mathbf{u}||_1$.*

Proof. To calculate the (x, y) entry in the matrix $(\hat{A})^t$, we should sum the probabilities of walks of length t that start at vertex x and end at y. In the same calculations for the matrix $(P\hat{A})^t P$ only walks confined to B are to be considered, since all other contributions are eliminated by the matrices P in the product. But $||(P\hat{A})^t P\mathbf{u}||_1$ is just $\frac{1}{n}$ of the total sum of these entries and the conclusion follows. $\square$

We also need the following lemma.

Lemma 3.8. *For any vector $\mathbf{v}$,*

$$||P\hat{A}P\mathbf{v}||_2 \leq (\beta + \alpha) \cdot ||\mathbf{v}||_2.$$

Proof. First note that there is no loss in assuming that $\mathbf{v}$ is supported on B, i.e. $P\mathbf{v} = \mathbf{v}$. Otherwise we may replace $\mathbf{v}$ by $P\mathbf{v}$. This leaves the left-hand side unchanged and does not increase the right-hand side, since P is a contraction in l_2. Similarly, we may assume that $\mathbf{v}$ is nonnegative. Also, by linearity of both sides we may assume that $\sum v_i = 1$ and so $\mathbf{v}$ can be expressed as: $P\mathbf{v} = \mathbf{v} = \mathbf{u} + \mathbf{z}$ where $\mathbf{z}$ is orthogonal to $\mathbf{u}$. It follows that

$$P\hat{A}P\mathbf{v} = P\hat{A}\mathbf{u} + P\hat{A}\mathbf{z} = P\mathbf{u} + P\hat{A}\mathbf{z}$$

and hence

$$||P\hat{A}P\mathbf{v}||_2 \leq ||P\mathbf{u}||_2 + ||P\hat{A}\mathbf{z}||_2.$$

We now prove that $||P\mathbf{u}||_2 \leq \beta \cdot ||\mathbf{v}||_2$ and $||P\hat{A}\mathbf{z}||_2 \leq \alpha \cdot ||\mathbf{v}||_2$, which together imply the claim.

Since $\sum v_i = 1$, and the support of $\mathbf{v}$ has at most βn coordinates, Cauchy-Schwartz yields $1 = \sum v_i \leq \sqrt{\beta n} \cdot ||\mathbf{v}||_2$. But, since $||P\mathbf{u}||_2 = \sqrt{\beta/n}$, we obtain

$$||P\mathbf{u}||_2 \leq \beta \cdot ||\mathbf{v}||_2.$$

As for the other term, $||\hat{A}\mathbf{z}||_2 \leq \alpha ||\mathbf{z}||_2$, since z is orthogonal to $\mathbf{u}$ and therefore is a combination of eigenvectors of $\hat{A}$ with eigenvalues $\leq \alpha$. But $||P\hat{A}\mathbf{z}||_2 \leq ||\hat{A}\mathbf{z}||_2$, since P is a contraction in l_2. Also $\mathbf{v}$ is the sum of $\mathbf{z}$ and the orthogonal vector $\mathbf{u}$, whence $||\mathbf{z}||_2 \leq ||\mathbf{v}||_2$. It follows that $||P\hat{A}\mathbf{z}||_2 \leq \alpha \cdot ||\mathbf{v}||_2$, as needed. □

Now we use the two lemmas to prove Theorem 3.6.

Proof. (Theorem 3.6)

$$\begin{aligned} ||(P\hat{A})^t P\mathbf{u}||_1 &\leq \sqrt{n} \cdot ||(P\hat{A})^t P\mathbf{u}||_2 \\ &= \sqrt{n} \cdot ||(P\hat{A}P)^t \mathbf{u}||_2 \\ &\leq \sqrt{n} \cdot (\beta + \alpha)^t ||\mathbf{u}||_2 \\ &= (\beta + \alpha)^t. \end{aligned}$$

□

The probability that $t+1$ uniformly and independently sampled vertices all land in a set B of density β is β^{t+1}. Is it true that $\Pr[(B,t)]$ is very close to this bound for a sufficiently good expander? Theorem 3.6 falls a factor of β short of this bound (as we have $t+1$ sample points). It also does not provide a comparable lower bound. The following result, which we do not prove here, fills this gap.

Theorem 3.9 ([AFWZ95]). *If $\beta > 6\alpha$, then*

$$\beta \cdot (\beta + 2\alpha)^t \geq Pr[(B,t)] \geq \beta \cdot (\beta - 2\alpha)^t.$$

It is also possible to derive "time dependent" versions of the upper bound through simple adaptations of the proof of Theorem 3.6. This will be useful for us a little later.

Theorem 3.10. *For every subset $K \subset \{0, \dots, t\}$ and vertex subset B of density β,*

$$Pr[X_i \in B \text{ for all } i \in K] \leq (\beta + \alpha)^{|K|-1}.$$

Occasionally we have to deal with a situation where the excluded set varies between time steps. This is addressed in the following theorem:

Theorem 3.11. *Let $B_0, \dots, B_t$ be vertex sets of densities $\beta_0, \dots, \beta_t$ in an (n, d, α)-graph G. Let $X_0, \dots, X_t$ be a random walk on G. Then*

$$\Pr[X_i \in B_i \text{ for all } i] \leq \prod_{i=0}^{t-1} (\sqrt{\beta_i \beta_{i+1}} + \alpha).$$

This follows by a small adaptation of the previous arguments. Let P_i be the projection on B_i, and note that $\Pr[X_i \in B_i \text{ for all } i] = ||\prod_{i=1}^{t} (P_i \hat{A}) P_0 \mathbf{u}||_1$, and $||P_{i+1} \hat{A} P_i \mathbf{v}||_2 \leq (\sqrt{\beta_i \beta_{i+1}} + \alpha) \cdot ||v||_2$.

As before, this simple approach seems to give away a factor of $\sqrt{\beta_0\beta_t}$, which is important for certain applications. For further discussion, see Alon-Feige-Wigderson-Zuckerman [AFWZ95] or Bilu-Hoory [BH04]. For a different analysis using perturbation theory which gives a Chernoff-like bound for expander walks, see [Gil98].

3.3. Applications.

3.3.1. *Efficient error reduction in probabilistic algorithms.* We now return to the computational problem raised in subsection 1.1.3 of reducing the error in probabilistic algorithms while trying to save on random bits used for this task.

Let A be a probabilistic algorithm for the language (=set of binary strings) $\mathcal{L}$. We first deal with the simpler case that the algorithm makes errors only on inputs outside $\mathcal{L}$ (so called *one-sided* error), in which case we say that $\mathcal{L}$ is in the complexity class **RP**. We then deal with the case that A may err both on inputs in and outside $\mathcal{L}$ (so called *two-sided* error), in which case $\mathcal{L}$ is in the corresponding complexity class **BPP**.

Again, recall that if we do not attempt to save random bits, an obvious reduction in error can be achieved by running the algorithm many (say t) times, each time with independent random bits. In the one-sided error case we would take the conjunction of the answers, and in the two-sided error case we would take the majority. In both cases the error probability will drop exponentially in t. However, the number of random bits used will increase by a *factor* of t. We now proceed to achieve the same error probability in both cases, with much fewer random bits, using expander walks.

One-sided error. Let A be an algorithm to decide membership in $\mathcal{L}$ that is randomized with a one-sided error. To decide whether a given input x belongs to $\mathcal{L}$, the algorithm samples a string $r \in \{0,1\}^k$ and computes in polynomial time a boolean function $A(x,r)$. If $x \in \mathcal{L}$, then $A(x,r) = 1$ for all r. If $x \notin \mathcal{L}$ the probability (over choices of r) that $A(x,r) = 1$ is at most β. Again our goal is to reduce the probability of error to below a given threshold without a substantial increase in the number of random bits that are required. To this end, choose an explicit (n,d,α)-graph $G = (V,E)$, with $V = \{0,1\}^k$, and a value of α sufficiently smaller than the bound β which is the error of the given algorithm. Note that the choice of α will put a lower bound on d (but as we shall see later d can be taken to be $O(\alpha^{-2})$).

For a given input x let $B_x = B \subseteq \{0,1\}^k$ be the set of all strings r for which the algorithm A errs on input x. We now introduce another algorithm A' for the membership problem.

(1) Pick a vertex $v_0 \in V$ uniformly at random.
(2) Start from it a length t random walk, say $(v_0, v_1, \ldots, v_t)$.
(3) Return $\bigwedge_{i=0}^{t} A(x, v_i)$.

We note that for the new algorithm A' to be efficient this walk has to be efficiently computed, hence the importance of having G explicitly described.

By Theorem 3.6

$$Pr[A' \text{ fails }] = Pr[\forall i \; v_i \in B] \le (\beta + \alpha)^t.$$

Compared with the algorithm from Section 1, the new algorithm achieves an exponential reduction in error probability, while the number of random bits used is only $m + t\log d = m + O(t)$.

Two-sided errors. When our basic algorithm can err on every input, not just when $x \notin \mathcal{L}$, we say that it makes a two-sided error. We show that the probability of success can be amplified as well in algorithms which make two-sided errors using the same trick. We say that the language $\mathcal{L}$ belongs to the complexity class **BPP** if there is a polynomial time randomized algorithm A to decide whether a given input x belongs to $\mathcal{L}$. It is assumed that for every x (either in or out of $\mathcal{L}$) A errs with probability $\leq \beta \leq \frac{1}{10}$. To reduce our probability of error we can run A on t independently sampled random strings and take a **majority** vote. It is a simple consequence of the Chernoff bound that the resulting error probability decreases exponentially with t. To save on randomness we again use expander walks.

As before we assume that A uses k random bits and we employ an (n, d, α)-graph on the vertex set $V = \{0,1\}^k$. Again let $B_x = B \subset V$ be the collection of all random strings for which the algorithm A errs on input x. Our modified algorithm A' works as follows:

(1) Pick a vertex $v_0 \in V$ uniformly at random.
(2) Start from it a length t random walk, say $(v_0, \ldots, v_t)$.
(3) Return $majority\{A(x, v_i)\}$.

The algorithm A' fails iff a majority of the v_i's belong to B. Fix a set of indices $K \subset \{0, 1, \ldots, t\}$ of cardinality $|K| \geq (t+1)/2$. By Theorem 3.10

$$\Pr[v_i \in B \text{ for all } i \in K] \leq (\beta + \alpha)^{|K|-1} \leq (\beta + \alpha)^{(t-1)/2}.$$

We will assume that $\alpha + \beta \leq 1/8$ and apply the union bound on the possible choices of K to deduce that

$$\Pr[A' \text{ fails }] \leq 2^t \cdot (\beta + \alpha)^{(t-1)/2} = O\left(2^{-t/2}\right).$$

So here too we achieve an exponential reduction of the error probability using only $m + O(t)$ random bits. In the following table we collect the main parameters of the various techniques presented for error reduction.

Method	**Error Probability**	**No. of random bits**
Randomized algorithm A	$1/10$	m
t independent repetitions of A	2^{-t}	$t \cdot m$
Sampling a point and its neighbors in an $(n, t, 1/\sqrt{t})$-graph.	$1/t$	m
A random walk of length t on an $(n, d, 1/40)$-graph	$2^{-t/2}$	$m + O(t)$

Further progress and reading.

The exact form of the exponential decay in error using expander walks and its dependence on the spectral gap was found by Gillman [Gil98] and is a natural optimal generalization of the Chernoff bound for independent sampling.

It is easy to see that a good approximation to the probability of hitting any set (event) in the space actually gives a good sampler for the average of any real function, and indeed expander walks are used for that purpose (for a good survey on randomness efficient samplers, see [Gol97]).

3.3.2. *Hardness of approximating maximum clique size.* We now turn to a different application of random walks on expanders to computational complexity. We show how they are used in enhancing hardness of approximation factors of the clique problem. We first give the necessary background on hardness of approximation.

Recall that a **clique** in a graph G is a subset of vertices $S \subseteq V(G)$ in which every two vertices are adjacent. The **clique number** $\omega(G)$ of a graph G is defined as the largest cardinality of a clique in G. An important computational problem is estimating this parameter. No efficient (or even subexponential time) algorithm is known for this problem. The discussion below explains why it is unlikely that such an algorithm exists.

Among the great achievements of theoretical computer science in the 1970's and 80's was the discovery of numerous natural decision problems that are **NP**-complete. The determination of the clique number is among these problems. We are given a graph G and an integer k and are asked to determine whether $\omega(G) \geq k$. The fact that this problem is **NP**-complete means that if there is a polynomial-time algorithm to solve the clique problem, then $\mathbf{P} = \mathbf{NP}$. This means that **every** problem in **NP** has a polynomial time algorithm. That conclusion is considered extremely unlikely, even though we seem far from proving it. One empirical reason for the belief that $\mathbf{P} \neq \mathbf{NP}$ is that the class **NP** is extremely rich. Literally thousands of important problems from many diverse branches of science and technology are known to be **NP**-complete. These problems have been attacked (independently) for decades by many scientists and engineers for their practical importance, and no efficient algorithm for any was found.

Assuming these problems, most of which are about finding the *optimal* solution to a given problem, are all hard, a natural relaxation is to seek an **approximate** solution. How hard are these problems? For a long time only a few hardness results for approximation were known. A breakthrough in the study of this fundamental problem was made with the proof of the **PCP** Theorem in [AS98, ALM98]. The connection between such theorems and hardness of approximation was established in [FGL91]. We are unable to go into this fascinating subject at any length and refer the reader to surveys by Sudan [Sud04] and by Arora-Lund [AL96].

As mentioned above, in a classical paper, Karp [Kar72] showed that it is **NP**-hard to determine exactly the clique number of a given graph. An early triumph of the **PCP** theory was the proof that it is **NP**-hard even to approximate the clique number to within any constant factor.[2]

Theorem 3.12 (Feige-Goldwasser-Lovász-Safra-Szegedy [FGL91]). *There are two constants $1 > \delta_1 > \delta_2 > 0$ such that it is* **NP***-hard to decide for a given n-vertex graph G whether $\omega(G) \leq \delta_2 n$ or $\omega(G) \geq \delta_1 n$.*

In this section we will show that even obtaining a **very rough** approximation for $\omega(G)$ is **NP**-hard. That is, we will show that it is **NP**-hard to approximate $\omega(G)$ even within a factor of n^ϵ for some fixed $\epsilon > 0$. Specifically we show the following theorem.

Theorem 3.13. *There exists a constant $\epsilon > 0$ with the following property. If there exists a polynomial-time algorithm A whose output on every n-vertex graph G satisfies $n^{-\epsilon} \leq A(G)/\omega(G) \leq n^\epsilon$, then* $\mathbf{NP} = \mathbf{P}$*.*

This theorem was proven by [ALM98]. Our proof will follow [AFWZ95]. It assumes an algorithm A as in the theorem and creates from it an efficient algorithm B for the problem of approximating the clique number to within a constant factor.

[2]In fact, the simplest form of the **PCP** Theorem is almost equivalent to this statement.

Such a conversion is called a **reduction**. If the resulting algorithm B is deterministic, then we can use Theorem 3.12 to conclude that the existence of algorithm A implies $\mathbf{P} = \mathbf{NP}$.

We will first present a **probabilistic reduction** (due to Berman and Schnitger [BS92]) between the two problems, namely an algorithm B which is probabilistic polynomial time. Note that with this weaker reduction, the assumption of Theorem 3.12 only implies a probabilistic polynomial time algorithm for all problems in $\mathbf{NP}$ (or $\mathbf{RP} = \mathbf{NP}$ in complexity theoretic lingo). After describing this probabilistic reduction, we will show how to eliminate the randomness from the algorithm B, resulting in a deterministic algorithm B'. For this we will again employ walks on expander graphs. This will prove the conclusion $\mathbf{P} = \mathbf{NP}$ and thus the fact that approximating the clique number to within n^ϵ is $\mathbf{NP}$-hard.

We should note that a much stronger hardness result is known about approximating the clique number, due to Håstad [Hås99] (whose proof requires much more advanced techniques). He showed that efficiently approximating $\omega(G)$ to within $n^{1-\delta}$, for any $\delta > 0$, implies that $\mathbf{NP} = \mathbf{RP}$ (via a probabilistic reduction). This was very recently derandomized by Zuckerman [Zuc05], again via more difficult techniques than mere expander walks, to yield that even such an approximation is actually $\mathbf{NP}$-hard. Note that since a factor-n approximation is trivial, this result is surprisingly tight.

The probabilistic reduction.

The randomized reduction of [BS92] yields a version of Theorem 3.13 where the same assumptions lead to a weaker conclusion:

Lemma 3.14 (Theorem 3.13, weak version)**.** *If there exists a polynomial-time algorithm A whose output on every n-vertex graph G satisfies $n^{-\epsilon} \leq A(G)/\omega(G) \leq n^{\epsilon}$, then $\mathbf{NP} \subseteq \mathbf{RP}$. Here $\epsilon > 0$ is some absolute constant.*

The proof follows by converting algorithm A into a probabilistic polynomial-time algorithm B that can solve the decision problem considered in Theorem 3.12. This shows that $\mathbf{NP} \subseteq \mathbf{RP}$.[3]

In order to apply algorithm B to a given n-vertex graph $G = (V, E)$, consider a graph H, with vertex set V^t, where $t = \log n$. The vertices $(v_1, \ldots, v_t)$ and $(u_1, \ldots, u_t)$ in H are adjacent if the subgraph of G induced by the set $\{v_1, \ldots, v_t\} \cup \{u_1, \ldots, u_t\}$ is a clique. Whether $\omega(G)$ is below $\delta_2 n$ or above $\delta_1 n$, this is significantly amplified in H. The amplification is so strong that a random subset of $m = \text{poly}(n)$ vertices in H tends to behave very differently with respect to the clique number of the induced graph.

Here is what algorithm B does on input $G = (V, E)$:

(1) Pick m random vertices from V^t and compute the subgraph H' of H induced on this set of vertices.
(2) Apply algorithm A to H'.
(3) Algorithm B returns 1 if $A(H') > \frac{1}{2}\delta_1^t m$, and otherwise it returns 0.

We need the following simple combinatorial observation.

Claim 3.14.1. *Every clique in H is contained in a clique of the form S^t where S is an inclusion-maximal clique in G. In particular, $\omega(H) = \omega(G)^t$.*

[3] This in fact shows only $\mathbf{NP} \subseteq \mathbf{BPP}$, but it is a standard fact (using the self-reducibility of $\mathbf{NP}$-complete problems) that the inclusion $\mathbf{NP} \subseteq \mathbf{BPP}$ actually implies the stronger conclusion $\mathbf{NP} \subseteq \mathbf{RP}$.

Proof. Clearly if S is a clique in G, then the set S^t is a clique in H, and in particular $\omega(H) \geq \omega(G)^t$. On the other hand, consider some clique S' in H. Let $S \subseteq V(G)$ be the set of those vertices in G that appear as an entry in any of the t-tuples in S'. Clearly S forms a clique in G, and $|S'| \leq |S|^t$, whence also $\omega(H) \leq \omega(G)^t$. $\square$

We need to show two things:

(1) If $\omega(G) \geq \delta_1 n$, then almost surely $\omega(H') \geq \frac{1}{2}\delta_1^t m$.
(2) If $\omega(G) \leq \delta_2 n$, then almost surely $\omega(H') \leq 2\delta_2^t m$.

With a proper choice of $m = \text{poly}(n)$ and $t = \log n$ this proves the lemma.

For the first claim, consider a clique Q in H of size $\omega(H) = \omega(G)^t \geq (\delta_1 n)^t$. The expected number of vertices from Q in H' is $|Q| \cdot \frac{|V(H')|}{|V(H)|} \geq \delta_1^t m$. By the Chernoff bound,[4] with high probability this intersection is at least $\frac{1}{2}\delta_1^t m$ as stated.

For the other claim we need to show that it is very unlikely that the m vertices we sample from H include a large clique. For this analysis it suffices to consider subsets of **inclusion-maximal** cliques Q in H, of which, by Claim 3.14.1, there are at most 2^n. The cardinality of Q does not exceed $(\delta_2 n)^t$, and so we expect to sample $\frac{|Q|m}{n^t} < \delta_2^t m$ vertices from Q. We consider it a **failure** if we sample more than $2 \cdot \delta_2^t m$ vertices from Q. Again by Chernoff's bound, the failure probability does not exceed $\exp(-\Omega(m\delta_2^t))$. As mentioned, there are at most 2^n inclusion-maximal cliques in H, and so the total probability of failure is still $o(1)$, provided that $m\delta_2^t \gg n$. This can be guaranteed by a proper choice of $m = \text{poly}(n)$.

The deterministic reduction. Again we assume there exists a polynomial-time algorithm A distinguishing between the two cases of Theorem 3.13. This time we use A to derive a **deterministic** polynomial-time algorithm B' that can distinguish between the two cases of Theorem 3.12, whence $\mathbf{NP} = \mathbf{P}$.

The only difference between algorithm B' and algorithm B is that algorithm B' uses a **derandomized sampling** to construct the graph H'. This is done as follows. Choose some (n, d, α)-expander $\mathcal{G}$ on the same vertex set as G. In order to select the vertices in H', we no longer take a random sample of t-tuples from $V(G)^t$. Rather we consider all t-tuples representing a **length** $(t-1)$ **walk** in the graph $\mathcal{G}$. The resulting graph H' has $m = nd^{t-1}$ vertices. Since d is fixed and $t = \Theta(\log n)$, it follows that m is polynomial in n. We are already familiar with the idea that length $(t-1)$ random walks on $\mathcal{G}$ should behave like random t-tuple in $|V|^t$. We need to establish this principle in the present context as well.

Claim 3.15. *If $\omega(G) \leq \delta_2 n$, then $\omega(H') \leq (\delta_2 + 2\alpha)^t m$.*

Proof. As before, a clique in H' corresponds to all length-$(t-1)$ walks in $\mathcal{G}$ that are confined to some clique in G. Consequently, $\omega(H')/m$ is the largest probability that such a walk is confined to some clique S in G (i.e., the maximum of such a probability over all choices of a clique S). By assumption $|S| \leq \omega(G) \leq \delta_2 n$, and our claim follows now by Theorem 3.9. $\square$

The complementary statement that we need is:

Claim 3.16. *If $\omega(G) \geq \delta_1 n$, then $\omega(H') \geq (\delta_1 - 2\alpha)^t m$.*

[4] This is an upper bound on the tail of a binomial distribution. Let Z be Binomial(N, p), i.e. the sum of N independent 0-1 random variables that are one with probability p. Then $\Pr[|Z - E[Z]| < \Delta \cdot E[Z]] < 2\exp(-Np\Delta^2/3)$, for all $0 < \Delta < 1$.

Proof. Let S be a clique in G of cardinality $|S| \geq \delta_1 n$. By Theorem 3.9 a random walk of length $(t-1)$ in $\mathcal{G}$ remains confined to S with probability $(\delta_1 - 2\alpha)^t$. The conclusion follows. $\square$

The rest of the proof follows as above.

4. A geometric view of expander graphs

An attractive feature of expander graphs is that they can be viewed from many different angles. As we saw already, these combinatorial objects have a variety of applications in the design of algorithms and in computational complexity. The relationship between expansion and the spectral gap adds an algebraic perspective to the picture. Probabilistic considerations arise in the study of rapidly mixing random walks on graphs. In this section we investigate some geometric aspects of expander graphs.[5]

4.1. The classical isoperimetric problem.

The fact that a given graph G is a good expander is equivalent to the statement that G satisfies a certain discrete isoperimetric inequality. Let us quickly recall the grandmother of all isoperimetric inequalities which goes back to the ancient Greeks.

Problem 4.1. Of all simple closed curves in the plane of a given length, which curve encloses the greatest area?

The ancient Greeks had no doubt about the correct answer, namely that the optimal curve is the circle. Proving this statement was a different matter altogether. The first rigorous proof was claimed by Jacob Steiner (1841), based on the so-called **Steiner Symmetrization**. A flaw in Steiner's argument was pointed out by Weierstass, but several valid proofs were found shortly afterwards. Some of these proofs do rely on Steiner's original ideas. As we show below, analogous ideas are very useful in the study of isoperimetric inequalities on graphs and in particular in recent work [LL06] on the Margulis expander (viz. subsection 2.2).

Here is the idea of Steiner Symmetrization. First observe that there is no loss of generality in considering only convex domains, for the convex hull of a nonconvex closed planar domain K has a larger area and a smaller circumference than K itself. So given that K is a compact convex set, let us symmetrize it as follows. We describe the symmetrization of K around the x-axis. For the general operation, one considers a rotated K. Let $[x_1, x_2]$ be the projection of K to the x-axis, and for $x \in [x_1, x_2]$ let $y_1(x)$ and $y_2(x)$ be the smallest and largest values of y attained by some point $(x, y) \in K$. The resulting set is

$$K' = \{(x, y) : x \in [x_1, x_2] \text{ and } |y| \leq (y_2(x) - y_1(x))/2\}.$$

This transformation preserves the area and does not increase the circumference. Consequently the circumference of an optimal[6] K must be invariant under this operation. With a little extra work, one shows that an optimal K must, in fact,

[5] Just for the record, this is only a partial list of topics that are related to expander graphs. For example, most of the known explicit constructions of expander graphs depend on deep ideas from number theory and representation theory.

[6] ... should the optimum exist. This subtle issue was pointed out by Weierstrass.

be invariant under such a symmetrization step. From here it is only a short route to proving that the optimal K is a disc. Let us point out that the subtle points concerning the existence of an optimum do not bother us in the realm of finite graphs. For a general survey of geometric isoperimetric problems, see the book by Burago and Zalgaller [BZ88]. Siegel [Sie] provides a historical review.

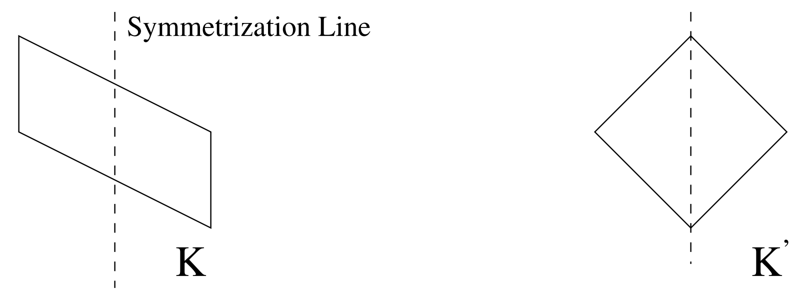


FIGURE 2. Steiner Symmetrization.

4.2. **Graph isoperimetric problems.** In the spirit of the classical isoperimetric problem, one can consider analogous problems in graphs rather than in Euclidean spaces. In this analogy the "area" of a set of vertices S is its cardinality $|S|$. There are two main natural discrete analogs for the "circumference" of S. We can consider the number of edges going out from S or the number of vertices outside S that have a neighbor in it. This leads us to define the two basic isoperimetric parameters for a given n-vertex graph G and an integer $n > k \geq 1$.

Definition 4.2. The edge isoperimetric parameter:

$$\Phi_E(G,k) = \min_{S \subset V} \{|E(S,\overline{S})| \ : \ |S| = k\}.$$

Definition 4.3. The vertex isoperimetric parameter:[7]

$$\Phi_V(G,k) = \min_{S \subset V} \{|\Gamma(S) \setminus S| \ : \ |S| = k\}.$$

It is desirable to completely understand these parameters for natural families of graphs (see below). It is also of great interest to determine or estimate these parameters for given G and k. In this generality, these computational problems are difficult (co-**NP**-hard) [BKV81]. Much (and still ongoing) research was dedicated to the search for efficient algorithms to approximate these quantities. Deep and surprising connections were found between these problems and the subject of Metric Embeddings; see Section 13.

We illustrate these concepts with an important family of graphs for which the two basic isoperimetric problems are completely solved. The d-dimensional discrete cube Q_d is a graph on vertex set $V(Q_d) = \{0,1\}^d$. Two vertices $v_1, v_2 \in \{0,1\}^d$ are adjacent if these two vectors differ in exactly one coordinate. The graph of the d-dimensional cube plays a major role in many parts of discrete mathematics, and

[7] This parameter counts the number of vertices in $\bar{S}$ with a neighbor in S. Other interesting variants may be defined, e.g. (i) the number of vertices in S with a neighbor in $\bar{S}$, or (ii) the more symmetric $|\{x \in S | x \text{ has some neighbors in } \bar{S}\}| + |\{y \in \bar{S} | y \text{ has some neighbors in } S\}|$. However, the present definition seems to be the most useful for applications.

both the vertex and the edge-isoperimetric inequalities on the d-cube are useful in a variety of applications.

4.2.1. *Example: The discrete cube.* As stated above, both the vertex and edge-isoperimetric parameters for the d-cube are known for every d and $2^d > k > 0$. To reduce the technicalities we state these results only for certain values of k and in an asymptotically tight form. A thorough survey of these and related problems can be found in [Bol86].

- $\Phi_E(Q_d, k) \geq k(d - \log_2 k)$. This bound holds with equality when k is a power of two, $k = 2^l$. In this case equality holds for S that is the set of vertices of an l-dimensional subcube.
- If $k = \binom{d}{0} + \binom{d}{1} + \ldots + \binom{d}{r}$ (for some integer r), then the vertex isoperimetric parameter $\Phi_V(Q_d, k) = \binom{d}{r+1}$. Equality is achieved for the set S that is a ball of radius r around some vertex v_0, namely $\{v \in \{0,1\}^d : d_H(v, v_0) \leq r\}$. (Recall that $d_H(u, v)$, the Hamming distance between $u, v \in \{0,1\}^d$, is the number of coordinates on which u, v differ.)

4.3. **The Margulis construction.** Margulis [Mar73] gave the first explicit construction of an infinite family of expander graphs. Section 8 is devoted to this construction and its analysis. Here we study it from a geometric point of view by considering an infinite analog. The vertex set of this infinite graph is the unit square (or torus) $I \times I$, where I is the half-open interval $[0, 1)$. The edges are defined by two linear transformations:

$$T(x, y) \to (x + y, y) \bmod 1, \quad S(x, y) \to (x, x + y) \bmod 1.$$

The neighbors of a point (x, y) are the points $T(x, y), S(x, y), T^{-1}(x, y), S^{-1}(x, y)$. Thus the graph is 4-regular. The expansion property of this graph is described by the following theorem:

Theorem 4.4 (Margulis [Mar73], Gabber-Galil [GG81])**.** *There exists an explicit $\epsilon > 0$ such that for any measurable set $A \subset I \times I$ of Lebesgue measure $\mu(A) \leq \frac{1}{2}$,*

$$\mu(\Gamma(A) \cup A) \geq (1 + \epsilon)\mu(A),$$

where $\Gamma(A) = S(A) \cup T(A) \cup S^{-1}(A) \cup T^{-1}(A)$ is the neighbor set of A.

It is natural to conjecture which sets are extremal for the isoperimetric problem in this graph.

Conjecture 4.5 (Linial)**.** For every $A \subset [0, 1]^2$ of Lebesgue measure $\mu(A) \leq 1/2$,

$$\mu(A \cup S(A) \cup T(A) \cup S^{-1}(A) \cup T^{-1}(A)) \geq 2\mu(A).$$

Also,

$$\mu(A \cup S(A) \cup T(A)) \geq \frac{4}{3}\mu(A).$$

If true, these bounds are clearly tight. If $A = \{(x, y) : |x| + |y| < t\}$, then $A \cup \Gamma(A) = \{(x, y) : |x|, |y| < t\}$, and $\mu(A \cup \Gamma(A)) = 2\mu(A)$. The second inequality is attained for the hexagon $A = \{(x, y) : |x|, |y|, |x + y| < t\}$.

The analogous (and weaker) version of these conjectures concerning the transformations $\tilde{T}(x, y) \to (x + y, y)$, and $\tilde{S}(x, y) \to (x, x + y)$ (no mod 1 here; namely the ground set is no longer the unit torus but rather the Euclidean plane $\mathbb{R}^2$) was recently proved by Linial and London [LL06]. Their proof is very brief and completely elementary. It is based on the idea of symmetrization described above.

4.3.1. *The discrete Laplacian.* In classical vector analysis, the Laplace operator is defined as $\Delta(f) = \mathrm{div}(\mathrm{grad}(f))$. It turns out that it has an analogue, the discrete Laplacian, which is natural and useful for several reasons. To present this analogy, let us begin by introducing the discrete analogs for the gradient, and divergence operators in graphs. The correct definition for the Laplacian will then be apparent. Given an undirected graph $G = (V, E)$, we fix an arbitrary orientation of the edges. (The specific choice of the orientation does not affect anything in our discussion.) Let K be the $V \times E$ incidence matrix of G where the entry

$$K_{u,e} = \begin{cases} +1 & \text{if the edge } e \text{ exits the vertex } u \\ -1 & \text{if the edge } e \text{ enters the vertex } u \\ 0 & \text{otherwise.} \end{cases}$$

We now define:

The gradient: Let $f : V \to \mathbb{R}$ be a function on the vertices of G which we view as a row vector indexed by V. The gradient operator maps f to fK, a vector indexed by E. The gradient measures the change of f along the edges of the graph. If e is the edge from u to v, then $(fK)_e = f_u - f_v$.

The divergence: Let $g : E \to \mathbb{R}$ be a function on the edges of G. The divergence operator maps g, considered as a column vector indexed by E, to Kg, a vector indexed by V. If we think of g as describing a flow, then its divergence at a vertex is the net outbound flow, namely,

$$(Kg)_v = \sum_{\text{e exits v}} g_e - \sum_{\text{e enters v}} g_e.$$

The Laplacian: In order to maintain the analogy with the real Laplacian, the discrete Laplacian should map f to $KK^T f$, where $f : V \to \mathbb{R}$. The matrix $L = L_G = KK^T$ is accordingly called the (discrete) Laplacian of G.[8] A simple calculation shows that L is the following symmetric matrix with rows and columns indexed by V:

$$L_{u,v} = \begin{cases} -1 & (u,v) \in E \\ \deg(u) & u = v. \end{cases}$$

One can easily deduce the following equality:

$$fLf^T = fKK^Tf^T = ||fK||^2 = \sum_{(u,v)\in E} (f(u) - f(v))^2, \tag{1}$$

where $||\cdot||$ denotes the l_2 norm.

In particular:

Proposition 4.6. *For every graph G, the matrix L_G is positive semidefinite. Its smallest eigenvalue is zero, and the corresponding eigenfunction is the constant function.*

Though it is possible to develop some of the theory for general (irregular) graphs, the regular case is simpler to state and analyze.

Lemma 4.7. *The Laplacian of a d-regular graph G satisfies:*

- *$L = L_G = dI - A_G$, where A_G is G's adjacency matrix.*

[8] In the graph theory literature it is also called the Tutte Matrix of G and appears in contexts such as the Matrix-Tree Theorem and randomized algorithms for matchings in graphs.

- *The spectrum of L is in $[0, +2d]$ (since the spectrum of A_G is in $[-d, +d]$).*
- *The smallest eigenvalue of L is zero.*
- *The spectral gap of G, namely $\lambda_1(A_G) - \lambda_2(A_G)$, equals the smallest positive eigenvalue of L.*

4.4. **The Cheeger constant and inequality.** Many of the things we do here for graphs had been studied previously in the geometric framework of Riemannian manifolds. In particular, the Cheeger constant that we introduce now captures a notion of "expansion" in this geometric context. The geometry underlying this theory is that of an n-dimensional Riemannian manifold. We skip any formal definitions here, but note that Riemannian geometry is a deep and technical setting; we refer the reader to Buser's book [Bus82]. It suffices at this high level to say that this is a space that looks locally like $\mathbb{R}^n$ and carries a differentiable structure with a smoothly varying notion of inner product among tangent vectors. This allows us to carry out the familiar operations from calculus, compute volumes and distances. We now define the continuous analog of edge expansion for manifolds: the Cheeger constant.

Definition 4.8. The **Cheeger constant** of a compact n-dimensional Riemannian manifold M is[9]

$$h(M) = \inf_A \mu_{n-1}(\partial A) / \min(\mu_n(A), \mu_n(M \setminus A)),$$

where A runs over all open subsets of M and ∂A is the boundary of A. The n and $n-1$ dimensional measures are denoted μ_n and μ_{n-1}.

The analogy with the definition of edge expansion should be obvious: We partition M into A and its complement $M \setminus A$, and consider the ratio between two quantities: (i) the ($(n-1)$-dimensional) measure of the boundary of A and (ii) the minimum between the measure of A and its complement. These measures are n-dimensional.

An intuitive demonstration of the definition is illustrated in Figure 3. Here is a "dictionary" to move between the theory for graphs and the geometric counterpart.

$$\begin{array}{ccc} M & & G \\ \partial A & & E(S, \overline{S}) \\ A, M \setminus A & \Longleftrightarrow & S, \overline{S} \\ \mu_{n-1}(\partial A) & & |E(S, \overline{S})| \\ \min(\mu_n(A), \mu_n(M \setminus A)) & & \min\{|S|, |\overline{S}|\} \end{array}$$

As hinted above, it is possible to develop everything in differential calculus in the broader context of Riemannian manifolds. In particular, associated with any Riemannian manifold M is the Laplacian, a linear differential operator defined on real functions $f : M \to \mathbb{R}$. It is defined in the familiar way via $\Delta(f) = \mathrm{div}(\mathrm{grad}(f))$. If $\Delta f = \lambda f$, we say that f is an eigenfunction of the Laplacian with eigenvalue λ.

[9]The original definition of Cheeger's constant was in a slightly different context and did not involve $M \setminus A$.

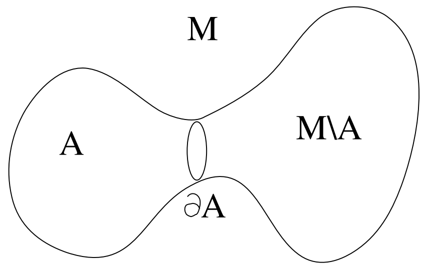


FIGURE 3. The sets A, $M \setminus A$ and ∂A for some Riemannian manifold M.

It can be shown that all eigenvalues of Δ are nonnegative and that its lowest eigenvalue is zero, corresponding to the constant eigenfunction.[10] A fundamental theme in this area is the connection between expansion (Cheeger constant h) and the spectrum of the Laplacian.

Theorem 4.9 (Cheeger [Che70]). *Let M be a compact Riemannian manifold, and let λ be the smallest positive eigenvalue of its Laplacian. Then $\lambda \geq h^2/4$.*

Below we will derive the discrete analog of this theorem.

4.5. **Expansion and the spectral gap.** We wish to prove a discrete analogue of Theorem 4.9, namely the qualitative equivalence between expansion and spectral gaps in graphs. This theorem was already stated without proof in Section 2. We restate the theorem and prove it here. Recall first the definition of (edge) expansion.

Definition 4.10. The edge expansion ratio of a graph $G = (V, E)$ is

$$h(G) = \min_{S \subseteq V, |S| \leq |V|/2} \frac{|E(S, \overline{S})|}{|S|}.$$

Theorem 4.11. *Let $G = (V, E)$ be a finite, connected, d-regular graph and let λ be its second eigenvalue. Then*

$$\frac{d - \lambda}{2} \leq h(G) \leq \sqrt{2d(d - \lambda)}.$$

This theorem was proved by Dodziuk [Dod84], and independently by Alon-Milman [AM85], and Alon [Alo86].

This theorem was also generalized in several ways, all beyond the scope of this manuscript. We just mention the most useful generalization, to general **reversible** Markov chains. For these, one can define a weighted analog of edge expansion called **conductance** and give similar bounds on it in terms of the spectral gap of the chain. This was done by Jerrum and Sinclair [JS89] and had a huge impact on the analysis of convergence of "Monte Carlo" algorithms, used extensively in statistical physics and optimization (see [JS96] for a survey).

The following two examples show that, up to a constant factor, both the upper and the lower bound in Theorem 4.11 can be tight.

---

[10] The standard definition of eigenvalues and eigenfunctions of the Laplacian is usually made through the variational principle. For compact manifolds, the infimum value of each Rayleigh quotient is attained by a corresponding eigenfunction.

(1) **The lower bound is tight:** Our discussion of $G = Q_d$, the d-dimensional cube, in subsection 4.2.1 yields that $h(G) = 1$ (the bound is attained for the $d-1$ dimensional subcube). On the other hand, the spectral gap is $d - \lambda = 2$.
(2) **The upper bound is tight:** The n-vertex cycle: This is the 2-regular graph C_n on the vertices $\{0, \ldots, n-1\}$, where vertex i is adjacent to $i+1$ and $i-1$ mod n. Here $h(C_n) = \Theta(1/n)$ is attained on a half cycle, while $d - \lambda = \Theta(1/n^2)$.

For the above estimates of the spectral gap see subsection 11.1 or the Lovász problem book [Lov93].

The proof of Theorem 4.11 is given in the next two subsections.

4.5.1. *Large spectral gap implies high expansion.* The proof is similar to that of the Expander Mixing Lemma 2.5, but using the fact that the two sets are complementary. The first eigenvector of a regular graph is the all-ones vector $\mathbf{1} = (1, \ldots, 1)$, so we can prove that $\lambda \geq d - 2h(G)$ by exhibiting a vector $f \perp \mathbf{1}$ with a large **Rayleigh quotient** $fAf^T/||f||^2 \geq d - 2h(G)$. The vector we consider is $f = |\overline{S}|1_S - |S|1_{\overline{S}}$, where 1_X denotes the characteristic vector of the set X. Here S is a set satisfying $h(G) = |E(S,\overline{S})|/|S|$, and $|S| \leq n/2$. Let us evaluate the Rayleigh quotient:

$$\begin{aligned} ||f||^2 &= |\overline{S}|^2|S| + |S|^2|\overline{S}| = |S||\overline{S}|(|S| + |\overline{S}|) = n|S||\overline{S}|, \\ fAf^t &= 2\left(|E(S)||\overline{S}|^2 + |E(\overline{S})||S|^2 - |S||\overline{S}||E(S,\overline{S})|\right). \end{aligned}$$

Since G is d-regular, we can substitute

$$\begin{aligned} 2|E(S)| &= d|S| - |E(S,\overline{S})| \\ 2|E(\overline{S})| &= d|\overline{S}| - |E(S,\overline{S})|. \end{aligned}$$

Putting it all together one concludes that

$$\lambda \geq \frac{fAf^t}{||f||^2} = \frac{nd|S||\overline{S}| - n^2|E(S,\overline{S})|}{n|S||\overline{S}|} = d - \frac{n|E(S,\overline{S})|}{|S||\overline{S}|} \geq d - 2h(G).$$

The last inequality follows from the fact that $h(G) = |E(S,\overline{S})|/|S|$, and $|\overline{S}| \geq n/2$.

4.5.2. *High expansion implies large spectral gap.* This is the more difficult (and interesting) direction. Let g be the eigenvector associated with λ_2. Based on the knowledge of g, we seek a cut with relatively few edges. Had it been the case that the function g takes only two values, the obvious thing is to partition the vertices according to the two values of g. In this idealized case, we need not give away a square in the inequality. For a general g, we partition according to the **sign** of g (since $g \perp \mathbf{1}$ it has both positive and negative entries). We then need to bound the edge expansion of the corresponding cut.

Define $f = g^+$ and $V^+ = \mathrm{supp}(f)$, i.e. $f_v = \max(g_v, 0)$ and $V^+ = \{v : f_v > 0\}$. Without loss of generality, V^+ contains at most $n/2$ vertices (or else we consider $-g$, which is also an eigenvector with the same eigenvalue λ). The result is obtained by considering the Rayleigh quotient $fLf^T/||f||^2$ and proving that (i) $fLf^T/||f||^2 \leq d - \lambda$ and (ii) $h^2/2d \leq fLf^T/||f||^2$.

We start by proving (i). First observe that for $x \in V^+$ we can write

$$\begin{aligned}(Lf)_x &= df_x - \sum_{y \in V} a_{xy} f_y = dg_x - \sum_{y \in V^+} a_{xy} g_y \\ &\leq dg_x - \sum_{y \in V} a_{xy} g_y = (Lg)_x = (d-\lambda) \cdot g_x.\end{aligned}$$

As $f_x = 0$ for $x \notin V^+$ we obtain that

$$fLf^T = \sum_{x \in V} f_x \cdot (Lf)_x \leq (d-\lambda) \sum_{x \in V^+} g_x^2 = (d-\lambda) \sum_{x \in V} f_x^2 = (d-\lambda)||f||^2.$$

It remains to prove (ii). To this end, we introduce yet another quantity:

$$B_f = \sum_{(x,y) \in E} |f_x^2 - f_y^2|,$$

and prove that

$$h \cdot ||f||^2 \leq B_f \leq \sqrt{2d} \cdot ||fK|| \cdot ||f||, \tag{2}$$

which clearly yields (ii), as $||fK||^2 = fLf^T$.

For ease of notation, we label the n vertices of G by $1, \ldots, n$, so that $f_1 \geq f_2 \geq \cdots \geq f_n$, and denote $[i] = \{1, \ldots, i\}$. The upper and lower bounds on B_f are proved in the following two lemmas. The upper bound is obtained using the Cauchy-Schwartz inequality, while the lower bound is obtained by considering the expansion of level sets, $[i]$ for $i \in V^+$.

Lemma 4.12. *The following inequality holds:* $B_f \leq \sqrt{2d} \cdot ||fK|| \cdot ||f||$.

Proof. Using the Cauchy-Schwartz inequality, we have

$$\begin{aligned}B_f &= \sum_E |f^2(x) - f^2(y)| = \sum_E |f(x) + f(y)| \cdot |f(x) - f(y)| \\ &\leq \sqrt{\sum_E (f(x)+f(y))^2} \cdot \sqrt{\sum_E (f(x)-f(y))^2}.\end{aligned}$$

The required result is obtained by evaluating the two factors:

$$\begin{aligned}\sqrt{\sum_E (f(x)-f(y))^2} &= ||fK|| \\ \sqrt{\sum_E (f(x)+f(y))^2} &\leq \sqrt{2\sum_E (f^2(x)+f^2(y))} = \sqrt{2d \sum_V f^2(x)} = \sqrt{2d} \cdot ||f||.\end{aligned}$$

□

Lemma 4.13. *The following inequality holds:* $B_f \geq h \cdot ||f||^2$.

Proof. This inequality intuitively says that the given eigenvector somehow approximates the optimal cut. We now make it precise. Rewrite B_f in terms of the values of f and the sizes of cuts $E([i], \overline{[i]})$ for $i \in V^+$. Then use expansion and the assumption that $|V^+| \leq n/2$ to give a lower bound on the number of edges in these cuts:

$$\begin{aligned} B_f &= \sum_{(x,y)\in E, x<y} (f_x^2 - f_y^2) = \sum_{(x,y)\in E, x<y} \sum_{i=x}^{y-1} (f_i^2 - f_{i+1}^2) \\ &= \sum_{i=1}^{n-1} (f_i^2 - f_{i+1}^2) \cdot |E([i], \overline{[i]})| = \sum_{i\in V^+} (f_i^2 - f_{i+1}^2) \cdot |E([i], \overline{[i]})| \\ &\geq h \sum_{i\in V^+} (f_i^2 - f_{i+1}^2) \cdot i = h \sum_{i\in V^+} f_i^2 = h \cdot \|f\|^2. \end{aligned}$$

The last equality is obtained by collapsing the telescopic sum and observing the $f_{i+1} = 0$ for $i = |V^+|$. □

4.6. **Expansion of small sets.** As mentioned before, there is great interest in the edge and vertex expansion of sets of varying sizes as captured by the parameters $\Phi_E(G,k)$ and $\Phi_V(G,k)$ and other parameters mentioned below. Typically, smaller sets exhibit better expansion, and this fact is crucial for certain applications (for more on this see Section 10). In this section we explore the expansion of small sets from two different perspectives: (i) connection with the spectral gap and (ii) typical behavior - expansion of small sets in random regular graphs.

4.6.1. *Connection with the spectral gap.* Theorem 4.11 reveals the connection between the spectrum of a graph (specifically $\lambda(G)$) and its expansion, $h(G)$. Here we present several variations on the theme of Theorem 4.11 that pertain to smaller sets.

Let us consider the following expansion parameters of a graph $G = (V, E)$ (compare with subsection 4.2):

$$\Psi_E(G,k) = \min_{\substack{S\subset V\\|S|\leq k}} \frac{|E(S,\overline{S})|}{|S|};\ \Psi_V(G,k) = \min_{\substack{S\subset V\\|S|\leq k}} \frac{|\Gamma(S)\setminus S|}{|S|};\ \Psi_V'(G,k) = \min_{\substack{S\subset V\\|S|\leq k}} \frac{|\Gamma(S)|}{|S|}.$$

The best known lower bound on vertex expansion for small sets is due to Kahale.

Theorem 4.14 (Kahale [Kah95])**.** *There is an absolute constant c such that an (n,d,α)-graph G satisfies the following inequality for all $\rho > 0$:*

$$\Psi_V'(G,\rho n) \geq (d/2)\cdot(1-\sqrt{1-4(d-1)/(d^2\alpha^2)})\cdot(1-c\log d/\log(1/\rho)).$$

How good a bound does this yield? As we'll see below (Theorem 5.3) the second eigenvalue of every (n,d)-graph is $\lambda(G) = d\alpha \geq 2\sqrt{d-1} - o(1)$. With this value of α and with ρ approaching zero, Kahale's bound yields vertex expansion of $d/2$ for small linear sized sets. The same paper [Kah95] contains a construction showing that the above bound is close to tight. The construction takes a Ramanujan graph G, i.e. a graph with $\lambda(G) \leq 2\sqrt{d-1}$, and by performing a small change, reduces the expansion without significantly increasing $\lambda(G)$. The resulting graph G' has two vertices with the same d-neighbors, i.e. expansion at most $d/2$, but $\lambda(G') \leq 2\sqrt{d-1} + o(1)$. It is not known if anything more substantial than that can happen in graphs with $\lambda \leq 2\sqrt{d-1} + o(1)$.

If we are willing to settle for vertex expansion of $d/4$ for small linear sized sets, then we can offer the following simple proof.

Theorem 4.15 (Tanner [Tan84])**.** *An (n,d,α)-graph G satisfies $\Psi_V'(G,\rho n) \geq 1/(\rho(1-\alpha^2)+\alpha^2)$ for all $\rho > 0$.*

Proof. Let S be some vertex set of cardinality ρn. The claim follows by comparing a lower and an upper bound on $||\hat{A}1_S||_2^2$, where 1_S is the characteristic function of S and $\hat{A}$ is the normalized adjacency matrix of G with eigenvalues $\hat{\lambda}_1 = 1, \ldots, \hat{\lambda}_n$.

Let the expansion of 1_S in the basis of the eigenvectors of $\hat{A}$ be $1_S = \sum_i a_i v_i$. Here $v_1 = (1 \cdots 1)/\sqrt{n}$, and $a_1 = |S|/\sqrt{n}$. Then:

$$\begin{aligned} ||\hat{A}1_S||_2^2 &= \sum_{i=1}^n \hat{\lambda}_i^2 a_i^2 \leq \frac{|S|^2}{n} + \sum_{i=2}^n \alpha^2 a_i^2 \leq \frac{|S|^2}{n} + \alpha^2(||1_S||_2^2 - a_1^2) \\ &= \rho n \cdot (\rho + \alpha^2(1-\rho)). \\ ||\hat{A}1_S||_2^2 &= \sum_{x\in\Gamma(S)} (|S\cap\Gamma(x)|/d)^2 \geq |S|^2/|\Gamma(S)| \\ &= \rho n \cdot |S|/|\Gamma(S)|, \end{aligned}$$

where the last inequality follows from Cauchy-Schwartz since $\sum_{x\in\Gamma(S)} |S\cap\Gamma(x)|/d = |S|$. The required result is obtained by putting together the two bounds. □

4.6.2. *Typical behavior.*

Theorem 4.16. *Let $d \geq 3$ be a fixed integer. Then for every $\delta > 0$ there exists $\epsilon > 0$ such that:*

(1) *For almost every (n, d)-graph G*

$$\Psi_E(G, \epsilon n) \geq d - 2 - \delta; \quad \Psi_V(G, \epsilon n) \geq d - 2 - \delta; \quad \Psi'_V(G, \epsilon n) \geq d - 1 - \delta.$$

(2) *For almost every d-regular bipartite graph G with n vertices on each side*

$$\Psi_V(G, \epsilon n) \geq d - 1 - \delta.$$

As observed below (subsection 5.1.1) these expansion parameters are best possible. By considering any connected subset S of s vertices we conclude that $\Psi_E(G, s) \leq d - 2 + \frac{2}{s}$, $\Psi_V(G, s) \leq d - 2 + \frac{2}{s}$, and $\Psi'_V(G, s) \leq d - 1 + \frac{2}{s}$. A similar argument shows that for bipartite graphs $\Psi_V(s) \leq d - 1 + \frac{1}{s}$.

The proof of the theorem is reminiscent of our proof for Lemma 1.9, though some extra care is needed here.

There is a nontrivial issue concerning the question of how to uniformly sample an (n, d)-graph, and this is done using the so-called "configuration model". In the proof below we say very little about this issue and refer the reader to subsection 7.2.1 for an additional brief discussion.

Proof. We start with the bipartite case, where G is a bipartite d-regular graph consisting of left and right vertex sets L, R, where $|L| = |R| = n$. To generate such a graph at random we let d half-edges emanate from each vertex and randomly match the dn left half-edges with the dn right half-edges to form the dn edges of the graph. Note that such a graph may have multiple edges.

Let $\eta = d - 1 - \delta$ denote the expansion we wish to prove. For sets $S \subset L$ and $T \subset R$, let $X_{S,T}$ be an indicator random variable for the event $\Gamma(S) \subset T$. It clearly suffices to prove that $\sum X_{S,T} = 0$ holds almost surely, where the sum is over all choices of S and T with $s = |S| \leq \epsilon n$ and $t = |T| = \lfloor \eta s \rfloor$. Note that smaller values of t can be safely ignored. Since for any sets S, T the probability of $X_{S,T} = 1$ is $\frac{(td)_{sd}\cdot(nd-sd)!}{(nd)!}$, where $(n)_k = n(n-1)\cdots(n-k+1)$, we obtain the following upper bound:

$$\Pr[\sum X_{S,T} > 0] \le \sum_{s=1}^{\epsilon n} \binom{n}{s}\binom{n}{t} \cdot \frac{(td)_{sd} \cdot (nd-sd)!}{(nd)!},$$

where $t = \lfloor \eta s \rfloor$. Using the inequality $\binom{n}{k} \le (en/k)^k$ yields:

$$\begin{aligned}\Pr[\sum X_{S,T} > 0] &\le \sum_{s=1}^{\epsilon n} \left(\frac{en}{s}\right)^s \cdot \left(\frac{en}{t}\right)^t \frac{td}{nd}\frac{td-1}{nd-1}\cdots\frac{td-sd+1}{nd-sd+1} \\ &\le \sum_{s=1}^{\epsilon n} \left[\frac{en}{s} \cdot \left(\frac{en}{\eta s}\right)^\eta \cdot \left(\frac{\eta s}{n}\right)^d\right]^s = \sum_{s=1}^{\epsilon n} \left[c(\delta, d) \cdot (s/n)^\delta\right]^s,\end{aligned}$$

where $c(\delta, d)$ is a function that depends only on δ and d. Given $\delta > 0$, we pick a sufficiently small $\epsilon > 0$ to make the expression in the square brackets smaller than $1/10$ (this computation yields $\epsilon(\delta, d) = c^{-d/\delta}$ for some absolute constant $c > 1$). It is easy to check by considering small and large values for s separately that the entire sum is $o(1)$.

We next turn to the case of general graphs. Our first step is to prove that for every $d \ge 3$ and $\delta > 0$ there is some $\epsilon > 0$ such that for almost every (n, d)-graph every vertex set S of cardinality at most ϵn does not have too many internal edges:

$$(1 + \delta/2) \cdot |S| > |E(S)|. \tag{3}$$

Clearly, this inequality implies the two lower bounds $\Psi_E(G, \epsilon n) \ge d - 2 - \delta$ and $\Psi_V(G, \epsilon n) \ge d - 2 - \delta$. The third lower bound $\Psi'_V(G, \epsilon n) \ge d - 1 - \delta$ requires some more work, which we defer to the end. It is instructive to consider inequality (3) in both the case where S is an independent set and the case where S is connected. In the first case $|E(S)| = 0$, while in the second case $|E(S)| \ge |S| - 1$ so the inequality is almost tight.

In the configuration model the graph G is generated by letting d half-edges emanate from each of the n vertices and picking a random perfect matching of the half-edges to form the $dn/2$ edges of the graph. To prove inequality (3) we define an indicator random variable $Y_{S,K}$, where S is a nonempty vertex set of cardinality $s \le \epsilon n$ and K is a set of half-edges from S of cardinality k. The variable $Y_{S,K}$ equals one if all half-edges in K are matched among themselves. Consequently, $\Pr[Y_{S,K} = 1]$ is the probability that a random matching of the dn half-edges of G matches the half-edges of K with themselves. Recalling the standard notation $l!! = (l-1)(l-3)\cdots 1$ for the number of perfect matchings of a set of size l, we obtain that $\Pr[Y_{S,K} = 1] = \frac{k!!\,(nd-k)!!}{(nd)!!}$. Therefore, the probability of failure is bounded by

$$
\begin{aligned}
\Pr[\sum Y_{S,K} > 0] &\leq \sum_{s=1}^{\epsilon n} \sum_{k=2(1+\delta/2)s}^{sd} \binom{n}{s}\binom{ds}{k} \frac{k!!\,(nd-k)!!}{(nd)!!} \\
&\leq \sum_{s=1}^{\epsilon n} \sum_{k=2(1+\delta/2)s}^{sd} \left(\frac{ne}{s}\right)^s \left(\frac{dse}{k}\right)^k \frac{k-1}{nd-1}\frac{k-3}{nd-3}\cdots\frac{1}{nd-k+1} \\
&\leq \sum_{s=1}^{\epsilon n} \sum_{k=2(1+\delta/2)s}^{sd} c(d,\delta)^s \left(\frac{s}{n}\right)^{-s+k/2} \\
&\leq \sum_{s=1}^{\epsilon n} sd\left[c(d,\delta)(s/n)^{\delta/2}\right]^s,
\end{aligned}
$$

which is $o(1)$ for a sufficiently small $\epsilon > 0$. Therefore, inequality (3) holds almost surely as claimed, and the required lower bounds on Ψ_E and Ψ_V follow.

We prove the lower bound on Ψ'_V by showing that in almost every (n,d)-graph, every vertex set S of cardinality at most ϵn satisfies

$$|\Gamma(S)\setminus S| + |E(S)| > (d-1-\delta/2)\cdot|S|. \tag{4}$$

The required lower bound on $|\Gamma(S)|$ can be obtained from inequalities (3) and (4) as follows. First apply inequality (3) to the set $S\cap\Gamma(S)$ to obtain

$$|S\cap\Gamma(S)| + (\delta/2)\cdot|S| \geq (1+\delta/2)\cdot|S\cap\Gamma(S)| > |E(S\cap\Gamma(S))| = |E(S)|.$$

Summing this inequality with (4) implies that $|\Gamma(S)| > (d-1-\delta)|S|$, yielding the desired lower bound on Ψ'_V. Therefore it remains to prove inequality (4).

We define the indicator variable $Z_{S,R,K}$ for two disjoint vertex sets S, R of cardinalities s, r and a subset K of cardinality k of the ds half-edges from S. Given some graph G picked by the configuration model, $Z_{S,R,K}$ is one if the half-edges K are matched among themselves and the other $ds-k$ half-edges emanating from S are matched with half-edges from R. Then it suffices to prove that $\sum Z_{S,R,K} = 0$ holds almost surely, where the sum is on triplets S, R, K with $0 < s \leq \epsilon n$ and $r + \frac{k}{2} = (d-1-\delta/2)s$. We have

$$\Pr[Z_{S,R,K} = 1] = \frac{k!!\,(rd)_{sd-k}\,(nd-2sd+k)!!}{(nd)!!},$$

where the three factors in the numerator are the number of ways to match the half-edges in K with themselves, the number of ways to match the other $ds-k$ half-edges from S with half-edges from R, and the number of ways to match the remaining $nd-2sd+k$ half-edges with themselves. Therefore,

$$
\begin{aligned}
&\Pr[\sum Z_{S,R,K} > 0] \\
&\leq \sum_{s=1}^{\epsilon n} \sum_{\substack{r,k \\ r+k/2=(d-1-\delta/2)s}} \binom{n}{s}\binom{n-s}{r}\binom{ds}{k} \frac{k!!\,(rd)_{sd-k}\,(nd-2sd+k)!!}{(nd)!!} \\
&= \sum_{s=1}^{\epsilon n} \sum_{\substack{r,k \\ r+k/2=(d-1-\delta/2)s}} \binom{n}{s}\binom{n-s}{r}\binom{ds}{k} \frac{k!!\cdot[(rd)(rd-1)\cdots(rd-sd+k+1)]}{(nd-1)(nd-3)\cdots(nd-2sd+k+1)}.
\end{aligned}
$$

Since each of the $sd-k/2$ factors in the numerator does not exceed d^2s, and the $sd-k/2$ factors in the denominator are at least $nd-2sd$ each, one obtains

$$\begin{aligned}\Pr[\sum Z_{S,R,K}>0] &\leq \sum_{s=1}^{\epsilon n} \sum_{\substack{r,k \\ r+k/2=(d-1-\delta/2)s}} \left(\tfrac{ne}{s}\right)^s \left(\tfrac{ne}{r}\right)^r \left(\tfrac{dse}{k}\right)^k \left(\tfrac{ds}{n-2s}\right)^{sd-k/2} \\ &= \sum_{s=1}^{\epsilon n} \sum_{\substack{r,k \\ r+k/2=(d-1-\delta/2)s}} \left(\tfrac{ne}{s}\right)^s \left(\tfrac{ne}{s}\right)^r \left(\tfrac{s}{r}\right)^r (de)^k \left(\tfrac{s}{k}\right)^k \left(\tfrac{ds}{n-2s}\right)^{sd-k/2}.\end{aligned}$$

Therefore, since $(s/k)^{k/s}$ and $(s/r)^{r/s}$ are bounded by a constant,

$$\begin{aligned}\Pr[\sum Z_{S,R,K}>0] &\leq \sum_{s=1}^{\epsilon n} \sum_{\substack{r,k \\ r+k/2=(d-1-\delta/2)s}} c(d,\delta)^s \cdot \left(\frac{s}{n}\right)^{-s-r+sd-k/2} \\ &\leq \sum_{s=1}^{\epsilon n} sd \cdot \left[c(d,\delta)\cdot (s/n)^{\delta/2}\right]^s .\end{aligned}$$

As before, given $d \geq 3$ and $\delta > 0$, we choose a sufficiently small $\epsilon > 0$ such that the above probability is $o(1)$. This completes the proof of inequality (4) and of the lower bound on Ψ'_V. □

4.7. **Expansion in hypergraphs?** A *hypergraph* $H=(V,E)$ consists of a collection E of subsets of a set V. The sets in E are called edges or *hyperedges*. When each hyperedge has cardinality r we say that H is *r-uniform*. From this perspective, a graph is a 2-uniform hypergraph. In the same way that graphs can be viewed as one-dimensional complexes, such hypergraphs represent $(r-1)$-dimensional complexes and are, therefore, interesting from the geometric perspective of combinatorics. For this reason it is very interesting to try and develop a parallel theory of expansion to the one described here that applies to hypergraphs. This idea turns out to be rather difficult to carry out when one is trying to extend the notion of spectral gap. We refer the reader to some papers where (different) initial steps in this direction are taken, e.g. [FW95] and [LSV05]. When one considers combinatorial expansion one can use the definition of [BH04] or the concrete setting of extractors [Sha04]. The latter provides a successful theory (in which r cannot be a constant). It should be noted that at present we do not even have a satisfactory way of generating uniform hypergraphs at random; see [Cam].

5. **Extremal problems on spectrum and expansion**

Much research in modern Discrete Mathematics revolves around the study of **extremal problems**, and this is the topic of the present section. Here are several natural extremal problems about expansion and spectra:

- What is the largest expansion ratio (or vertex expansion, etc.) of an (n,d) graph?
- More generally, recall the edge and vertex isoperimetric parameters of a graph (subsections 4.2, 4.6):

$$\Phi_E(G,k)=\min_{S\subset V}\{|E(S,\overline{S})| \ : \ |S|=k\} \text{ and } \Phi_V(G,k)=\min_{S\subset V}\{|\Gamma(S)\setminus S| \ : \ |S|=k\}.$$

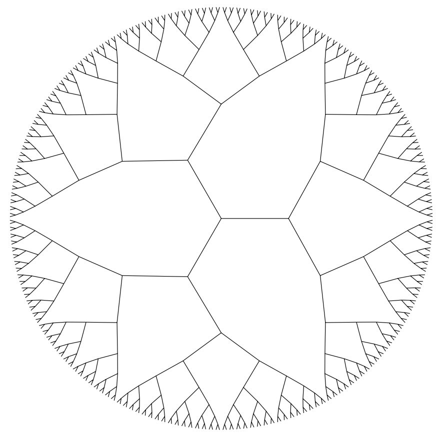


FIGURE 4. The 3-regular infinite tree.

We ask, given d, k and n, how large these parameters can be in an (n, d)-graph.

- How large can the spectral gap be in an (n, d) graph?

In the study of an extremal problem it is often beneficial to try to identify the problem's extremal instances. This is difficult for the present questions, but if we are willing to consider infinite graphs as well, then the extremal case is clearly the d-regular infinite tree $\mathbf{T_d}$, the ultimate expander. The question is how close one can get to this level of expansion with *finite* d-regular graphs.

In this light we start our discussion with an analysis of the d-regular tree. We observe that its edge expansion is $d-2$ and show that its spectrum is the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$. This sheds both some light and some shade (it is a tree after all...) on the analogous questions in finite graphs: (i) If G is an (n, d) graph, then even if $k \ll n$, the relative expansion of sets of k vertices cannot exceed $d-2+o_k(1)$, namely, $\Phi_E(G, k) \leq k(d-2+\frac{2}{k})$. (ii) As mentioned already, the Alon-Boppana theorem says that the second largest eigenvalue is at least $2\sqrt{d-1} - o(1)$.

5.1. **The d-regular tree.**

5.1.1. *The expansion of* $\mathbf{T_d}$. Consider the edge expansion function $\Phi_E(\mathbf{T_d}, k)$ of $\mathbf{T_d}$ (sections 4.2, 4.6). The minimizing set S must clearly be connected, i.e. a subtree. Therefore $|E(S)| = |S| - 1$, and (recalling that we are counting directed edges) $\Phi_E(\mathbf{T_d}, k) = kd - 2(k-1) = k(d-2) + 2$. Consequently, the expansion ratio of $\mathbf{T_d}$ is

$$h(\mathbf{T_d}) = \inf_{\text{finite } S \subseteq V} |E(S, \overline{S})|/|S| = d - 2.$$

The above argument implies that $\Phi_E(G, k) \leq k(d-2) + 2$ for every (n, d)-graph G and every k. However, $h(G)$ is necessarily smaller than $h(\mathbf{T_d}) = d - 2$. To see this, consider a random subset $S \subseteq V(G)$ of size $n/2$. The expected value of $|E(S, \overline{S})|/|S|$ is $d/2 + o(1)$, since every edge from G belongs to the random cut $E(S, \overline{S})$ with probability $1/2 + o(1)$. Therefore, there exists some set S of cardinality $n/2$ with $|E(S, \overline{S})|/|S| \leq d/2 + o(1)$, whence $h(G) \leq d/2 + o(1)$. A more refined analysis (see Alon [Alo97]) yields that $h(G) \leq d/2 - c\sqrt{d}$ for every

$d \geq 3$ and sufficiently large n. Here $c > 0$ is an absolute constant. This result is tight, up to the value of c. As we shall see later in Theorems 5.12 and 7.10, there are (n, d, α) graphs with $\alpha = O(d^{-1/2})$. The lower bound in Theorem 4.11 yields $h(G) \geq d/2 - c\sqrt{d}$ in this case.

Open problem 5.1. What is the smallest $f = f(d, k, n)$ such that every (n, d)-graph has a set S of k vertices and $|E(S, \overline{S})| \leq f$?

In full generality this question is exceedingly difficult and includes e.g. the girth problem. This question asks, for given integers $d, g \geq 3$, for the smallest n such that there is an (n, d) graph with no cycles of length shorter than g; see e.g. [Hoo02]. Many other instances of this question are interesting and may not be as difficult as the full problem.

5.1.2. *The spectrum of* $\mathbf{T_d}$. Let A_T be the (infinite) adjacency matrix of $\mathbf{T_d}$. We consider it as a linear operator on $l_2(V(\mathbf{T_d}))$, the space of real square summable functions on the vertices of the tree, and seek its spectrum. We recall some of the basic definitions from operator theory and refer the reader to Rudin's book [Rud91] for a comprehensive discussion of spectra of general linear operators. As usual, we define the spectrum of A_T via $\mathrm{spec}(A_T) = \{\lambda : (A_T - \lambda I) \text{ is noninvertible}\}$. That is, λ is in the spectrum if $(A_T - \lambda I)$ has a nontrivial kernel **or** is not onto. For finite matrices the two conditions are obviously equivalent, and we determine whether λ is in the spectrum by seeking a nonzero eigenvector u satisfying $(A - \lambda I)u = 0$. In contrast, the matrix A_T has no eigenvectors and its entire spectrum follows from the second reason.

One can go a step further and compute the **spectral measure** of A_T that corresponds to the eigenvalue distribution for finite graphs. We will say more about it in Section 7. But, for now, we return to the problem of computing the spectrum of $\mathbf{T_d}$.

Theorem 5.2 (Cartier [Car72]). *The spectrum of the infinite tree* $\mathbf{T_d}$ *is*

$$spec(A_T) = [-2\sqrt{d-1}, 2\sqrt{d-1}].$$

Partial Proof. (Friedman [Fri91])

Fix some vertex $v \in V(\mathbf{T_d})$ as the root of the tree. Then

$$\lambda \in \mathrm{spec}(A_T) \iff \delta_v \notin Range(\lambda I - A_T),$$

where δ_v is the characteristic function of v,

$$\delta_v(u) = \begin{cases} 1 & u = v \\ 0 & u \neq v. \end{cases}$$

The necessity of this condition is obvious. Sufficiency is not hard either, but it will not be proven here.

We wish to find out for which values of λ there is a function $f \in l_2$ satisfying

$$\delta_v = (\lambda I - A)f. \tag{5}$$

We say that a function f on $V(\mathbf{T_d})$ is **spherical around the vertex** v if $f(u)$ depends only on the distance (in the graph metric of $\mathbf{T_d}$) from u to v. The *spherical symmetrization* of $g \in l_2$ around v is a spherical function f such that $\sum_{\mathrm{dist}(u,v)=r} f(u) = \sum_{\mathrm{dist}(u,v)=r} g(u)$ for every $r \geq 0$. It is easy to observe that if $g \in l_2$ is a solution to (5), then g's spherical symmetrization f is in l_2 and satisfies the same equation. We may therefore assume without loss of generality that f is spherical. A spherical f is determined by a sequence of numbers $\{x_0, x_1, \ldots\}$ such that $f(u) = x_i$ whenever $d_T(u, v) = i$. This reduces (5) to the following recurrence:

$$\begin{aligned} \lambda x_0 &= dx_1 + 1 \\ \lambda x_i &= x_{i-1} + (d-1)x_{i+1} \quad \text{for } i \geq 1. \end{aligned} \tag{6}$$

The solution to such a recurrence is $x_i = \alpha\rho_1^i + \beta\rho_2^i$, where $\rho_{1,2} = \frac{\lambda \pm \sqrt{\lambda^2 - 4(d-1)}}{2(d-1)}$ are the roots of the quadratic equation $\lambda\rho = 1 + (d-1)\rho^2$.

If $|\lambda| < 2\sqrt{d-1}$ the roots are complex with absolute value $1/\sqrt{d-1}$. In this case, $\lambda \in \operatorname{spec}(A)$ since f is not in l_2. Indeed, $|x_i| = \Theta((d-1)^{-i/2})$, and since there are $\Theta((d-1)^i)$ vertices at distance i from v, it follows that $\|f\|_2 = \infty$.

We claim that when $|\lambda| > 2\sqrt{d-1}$ Equation (6) has a solution in l_2, implying that λ is not in the spectrum of A. To see that, observe that $|\rho_1| < 1/\sqrt{d-1}$ and so if $x_i = \alpha\rho_1^i$, then the resulting f belongs to l_2 and satisfies Equation (6) for all $i > 0$. It is left to verify that there is a value of α for which Equation (6) holds for $i = 0$ as well. This additional condition reads $\lambda\alpha = d\alpha\rho_1 + 1$, which is possible iff $\lambda \neq d\rho_1$. Indeed, $|\rho_1| < |\lambda|/2(d-1) \leq |\lambda|/d$ for all $d \geq 2$. □

5.2. **The Alon-Boppana lower bound.**

5.2.1. *Statement of the theorem.* In this section we return to the question how small λ_2 can be in a large d-regular graph. A weak bound of this form was given in subsection 2.5, and here we prove the Alon-Boppana lower bound $\lambda_2 \geq 2\sqrt{d-1} - o_n(1)$.

Theorem 5.3 (Alon-Boppana, Nilli [Nil91], Friedman [Fri93]). *There exists a constant c such that for every (n, d)-graph G of diameter Δ:*

$$\lambda_2(G) \geq 2\sqrt{d-1} \cdot (1 - c/\Delta^2).$$

Since the diameter of an (n, d)-graph is $\Omega(\log_{d-1} n)$, it follows that

Corollary 5.4. *For every (n, d)-graph*

$$\lambda_2 \geq 2\sqrt{d-1} \cdot (1 - O(1/\log^2 n)).$$

We present two proofs of this theorem. The first proof illustrates the **moment method**, which we will encounter later in Section 7. We first define the **tree number** t_{2k} as the number of closed walks of length $2k$ that start and end at some given vertex in $\mathbf{T_d}$. The key fact of the proof is that the analogous quantity in any regular graph is $\geq t_{2k}$. We also use the fact that **the spectral radius** $\rho(\mathbf{T_d}) = \max\{|\lambda| | \lambda \in \operatorname{spec}(\mathbf{T_d})\} = 2\sqrt{d-1}$. This proof achieves a little less than we desire: (i) it only bounds $\lambda(A) = \max_{i \geq 2} |\lambda_i(A)|$, and (ii) the resulting error term is slightly weaker than in Theorem 5.3. The second proof yields tighter estimates by computing the Rayleigh quotient of a certain function. This function is an eigenfunction of a truncation of the tree.

5.2.2. *Proof I: Counting closed walks in $\mathbf{T_d}$.* Let A be the adjacency matrix of G. Clearly, $\lambda(A^{2k}) = (\lambda(A))^{2k}$ for every integer k, where $\lambda(A) = \max_{i \geq 2} |\lambda_i(A)|$. We give a lower bound on $\lambda(A^{2k})$ by estimating the Rayleigh quotient of the function $f = \delta_s - \delta_t$, where s, t are two vertices at distance Δ in G. That is $f(s) = 1$,

$f(t) = -1$, and $f(u) = 0$ for any other vertex u. Since f is orthogonal to the first eigenfunction, the constant function,

$$\lambda^{2k} \geq \frac{fA^{2k}f^T}{||f||^2} = \frac{(A^{2k})_{ss} + (A^{2k})_{tt} - 2(A^{2k})_{st}}{2}.$$

Choose $k = \lfloor \frac{\Delta-1}{2} \rfloor$, so the negative term in the numerator vanishes. The positive terms in the numerator count closed walks of length $2k$ that start and end at s and t respectively and are therefore $\geq t_{2k}$. Consequently,

$$\lambda^{2k} \geq t_{2k}.$$

The tree numbers t_{2k} have been studied in great detail; see [McK81], and [Lub94]. Good estimates, a recursion, and their generating function are known, but all we need here is a rough estimate. (Slightly more will be needed below, Lemma 7.3.) Associated with every walk that starts and ends at the same vertex v in $\mathbf{T_d}$ is a **sign pattern**. Each step is associated a $+1$ or -1 according to its being directed either away or toward v. Clearly such a sign pattern is characterized by two conditions: (i) it sums up to zero, and (ii) the sum of each prefix is nonnegative. It is well known that the number of such sequences of length $2k$ is the k-th Catalan number $C_k = \binom{2k}{k}/(k+1)$ (see [vLW01]). Corresponding to every such sign pattern are at least $(d-1)^k$ walks, since there are exactly k occurrences of $+1$ in the sequence, and at least $d-1$ choices for moving away from v regardless of the current position of the walk. (For an accurate estimate of t_{2k} note that a prefix that sums to zero corresponds to a time when the walk reaches the vertex v, at which time the number of choices for the next step is d, not $d-1$). Therefore

$$\lambda^{2k} \geq t_{2k} \geq C_k \cdot (d-1)^k = \Theta((2\sqrt{d-1})^{2k} \cdot k^{-3/2}).$$

Taking the $2k$-th root and recalling that $k = \lfloor(\Delta-1)/2\rfloor$ yield:

$$\lambda(A) \geq 2\sqrt{d-1} \cdot (1 - O(\log \Delta \,/\, \Delta)).$$

5.2.3. *Proof II: Using spherical functions.* This argument follows Friedman [Fri93]. Here we derive a lower bound on $\lambda_2(A) = \max_{f \perp \mathbf{1}} fAf^T/||f||^2$ through a proper choice of a **test function** f. The function we use is an adaptation of an eigenfunction for a truncation of $\mathbf{T_d}$. Given two vertices s, t at distance Δ, we construct a function f that is positive on vertices at distance $\leq k = \lfloor \frac{\Delta}{2} \rfloor - 1$ from s, negative on vertices at distance $\leq k$ from t, and zero elsewhere. The values of f are derived from those of the eigenfunction g with maximal eigenvalue μ for the d-regular tree of height k. We view s and t as roots of (separate) k-tall trees. We show that f^+, the positive part of f, satisfies $Af^+ \geq \mu f^+$, and likewise for the negative part $Af^- \leq -\mu f^-$, so that $fAf^T \geq \mu||f||^2$ and $fAf^T/||f||^2 \geq \mu$. Finally, the positive and negative parts of f are normalized so that $\sum f(x) = 0$, so we can conclude that $\lambda_2(A) \geq \mu$.

We now get down to business. Let $k = \lfloor \frac{\Delta}{2} \rfloor - 1$ and select two vertices s, t in G at distance Δ. Classify the vertices according to their distance from s or t, along with a no-man's-land of the vertices that are far from both vertices:

$$\begin{aligned} S_i &= \{v : d(s,v) = i\} && \text{for } i = 0, \ldots, k, \\ T_i &= \{v : d(t,v) = i\} && \text{for } i = 0, \ldots, k, \\ Q &= V(G) \setminus \bigcup_{0 \leq i \leq k} (S_i \cup T_i). \end{aligned}$$

There are, of course, no edges between any S_i and T_j.

Let $\mathbf{T_{d,k}}$ denote the d-regular tree of height k, and let A_{T_k} be its adjacency matrix.

Claim 5.5. *Let μ be the largest eigenvalue of A_{T_k}. There is a unique function $g : V(\mathbf{T_{d,k}}) \to \mathbb{R}$ satisfying $A_{T_k} g = \mu g$. The function g is nonnegative and spherically symmetric.*

Proof. This can be verified either directly or by appealing to the Perron Frobenius theorem. The spherical symmetry can be verified as in the proof of Theorem 5.2. $\square$

Let g_i be the value that g takes on vertices at the i-th level. These numbers clearly satisfy the following recursion and boundary conditions:

$$\begin{aligned}
\mu g_0 &= d g_1, \\
\mu g_i &= g_{i-1} + (d-1) g_{i+1} \quad \text{for } i = 1, \ldots, k, \\
g_{k+1} &= 0.
\end{aligned} \tag{7}$$

Define $f : V(G) \to \mathbb{R}$ as follows:

$$f(v) = \begin{cases} c_1\, g_i & v \in S_i \\ -c_2\, g_i & v \in T_i \\ 0 & otherwise \end{cases},$$

where c_1, c_2 are nonnegative constants to be determined later. We next prove that f gives the desired properties:

Lemma 5.6. *If g is nonincreasing (as indeed will be shown below), then*

$$\begin{aligned}
(Af)_v \geq \mu f_v \quad & \textit{for} \quad v \in \cup_i S_i, \\
(Af)_v \leq \mu f_v \quad & \textit{for} \quad v \in \cup_i T_i.
\end{aligned}$$

Proof. Let $v \in S_i$ for some $i > 0$. Then of its d neighbors, $p \geq 1$ belong to S_{i-1}, q neighbors to S_i, and $d-p-q$ to S_{i+1}. Therefore,

$$(Af)_v \quad = \quad p \cdot c_1\, g_{i-1} + q \cdot c_1\, g_i + (d-p-q) \cdot c_1\, g_{i+1}.$$

Comparing with (7) and using the fact that g is nonnegative and nonincreasing, we obtain

$$\begin{aligned}
(Af)_v &= c_1 \cdot (p g_{i-1} + q g_i + (d-p-q) g_{i+1}) \\
&\geq c_1 \cdot (g_{i-1} + (d-1) g_{i+1}) \\
&= c_1 \cdot (A_{T_k} g)_i = c_1 \mu g_i = \mu f_v.
\end{aligned}$$

A similar argument works for $v = s$ and for $v \in \cup_i T_i$ to yield the required result. $\square$

Corollary 5.7. $\lambda_2(A) \geq \mu = \lambda_2(T_k)$.

Proof. First observe that the previous lemma implies that

$$\begin{aligned}
f A f^T &= \sum_{v \in V(G)} f_v (Af)_v \\
&= \sum_{v \in \cup_i S_i} f_v (Af)_v + \sum_{v \in \cup_i T_i} f_v (Af)_v + \sum_{v \in Q} f_v (Af)_v \\
&\geq \sum_{v \in \cup_i S_i} f_v \mu f_v + \sum_{v \in \cup_i T_i} f_v \mu f_v = \mu f f^T.
\end{aligned}$$

Also, a proper choice of c_1, c_2 gives $\sum_{v\in\cup_i S_i} f_v = -\sum_{v\in\cup_i T_i} f_v$ and hence $f \perp \mathbf{1}$. Therefore, $\lambda_2(A) \geq fAf^T/||f||^2 \geq \mu$ as claimed. $\square$

We still need to show that g is nonincreasing and prove a lower bound on μ. We do that by giving an explicit solution $h : \{0, \ldots, k+1\} \to \mathbb{R}$ to the recursion (7). By Claim 5.5, h must coincide with g. We then show that h is nonincreasing and derive the required lower bound on μ. Let

$$h_i = (d-1)^{-i/2} \cdot \sin((k+1-i)\theta)$$

where the parameter θ will be determined below. It is easy to check that $h_{k+1} = 0$ and that h is nonnegative and decreasing provided that $0 < \theta < \pi/(k+1)$. We claim that it satisfies the recursion (7), with $\mu = 2\sqrt{d-1}\cos\theta$. For $0 < i \leq k$ we have:

$$\begin{aligned} &h_{i-1} + (d-1)h_{i+1} \\ &\quad = (d-1)^{-(i-1)/2} \cdot [\sin((k+1-(i-1))\theta) + \sin((k+1-(i+1))\theta)] \\ &\quad = \sqrt{d-1} \cdot (d-1)^{-i/2} \cdot [\sin((k+2-i)\theta) + \sin((k-i)\theta)] \\ &\quad = 2\sqrt{d-1} \cdot (d-1)^{-i/2} \sin((k+1-i)\theta)\cos\theta = \mu \cdot h_i. \end{aligned}$$

The condition for $i = 0$ is that $\mu h_0 = dh_1$, or equivalently:

$$(2d-2) \cdot \cos\theta \cdot \sin((k+1)\theta) = d \cdot \sin(k\theta). \tag{8}$$

The smallest positive root of this equation, θ_0, satisfies $0 < \theta_0 < \pi/(k+1)$. This is because the difference between the two sides of (8) changes sign between 0 and $\pi/(k+1)$. Set $\theta = \theta_0$, and so the recursion (7) is satisfied and $h \geq 0$. Now $\theta_0 < \pi/(k+1) \approx 2\pi/\Delta$, since $k = \lfloor \Delta/2 \rfloor - 1$. By the Taylor expansion of the cosine,

$$\cos(\theta_0) > 1 - c/\Delta^2,$$

yielding the required lower bound on μ with constant $c \approx 2\pi^2$.

5.2.4. *Extensions of the Alon-Boppana theorem.* A quantitative variation of the Alon-Boppana theorem states that a constant fraction of the n eigenvalues must exceed $2\sqrt{d-1} - \epsilon$ for any fixed $\epsilon > 0$.

Theorem 5.8 (Serre). *For every integer d and $\epsilon > 0$ there is a constant $c = c(\epsilon, d)$, such that every (n, d) graph G has at least $c \cdot n$ eigenvalues greater than $2\sqrt{d-1} - \epsilon$.*

There are several available proofs of this theorem, e.g., [DSV03], [Fri93], [Nil04]. These papers also give the proper credit to the theorem's originator, J.-P. Serre. The best known lower bound on $c(\epsilon, d)$, by [Fri93, Nil04], is approximately $(d-1)^{-\pi\sqrt{2/\epsilon}}$. We present here an elegant proof by Cioabă [Cio06] that yields a slightly inferior bound.

Proof. Let A be the adjacency matrix of G. We seek a lower bound on n_ϵ, the number of eigenvalues larger than $2\sqrt{d-1} - \epsilon$. Consider the matrix $(A + dI)^k$, where the positive integer k will be specified below. On one hand

$$\begin{aligned} \operatorname{trace}(A + dI)^k &= \sum_{i=1}^{n} (\lambda_i + d)^k \\ &\leq (2d)^k \cdot n_\epsilon + (d + 2\sqrt{d-1} - \epsilon)^k \cdot n. \end{aligned} \tag{9}$$

On the other hand,

$$\operatorname{trace}(A+dI)^k = \sum_{j=0}^{k} \binom{k}{j} \cdot \operatorname{trace}(A^j) \cdot d^{k-j} \geq \sum_{l=0}^{\lfloor k/2 \rfloor} \binom{k}{2l} n \cdot t_{2l} \cdot d^{k-2l}.$$

Here we have eliminated the (positive) terms for odd j. We can use the estimates of the tree numbers t_j from subsection 5.2.2 to conclude

$$\begin{aligned}\operatorname{trace}(A+dI)^k &\geq (c'/k^{3/2}) \cdot \sum_{l=0,}^{\lfloor k/2 \rfloor} \binom{k}{2l} n \cdot (2\sqrt{d-1})^{2l} \cdot d^{k-2l} \\ &= (c'/2k^{3/2}) \cdot n \cdot [(d+2\sqrt{d-1})^k + (d-2\sqrt{d-1})^k] \\ &\geq (c'/2k^{3/2}) \cdot n \cdot (d+2\sqrt{d-1})^k\end{aligned}$$

for some absolute constant $c' > 0$. In combination with (9), this yields

$$\frac{n_\epsilon}{n} \geq \frac{(c'/2k^{3/2}) \cdot (d+2\sqrt{d-1})^k - (d+2\sqrt{d-1}-\epsilon)^k}{(2d)^k}.$$

This expression is positive for $k \geq \Omega(\frac{d}{\epsilon}\log(\frac{d}{\epsilon}))$, and the theorem follows. □

Open problem 5.9. What is the largest function $c(\epsilon, d)$ for which Theorem 5.8 holds?

Much less is known about the spectrum of **irregular graphs**. For example, the largest eigenvalue satisfies $\lambda_1 \geq d$, where d is the **average degree** of the graph. This is easily seen by considering the Rayleigh quotient of the constant function. But, is it true that the second eigenvalue satisfies $\lambda(G) \geq 2\sqrt{d-1} - o(1)$? Not necessarily. The **lollipop graph** L_n has $2n$ vertices and is obtained from an n-clique K_n and a path P_{n+1} on $n+1$ vertices by identifying some vertex of K_n with an end vertex of P_{n+1}. It is not hard to check that unlike a regular graph, the diameter and average degree of L_n are $\Theta(n)$, but $\lambda(L_n) \leq 2$. The following theorem shows that a simple additional condition yields the Alon-Boppana bound for irregular graphs.

Theorem 5.10 (Hoory [Hoo05]). *Let $d, r \geq 2$ be integers. Suppose that the average degree in the graph G is $\geq d$ whenever a ball of radius r is deleted from the graph. Then $\lambda(G) \geq 2\sqrt{d-1} \cdot (1 - c \cdot \log r/r)$, for some absolute constant $c > 0$.*

5.3. **Ramanujan graphs.** In light of the Alon-Boppana bound (Theorem 5.3) we define:

Definition 5.11. A d-regular graph G is **Ramanujan** if $\lambda(G) \leq 2\sqrt{d-1}$.

It is a major result discovered by Lubotzky-Phillips-Sarnak [LPS88] (who also coined this term) and independently by Margulis [Mar88] that arbitrarily large d-regular Ramanujan graphs exist when $d-1$ is prime, and moreover they can be explicitly constructed. Morgenstern [Mor94] extended this to the case when $d-1$ is a prime power. Here we only state the result and describe the construction. The book by Davidoff, Sarnak, and Valette [DSV03] offers a self-contained description of the beautiful mathematics around it. Lubotzky's book [Lub94] should be consulted as well.

Theorem 5.12 (Lubotzky-Phillips-Sarnak [LPS88], Margulis [Mar88], Morgenstern [Mor94]). *For every prime p and every positive integer k there exist infinitely many d-regular Ramanujan graphs with $d = p^k + 1$.*

The following suggests itself:

Conjecture 5.13. For every integer $d \geq 3$ there exist arbitrarily large d-regular Ramanujan graphs.

We will review in the following sections some recent attempts at solving this problem using combinatorial and probabilistic methods.

We conclude this section with a description of the Ramanujan graphs $X^{p,q}$ from [LPS88]. Let p, q be distinct primes that are congruent to 1 mod 4. Then $X^{p,q}$ is a $p+1$-regular graph of size $\Theta(q^3)$. Their explicitness level depends on the complexity of finding large primes.

Let us recall the definition of a **Cayley graph**. Let G be a group and let S be a subset of G that is closed under inversion. The corresponding Cayley graph $C(G, S)$ is a graph with vertex set G and edge set $\{(x, xs) : x \in G \text{ and } s \in S\}$. In the present case, $G = \mathrm{PGL}(2, q)$, the group of 2 by 2 nonsingular matrices over $\mathbb{F}_q$, where we identify two matrices that are proportionate to each other. Fix some integer i with $i^2 \equiv -1 \pmod q$. We define S as

$$S = \left\{ \begin{pmatrix} a_0 + ia_1 & a_2 + ia_3 \\ -a_2 + ia_3 & a_0 - ia_1 \end{pmatrix} : a_0^2 + a_1^2 + a_2^2 + a_3^2 = p, \right.$$

$$\left. \text{with odd } a_0 > 0 \text{ and even } a_1, a_2, a_3 \right\}.$$

By a theorem of Jacobi, there are exactly $p+1$ such solutions $(a_0.a_1, a_2, a_3)$ (over the integers!), so $|S| = p+1$. It can be verified that S is closed under inversion, as needed. The graph $X^{p,q}$ is obtained by taking the connected component of the identity of $C(G, S)$. (It can be shown that $C(G, S)$ is either connected or has exactly two equal connected components, depending on the quadratic residue symbol $(\frac{q}{p})$). In both cases, every connected component of the 2nd largest eigenvalue is bounded as in Theorem 5.12.

6. Spectrum and expansion in lifts of graphs

6.1. **Covering maps and lifts.** From a topological perspective graphs are one-dimensional simplicial complexes. The notion of covering maps from topology turns out to be quite useful for our subject. For one thing, $\mathbf{T_d}$ is the universal covering space of any d-regular graph. This explains the role of $\mathbf{T_d}$ in Section 5. Also if there is a covering map from a graph H onto a graph G, then every eigenvalue of G is also an eigenvalue of H. We also show how lifts allow us to construct regular graphs with near-optimal spectral gaps. We do not need anything substantial from topology, but we do borrow some of the terminology and several of our observations are special cases of more general phenomena in the theory of covering maps in topology.

We start by defining the notion of coverings and lifts.

Definition 6.1. Let G and H be two graphs. We say that a function $f : V(H) \to V(G)$ is a **covering map** if for every $v \in V(H)$, f maps the neighbor set $\Gamma_H(v)$ of

v one to one and onto $\Gamma_G(f(v))$. If there exists a covering function from H to G, we say that H is a **lift** of G or that G is a **quotient** of H.

This definition is appropriate for simple graphs (with no parallel edges and self loops) but can be easily extended to all graphs.

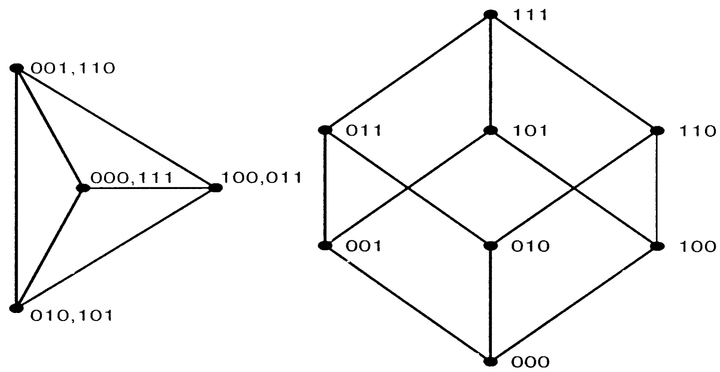


FIGURE 5. The three-dimensional cube is a 2-lift of the clique on four vertices. The cover map identifies antipodal vertices in the cube.

If f is a covering map onto G and v is a vertex in G, we call the set $f^{-1}(v)$ the **fiber** of v. Similarly, if $e \in E(G)$, we say that $f^{-1}(e)$ is the fiber of e. It is easily verified that if G is connected, then every covering map of a finite graph onto G has a well defined **covering number** n, such that all vertex and edge fibers have fixed size n.

Let G be a connected graph. We denote by $L_n(G)$ the set of all lifts of G with covering number n, which we call n-lifts of G. There is a simple and convenient description of the members of $L_n(G)$. If $H \in L_n(G)$, then $V(H) = V(G) \times [n]$. That is, the fiber of every $v \in V(G)$ consists of the vertices $(v,1), \cdots, (v,n)$ in $V(H)$. To define the edges of H, we associate with every edge $e = (v,u) \in E(G)$ a permutation π_e from S_n. The fiber of the edge (u,v) consists of the edges $((u,i),(v,\pi_e(i)))$ for all $i \in [n]$. (Note that here we consider $e = (u,v)$ as a **directed** edge, and the permutation corresponding to (v,u) is the inverse π_e^{-1}.) Thus every choice of permutations in S_n, one for each edge in G, defines a member in $L_n(G)$. This also gives us a natural way to sample a *random* n-lift of G, as we elaborate in subsection 7.3.3.

6.2. **Eigenvalues - old and new.** What can we say about the eigenvalues of lifts of G? The eigenvalues of G are eigenvalues of its lifts as well. To see this, let $h : V(G) \to \mathbb{R}$ be an eigenfunction of G with eigenvalue λ, and let H cover G via the map $f : V(H) \to V(G)$; then $h \circ f$ is an eigenfunction of H with the same eigenvalue. Such an eigenfunction and the corresponding eigenvalue are considered **old**. Thus if H is a lift of G, we talk about its **old** eigenvalues and eigenfunctions which are inherited from G, and the rest of the eigenvalues and eigenfunctions which are considered **new**.

Example 6.2. In the above example of the 3-dimensional cube and K_4, the spectrum of the cube is composed of: the old spectrum $\{3,-1,-1,-1\}$ inherited from K_4, and the new spectrum $\{1,1,1,-3\}$.

This leads us to two simple but useful observations from [BL]. Since the eigenfunctions of G span the space of real functions on $V(G)$ and since distinct eigenfunctions can always be chosen to be mutually orthogonal, we conclude:

Proposition 6.3. *Let $f : V(H) \to V(G)$ be a covering map and let ψ be a new eigenfunction of H. Then $\sum_{f(x)=v} \psi(x) = 0$ for every $v \in V(G)$. In words, ψ sums to zero on every fiber.*

This leads to a particularly pleasing description of the spectrum for 2-lifts of G. Let $A = A_G$ be the adjacency matrix of G. A **signing** $\tilde{A}$ of A is a symmetric matrix that is obtained by replacing some of the 1-entries in A by -1. There is a natural $1:1$ correspondence between 2-lifts of G and signings of A_G; namely, for every edge $e = (u, v) \in E(G)$ we define $\tilde{A}(u, v)$ to be ± 1 according to the permutation π_e being the identity or $(2, 1) \in S_2$. It follows that there is a $1:1$ correspondence between the eigenfunctions ϕ of $\tilde{A}$ and new eigenfunctions ψ of the 2-lift. The correspondence is given by $\psi(v, 1) = -\psi(v, 2) = \phi(v)$ for every $v \in V(G)$.

In particular:

Proposition 6.4. *Let H be a 2-lift of G that is encoded by the matrix $\tilde{A}$. The new eigenvalues of H are the eigenvalues of $\tilde{A}$.*

These observations are used below (subsection 6.4) to construct graphs with a near-optimal spectral gap.

6.3. **The universal covering tree.** Associated with every connected graph G is its **universal covering space** $\hat{G}$. This is an infinite tree that is uniquely defined through the condition that every connected lift of G is a quotient of $\hat{G}$. It is not hard to explicitly construct $\hat{G}$. Fix some vertex v_0 in G. The vertices of $\hat{G}$ are in $1:1$ correspondence with all non-backtracking walks in G starting at v_0. That is, all finite sequences $v_0, v_1, \ldots,$ such that v_i is adjacent to v_{i+1} and $v_i \neq v_{i+2}$ for all i. Two vertices of $\hat{G}$ are adjacent if one walk is a single-step extension of the other. It is easy to verify that:

Example 6.5. The infinite d-regular tree $\mathbf{T_d}$ is the universal covering space of every d-regular graph.

6.3.1. *Irregular Ramanujan graphs?* As we have already mentioned, the counterpart of the present theory for irregular graphs is still quite poorly understood. It is not even clear how to define an irregular Ramanujan graph. In particular, is there a natural lower bound on $\lambda(G)$ for an irregular graph? We recall that the spectral radius of a (finite or infinite) graph H is defined via $\rho(H) = \sup\{|\lambda| | \lambda \in \mathrm{spec}(H)\}$. The following generalization for the Alon-Boppana lower bound is due to Greenberg and Lubotzky [Gre95].

Theorem 6.6 (Greenberg-Lubotzky [Gre95]). *Let $\{G_i\}$ be a family of graphs covered by the same universal cover T. Then $\lambda(G_i) \geq \rho(T) - o(1)$.*

The proof is similar to the first proof we gave in the regular case, subsection 5.2.2. One considers $fA^{2k}f^T/||f||^2$, where f is ± 1 on two faraway vertices and zero elsewhere. However, the numbers t_{2k} of the d-regular tree are replaced by the number of closed walks of length $2k$ from v to v in the universal cover T, which is at least $(\rho - o(1))^{2k}$.

This naturally suggests a definition of (not necessarily regular) Ramanujan graphs.

Definition 6.7. A graph G is Ramanujan if $\lambda(G) \leq \rho(\hat{G})$, in other words, if the absolute value of all of G's eigenvalues excluding λ_1 are bounded by the spectral radius of its universal cover.

If G is d-regular, then $\hat{G} = \mathbf{T_d}$, so that $\rho(\hat{G}) = 2\sqrt{d-1}$, and this definition coincides with the definition of d-regular Ramanujan graphs.[11]

A possible restatement of Theorem 5.12 is that if $d-1$ is a prime power, then infinitely many quotients of $\mathbf{T_d}$ are Ramanujan graphs. Conjecture 5.13 posits that the same holds for every $d \geq 3$. Might it be that a similar statement holds for every infinite tree T with infinitely many finite quotients? This is not so, as pointed out by Lubotzky and Nagnibeda [LN98]. They constructed a tree T such that there is a single graph G^* that is covered by every finite quotient of T. Moreover, the second eigenvalue of G^* is larger than $\rho(T)$. Since this large eigenvalue is inherited by all graphs covering G^*, none of them can be Ramanujan.

6.4. **Nearly-Ramanujan graphs by way of 2-lifts.** It seems plausible that unlike the currently known proof of Theorem 5.12, a proof of Conjecture 5.13 would have to resort to methods outside number theory. For this and for several other reasons, it is a major challenge to develop combinatorial and probabilistic arguments that can establish such results. Such an attempt was recently made by Bilu and Linial [BL], as we now describe. We start from a small d-regular Ramanujan graph, such as the complete graph K_{d+1}, and attempt to construct large d-regular Ramanujan graphs by applying a series of 2-lifts. In view of Proposition 6.4, the following conjecture would guarantee the success of this plan, and in particular it would imply the correctness of Conjecture 5.13.

Conjecture 6.8. Every d-regular *Ramanujan* graph G has a 2-lift such that all new eigenvalues are in the range $[-2\sqrt{d-1}, 2\sqrt{d-1}]$.

Extensive computer experiments tend to suggest an even more daring conjecture.

Conjecture 6.9. Every d-regular graph G has a 2-lift such that all new eigenvalues are in the range $[-2\sqrt{d-1}, 2\sqrt{d-1}]$.

Equivalently,

Conjecture 6.10. Let A be the adjacency matrix of some d-regular graph. Then A has a signing $\tilde{A}$ with spectral radius $\rho(\tilde{A}) \leq 2\sqrt{d-1}$.

A somewhat weaker statement can be proven. It uses the Lovász Local Lemma, a standard tool in the probabilistic method (see [AS00]) and an enumerative argument to show:

Lemma 6.11. *Let A be the adjacency matrix of some d-regular graph. Then A has a signing $\tilde{A}$ such that*

$$x\tilde{A}x \leq O\left(\sqrt{d}\log d \cdot \|x\|^2\right)$$

for every vector x in which each coordinate is $-1, 0$ or 1.

The main result of that paper is obtained now using Lemma 2.6:

[11] Definition 6.7 is not the only one possible. One can give a similar definition based on the smallest positive eigenvalue of the Laplacian of G or based on the largest non-Perron eigenvalue of the normalized adjacency matrix. A different direction is to consider the spectrum of the non-backtracking adjacency operator or, equivalently, the Ihara zeta function of the graph; see [ST, AFH] for a more detailed discussion.

Theorem 6.12 (Bilu-Linial [BL]). *For every $d \geq 3$ there is a construction of a mildly explicit family of (n, d, α)-graphs, with $\alpha d = O(\sqrt{d \log^3 d})$.*

The algorithmic aspects of this construction need some further elaboration, but we refer the interested reader to the original paper.

7. The spectrum of random graphs

The **probabilistic method** is one of the most fruitful ideas in modern Discrete Mathematics. The basic tenet of this method can be phrased as follows: It is in general essentially impossible for a human to investigate a large graph (or other discrete mathematical objects). However, it is possible and very beneficial to craft certain **ensembles** of large graphs and investigate their **typical** properties. Specifically, it is possible to introduce probability spaces whose elements are graphs. Any graph parameter then becomes a random variable which can be investigated using tools from probability theory. These probability spaces and these random variables are being investigated in their own right (see e.g. the book [JŁR00]) or as means for solving other problems (here [AS00] is the standard reference). This important idea was first crystallized nearly a half century ago in a series of ground-breaking papers by Erdős and Rényi, starting with [ER59]. These authors have introduced $G(n, p)$, which is a basic model of random graphs.[12] To sample a graph from this distribution, start with n vertices. Independently, for every pair of vertices, introduce an edge with probability $0 < p < 1$. This important model is, however, not very useful for our purposes, since we are interested mostly in regular graphs. Regular graphs have only a tiny probability in the $G(n, p)$ model, and different random models are needed. Indeed it took about two decades of research in random graphs until it was discovered how to sample uniformly (n, d) graphs (see [JŁR00], Section 9, and [Wor99]).

In this section we study what the typical eigenvalues of random (n, d) graphs look like. The first subsection deals with the bulk of the spectrum (where most eigenvalues tend to be), and the second with the extreme eigenvalues (those that define expansion). In both cases, this study parallels earlier investigations of the spectrum of random matrices, as we mention. A central tool in both parts is the trace method (or moment method) which we have already seen. This method is particularly appropriate for the random setting. In the third subsection we consider some variations of this study of random sparse graphs.

7.1. The bulk of the spectrum.

Our focus has so far been the extreme eigenvalues of graphs, namely λ_2 and λ_n, the study of which is rather challenging, as we saw. It turns out that the bulk of the spectrum is more amenable to analysis. Here is some of what we know about these parts of the spectrum. The adjacency matrix of a graph is a real symmetric matrix, and a large body of work exists about the typical spectra of such matrices. The grandfather of this whole area is

[12] This model has many important connections to statistical physics, some of which are related to expansion in both the random graphs thus generated as well as in certain natural dynamics on them, but these are beyond the scope of this article.

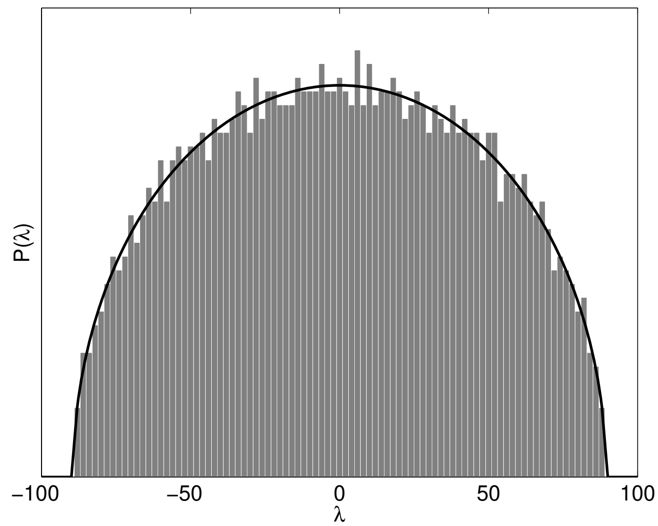


FIGURE 6. Eigenvalue distribution, 2000×2000 symmetric matrix with independent standard normal entries. The gray area is a 100 bin histogram, where the height is the portion of eigenvalues in the respective bin. The black line is the prediction of Theorem 7.1.

"Wigner's semicircle law", which states that, under some conditions, the eigenvalues of a large random symmetric matrix with independent entries are distributed close to a semicircle.

Theorem 7.1 (Wigner [Wig58]). *Let A_n be an $n \times n$ real symmetric matrix, where off-diagonal entries are sampled independently from the distribution F, and the diagonal entries from the distribution G. Furthermore, assume that var(F) = var$(G) = \sigma^2$ and that F and G have finite moments; i.e. $\int |x|^k dF(x)$ and $\int |x|^k dG(x)$ are finite for all k. Define the empirical eigenvalue distribution as*

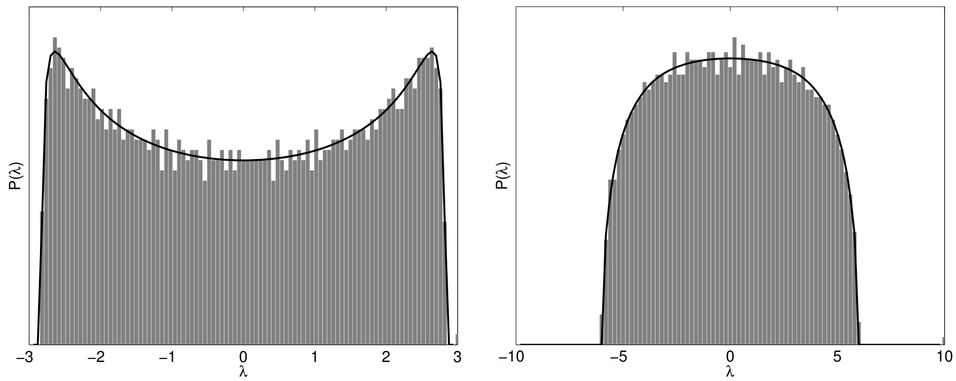
$$W_n(x) = \frac{1}{n}|\{i : \lambda_i(A_n) \leq x\}|,$$

where $\lambda_1(A_n) \geq \cdots \geq \lambda_n(A_n)$ are the eigenvalues of A_n. Then for every x,

$$W(x) = \lim_{n\to\infty} W_n(2x\sigma\sqrt{n}) = \frac{2}{\pi}\int_{-1}^{x} \sqrt{1-z^2}\, dz.$$

What about the analogous questions when we deal with the adjacency matrices of random d-regular graphs? As can be seen in Figure 7, the eigenvalue distribution no longer resembles a semicircle. Nevertheless, for large d, the distribution does approach a semicircle. This behavior is explained by the following theorem of McKay. In fact, this theorem applies not only to random graphs. It assumes only that the graph in question has few short cycles. Specifically, let $C_k(G)$ be the number of cycles in G of length k. The theorem applies whenever $C_k(G) = o(|V(G)|)$ for every fixed $k \geq 3$, a property that holds almost surely for random d-regular graphs.

Theorem 7.2 (McKay [McK81]). *Let G_n be an infinite sequence of d-regular graphs such that $C_k(G_n) = o(|V(G_n)|)$ for all $k \geq 3$, where $C_k(G_n)$ is the number of length*



$d = 3$ $d = 10$

FIGURE 7. Eigenvalue distribution, d-regular graph with 2000 vertices. The gray area is a 100 bin histogram, where the height is the portion of eigenvalues in the respective bin. The black line is the prediction of Theorem 7.2.

k cycles in G_n. Define the empirical eigenvalue distribution as

$$F(G_n, x) = \frac{1}{|V(G_n)|} |\{i : \lambda_i(G_n) \le x\}|.$$

Then for every x,

$$F(x) = \lim_{n\to\infty} F(G_n, x) = \int_{-2\sqrt{d-1}}^{x} \frac{d\sqrt{4(d-1)-z^2}}{2\pi(d^2 - z^2)}\, dz.$$

Note that the limit distribution $F(x)$ is supported on $[-2\sqrt{d-1}, 2\sqrt{d-1}]$, which is the spectrum of $\mathbf{T_d}$ (Theorem 5.2). Here is the main idea: Since there are only a few short cycles, the neighborhood of most vertices is (nearly) a tree. Consequently, for most vertices v, the number of walks of length k that start and end at v roughly equals t_k, the analogous quantity for $\mathbf{T_d}$. To proceed, we need a good estimate for t_k (compare with the lower bound in subsection 5.2.2).

Lemma 7.3. *For the infinite tree $\mathbf{T_d}$ the number described above is*

$$t_{2s+1} = 0, \qquad t_{2s} = \sum_{j=1}^{s} \binom{2s-j}{s} \frac{j}{2s-j}\, d^j\, (d-1)^{s-j}.$$

Proof. The first claim follows from the fact that a tree is a bipartite graph and contains no odd length cycles. To determine t_{2s}, consider any path of length $2s$ from v to v. Associate with it a sequence $0 = \delta_0, \delta_1 \cdots \delta_{2s} = 0$, where δ_i is our distance from v at time i. Clearly, δ_i are nonnegative integers and $|\delta_i - \delta_{i-1}| = 1$ for all i. The number of such sequences in which exactly j of the terms $\delta_0 \cdots \delta_{2s-1}$ are 0 is

$$\binom{2s-j}{s} \frac{j}{2s-j}.$$

This is a simple generalization of Catalan numbers (see [Fel68]). How many paths correspond to such a sequence? Each such path takes s steps away from v and s steps towards v; namely there are s indices each with $\delta_{i+1}-\delta_i = 1$ or -1 respectively.

In a step towards v, the next vertex is uniquely determined. On steps away from v we have d choices whenever $\delta_i = 0$ (and the path resides at v), and $d-1$ choices at each of the remaining $s-j$ times. The conclusion follows. $\square$

If $C_k(G_n) = o(|V(G_n)|)$ for every $k \geq 3$, then for every constant r, almost every vertex has a cycle-free r-neighborhood. Therefore there are $(1+o_n(1))|V(G_n)| \cdot t_r$ closed paths of length r in G_n. It follows that the function $F(x)$ satisfies $\int x^r dF(x) = t_r$ for all r. In order to finish the proof of the theorem we need to explicitly recover $F(x)$ from its moments. This is accomplished by expanding F in the basis of Chebyshev polynomials. The details are omitted.

7.2. **The extreme eigenvalues.** Quite a lot is known about the extreme eigenvalues of random symmetric matrices as well.

Theorem 7.4 (Füredi-Komlós [FK81], Vu [Vu05])**.** *Let $A = A_n$ be an $n \times n$ real symmetric matrix with independent entries from a distribution F that has zero expectation, variance σ^2, and is supported on $[-K, K]$ for some constant K. Then with probability $1 - o_n(1)$, all eigenvalues of A satisfy $|\lambda_i| < 2\sigma\sqrt{n} + O(n^{\frac{1}{3}} \log n)$.*[13]

The proof of this theorem is based on the **trace method** (a.k.a. the moment method), which we have already encountered in subsection 5.2.2. This old workhorse of the theory stands as well behind Friedman's Theorem 7.10 and many other results. One estimates first, by combinatorial reasoning, the number trace(A^{2k}), which is a count of closed paths of length $2k$ in a certain (possibly edge-weighted) graph. We subtract from it λ_1^{2k} and claim that if k is large enough, then the difference is dominated by λ_2^{2k}, whence we can estimate $|\lambda_2|$. On the one hand, the larger k is, the more dominant the contribution of λ_2^{2k} and the better our estimates become. However, as k grows, the enumeration of closed walks becomes hairy, so one needs to strike a balance between these two conflicting needs.

Before we proceed to exemplify this method, let us mention an alternative approach to the same problem due to Kahn and Szemerédi [FKS89]. They show that if A is the adjacency matrix of a random d-regular graph, then A's Rayleigh quotient is likely to be small on all points of an appropriate ϵ-net on the unit sphere. Since the Rayleigh quotient of a matrix is a Lipschitz function on unit vectors, the eigenvalues of A can be bounded. Lemma 2.6 and Lemma 6.11 suggest some ways in which this method can be further advanced.

7.2.1. *An illustration of the trace method.* We now illustrate the power of the trace method by proving the following upper bound on $\lambda(G) = \max_{i \geq 2} |\lambda_i(G)|$ for a random (n,d)-graph. A much stronger (possibly tight) result (Theorem 7.10) is now known, but its proof is much too long and complex to be reviewed here.

Theorem 7.5 (Broder-Shamir [BS87])**.** *The largest nontrivial eigenvalue of almost every $2d$-regular graph G satisfies $\lambda(G) = O(d^{3/4})$.*

There is a technicality that has to be addressed here on how to sample from the space of $2d$-regular graphs. In the proof below we sample such a graph using the **permutation model**. A $2d$-regular graph on n vertices in the permutation model is constructed by independently choosing at random d permutations $\pi_1, ..., \pi_d$ in

[13]There is a version where F has expectation $\mu > 0$. In this case, the bound of the theorem still holds for $i \geq 2$. The remaining eigenvalue, λ_1, asymptotically has normal distribution with expectation $n\mu + \sigma^2/\mu$ and variance $2\sigma^2$.

the symmetric group S_n and introducing an edge $(v, \pi_i(v))$ for every $v \in [n]$ and $i \in [d]$.

It turns out that this does **not** yield a uniform distribution on all $(n, 2d)$-graphs. However the theorem is valid as stated, since it is known (see e.g. [JŁR00], Section 9, and Wormald [Wor99]) that the distribution induced by the permutation model is **contiguous** with the uniform distribution. Namely a family of events has probability $1 - o(1)$ in one distribution iff it has probability $1 - o(1)$ in the other. Put differently, both distributions agree on the notion of "asymptotically almost sure events". It is possible to sample uniformly from among the $(n, 2d)$ graphs using the so-called *configuration model*, but that would complicate matters significantly. By contiguity this additional complication is unnecessary.

Proof. Let G be a random $2d$-regular graph on n vertices in the permutation model. Let P be the transition matrix of the random walk on G, i.e. the adjacency matrix of G divided by $2d$. Let $1 = \mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$ be the eigenvalues of P, and $\rho = \max\{|\mu_2|, |\mu_n|\}$. Since the eigenvalues of P^k are $\{\mu_i^k\}_{i=1}^n$, we have $\rho^{2k} \leq \operatorname{trace}(P^{2k}) - 1$ for any k. Therefore, by Jensen's inequality, $\mathbb{E}[\rho]$, the expected value of ρ satisfies

$$\mathbb{E}[\rho] \leq (\mathbb{E}[\rho^{2k}])^{1/2k} \leq (\mathbb{E}[\operatorname{trace}(P^{2k})] - 1)^{1/2k}. \tag{10}$$

This basic inequality is at the heart of the trace method. It bounds the eigenvalues by estimating the trace of powers of the matrix. These traces are combinatorial quantities, counting the number of closed paths of length $2k$ in our (random) graph.

Observe that the paths in G starting at vertex 1 are in one to one correspondence with words over the alphabet $\Sigma = \{\pi_1, \pi_1^{-1}, \ldots, \pi_d, \pi_d^{-1}\}$. Just think of the directed edge $(v, \pi_i(v))$ (respectively $(\pi_i(v), v)$) as being labeled by π_i (respectively π_i^{-1}). Now interpret a word as a sequence of directed edge labels to follow. Therefore

$$\mathbb{E}[\operatorname{trace}(P^{2k})] = \mathbb{E}[\text{f.p.}(\omega)] = n \cdot \Pr[\omega(1) = 1] \tag{11}$$

where ω is a uniformly chosen word in Σ^{2k} and f.p.(ω) is the number of fixed points of the permutation ω, i.e., the number of i with $\omega(i) = i$.

Our analysis is in two parts. We first consider the structure of the word ω as an element of the free group in d generators. We then investigate the actual walk corresponding to ω when these d generators take on particular (random) permutations from S_n. For the first part it is natural to **reduce** this word: namely repeatedly remove every two consecutive letters, one of which is π_j and the other π_j^{-1}. Let $\omega' = \operatorname{red}(\omega)$ be the reduction of ω. Clearly, $\Pr[\omega(1) = 1] = \Pr[\omega'(1) = 1]$. We show that such a random reduced word is very unlikely to exhibit a nontrivial periodicity of a type specified below. A word that is periodic in this sense is called a **bad** reduced word.

For the second part we examine a fixed reduced **good** word ω' and derive an upper bound on $\Pr[\omega'(1) = 1]$. As mentioned, ω' is fixed and the probability is considered under the random choice of the d permutations $\pi_1, \ldots, \pi_d \in S_n$. This is relatively simple when ω' is good and reduced.

A reduced word ω' is **bad** if it has the form $\omega' = \omega_a \omega_b^j (\omega_a)^{-1}$ for some words ω_a, ω_b and some $j \geq 2$. Note that the empty word is bad. We estimate two probabilities, whose sum is an upper bound on $\Pr[\omega(1) = 1]$:

(1) The probability that ω' is a bad word.

(2) The probability $\Pr[\omega'(1) = 1]$ for some arbitrary **fixed** good reduced word ω'.

The first probability is estimated in the next lemma.

Lemma 7.6. *Let $\omega \in \Sigma^{2k}$ be a random uniformly drawn word and let ω' be its reduced form. The probability that ω' is bad is at most $O(k^2 \cdot (2/d)^k)$.*

Proof. The idea is that for ω to have a bad reduction, at least half of its letters are determined by the rest (either to get the cancellations or to guarantee periodicity). Implementing this idea takes a bit of care.

Observe that all words ω of length $2k$ which reduce to a word of length $2l$ can be generated as follows. Start with a string that consists of $k-l$ left brackets and $k-l$ right brackets where each initial segment contains at least as many left brackets as right brackets. The brackets in such a sequence can be paired up in a unique allowable way. The *level* of an initial segment in this sequence is the (nonnegative) difference between the number of left and right brackets that it contains. At every point where the level is zero (including the positions to the left and to the right of the whole sequence) place an arbitrary number of $\star$ symbols with a total of $2l$ $\star$'s. Note that the positions of the $k-l$ left brackets within the length $2k$ word uniquely determine the positions of the $k-l$ right brackets and the $2l$ $\star$'s. This is shown by the following simple procedure: Scan the string from left to right and initially set a variable *level* to zero. If the next position is occupied by a left bracket, increase *level* by one. Otherwise, do as follows: If *level* $= 0$ place a $\star$ in this position. If *level*> 0, place a right bracket in this position and reduce *level* by one.

It follows that the number of such sequences is at most $\binom{2k}{k-l}$. Now, we say that a word $\omega \in \Sigma^{2k}$ matches such a sequence of brackets and stars if two conditions hold. The first is that every pair of matched brackets is assigned a letter and its inverse. This specifies the cancellation steps that reduce ω to a word of length $2l$. The second requirement is that no additional cancellations are possible. We ignore the second requirement all together.

The event we consider is that ω matches a specific bracket-star sequence with $2l$ stars and that the reduced word ω' is bad and has the form $\omega_a\omega_b^j(\omega_a)^{-1}$ for some $j \geq 2$. The probability of the first event is clearly bounded by $(2d)^{-k+l}$ (the letter corresponding to each right bracket is uniquely specified). To bound the probability of the second event, we specify the lengths of ω_a and ω_b. These already determine the value of j, which, by assumption, is ≥ 2. Having specified the lengths $|\omega_a|$ and $|\omega_b|$, this probability is clearly bounded by $(2d)^{-l}$ since the first half of the reduced word ω' uniquely determines the second half. Finally, we observe that these two events are independent since they concern a disjoint set of positions in the word ω. Putting everything together yields the required bound:

$$\Pr[\omega' \text{ is bad}] \leq k^2 \sum_l \binom{2k}{k-l}(2d)^{-k+l}(2d)^{-l} \leq k^2 \cdot (2/d)^k.$$

(One factor of k is for the choice of $|\omega_a|$ and one for $|\omega_b|$.) □

Now we fix a good reduced word ω' of length $s \leq 2k$ and estimate $\Pr[\omega'(1) = 1]$. The main conceptual idea of Broder and Shamir is that we expose the path that we follow only "as we go". It is here that the value of the permutation model in the analysis becomes apparent. We gradually reveal the relevant information about the random permutations $\pi_1, \cdots, \pi_d$ and their inverses as follows. Start with the vertex $v_0 = 1$ and compute the path $v_0, v_1, v_2, \cdots, v_s$ one step at a time. For $i \geq 1$ we compute v_i as follows. If $\sigma \in \Sigma$ is the i-th letter in ω', then we should set $v_i = \sigma(v_{i-1})$. Call step i **free** if the value of $\sigma(v_{i-1})$ was not "revealed" in a previous step and is still undetermined. If this is the case and if t values of σ were previously revealed, then v_i is selected from among the $n-t$ vertices not yet assigned to the range of σ. A step that is not free is called **forced**.

Call step i a **coincidence** if it is free and moreover if (the randomly selected vertex) v_i coincides with a previous vertex on the path. Namely $v_i \in \{v_0, \cdots, v_{i-1}\}$. Let C_i denote the event of a coincidence at step i. By what we said above $\Pr[C_i | v_0 = u_0, \ldots, v_{i-1} = u_{i-1}] \leq s/(n-s)$ for every $u_0 = 1, u_1, \ldots, u_{i-1}$. We see that the probability of a coincidence at step i is at most $s/(n-s) \leq 2k/(n-2k)$, regardless of preceding history.

Clearly for the event that $\omega'(1) = 1$ to hold, at least one coincidence must occur, so we bound the probability of this event by the sum of the probabilities of two events:

(1) At least two coincidences have occurred along the path.
(2) Exactly one coincidence has occurred and $v_s = 1$.

The first probability is easily bounded as follows. Given the positions i, j of the first two coincidences, as mentioned above, the probability that these coincidences occur is at most $\Pr[C_i] \cdot \Pr[C_j | C_i] \leq (s/(n-s))^2$. There are at most s^2 such positions, and so the total probability is at most $s^4/(n-s)^2 = O(k^4/n^2)$. The last inequality holds, since our choice will be that $k = o(n)$.

The second probability is bounded in the following lemma.

Lemma 7.7. *Let ω' be a good reduced word of length $s \leq 2k$, and consider the path starting at* 1 *and following ω'. Then the probability of having exactly one coincidence and ending at vertex* 1 *is at most* $\frac{1}{n-s+1} \leq \frac{1}{n-2k}$.

Proof. A **realization** of ω' is any walk corresponding to ω' that is determined by a choice of the permutations in Σ. Since ω' is reduced, any initial segment of such a realization preceding the (single) coincidence is a simple path. The step at which the coincidence occurs turns this path into a cycle with a (possibly empty) tail (see Figure 8). Since no more coincidences take place and since ω' is reduced, no additional edges will be visited, and all future steps are forced. In order to eventually reach vertex 1, the walk could proceed in one of two ways: (i) revolve around the cycle any (positive) number of times and then proceed to vertex 1 or (ii) turn immediately to the end of the tail and terminate the walk at vertex 1 there. The former possibility is ruled out since the word ω' is good.

It follows that $\omega' = \omega_a \omega_b (\omega_a)^{-1}$. Here ω_a may be empty and corresponds to the walk along the tail and $(\omega_a)^{-1}$ to the way back. The word ω_b is nonempty and

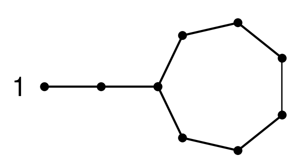


FIGURE 8. A cycle with a tail.

corresponds to the walk around the cycle. Moreover, ω_b must be cyclically reduced, and this condition uniquely determines the decomposition $\omega' = \omega_a\omega_b(\omega_a)^{-1}$. To see this, let v be the vertex at which this first coincidence occurs. If the walk left v on the first time with an edge labeled π, then this (first) coincidence cannot occur on a π^{-1} edge. Let $r = |\omega_a| + |\omega_b|$. It follows that the event $\omega'(1) = 1$ is included in the event that the $(r-1)$-th step is a free move to a **specific** previously visited vertex, $v_{|\omega_a|}$. The probability of the latter is $\leq \frac{1}{n-r} \leq \frac{1}{n-s+1}$, and the claim follows.

Adding the two bounds, the proof of the lemma follows. □

We now put the bounds of the two lemmas together, obtaining

$$\Pr[\omega(1) = 1] \leq 1/(n-2k) + O(k^2 \cdot (2/d)^k) + O(k^4/n^2).$$

Choose $k = 2\log_{d/2} n$ to minimize the upper bound. Substituting it back into (10) and (11) yields $\mathbb{E}[\rho] \leq (2/d)^{1/4} \cdot (1 + o(1))$. We finish up the proof by Markov's inequality, stating that a random variable is not likely to take on values much larger than its expectation.

Broder and Shamir [BS87] have also observed that the random variable $\lambda(G)$ considered for the probability space of d-regular graphs G is highly concentrated. Namely, for a large d-regular graph G, almost surely, $\lambda(G)$ is within $O(\sqrt{d})$ of its expected value. This is done using martingales and Azuma's inequality. For more on this very useful technique see e.g. [AS00, JŁR00].

7.3. **Variations on a theme.**

7.3.1. *Back to the irregular case.* Let us come back to Definition 6.7, which extends the concept of Ramanujan graphs to the irregular case. It is tempting to consider problems such as Conjecture 5.13, in this more general context. Is it true that for every graph G, almost all sufficiently high lifts are Ramanujan?

To address this problem, Friedman [Fri03] rephrased the conjecture:

Conjecture 7.8. For almost all sufficiently high lifts of any graph G, the absolute value of all **new** eigenvalues is bounded by $\rho + o(1)$, where ρ is the spectral radius of the universal cover of G.

To support his conjecture, Friedman proved the following theorem by generalizing the Broder Shamir argument of Theorem 7.5.

Theorem 7.9 (Friedman [Fri03]). *Let G be a graph with a largest eigenvalue λ_1, and let ρ be the spectral radius of its universal cover. Then for almost all sufficiently high lifts of G, the absolute value of all new eigenvalues is bounded by $\sqrt{\lambda_1\rho} + o(1)$.*

Indeed, this generalizes Theorem 7.5, since for $2d$-regular graphs, the permutations model coincides with random n-lifts of a $2d$-regular graph consisting of one vertex with d loops. In this case, the old spectrum is the single eigenvalue $\lambda_1 = 2d$ and $\rho = 2\sqrt{d-1} = O(\sqrt{d})$.

As mentioned before, even in the d-regular case, Conjecture 7.8 is not settled yet.

7.3.2. *Are most regular graphs Ramanujan?* A major result due to Friedman is a relaxed version of Conjecture 5.13 and the regular case of Conjecture 7.8:

Theorem 7.10 (Friedman [Fri]). *For every $\epsilon > 0$,*

$$Pr\left(\lambda(G) \leq 2\sqrt{d-1} + \epsilon\right) = 1 - o_n(1)$$

where G is a random (n, d)-graph.

To arrive at an educated guess whether most d-regular graphs are Ramanujan, the most natural approach is to carry out some computer experiments. Such experiments where conducted by Novikoff [Nov02] (see also [Sar04]) and by Hoory. We show in the following table and in Figure 9 the results obtained by taking 1000 random 4-regular graphs in the permutation model for varying graph sizes up to $n = 400000$.[14]

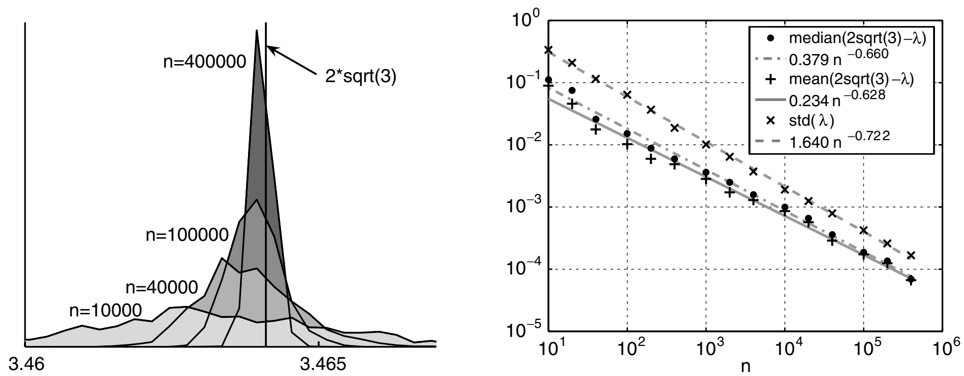
n	100	400	1000	4000	10000	40000	100000	400000
$\Pr(\lambda(G) < 2\sqrt{3})$	0.62	0.64	0.63	0.66	0.68	0.67	0.67	0.68
$2\sqrt{3} - \text{median}(\lambda(G))$	0.01519	0.00593	0.00361	0.00157	0.00100	0.00036	0.00019	0.00007
$2\sqrt{3} - \text{mean}(\lambda(G))$	0.01026	0.00487	0.00283	0.00129	0.00086	0.00029	0.00017	0.00007
$\text{std}(\lambda(G))$	0.06376	0.01872	0.01012	0.00372	0.00190	0.00079	0.00042	0.00017

It is evident that the median and expected value of λ tend to $2\sqrt{d-1}$ from below and that its standard deviation tends to zero with growing graph size. In [Nov02], it is conjectured that, after normalization, the distribution of λ tends to a Tracy-Widom distribution, [TW96]. Furthermore, they conjecture that the mean approaches $2\sqrt{d-1}$ faster than the standard deviation tends to zero and that the probability of being Ramanujan tends to a constant *strictly* between zero and one.

7.3.3. *More on random lifts.* As pointed out in subsection 6.1, n-lifts of a given connected graph can be generated at random. This is an entirely different source of random graphs whose vertex degrees can be completely controlled. A recent application for these "crafted" random graphs is the construction of error correcting codes where belief propagation decoding outperforms codes generated by the standard configuration model; see [Tho03, RU]. We illustrate below some of what is known about combinatorial properties of random lifts. The following statements refer to graphs in $L_n(G)$, where G is a fixed connected graph with more edges than vertices and where n is large. In particular "almost all" means that the statement holds in the above space with probability $1 - o_n(1)$.

- For every G there is an $\epsilon_G > 0$ such that almost all lifts of G have expansion ratio $\geq \epsilon_G$, [AL06].

[14] $\lambda(G)$ was calculated using the Matlab function eigs. The 3 smallest and largest values were excluded in the calculation of the mean and the standard deviation.



(a) (b)

FIGURE 9. (a) Distribution of $\lambda(G)$ for 1000 random 4-regular graphs in the permutation model. Four 40 bin histograms of $\lambda(G)$ for graph sizes 10000, 40000, 100000, 400000. (b) Median, mean and standard deviation of $2\sqrt{d-1} - \lambda(G)$ as a function of the graph size n. A log-log graph, along with the best linear interpolations.

- If $\delta \geq 3$ is the smallest vertex degree in G, then no graph in $L_n(G)$ has connectivity $> \delta$ and almost all graphs there are δ-connected, [AL02].
- There is a zero/one law about the existence of perfect matching. Here n is even and large. All graphs G fall into two categories: one where almost all n-lifts of G have a perfect matching, and the other where almost none of G's n-lifts have a perfect matching, [LR05].

There are also some results about chromatic numbers [ALM02], graph minors [DL], and Hamiltonian circuits in lifts [BCCF05].

In analogy with Conjecture 5.13, it seems natural to ask questions such as:

Open problem 7.11. Fix a d-regular base graph, e.g., $G = K_{d+1}$. How likely is it that all new eigenvalues of an n-lift of G fall in the range $[-2\sqrt{d-1}, 2\sqrt{d-1}]$?

7.3.4. *The eigenvectors.* There is very little we have to say here, except to point out what seems at present like a total mystery worth investigating. Can anything be said about the distribution of the coordinates in eigenvectors for any of the above classes of random matrices? In particular:

Open problem 7.12. Fix an integer $d \geq 3$ and consider large random (n, d)-graphs. Clearly, all the coordinates of the first normalized eigenvector v_1 equal $1/\sqrt{n}$. What can be said about the distribution of the coordinates of v_2? Specifically, does v_2 tend to be uniformly distributed on the unit sphere (in which the distribution of these coordinates is nearly normal)?

We hesitate to say much here, but some preliminary numerical experiments carried out by N.L. and S. H. suggest that the answer is negative. What is perhaps even more intriguing is that there seems to be another (nonuniform) limit distribution involved. At present we do not know much about these fascinating issues.

8. The Margulis construction

In this section we return to the oldest explicit construction of a family of expander graphs. This construction is still among the most elegant and most easy to generate of all known constructions. Nevertheless, the simplest known proof that it is an expander (only a few pages of basic linear algebra), which we present here, is still subtle and mysterious. We recall an elementary approach to these graphs that is mentioned in subsection 4.3. The concrete conjecture stated there suggests what a more intuitive proof may look like.

We now turn to the construction. Recall from subsection 4.3 an infinite analog of this construction. There we started with the action of two linear transformations on the unit torus. Here, in the finite setting, we take essentially the same linear transformations, together with their affine shifts (which serve as a discrete substitute for continuity). These act on the finite torus $(\mathbb{Z}_n)^2$. In Section 11 we will see that the linear transformations themselves suffice for expansion, at least when n is a prime. It will be viewed as the action of the group $SL_2(n)$ of 2×2 matrices of determinant 1, which is generated by these two linear transformations (without the affine shifts). That proof of expansion is, however, far from elementary.

Construction 8.1. Define the following 8-regular graph $G_n = G = (V, E)$ on the vertex set $V = \mathbb{Z}_n \times \mathbb{Z}_n$. Let

$$T_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, T_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Each vertex $v = (x, y)$ is adjacent to the four vertices $T_1 v$, $T_2 v$, $T_1 v + e_1$, $T_2 v + e_2$, and the other four neighbors of v obtained by the four inverse transformations. Note that all calculations are mod n and that this is an undirected 8-regular graph (that may have multiple edges and self loops).

Theorem 8.2 (Gabber-Galil [GG81]). *The graph G_n satisfies $\lambda(G_n) \leq 5\sqrt{2} < 8$ for every positive integer n.*

We will, in fact, prove in full a slightly weaker bound on λ that is still smaller than 8. As we saw in subsection 4.5, this already implies that the graphs G_n form a family of expanding graphs.

Margulis [Mar73] proved the same conclusion for a closely related family of graphs in 1973. However, his method was inherently existential (using the machinery of Kazhdan's property (T)) and could not give an explicit lower bound on the spectral gap. In 1981 Gabber and Galil [GG81] were able to derive a lower bound on the gap. Their argument uses classical harmonic analysis. In 1987 Jimbo and Marouka [JM87] improved it further using Fourier analysis on the group $\mathbb{Z}_n^2$. We present here a slight simplification of their proof, due to Boppana.

By the variational formula for the eigenvalues (subsection 4.5) we are trying to prove that if $f : \mathbb{Z}_n^2 \to \mathbb{R}$ satisfies $\sum_x f(x) = 0$, then

$$\sum_{(x,y) \in E} f(x) f(y) \leq \frac{5\sqrt{2}}{2} \sum f^2(x).$$

Taking into account the definition of the graph G, Theorem 8.2 can therefore be restated as:

Theorem 8.3. *For any $f : \mathbb{Z}_n^2 \to \mathbb{R}$ satisfying $\sum_z f(z) = 0$, the following inequality holds:*

$$\sum_{z \in \mathbb{Z}_n^2} f(z) \cdot \left[f(T_1 z) + f(T_1 z + e_1) + f(T_2 z) + f(T_2 z + e_2) \right] \leq \frac{5\sqrt{2}}{2} \sum f^2(x). \tag{12}$$

8.1. **A detour into harmonic analysis.** The intensifying connections between harmonic analysis and discrete mathematics are among the most wonderful recent developments of both fields. We are unable to do here any justice to this domain of research. The reader can learn the fundamentals of harmonic analysis from excellent texts such as Körner [Kör89]. Certain parts of this area are covered by lecture notes available at `www.cs.huji.ac.il/~nati/PAPERS/uw/`.

Below we collect some of the basic tools that we need to proceed with our analysis. None of this is hard, and the proofs can be found in the above references (or be carried out on the reader's own). We will consider similar issues in a more general context in Section 11, namely, Fourier analysis on general groups, a.k.a representation theory. Here we confine ourselves to Abelian groups. It is remarkable, though, that we are able to investigate the action of the (non-Abelian) group of 2×2 matrices using "only" harmonic analysis.

8.1.1. *Characters.*

Definition 8.4. A **character** of a group H is a homomorphism $\chi : H \to \mathbb{C}^*$; that is $\chi(gh) = \chi(g) \cdot \chi(h)$ for all $h, g \in H$.

Note that this definition implies that for finite groups H the range of any character is actually contained in the unit circle, namely complex numbers of absolute value 1. Also, when H is Abelian (as is the case in the present section), we denote, as customary, its group operation by $+$.

Here are some examples of characters.

- The **trivial character** maps all elements of H to 1.
- For H the cyclic group $\mathbb{Z}_n$, the characters are $\chi_k(h) = e^{2\pi i k h/n}$. (The trivial character corresponds to $k = 0$.)
- For $H = (\mathbb{Z}_2)^d$: The characters are $\chi_a(x) = (-1)^{<a,x>}$, where $a = (a_1, \ldots, a_d) \in (\mathbb{Z}_2)^d$ and $< x, y > = \sum_i x_i y_i$ is the inner product. (The trivial character corresponds to $a = 0^d$.)
- The group that is most relevant for the present section is $H = \mathbb{Z}_n^2$. It has a character χ_b for each $b = (b_1, b_2) \in \mathbb{Z}_n^2$, where $\chi_b(a_1, a_2) = \omega^{a_1 b_1 + a_2 b_2}$. Here $\omega = e^{2\pi i/n}$ is the primitive n-th root of unity.

The collection $\mathcal{F}$ of all complex functions on H is a linear space with inner product $\langle f, g \rangle = \sum_{x \in H} \overline{f(x)} g(x)$. Fourier analysis on H entails expanding functions in $\mathcal{F}$ as linear combinations of characters.

Proposition 8.5. *Every finite Abelian group H has $|H|$ distinct characters which can be naturally indexed as $\{\chi_x\}_{x \in H}$. They form an orthonormal basis of $\mathcal{F}$. Thus every $f : H \to \mathbb{C}$ can be uniquely expressed as $f = \sum_{x \in H} \widehat{f}(x) \chi_x$, where $\widehat{f} : H \to \mathbb{C}$ is the discrete Fourier transform of f,*

$$\widehat{f}(x) = \langle f, \chi_x \rangle = \sum_{y \in H} \overline{f(y)} \cdot \chi_x(y).$$

Therefore, in the case of interest here, $H = \mathbb{Z}_n^2$, the discrete Fourier transform of f takes the form

$$\widehat{f}(x) = \sum_b \overline{f(b)} \cdot \omega^{b_1 x_1 + b_2 x_2}.$$

In the following proposition we collect some of the basic properties of the Fourier transform. We state the results for the group $\mathbb{Z}_n^2$, though most of these claims apply just as well for any Abelian group.

Proposition 8.6. *Let $f, g \in \mathcal{F}$. Then the following statements hold:*

(a) $\sum_{a \in H} f(a) = 0 \Leftrightarrow \widehat{f}(0) = 0$.

(b) $\langle f, g \rangle = \frac{1}{n^2} \langle \widehat{f}, \widehat{g} \rangle$ *for any* $f, g \in \mathcal{F}$.

(c) *A special case of (b) when $f = g$, is Parseval's identity:*

$$\sum_{a \in H} |f(a)|^2 = \frac{1}{n^2} \sum_{a \in H} |\widehat{f}(a)|^2.$$

(d) *The inverse formula:*

$$f(a) = \frac{1}{n^2} \sum_{b \in H} \widehat{f}(b) \omega^{-\langle a, b \rangle}.$$

(e) *The shift property:*[15] *If A is a nonsingular 2×2 matrix over $\mathbb{Z}_n$, $b \in \mathbb{Z}_n^2$ and $g(x) = f(Ax + b)$, then*

$$\widehat{g}(y) = \omega^{-\langle A^{-1} b, y \rangle} \widehat{f}((A^{-1})^T y).$$

8.2. **Back to the proof.** To prove Theorem 8.3, we express inequality (12) using the Fourier coefficients of f . The condition $\sum_z f(z) = 0$ is restated as $\widehat{f}(0,0) = 0$. Using Parseval's identity and the behavior of the transform under composition with an affine transformation, we can rewrite Theorem 8.3 in the form below. It would suffice to show this with F the Fourier transform of f, but we prove this claim in its generality.

Theorem 8.7. *For every $F : \mathbb{Z}_n^2 \to \mathbb{C}$ with $F(0,0) = 0$,*

$$\left| \sum_{z=(z_1,z_2) \in \mathbb{Z}_n^2} \overline{F(z)} \cdot \left[F(T_2^{-1} z)(1 + \omega^{-z_1}) + F(T_1^{-1} z)(1 + \omega^{-z_2}) \right] \right| \leq \frac{5\sqrt{2}}{2} \sum_{z \in \mathbb{Z}_n^2} |F(z)|^2.$$

We observe that using Cauchy-Schwartz to upper bound the inner product would give a trivial upper of 4 times the L_2 norm of F, and the whole point of course is that $\frac{5\sqrt{2}}{2}$ is strictly less than 4. We also note that this can happen only if the two vectors in this inner product are not collinear, which is the main thing to prove.

Define $G : \mathbb{Z}_n^2 \to \mathbb{R}$ via $G = |F|$. By the triangle inequality and the identity $|1 + \omega^{-t}| = 2\,|\cos(\pi t/n)|$, it suffices to prove:

---

[15] This shows how the Fourier Transform behaves under a composition with a nonsingular affine transformation. It is this relation which makes it possible to analyze the non-Abelian group action of 2×2 matrices based on Fourier analysis of the Abelian group $(\mathbb{Z}_n)^2$.

Theorem 8.8. *For every nonnegative function $G : \mathbb{Z}_n^2 \to \mathbb{R}$ with $G(0,0) = 0$:*

$$\sum 2G(z) \cdot \big[\, G(T_2^{-1}z) \cdot |\cos(\pi z_1/n)| + G(T_1^{-1}z) \cdot |\cos(\pi z_2/n)| \,\big] \leq \frac{5\sqrt{2}}{2} \sum G^2(z) \tag{13}$$

where the summations are over $z = (z_1, z_2) \in \mathbb{Z}_n^2$.

We seek to bound the terms on the left by squares that match the terms on the right. We will do this by means of the elementary inequality $2\alpha\beta \leq \gamma\alpha^2 + \gamma^{-1}\beta^2$ that holds for any nonnegative α, β, and γ. The key will be to choose γ in a clever way; it will not be constant, but rather a mapping $\gamma : (\mathbb{Z}_n^2)^2 \to \mathbb{R}$.

How should we choose γ? First note that setting γ to be identically 1 is problematic for small z_1, z_2 where the cosines are near 1. That would yield the inequality with a coefficient of 8, which is just what we wanted to avoid. This suggests, though, that we let γ take only values close to 1. The idea is to define a partial order on $\mathbb{Z}_n^2$ and let γ satisfy:

$$\gamma\big(\,(z_1, z_2),\, (z_1', z_2')\,\big) = \begin{cases} \alpha & \text{if } (z_1, z_2) > (z_1', z_2') \\ 1/\alpha & \text{if } (z_1, z_2) < (z_1', z_2') \\ 1 & \text{otherwise.} \end{cases}$$

Here α should be a constant slightly bigger than 1. Indeed we set $\alpha = 5/4$, so that γ takes only the values $1, \alpha = 5/4$ and $1/\alpha = 4/5$. This definition implies that

$$\gamma(x, y) \cdot \gamma(y, x) = 1, \tag{14}$$

for every $x, y \in \mathbb{Z}_n^2$. We write

$$2G(x)G(y) \leq \gamma(x, y)G^2(x) + \gamma(y, x)G^2(y)$$

to derive the following upper bound on the left side of (13):

$$\sum_{z \in \mathbb{Z}_n^2} \begin{aligned} & |\cos(\pi z_1/n)| \cdot \big[\, \gamma(z, T_2^{-1}z)G^2(z) + \gamma(T_2^{-1}z, z)G^2(T_2^{-1}z) \,\big] \\ + \quad & |\cos(\pi z_2/n)| \cdot \big[\, \gamma(z, T_1^{-1}z)G^2(z) + \gamma(T_1^{-1}z, z)G^2(T_1^{-1}z) \,\big]. \end{aligned}$$

So far everything applies to any linear transformations T_1, T_2. We now use the fact that they are triangular. The exact choice of the matrices will come in only later. We use the fact that z_2 is invariant under T_1 and likewise with z_1 and T_2. Inequality (13) would follow from:

$$\sum_{z \in \mathbb{Z}_n^2} G^2(z) \cdot \Big(\, |\cos(\pi z_1/n)| \cdot \big[\, \gamma(z, T_2 z) + \gamma(z, T_2^{-1}z) \,\big]$$
$$+ |\cos(\pi z_2/n)| \cdot \big[\, \gamma(z, T_1 z) + \gamma(z, T_1^{-1}z) \,\big] \,\Big) \leq \frac{5\sqrt{2}}{2} \sum_{z \in \mathbb{Z}_n^2} G^2(z).$$

Since this should hold for every G, let us examine the case where G is nonzero on a single point $z \neq (0,0)$. If the present approach to proving Theorem 8.2 is viable, then it should work **term by term**. Thus our proof will succeed if our function γ

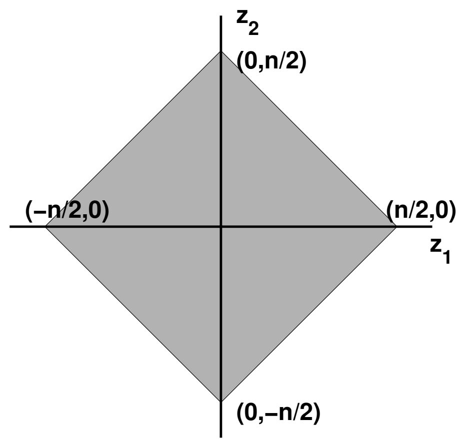


FIGURE 10. The diamond.

satisfies

$$|\cos(\pi z_1/n)| \cdot \left[\,\gamma(z, T_2 z) + \gamma(z, T_2^{-1} z)\,\right] + |\cos(\pi z_2/n)| \cdot \left[\,\gamma(z, T_1 z) + \gamma(z, T_1^{-1} z)\,\right] \le \frac{5\sqrt{2}}{2} \tag{15}$$

for every $z = (z_1, z_2) \in \mathbb{Z}_n^2 \setminus (0,0)$.

We split the proof into two domains. **Outside the diamond** in Figure 10 we overestimate all the γ terms by $\alpha = \frac{5}{4}$. As we verify below, in this range $|\cos(\pi z_1/n)| + |\cos(\pi z_2/n)| \le \sqrt{2}$, which implies the necessary inequality. Let us assume z_1, z_2 are in the first quadrant. The other cases follow similarly. Since $\cos(\pi z_2/n)$ is decreasing and since we are outside the diamond, this expression is maximized on the boundary of the diamond $z_2 = n/2 - z_1$, where $\cos(\pi z_2/n) = \sin(\pi z_1/n)$. Consequently, $\cos(\pi z_1/n) + \cos(\pi z_2/n) = \cos(\pi z_1/n) + \sin(\pi z_1/n) \le \sqrt{2}$, as needed.

When (z_1, z_2) is **inside the diamond**, we just bound the cosines by 1 and wish to prove that

$$\gamma(z, T_1 z) + \gamma(z, T_1^{-1} z) + \gamma(z, T_2 z) + \gamma(z, T_2^{-1} z) \le \frac{5\sqrt{2}}{2}. \tag{16}$$

That will follow from the following claim:

Proposition 8.9. *There is a partial order on $\mathbb{Z}_n^2$ such that for every (z_1, z_2) inside the diamond, either:*
Three of the four points $T_1 z$, $T_2 z$, $T_1^{-1} z$ and $T_2^{-1} z$ are $> z$ and one is $< z$
or
Two of the four points $T_1 z$, $T_2 z$, $T_1^{-1} z$ and $T_2^{-1} z$ are $> z$ and two are incomparable with z.

In the first case, the left-hand side of (16) is $3/\alpha + \alpha$, while in the second case it is $2/\alpha + 2$. Since $\alpha = 5/4$, the left-hand side of (15) is bounded by 3.65. This is not as good as $5\sqrt{2}/2 = 3.53\ldots$, but it does prove that the graphs are a family of good expanders. A different choice for α and a more careful analysis of the inequality inside the diamond yield the $5\sqrt{2}/2$ bound.

To define the partial order on $\mathbb{Z}_n^2$, we use the convention that our variables take values in the interval $[-n/2, n/2)$. Say that $(z_1, z_2) > (z_1', z_2')$ if $|z_1| \geq |z_1'|$ and $|z_2| \geq |z_2'|$ and at least one of the inequalities is strong.

We know how to verify Proposition 8.9 only by a case analysis. It is easy to verify that if $|z_1| = |z_2|$, then the second case of the proposition holds, so we assume without loss of generality that $|z_1| > |z_2|$. As mentioned, the second coordinate of $T_1^{\pm 1} z$ is z_2 and likewise with $T_2^{\pm 1}$ and z_1. So in each case there is only one inequality to analyze. By symmetry we may assume that $z_1 > z_2 \geq 0$, and since we are inside the diamond, $z_1 + z_2 \leq n/2$. It follows that $|z_1 - 2z_2| < |z_1|$, so $T_1^{-1} z < z$. The other three points $T_1 z$, and $T_2^{\pm 1} z$ are $> z$ since $|z_1 + 2z_2| > |z_1|$ and $|z_2 \pm 2z_1| > |z_2|$. Consequently the first case of the proposition holds.

9. **The zig-zag product**

In this section we introduce a new kind of graph product called the **zig-zag product**. We show that the zig-zag product of two expanders is a (not much worse) expander as well. We start by showing how this fact leads to an iterative construction of an explicit family of expanders. We then present the proof of this fact.

We already know that a given graph is an expander iff the random walk on that graph mixes rapidly. As we discussed in subsection 3.1.2, it is possible to estimate the expansion ratio of an (n, d)-graph G by considering how rapidly the entropy of the random walk on G converges to $\log n$. This suggests an interesting perspective of zig-zag products which we discuss as well.

This combinatorial method for constructing expanders was suggested by Reingold, Vadhan and Wigderson in [RVW02]. It has led to other constructions, such as expanders which beat the eigenvalue bound (i.e. have better expansion than implied by their spectral gap) and new Cayley expanders. Both of these will be discussed in later sections. We conclude this section with a remarkable recent application of the zig-zag product to complexity theory, namely, Reingold's result $SL = L$. This means that there is a deterministic algorithm using only S units of working memory that can explore every graph of size $e^{O(S)}$.

9.1. **Introduction.**

As in earlier sections we consider an (n, d)-graph $G = (V, E)$ with adjacency matrix $A = A(G)$. We also recall the notation $\hat{A} = \frac{1}{d} A(G)$, the transition matrix of the random walk on G. Also, if G is an (n, d, α)-graph, then Theorem 4.11 gives a lower bound $(1 - \alpha) d/2 \leq h(G)$ on $h(G)$, the expansion ratio of G.

The k-th power $G^k = (V, E')$ is a graph on the same vertex set where we put an edge $(u, w) \in E'$ for **every** path of length k in G from u to w. The adjacency matrix of G^k is just the k-th power of the adjacency matrix of G, whence G^k is an (n, d^k, α^k)-graph.

The zig-zag product, which is denoted by Ⓩ, is an asymmetric binary operation. The product of an (n, m)-graph and an (m, d) graph is an (mn, d^2)-graph. The main result is that the product of two expanders is an expander. We can use this result as a "black box" even before we actually define the zig-zag product.

Theorem 9.1 (The Zig-Zag Theorem, Reingold-Vadhan-Wigderson [RVW02]). *Let G be an (n, m, α)-graph and H be an (m, d, β)-graph. Then $G \textcircled{z} H$ is an $(nm, d^2, \varphi(\alpha, \beta))$-graph where the function φ satisfies the following:*

(1) *If $\alpha < 1$ and $\beta < 1$, then $\varphi(\alpha, \beta) < 1$.*
(2) $\varphi(\alpha, \beta) \leq \alpha + \beta$.
(3) $\varphi(\alpha, \beta) \leq 1 - (1 - \beta^2) \cdot (1 - \alpha)/2$.

Since these bounds depend only on the spectral gaps of G and H, the first bound indeed has the meaning alluded to above: zig-zag takes two expanders into another expander.[16]

The quantitative bounds (2) and (3) are crucial for applications. The former is useful when α, β are small, and the latter when they are large. We show below how to use bound (2) on φ for the explicit construction of a family of expander graphs, and in the last section we show how Reingold used bound (3) on φ for his proof that $SL = L$. The definition and analysis of the zig-zag product appear in subsection 9.3.

9.2. **Construction of an expander family using zig-zag.** Before we proceed to the definition, let us show that zig-zag products can be used to explicitly construct a family of constant degree expanders. To generate that infinite family we need one building block: a fixed size expander of certain parameters.

Let H be a $(d^4, d, 1/4)$-graph for some constant d. Note that there is a probabilistic proof (for example, an adaptation of Theorem 7.5) that such an expander exists, and since d is constant, one can find it by an exhaustive search in constant time. This is a brute-force argument. For a more efficient construction that uses the fact that the degree is quite large in terms of the number of vertices, see [RVW02].

Using the building block H, we inductively define the infinite sequence G_n by:

$$G_1 = H^2, \qquad G_{n+1} = (G_n)^2 \textcircled{z} H \quad \text{for } n \geq 1. \tag{17}$$

We claim that this sequence is an expander family:

Proposition 9.2. *The graph G_n is a $(d^{4n}, d^2, 1/2)$-graph for all n.*

Proof. For $n = 1$ this follows from the definition. We proceed by induction and assume that $(G_n)^2$ is a $(d^{4n}, d^4, 1/4)$-graph. When we zig-zag it with H (note that the degree of $(G_n)^2$ equals the size of H, as required) it follows from the second bound on φ in Theorem 9.1 that G_{n+1} is a $(d^{4(n+1)}, d^2, 1/2)$-graph. ☐

The observant reader must have noticed that this construction is only **mildly** explicit in the sense of Definition 2.3. To make it **strongly** explicit one can interleave the construction in every iteration with another operation: tensoring of G_n with itself. This provides a much faster growth of these graphs. For more details see [RVW02].

[16] The converse is also true: it is easy to see that $\varphi(\alpha, \beta) \geq \max\{\alpha, \beta\}$ and so zig-zag cannot improve the expansion of the input graphs.

9.3. **Definition and analysis of the zig-zag product.** Let G be an (n, m, α)-graph and H be an (m, d, β)-graph. For every vertex $v \in V(G)$ we fix some numbering $e_v^1, \ldots e_v^m$ of the edges incident with v. Also, we regard the vertex set of H as the set $[m] = \{1, \ldots, m\}$. The vertex set of $G ⓩ H$ is the Cartesian product $V(G) \times V(H)$. It is convenient to think of this vertex set as being created by replacing every vertex v in G with a "cloud" of m vertices $(v, 1), \ldots (v, m)$, one for every edge incident with v. To describe the edges of the product graph, it is easier to first describe the edges of another graph, $G ⓡ H$ (called the **Replacement product**) on the same vertex set. Figure 11 is useful for the purpose of explaining both constructions.

The edges of $G ⓡ H$ (shaded lines in the figure) are simply the union of the original edges of G (these are the wiggly edges, now going between clouds) and n copies of the edges of H, one copy per cloud (these are the dashed edges). The edges of $G ⓩ H$ arise from walks of length three of a "zig-zag" nature in $G ⓡ H$: dashed-wiggly-dashed. These edges are drawn as solid lines in Figure 11. More formally:

Definition 9.3. $G ⓩ H = (V(G) \times [m], E')$, where $((v, i), (u, j)) \in E'$ iff there are some $k, l \in [m]$ such that $(i, k), (l, j) \in E(H)$ and $e_v^k = e_u^l$.

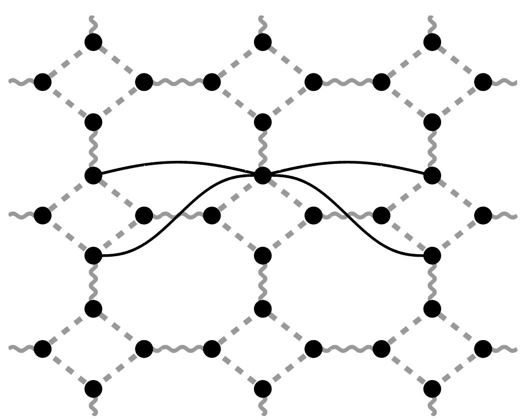


FIGURE 11. The zig-zag product of the grid $\mathbb{Z}^2$ with the 4-cycle.

A few remarks are in order about these constructions. The replacement product described above (when H is a cycle) is well known in graph theory. It is being used often for the purpose of reducing vertex degrees without losing connectivity. Such an argument is used to show for quite a few open problems in graph theory that it suffices to solve the problem for 3-regular connected graphs. (See an example of this below in subsection 9.5.) Gromov [Gro83] has studied the expansion of this product of a d-dimensional cube with an appropriate lower-dimensional cube. In a different context (analysis of expansion in Markov chains via decomposition, which may be viewed as reversing the replacement product), Martin and Randal [MR00] proved the third bound on φ in Theorem 9.1 for the replacement product (the second bound need not hold for the replacement product).

Here is an interesting special case of the zig-zag product. Vizing's theorem says that the edges of an m-regular graph can be colored with $m+1$ colors so that incident edges are colored with different colors. Many families of graphs are known where already m colors suffice. Notice that the graph $G \circledz H$ depends on the arbitrary labels we assign to the edges at each vertex. When G's edges can be m-colored, then we can label each edge by its color. It is easy to verify that in this case $G \circledz H$ is a **lift** of H^2 in the sense of Section 6. So far we have not been able to take advantage of the similarity between these two methods of constructing graphs.

Proof of the zig-zag theorem. We prove here only a weaker version of the second bound on φ. In particular this proof gives $\varphi \leq \alpha + \beta + \beta^2$, which suffices for the expander construction. The stronger bounds stated in Theorem 9.1 have a complex proof in [RVW02] but were greatly simplified in [RTV05].

Clearly $G \circledz H$ is an (mn, d^2)-graph. It is convenient to estimate the spectral gap by considering the random walk on $G \circledz H$. Each step in this walk can be conveniently split into three parts: (i) a random step on an edge within a cloud, (ii) a **deterministic** step on an edge connecting two clouds and (iii) another random step within a cloud. We can now write down the transition matrix Z of the random walk on $G \circledz H$. Let B and $\hat{B}$ be the adjacency matrix of H and the transition matrix of the corresponding random walk respectively. The random steps (i) and (iii) are done on n disjoint copies of H, so the corresponding transition matrix is $\tilde{B} = \hat{B} \otimes I_n$. In the deterministic step (ii) we move from a vertex (v, k) to the unique vertex (u, l) for which $e_v^k = e_u^l$. Consequently, this step is carried out by multiplying with the matrix P which is a permutation matrix of an involution:

$$P_{(v,k),(u,l)} = \begin{cases} 1 & \text{if } e_v^k = e_u^l \\ 0 & \text{otherwise.} \end{cases}$$

It follows that $Z = \tilde{B} P \tilde{B}$. The graph $G \circledz H$ is regular, so the constant vector 1_{mn} is an eigenvector. Therefore what we are claiming is that $|fZf|/||f||^2 \leq \alpha + \beta + \beta^2$ for all $f \perp 1_{mn}$.

We next seek a way to decompose f so as to reflect the fact that $V(G \circledz H) = V(G) \times [m]$. Define $f^{||}$ as the average of f on clouds, viz., $f^{||}(x, i) = \frac{1}{m} \sum_{j \in [m]} f(x, j)$. Define $f^{\perp}$ via $f^{\perp} = f - f^{||}$. Clearly $f^{\perp}$ sums up to zero on each cloud. Let us expand

$$\begin{aligned} |fZf| &= |f\tilde{B}P\tilde{B}f| \\ &\leq |f^{||}\tilde{B}P\tilde{B}f^{||}| + 2|f^{||}\tilde{B}P\tilde{B}f^{\perp}| + |f^{\perp}\tilde{B}P\tilde{B}f^{\perp}|. \end{aligned}$$

Two simplifications follow from the fact that $\tilde{B}$ is a direct sum of n copies of $\hat{B}$.

- Since $\hat{B}1_m = 1_m$, it follows that $\tilde{B}f^{||} = f^{||}$.
- By assumption, $||\hat{B}u|| \leq \beta||u||$ whenever $u \perp 1_m$ and $f^{\perp}$ sums to zero on every cloud. Therefore, $||\tilde{B}f^{\perp}|| \leq \beta||f^{\perp}||$.

In order to deal with $f^{||}$, define the real function g on $V(G)$ via $g(v) = \sqrt{m} f^{||}(v, i)$ (this does not depend on i, of course). Note that with this definition we have $||g||^2 = ||f||^2$. The definition of P implies that $f^{||}Pf^{||} = g\hat{A}g$ where $\hat{A} = \hat{A}_G$ is the transition matrix of the random walk on G. But $f^{||} \perp 1_{mn}$ implies that $g \perp 1_n$ and hence that $g\hat{A}g \leq \alpha||g||^2$. Consequently, $|f^{||}Pf^{||}| \leq \alpha||f^{||}||^2$. Also, both $\tilde{B}$ and P are stochastic matrices and are therefore contractions in l_2. Putting all of this together we conclude

$$|fZf| \leq \alpha ||f^{||}||^2 + 2\beta ||f^{||}|| \cdot ||f^{\perp}|| + \beta^2 ||f^{\perp}||^2.$$

But $||f||^2 = ||f^{||}||^2 + ||f^{\perp}||^2$, so the maximum of this quadratic form is the large eigenvalue of the matrix

$$\begin{pmatrix} \alpha & \beta \\ \beta & \beta^2 \end{pmatrix}.$$

The conclusion follows. □

9.4. **Entropy analysis.** We return to the perspective suggested in subsection 3.1.2 and consider the graph $G \textcircled{z} H$ when both G and H are good expanders. Why is it that entropy grows with each random step on $G \textcircled{z} H$? The intuition below really explains that very informally. At issue is the fact that entropy can no longer grow when the distribution is uniform. As above, we view each step as made up of three substeps: A random step ("zig") in one copy of H, then a deterministic step to a neighboring cloud, and another random step in the new copy of H. Note that steps 1 and 3 are **independent** random steps on H. If the conditional distribution, restricted to one of these two clouds, is far from uniform, then the entropy grows by virtue of H's expansion. The other two steps cannot harm this, since the entropy never decreases when we multiply by a stochastic matrix.

What is more curious is the situation where on a typical cloud the distribution is nearly uniform. In this case Step 1 cannot increase the entropy since the restriction of the distribution to the cloud remains nearly uniform. To see what happens, recall that $V(G \textcircled{z} H) = V(G) \times V(H)$ and consider the two marginal distributions p_G and p_H for p. Since the distribution on most clouds is near uniform, the (deterministic) middle step on $G \textcircled{z} H$ (which uses the cloud value to determine which edge to follow to a neighboring cloud) is then like a real random step on G. Consequently, and since G is an expander, the entropy of p_G increases. But this middle step is a permutation on $V(G \textcircled{z} H)$, so the entropy of the whole distribution remains unchanged. It follows that the entropy of p_H must have decreased. That means that in Step 3 we are in the good case, where the conditional distribution on clouds is far from uniform, and the entropy increases due to the expansion of H. Thus the key is that Step 2 is simultaneously a permutation (so that any progress made in Step 1 is preserved) and an operation whose G-marginal is simply a random step on G.

The linear algebra proof we gave for the zig-zag theorem can be seen as a formal statement of this intuition. Other papers in the area such as [RTV05] follow this approach even more closely. This is particularly helpful in establishing the third bound in Theorem 9.1.

9.5. **An application to complexity theory: $SL = L$.** Assume you arrive in an unfamiliar city and you do not have a map. You would like to get to your hotel (whose street name you know). It is natural to start exploring, hoping to cover every street including the one your hotel is on (an equivalent problem is finding the exit from a maze). Naturally, you can create your own map of the city, which will guarantee that you do not go around in loops. But that takes a lot of memory, linear in the size of the city. Suppose you don't have that. All you can remember is where you are, the street name of your hotel, and perhaps a few more items of that magnitude - in short, only **logarithmic** space in the size of the city.

Formally, this is cast as the problem of exploring a given undirected graph G (or even determining if there is a path between two given vertices s and t).[17]

An important step towards the solution was made by Aleliunas, Karp, Lipton, Lovász, and Rackoff [AKL79]. They showed that this problem can be solved by a **probabilistic** logspace algorithm.[18] To determine if s and t are connected, one simply performs a polynomial length random walk starting at s and checks if the walk ever reaches t.

With all the background we have covered so far, the analysis of this result is simple, so let us sketch it. First observe that the algorithm uses only logarithmic memory: all it needs to remember is its current position and the goal t. As for the time bound, let us first note that we can assume without loss of generality that the graph is regular. (E.g. replace each vertex v by a cycle of length $d(v)$, the degree of v, as in the replacement product described above. This makes the graph 3-regular, maintaining connectivity in the natural way.) Now using the connection between edge expansion $h(G)$ (which is at least $\Omega(1/n)$) in every connected 3-regular graph with n vertices) and the eigenvalue, we conclude via Theorem 4.11 that it must be an $(n, 3, \alpha)$-graph, with $\alpha < 1-\Omega(1/n^2))$. Thus a random walk of length $O(n^3)$ will get exponentially close to the uniform distribution, and if we repeat it n^2 times, resulting in a walk of length n^5, we will not miss a single vertex in this connected component, except with exponentially small probability.

A natural approach to this problem is to try to **derandomize** this probabilistic algorithm: namely, to generate deterministically a walk which explores all vertices in every connected graph. This has, indeed, been tried often (see [Rei05] for background). A key to the success of Reingold's approach was an ingenious application of the zig-zag product.

As mentioned two paragraphs earlier, we can assume the input graph G has constant degree D, of some fixed D of our choice (e.g. we can add self loops to the 3-regular graph mentioned there). The idea of the proof is the following. If each connected component of the graph G is an expander graph, then each component has a logarithmic diameter. Then one can enumerate over all the logarithmically long paths starting at s and check if one of them arrives at t. In short, for graphs whose components are expanders, the logspace algorithm is trivial.

But this is indeed a big "if". The question is how to convert efficiently in logspace an arbitrary graph into an expander. The answer is by using the zig-zag product. Consider the D-regular input graph G, and assume it is connected (otherwise we apply the same argument for each component). The connectivity of G already implies (with an argument identical to the one above for 3-regular graphs) that it is somewhat expanding, i.e. is an (n, D, α)-graph for some $\alpha < 1 - \Omega(1/n^2))$.

To turn G into an expander, assume that $D = d^{16}$ and that we have a $(d^{16}, d, 1/2)$-graph H. Inductively construct the graphs G_i in a similar manner to the expander construction in subsection 9.1, but rather than starting from a fixed size expander as the first graph, start with G itself:

$$G_1 = G, \qquad G_{i+1} = (G_i \,ⓩ\, H)^8 \quad \text{for } i \geq 1.$$

We terminate the sequence after only $k = O(\log n)$ iterations. In each iteration, the size of G increases by a constant factor which is the size of the graph H until

[17] A word to the maven: this problem is complete for the class SL, which simply means that solving it in logarithmic space (in the class L) would imply $SL = L$.

[18] Thus proving $SL \subseteq RL$.

the final iteration yields the graph G_k. The required proposition follows from the following two claims.

Claim 9.4. *The graph G_k is an $(nd^{16k}, d^{16}, 3/4)$-graph.*

Claim 9.5. *Neighborhood queries for G_k can be answered in logspace.*

Note that the second claim is not obvious. In logspace one cannot keep copies of the graphs G_i in memory, and therefore one has to evaluate the recursion anew for each query. This means that large expander graphs constructed by the zig-zag product are **very explicit**, even in a stricter sense than the one required by Definition 2.3. But indeed the intuition that every one of these steps can be performed with only additional **constant** space is correct. This requires a clever construction of an appropriate data structure. We skip the proof of this claim and refer the reader to Reingold's paper [Rei05].

To prove the first claim, we use the third bound in Theorem 9.1. Using it we see that (roughly speaking) the spectral gap doubles in each iteration and thus reaches a constant in logarithmically many iterations:

$$1 - \varphi(\alpha, \beta) \geq (1 - \beta^2) \cdot (1 - \alpha)/2. \tag{18}$$

Let us denote by λ_i, μ_i the normalized second eigenvalue of G_i and $G_i ⓩ H$ respectively. Then, using the parameters of the graph H in Equation (18) yields:

$$1 - \mu_i \geq \frac{3}{8} \cdot (1 - \lambda_i).$$

Therefore,

$$\lambda_{i+1} = \mu_i^8 \leq \left[1 - \frac{3}{8}(1 - \lambda_i)\right]^8 \leq \max\left(\lambda_i^2, \frac{1}{2}\right).$$

Therefore, for $k = O(\log n)$, we obtain $\lambda_k \leq 1/2$ as needed.

We end by noting that a recent paper of Rozenman and Vadhan [RV05] gives a different proof of Reingold's result using a new notion of **derandomized** graph squaring. This product is closely related to zig-zag (so the proof has the same spirit), but the analysis of the space required in their implementation is more straightforward.

10. **Lossless conductors and expanders**

Early in our discussion we introduced the notion of expansion ratio $h(G) = \min_{\{S \mid |S| \leq \frac{n}{2}\}} \frac{|E(S, \overline{S})|}{|S|}$. Indeed, quite a few of the important properties of expander graphs can already be derived from the fact that $h(G)$ is bounded away from zero. For more refined analysis it becomes necessary to consider the edge isoperimetric parameter $\Phi_E(G, k) = \min_{S \subset V}\{|E(S, \overline{S})| \ : \ |S| = k\}$. We saw that a comprehensive analysis of the extremal properties of this parameter would be both interesting and very challenging, though some very good explicit constructions are known (see subsection 5.3). The analogous questions about the **vertex isoperimetric parameter**: $\Phi_V(G, k) = \min_{S \subset V}\{|\Gamma(S) \setminus S| \ : \ |S| = k\}$ seem even more challenging. Simple considerations from subsection 5.1.1 yield that every large set of vertices in a d-regular graph has vertex expansion at most $d - 2 + o(1)$. As mentioned in subsection 4.6, results of Kahale [Kah95] show that d-regular Ramanujan graphs are guaranteed to have vertex expansion of about $d/2$ and that this is essentially tight.

To find out how good these bounds are, we already know that it is a good idea to look at random graphs first. Indeed, as seen in subsection 4.6, for every $\delta > 0$, a random (n, d)-graph almost surely satisfies $\Psi_V(G, \epsilon n) \geq d-2-\delta$ for some sufficiently small constant $\epsilon > 0$. Since we know how to construct (n, d)-graphs in which small sets have a vertex expansion of $d/2$, it is natural to enquire why we bother about increasing this factor to $d-2$ or so. Indeed, this question is motivated by more than sheer curiosity. There is a variety of applications that require (n, d)-graphs in which vertex sets of size ϵn have vertex expansion γd for $\gamma > 1/2$. Such applications include networks that can implement fast distributed, routing algorithms, e.g. [PU89, ALM96, BFU99, HMP06], as well as the construction of expander-based linear codes, e.g. [SS96, Spi96, LMSS01]. Analogous, irregular, even highly unbalanced graphs are used in various storage schemes [UW87, BMRV02] and in proving lower bounds on computational complexity (the generation of hard tautologies for various proof systems [BSW01, ABSRW04, AR01, BOGH03]).

It is still a major challenge to construct families of (n, d)-graphs in which vertex sets of size ϵn have vertex expansion γd for $\gamma > 1/2$. However, in this section we present a construction due to Capalbo, Reingold, Vadhan, and Wigderson [CRVW02] that takes a substantial step in this direction. For every $\delta > 0$ and sufficiently large d, this is an explicit construction of families of bipartite expander graphs whose left degree is d and whose left expansion is $(1-\delta)d$ for small sets of linear size. The construction is based on a generalization of the zig-zag product to **conductors**. This will offer us a glimpse into the realm of randomness-enhancing objects, which, aside from expanders include creatures such as conductors, extractors, and dispersers (we recommend again the excellent survey [Sha04]).

10.1. Conductors and lossless expanders.

10.1.1. *Conductors.* The main difficulty we face here is that the spectral gap which has served us very well so far seems, by the aforementioned result of Kahale, no longer adequate for our purposes. We therefore revert to the language of entropy. A natural choice would be to use the min-entropy H_∞ defined in Section 3. A lower bound on the min-entropy is a rather strong condition. The inequality $H_\infty(p) \geq k$ means that no point has probability exceeding 2^{-k}. (In contrast, a lower bound on the Shannon entropy means that this holds only on average.) In this light we consider a weaker condition and ask only how close, in **total variation distance** (or equivalently in l_1 norm), our distribution is to one with high min-entropy. We introduce now some terminology.

Definition 10.1. A **k-source** is a distribution with min-entropy at least k. A distribution is called a **(k, ϵ)-source** if there is a k-source at l_1 distance $\leq \epsilon$ from it.

A main aspect of the present section is to emphasize an idea that has already appeared in Sections 3 and 9: namely, that it is useful to view the entropy of a distribution as a physical entity which we transfer around the graph as necessary. We develop appropriate terminology to advance this idea. In particular, since our main concern here is with bipartite graphs, we will view a bipartite graph as a function. Our bipartite graphs have vertex sets Left and Right, with all Left vertices of the same degree. This function associates with a given Left vertex x and an edge label i, the Right vertex that is the i-th neighbor of x. To facilitate our analysis of entropy, we use bit strings to name the vertices and edge labels. As in subsections 3.1.2 and 9.4, we are concerned with the following question. Consider a distribution on the Left vertices of known entropy. Now take a random step along an edge to the Right. This induces a distribution on the Right vertices, the entropy of which we seek to bound from below. This problem is being attacked by means of several classes of graphs which serve as building blocks for the final construction. The common theme is this: Given a bound on the "incoming" entropy, we seek a lower bound on the amount of entropy which is coming out (up to a small l_1 distance). In keeping with the common terminology in the area of pseudorandom number generators and extractors [RVW02, Sha04], we refer to the choice of an edge to be taken in the next step as the randomness "injected" into the process or as the "seed" being used. Here is our toolbox.

Denote the uniform distribution over $\{0,1\}^d$ by U_d.

Definition 10.2. A function $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k_{\max}, a, \epsilon)$-**conductor** if for any $k \le k_{\max}$ and any k-source X over $\{0,1\}^n$, the distribution $E(X, U_d)$ is a $(k+a, \epsilon)$-source.

The analysis of the type conducted in subsection 9.4 must always be adapted to situations where the entropy reaches its maximum either locally or globally. Recall (subsection 3.1.2) that a distribution on an M-element set has entropy $\le \log M$. Most of the tools we describe below are geared to handle this problem.

(1) A function $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is an (a,ϵ)-**extracting conductor** if it is an $(m-a, a, \epsilon)$-conductor.
In particular, if the input entropy is $m-a$, the output will be ϵ-close to uniform. We mention that an explicit construction of an (a,ϵ)-extracting conductor can be derived from known constructions (namely that of an $(m-a,\epsilon)$-extractor; see [Sha04]. This is a well-known and easy to construct object).

(2) A function $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a $(k_{\max}, \epsilon)$-**lossless conductor** if it is a $(k_{\max}, d, \epsilon)$-conductor.
We view the specification of the edge label as an "injection of randomness" and wish **none** of this additional entropy to be lost. The construction of these graphs is our ultimate goal.

(3) A pair of functions $\langle E, C\rangle\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m \times \{0,1\}^b$ is a $(k_{\max}, a, \epsilon)$-**buffer conductor** if E is an (a,ϵ)-extracting conductor and $\langle E, C\rangle$ is an $(k_{\max}, \epsilon)$-lossless conductor.
This notion is intended to assure that no entropy is lost when the input randomness is too high. Whatever entropy is lost by the first function is to be saved completely in the second, which may be viewed as an overflow buffer or bucket.

(4) A pair of functions $\langle E, C\rangle\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m \times \{0,1\}^b$, where $n+d = m+b$ is an (a,ϵ)-**permutation conductor** if E is an (a,ϵ)-extracting conductor and $\langle E, C\rangle$ is a permutation over $\{0,1\}^{n+d}$.
This is a special case of a buffer conductor, where the requirement that $\langle E, C\rangle$ is a $(k_{\max}, \epsilon)$-lossless conductor follows since $\langle E, C\rangle$ is a permutation.

As we shall see, lossless conductance implies lossless expansion, which was our motivating problem to begin with.

10.1.2. *Lossless expanders.* Let $G = (V_L, V_R; E)$ be a bipartite graph such that $|V_L| = N$, $|V_R| = M$, and all left vertices have degree D.

Definition 10.3. The graph G is a $(K_{\max}, \epsilon)$-lossless expander if every set of $K \leq K_{\max}$ left vertices has at least $(1-\epsilon)DK$ neighbors.

That is, in a lossless expander sufficiently small, left vertex sets have almost the maximal expansion possible. An alternative view is that most of the neighbors of a small left set are unique neighbors, i.e. neighboring a single vertex of the set. Naturally, $K_{\max}$ should be somewhat smaller than M/D for this to be possible.

As mentioned above, we view a conductor $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ as a bipartite graph with $N = 2^n$ left vertices, $M = 2^m$ right vertices, where each left vertex x is connected to $D = 2^d$ right vertices $E(x, \cdot)$. It is easy to check that from this perspective, a $(k_{\max}, \epsilon)$-lossless conductor is a $(K_{\max}, \epsilon)$-lossless expander, where $K_{\max} = 2^{k_{\max}}$.

Theorem 10.4. *For any $\epsilon > 0$ and $M \leq N$, there is an explicit family of left D-regular bipartite graphs that are $(\Omega(\epsilon M/D), \epsilon)$-lossless expanders, where $D \leq (N/\epsilon M)^c$ for some constant c.*

We note that in the (useful) case where M/N and ϵ are bounded below by a constant, the degree D will be constant as well. As can be expected, our approach to this theorem is to construct an explicit family of $(m - d - \log(1/\epsilon) - O(1), \epsilon)$-lossless conductors, where $d = O(n - m - \log(1/\epsilon))$.

10.2. **The construction.** The required lossless conductors are constructed using a zig-zag product for conductors. However, before we define this product, we need to recall the definition of the zig-zag product given in Section 9 and adapt it slightly to the case of bipartite graphs. Figure 12 should help the formal definition.

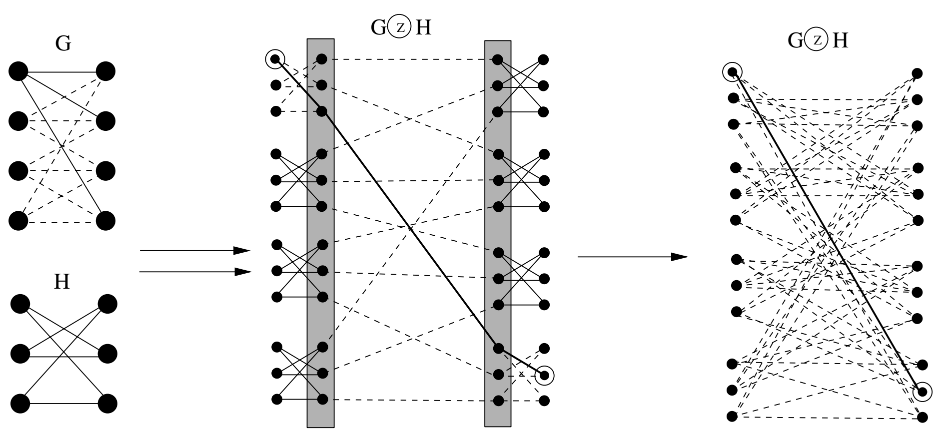


FIGURE 12. Zig-zag product of bipartite graphs.

Definition 10.5 (Zig-zag product for bipartite graphs)**.** Let H be a d-regular bipartite graph with s vertices on each side, and let G be an s-regular bipartite graph with n vertices on each side. The zig-zag product $G ⓩ H$ is a d^2-regular bipartite graph with sn vertices on each side, where the left and right sides are arranged as n copies of H, one per each vertex of G. The edges emanating from a left vertex $(x, y) \in [n] \times [s]$ are labeled by $[d] \times [d]$. The edge labeled (a, b) is determined as follows:

(1) Take a left to right step in the local copy of H, using a to choose an edge.
(2) Take a left to right step along an edge of G, between copies of H. More precisely, suppose we are at (x, y'). Let $x' \in G$ be the y'-th neighbor of x. Suppose that on x' neighbor list, x is the z-th neighbor of x'. Then the second step takes us from (x, y') to (x', z).
(3) Take a left to right step in the new local copy of H, using b to choose an edge.

Recall that by the zig-zag theorem from Section 9, $Gⓩ H$ is an expander if both G and H are such. Moreover, the degree of $G ⓩ H$ is controlled by the degrees in H, while its size and expansion are related to both G and H. Unfortunately, while $\deg(G ⓩ H) = \deg^2(H)$, the vertex expansion of $G ⓩ H$ cannot be better than either the expansion of H or the expansion of G. In particular, it will never exceed $d = \sqrt{deg(Gⓩ H)}$. This can be seen by considering the expansion of a set consisting of a single copy of H on the left. This is a far cry from the lossless expanders we seek. Clearly, in this example the "injected" entropy is wasted since we have begun with a uniform distribution on a copy of H, which entropy cannot be increased by a walk inside H. This problem will be solved by means of a buffer in which we save the injected entropy in such cases. We now formalize this idea and define the zig-zag product for conductors.

10.2.1. *The zig-zag product for conductors.* Our strategy to avoid entropy loss is to save (or buffer) leftover entropy from the random choices made at each step. Then, we use a lossless conductor with some truly random bits to "condense" the leftover entropy. As suggested by the name "conductor", we make an analogy with the flow of electricity or water. Indeed, we think of pouring randomness (water) into a conductor and collecting the leftovers (unused randomness beyond the $k_{\max}$ bound) into a bucket for later use.

The zig-zag product for conductors must be carried out with carefully selected parameters to yield the constant degree lossless expanders/conductors required by Theorem 10.4. To define the product, we need three objects $\langle E_1, C_1\rangle$, $\langle E_2, C_2\rangle$, and E_3 (which respectively replace the roles of the three steps H, G, and H in the original zig-zag product. We avoid specifying their most general parameters here and pick a certain set of parameters which suffices for our purposes. The general achievable sets of parameters for an explicit construction can be found in [CRVW02] (E_1 from Lemma 4.4, and E_2, E_3 from Lemma 4.13). So, let us assume we have in our hands:

(1) $\langle E_1, C_1\rangle\colon \{0,1\}^{n_1} \times \{0,1\}^{d_1} \to \{0,1\}^{m_1} \times \{0,1\}^{b_1}$, a permutation conductor.
(2) $\langle E_2, C_2\rangle\colon \{0,1\}^{n_2} \times \{0,1\}^{d_2} \to \{0,1\}^{d_1} \times \{0,1\}^{b_2}$, a buffer conductor.
(3) $E_3\colon \{0,1\}^{b_1+b_2} \times \{0,1\}^{d_3} \to \{0,1\}^{m_3}$, a lossless conductor.

The zig-zag product for conductors produces the conductor $E\colon \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$, where $n = n_1 + n_2$, $d = d_2 + d_3$, and $m = m_1 + m_3$. Let $x_1, x_2, r_2, r_3, y_1, y_2,$

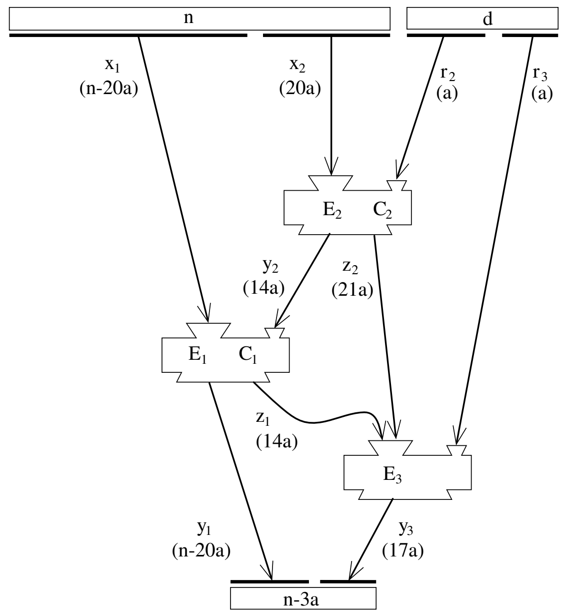


FIGURE 13. Entropy flow in a lossless conductor.

and y_3 be binary strings of respective lengths $n_1, n_2, d_2, d_3, m_1, d_1$, and m_3. Then, as depicted in Figure 13, we evaluate $y_1y_3 = E(x_1x_2, r_2r_3)$ by the following three steps:[19]

- $(y_2, z_2) = \langle E_2, C_2\rangle(x_2, r_2)$
- $(y_1, z_1) = \langle E_1, C_1\rangle(x_1, y_2)$
- $y_3 = E_3(z_1z_2, r_3)$

Recall that the zig-zag product for bipartite graphs $G ⓩ H$ uses H twice. Here, the first use is replaced with $\langle E_2, C_2\rangle$ to ensure that when x_2 has high min-entropy, y_2 is close to uniform and is a good seed for $\langle E_1, C_1\rangle$. The second use of H is replaced with lossless conductor E_3 that transfers entropy lost in $\langle E_1, C_1\rangle$ and $\langle E_2, C_2\rangle$ to the output. The deterministic step of the zig-zag product using the graph G is replaced with $\langle E_1, C_1\rangle$, which as before doesn't use any new random bits and whose output is just a permutation of its input (which, however, moves entropy about to allow more to come in later).

To simplify subsequent discussion, we focus on a specific example of achievable parameters for the three building blocks $\langle E_1, C_1\rangle$, $\langle E_2, C_2\rangle$ and E_3, whose zig-zag product, depicted in Figure 13, results in a lossless conductor.

10.2.2. *Proof of the main theorem.* We now need a technical lemma, showing how to partition a joint distribution according to conditional min-entropy. Recall that two distributions are called ϵ-close if their l_1 distance is at most ϵ.

Lemma 10.6. *Let (X_1, X_2) be a probability distribution on a finite product space. Given $\epsilon > 0$ and a, there exists a distribution (Y_1, Y_2) on the same space such that*

- *The distributions (X_1, X_2) and (Y_1, Y_2) are ϵ-close.*

[19] If u and v are bit strings, then uv denotes their concatenation.

- *The distribution* (Y_1, Y_2) *is a convex combination of two other distributions* $(\hat{Y}_1, \hat{Y}_2)$ *and* $(\check{Y}_1, \check{Y}_2)$*, each having min-entropy at least* $H_\infty(X_1, X_2) - \log(1/\epsilon)$.
- *For all* $x \in \mathrm{Supp}(\hat{Y}_1)$ *we have* $H_\infty(\hat{Y}_2 | \hat{Y}_1 = x) \geq a$.
- *For all* $x \in \mathrm{Supp}(\check{Y}_1)$ *we have* $H_\infty(\check{Y}_2 | \check{Y}_1 = x) < a$.

Proof. First, split $\mathrm{Supp}(X_1)$ according to $H_\infty(X_2|X_1 = x)$:

$$\hat{S} = \{z : H_\infty(X_2|X_1 = z) \geq a\}, \quad \check{S} = \{z : H_\infty(X_2|X_1 = z) < a\}.$$

Then, define $(\hat{Y}_1, \hat{Y}_2)$ and $(\check{Y}_1, \check{Y}_2)$, so that $\hat{Y}_1$ and $\check{Y}_1$ have disjoint supports, $\hat{S}$ and $\check{S}$ respectively:

$$\begin{aligned} \Pr\left[(\hat{Y}_1, \hat{Y}_2) = (z_1, z_2)\right] &= \Pr\left[(X_1, X_2) = (z_1, z_2) \,\middle|\, X_1 \in \hat{S}\right], \\ \Pr\left[(\check{Y}_1, \check{Y}_2) = (z_1, z_2)\right] &= \Pr\left[(X_1, X_2) = (z_1, z_2) \,\middle|\, X_1 \in \check{S}\right]. \end{aligned}$$

Let $p = \Pr[X_1 \in \hat{S}]$. Then the probability of each value in $(\hat{Y}_1, \hat{Y}_2)$ is multiplied by $1/p$, and the probability of each value in $(\check{Y}_1, \check{Y}_2)$ is multiplied by $1/(1-p)$. Therefore, if $\epsilon \leq p \leq 1 - \epsilon$ then the min-entropy of $(\hat{Y}_1, \hat{Y}_2)$ and $(\check{Y}_1, \check{Y}_2)$ is reduced by at most $\log(1/\epsilon)$. In this case, we define $(Y_1, Y_2) = (X_1, X_2)$, and we are done, since:

$$(Y_1, Y_2) = (X_1, X_2) = p(\hat{Y}_1, \hat{Y}_2) + (1-p)(\check{Y}_1, \check{Y}_2).$$

Otherwise, assume $p < \epsilon$ (the case $p > 1 - \epsilon$ is similar). In this case since $(\check{Y}_1, \check{Y}_2)$ still has sufficiently high min-entropy, we take $(Y_1, Y_2) = (\check{Y}_1, \check{Y}_2)$. This distribution is ϵ-close to (X_1, X_2), since:

$$\sum_{z_1 \in \hat{S}, z_2} \left| \Pr\left[(X_1, X_2) = (z_1, z_2)\right] - \Pr\left[(\check{Y}_1, \check{Y}_2) = (z_1, z_2)\right] \right| \leq p < \epsilon,$$

$$\begin{aligned} &\sum_{z_1 \in \check{S}, z_2} \left| \Pr\left[(X_1, X_2) = (z_1, z_2)\right] - \Pr\left[(\check{Y}_1, \check{Y}_2) = (z_1, z_2)\right] \right| \\ &\qquad \leq \left(\frac{1}{1-p} - 1\right)(1-p) = p < \epsilon. \end{aligned}$$ □

We now pick specific parameters for our building blocks, which can be explicitly constructed, and prove that combining them using the zig-zag product for conductors yields a lossless expander as stated in Theorem 10.4.

Fix the parameters $a = 1000 \log(1/\epsilon)$ and $d = 2a$. Then:

- $\langle E_1, C_1 \rangle \colon \{0,1\}^{n-20a} \times \{0,1\}^{14a} \to \{0,1\}^{n-20a} \times \{0,1\}^{14a}$ is an $(n - 30a, 6a, \epsilon)$-permutation conductor.
- $\langle E_2, C_2 \rangle \colon \{0,1\}^{20a} \times \{0,1\}^{a} \to \{0,1\}^{14a} \times \{0,1\}^{21a}$ is a $(14a, 0, \epsilon)$-buffer conductor.
- $E_3 \colon \{0,1\}^{35a} \times \{0,1\}^{a} \to \{0,1\}^{17a}$ is a $(15a, a, \epsilon)$-lossless conductor.

We state again the consequence for these choices, which suffices to prove the main Theorem 10.4.

Claim 10.7. *The resulting conductor* $E \colon \{0,1\}^n \times \{0,1\}^{2a} \to \{0,1\}^{n-3a}$ *is an* $(n - 30a, 2a, 4\epsilon)$*-lossless conductor.*

We first deal with the explicitness of our building blocks (more details can be found in [CRVW02]). Note the sizes of E_2 and E_3 are fixed constants, and so there is no issue concerning their explicit construction. They can be shown to exist by a simple probabilistic argument and can then be found by exhaustive search. The conductor E_1 needs to be arbitrarily large, but its explicit construction follows known explicit expanders. Indeed, any graph with sufficiently small second eigenvalue (close to Ramanujan for these parameters) is such a conductor E_1, and the proof follows, due to the equivalence of min-entropy and the Rényi entropy: $H_\infty(X) \leq H_2(X) \leq 2H_\infty(X)$ as shown in Proposition 3.5.

Let us follow the entropy flow from the input (x_1x_2, r_2r_3) to the output y_1y_3. We would like to prove that if $H_\infty(X_1, X_2) = k$, then y_1y_3 is a $(k+2a, 4\epsilon)$-source, as long as $k \leq n - 30a$. For ease of discussion, we first ignore the small l_1-errors in the outputs of all conductors (in other words, assume for simplicity that in our building blocks $\epsilon = 0$). These errors will simply be added at the end to give the final error of the lossless conductor E.

We prove first that E_1 and E_2 together transfer enough entropy into Y_1, namely:

$$H_\infty(Y_1) \geq k - 14a. \tag{19}$$

By Lemma 10.6 it suffices to prove this bound **only** in the two extreme cases when $H_\infty(X_2|X_1 = x_1)$ are **all** large or **all** small, as we vary all attainable values for x_1.

Case 1 : For all $x_1 \in \mathrm{Supp}(X_1)$, we have $H_\infty(X_2|X_1 = x_1) \geq 14a$.
In this case, since E_2 is an $(0, \epsilon)$-extracting conductor, $H_\infty(Y_2|X_1 = x_1) = 14a$, for any $x_1 \in \mathrm{Supp}(X_1)$. Therefore Y_2 is uniform and can be used as a seed for $\langle E_1, C_1\rangle$ for any $x_1 \in \mathrm{Supp}(X_1)$. Since $H_\infty(X_1) \geq k - 20a$, and since E_1 is a $(6a, \epsilon)$-extracting conductor, E_1 conducts $6a$ bits of entropy from the seed into Y_1, and we obtain $H_\infty(Y_1) \geq k - 14a$.

Case 2 : For all $x_1 \in \mathrm{Supp}(X_1)$, we have $H_\infty(X_2|X_1 = x_1) \leq 14a$.
Since $H_\infty(X_1, X_2) = k$, it follows that $H_\infty(X_1) \geq k - 14a$. Therefore, since E_2 is a lossless extractor, $H_\infty(Y_2|X_1 = x_1) \geq H_\infty(X_2|X_1 = x_1)$ for any $x_1 \in \mathrm{Supp}(X_1)$. It follows that $H_\infty(X_1, Y_2) \geq H_\infty(X_1, X_2) = k$. Since $\langle E_1, C_1\rangle$ is a permutation, also $H_\infty(Y_1, Z_1) \geq k$, and again we get that $H_\infty(Y_1) \geq k - 14a$.

Observe that both $\langle E_1, C_1\rangle$ and $\langle E_2, C_2\rangle$ conserve entropy, since $\langle E_1, C_1\rangle$ is a permutation conductor and $\langle E_2, C_2\rangle$ is a buffer conductor. Therefore:

$$k + a = H_\infty(X_1, X_2, R_2) = H_\infty(X_1, Y_2, Z_2) = H_\infty(Y_1, Z_1, Z_2). \tag{20}$$

To conclude the argument, consider any $y_1 \in \mathrm{Supp}(Y_1)$, and note that by the bound of $k+a$ on the total entropy and the lower bound proved above for $H_\infty(Y_1)$, we must have $H_\infty(Z_1, Z_2|Y_1 = y_1) \leq 15a$.

Thus, E_3, which is a $(15a, a, \epsilon)$-conductor, conducts a bits of entropy from R_3 to Y_3. That is, all the entropy of Z_1, Z_2 is transferred to the output Y_3 without any entropy loss, $H_\infty(Y_3|Y_1 = y_1) = H_\infty(Z_1, Z_2|Y_1 = y_1) + a$. Together with (19) and (20) this implies that

$$\begin{aligned} H_\infty(Y_3|Y_1 = y_1) - \log \Pr[Y_1 = y_1] &= H_\infty(Z_1, Z_2|Y_1 = y_1) - \log \Pr[Y_1 = y_1] + a \\ &\geq H_\infty(Z_1, Z_2, Y_1) + a = k + 2a. \end{aligned}$$

Therefore, $H_\infty(Y_1, Y_3) = k + 2a$, as claimed.

Finally to see the dependence on ϵ, note that these l_1-errors on the extractor outputs add up.[20] The probability of falling in this error event is at most ϵ each time we have such l_1-error. In the above analysis of the function E, we make four moves from a variable to its ϵ-close counterpart, one for each of the building blocks $\langle E_1, C_1\rangle$, $\langle E_2, C_2\rangle$, E_3, and one in the use of Lemma 10.6. Thus, the actual value of Y_1Y_3 is 4ϵ-close to the value obtained in the error-free analysis. This completes the proof that E is an $(n-30a, 2a, 4\epsilon)$-lossless conductor.

10.2.3. *Final comments.* We note again that there is no known algebraic way for constructing lossless expanders. Such strong vertex expansion does not seem to be implied by simple algebraic parameters of the graph. Indeed, the construction in this section provides only bipartite graphs which are losslessly expanding in **one** direction. Although this is sufficient for most known applications, there are exceptions such as [HMP06] that require a better expansion guarantee. It is a very interesting challenge to explicitly construct simple graphs that exhibit similar expansion to that of a random (n, d)-graph, as seen in Theorem 4.16. This is formally stated in the following open problem. In particular, we do not know whether the bounds in Theorem 4.16 apply to Cayley graphs as well.

Open problem 10.8. For any $\delta > 0$ and sufficiently large d, give an explicit construction of an arbitrarily large (n, d)-graph G satisfying $\Psi_V(G, \epsilon n) \geq d-2-\delta$, where $\epsilon = \epsilon(\delta, d)$.

We note that a somewhat easier problem, unique vertex expansion (which is implied by lossless expansion, but likewise eludes algebraic attacks), was studied by Alon and Capalbo [AC02], who gave explicit constructions of such graphs.

11. **Cayley expander graphs**

Cayley graphs offer a combinatorial depiction of groups and their generators. In this section we describe some methods to compute the second eigenvalue of Cayley graphs, or a relative of the spectral gap, the **Kazhdan constant**. Our methods include the Fourier Transform (as featured already in Section 8) and Group Representations. Serre's classical book [Ser77] is a good introduction to this beautiful theory. We explain how to use these methods to establish expansion for some examples seen in earlier sections and mention the remarkable recent progress in understanding expansion in Cayley graphs of **simple** groups.

We then describe a beautiful connection between a well-known group operation, the **semidirect product**, and the zig-zag product of graphs (described in Section 9). This connection is a basis of two elementary constructions of Cayley expander graphs for nonsimple groups with bounded, or very slowly growing, degrees. This follows the work of Alon, Lubotzky, and Wigderson [ALW01], Meshulam and Wigderson [MW02], and Rozenman, Shalev and Wigderson [RSW04].

Definition 11.1. Let H be a group and $S \subseteq H$ a subset thereof. The corresponding **Cayley graph** $\mathrm{C}(H, S)$ has H as a vertex set and (g, h) is an edge if $g \cdot s = h$ for some $s \in S$. On occasion we consider S as a multiset, with several copies of each

[20]Since if two random variables are ϵ-close, one can be written as a convex combination in which the other has weight at least $1-\epsilon$, and another arbitrary distribution (capturing the error).

element. In this case the Cayley graph has (the appropriate number of) parallel edges between vertices.

This definition yields a directed graph. However, we will usually assume that the set S is **symmetric**; i.e. $s \in S$ implies $s^{-1} \in S$. In this case $\mathrm{C}(H,S)$ is undirected and $|S|$-regular. We say that S **generates** H if every $h \in H$ can be written as a word in S; i.e. $h = s_1 \cdot s_2 \ldots \cdot s_k$ with $s_i \in S$. Note that S generates H iff the Cayley graph $C(H,S)$ is connected.

Definition 11.2. We will use the shorthand notation $\lambda(H,S)$ for $\lambda(\mathrm{C}(H,S))$, the second eigenvalue (in absolute value) of the normalized adjacency matrix of this Cayley graph. The same for all eigenvalues, in particular the second (without absolute value), $\lambda_2(H,S)$. We also define its *spectral gap* $\mathrm{g}(H,S) = 1 - \lambda_2(H,S)$.

Here are some very simple examples of Cayley graphs:

- The additive cyclic group $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ of integers modulo n is generated by $S = \{+1,-1\}$. The graph $\mathrm{C}(\mathbb{Z}_n,S)$ is the cycle on n vertices.
- The additive group $\mathbb{Z}_2^d$ of the d-dimensional vector space over the field $\mathbb{Z}_2$ with its standard set of generators $S = \{e_1 = (1,0,\ldots,0), e_2, \ldots, e_d\}$. Here $\mathrm{C}(\mathbb{Z}_2^d, S)$ is the graph of the d-cube (see subsection 4.2.1).
- For any group H, when $S = H$ the Cayley graph $\mathrm{C}(H,S)$ is the complete graph on vertex set H (plus a loop at each vertex).

Cayley graphs form a rather special class of graphs. For example, as we now show they are all vertex-transitive.

Definition 11.3. An **automorphism** of a graph is a permutation π of the vertices such that (v,w) is an edge if and only if $(\pi(v),\pi(w))$ is an edge. A graph is called **vertex transitive** if for any ordered pair of vertices v,w there is some automorphism of the graph that sends v to w.

Claim 11.4. *Every Cayley graph is vertex transitive.*

Proof. Let $G = \mathrm{C}(H,S)$ be a Cayley graph, and let g,h be two vertices. The mapping $x \to hg^{-1}x$ is an automorphism of G that sends g to h. □

Constructions of Cayley graph expanders - an overview. The special structure of Cayley graphs makes it easier to bound their second eigenvalues in many cases, or at least derive such bounds from known theorems in other areas.

The first construction of an explicit expander graph family was given by Margulis [Mar73] in 1973. We described his construction in Section 8 and noted that while it is given in terms of affine action of $SL_2(p)$, it is actually derived from Cayley graphs on the groups $SL_3(p)$ of 3×3 - matrices with determinant 1 over the field with p elements (where p varies over all primes). In these Cayley graphs the generating sets are all the **elementary matrices**, namely those with 1's on the main diagonal and an extra ± 1 at one other entry. This yields 12-regular Cayley graphs. They are shown to be expanders with second eigenvalue at most $1-\varepsilon$ for some fixed $\varepsilon > 0$, independent of p. The proof relies on a theorem of Kazhdan that the "mother group" of these finite groups, namely $SL_3(\mathbb{Z})$, has the Kazhdan property (T). From this the spectral gap follows simultaneously for all these quotient groups. Margulis' reduction from the Cayley graphs on SL_3 to the affine action of SL_2 follows Kazhdan's proof of property (T) for these groups.

Lubotzky [Lub94] proved a similar result directly for Cayley graphs on $SL_2(p)$. While the "mother group" here, $SL_2(\mathbb{Z})$, fails to have Kazhdan property (T), Lubotzky observes that Selberg's celebrated 3/16 theorem suffices to establish the spectral gap on all these quotients in an identical way. He calls this weaker property (τ), and his book [Lub94] (as well as his forthcoming book with Zuk [LZ]) elaborates on the usefulness of this property in a variety of contexts.

The related family of groups $PSL_2(p)$ features in the celebrated Ramanujan graphs [LPS88, Mar88], as described in subsection 5.3. Here different generators are used. The optimal eigenvalue bounds are based on deep results of Deligne solving the Ramanujan conjecture for varieties over finite fields, as well as subsequent estimates.

Quite surprisingly, about 20 years elapsed until the next slew of constructions for Cayley expanders appeared. These have a very diverse nature and use a variety of new techniques. The zig-zag product and its connection to the semidirect product were used [MW02, RSW04] to construct expander Cayley graphs. These constructions work iteratively and apply for some nonsimple groups. This is described at the end of this section.

We say that an infinite family of groups $\{H_i\}$ **can be made into a family of expanders** if there is a constant d and a generating set S_i of size d in each H_i so that the family $\mathrm{C}(H_i, S_i)$ is a family of expanders as in Definition 2.2.

We recall that a finite group is **simple** if it has no nontrivial normal subgroups. Such groups can be viewed as the building blocks from which all finite groups can be constructed. The complete classification of finite simple groups is one of the major achievements of modern mathematics. An understanding of expansion in almost all finite simple groups has recently evolved. Kassabov [Kas05a], following previous work of Shalom [Sha99] and others, showed that the family of simple groups $SL_n(p^m)$ for all $n \geq 3, m \geq 1$ and p prime can be made into a family of expanders. Lubotzky [Lub] derived similar results for the family $SL_2(p^m)$.

In another breakthrough [Kas05b], Kassabov showed that the family of alternating and symmetric groups A_n and S_n can also be made into a family of bounded-degree expanders. His construction combines ingenious combinatorial arguments with estimates on characters of the symmetric group due to Roichman [Roi96].

Using results by Nikolov [Nik05], Lubotzky [Lub] showed that a family consisting of most simple groups can be made into a family of expanders. (We only know this for "most" simple groups, since the claim is not known to hold for the so-called simple groups of Suzuki type). All these exciting developments are explained in [KLN05].

It is natural to seek other families of groups which can be made into a family of expanders. The next proposition shows that this is not always possible.

Proposition 11.5. *Let H be an Abelian group, and let S be a generating set such that $\lambda(H, S) \leq 1/2$. Then $|S| \geq \log |H|/3$.*

A deeper result of Lubotzky and Weiss [LW93] shows that solvable groups of bounded derived length cannot be made expanders with generating sets of bounded size.

Proof. As mentioned in subsection 2.4, an (n, d, α)-graph has a logarithmic diameter. Specifically, if the cardinality of H is n, the Cayley graph $\mathrm{C}(H, S)$ has diameter smaller than $3 \log n$. In particular, every element of H can be written as a product of at most $3\log n$ elements of S. In an Abelian group, the number of different products of length l of elements of S is equal to the number of partitions of l to $|S|$ parts. Define $m = |S|$. The number of such partitions is then $\binom{l+m-1}{l}$. Consequently, $\sum_{j\leq l}\binom{j+m-1}{j} \geq n$. Substitute $l = 3\log n$ to conclude that $|S| = m \geq \frac{1}{10}\log n$. This calculation uses a standard estimate $\binom{k}{l} \leq (ke/l)^l$. □

The following theorem of Alon and Roichman [AR94] proves that Abelian groups are extreme in the sense considered in Proposition 11.5. Indeed **any** group, Abelian or not, has a choice of logarithmically many generators which yield an expanding Cayley graph. In fact, a **random** subset of this size suffices.

Theorem 11.6 (Alon-Roichman [AR94]). *Let H be a group and let S be a subset chosen randomly uniformly in H with size $100\log|H|$. Then $\lambda(C(H,S)) < 1/2$ with probability at least 0.5.*

11.1. **Representations of finite groups.** Here we give a short introduction to necessary parts of representation theory and explain its usefulness for understanding the eigenvalues of Cayley graphs (and Schreier graphs, which we define below). For an excellent text on this theory we refer the reader to [Ser77].

The simplest representations of a group are **characters**, whose definition and basic properties were given in subsection 8.1.1. The important connection of characters to eigenvectors and eigenvalues of Cayley graphs is given below.

Proposition 11.7. *Let M be the normalized adjacency matrix of a Cayley graph $C(H,S)$. Let χ be a character of H. Then the vector $(\chi(h))_{h\in H}$ is an eigenvector of M, with eigenvalue $1/|S|\cdot\sum_{s\in S}\chi(s)$. In particular, the trivial character corresponds to the trivial eigenvalue 1.*

Proof. The proof follows from the following simple calculation:

$$(M\cdot\chi)(x) = \frac{1}{|S|}\cdot\sum_{s\in S}\chi(xs) = \frac{1}{|S|}\sum_{s\in S}(\chi(x)\cdot\chi(s)) = \frac{1}{|S|}\left(\sum_{s\in S}\chi(s)\right)\cdot\chi(x). \qquad \square$$

Together with proposition 8.5 we obtain a simple method to determine the eigenvalues of C(H,S) when H is Abelian. We simply compute $1/|S|\sum_{s\in S}\chi(s)$ for all the characters χ of H. Here are two examples, which have already appeared in subsection 4.5.

Eigenvalues of the discrete cube. We mentioned in subsection 8.1.1 the characters of $\mathbb{Z}_2^d$. For the character χ_a it holds that $\chi_a(e_i) = -1$ if $a_i = 1$ and $\chi_a(e_i) = 1$ otherwise. The corresponding eigenvalue is $(1/d)\cdot\sum_i(\chi_a(e_i)) = 1-2w(a)/d$, where $w(a)$ is the (Hamming) weight of a, namely the number of 1's in it. The eigenvalue set is therefore $\{1-\frac{2k}{d}\,|\,k = 0,\ldots,d\}$, where the eigenvalue $1-\frac{2k}{d}$ appears with multiplicity $\binom{d}{k}$. The largest nontrivial eigenvalue is thus $1-2/d$, which corresponds to every character χ_a where $a = e_j$ for some $d \geq j \geq 1$.

Eigenvalues of the cycle. From the remarks in subsection 8.1.1 it follows that the eigenvalues corresponding to the character χ_k is $(e^{2\pi k} + e^{-2\pi k})/2 = \cos(2\pi k)$. Thus the eigenvalues of the n-cycle are the numbers $\cos(2\pi k)$ for $k = 0,\ldots,n-1$. The largest nontrivial eigenvalue is $\cos(2\pi/n) = 1-\Theta(1/n^2)$.

11.1.1. *Representations and irreducible representations.* The method just discussed that works so well for Abelian groups does not, in general, apply to non-Abelian groups, for those may fail to have any nontrivial characters. To deal with this more complicated case we need to introduce a generalization of characters that exists for every group, namely, **group representations**. Rather than characters which are homomorphisms into the complex plane, we consider homomorphisms into matrix groups. We describe only the basic tools we need of this beautiful theory, and refer the reader to [Ser77] for more.

Recall that a linear operator A acting on a complex Hermitian inner-product space[21] V is called **unitary** if $< Av, Aw >=< v, w >$ for every $v, w \in V$ (namely A is a rotation). The set $U(V)$ of all unitary operators forms a group under operator composition.

Definition 11.8. A **unitary representation** of a group H is a pair (V, ρ) where V is a complex linear space with a Hermitian inner product and ρ is a group homomorphism $\rho : H \to \mathrm{U}(V)$. We often omit V when it is obvious or unimportant and denote the representation by ρ alone. The **dimension** of ρ is $\dim(V)$.

Thus, representations are the natural extension of characters, which are simply 1-dimensional representations. To be more concrete, a d-dimensional representation of a group H is an assignment of a $d \times d$ unitary matrix B_h to every $h \in H$, so that $B_g B_h = B_{gh}$ for every pair of elements $g, h \in H$.

We note that representation theory deals as well with nonunitary representations, but for finite groups, which are our main concern here, there is no loss of generality in restricting ourselves to unitary representations, as we do.

Now we define the main representation of a group: **regular representation**. We will later see that it "captures" all representations of the group in a well defined way.

Definition 11.9. Let H be a finite group and let V be the $|H|$-dimensional vector space of all complex functions $f : H \to \mathbb{C}$, with the standard Hermitian inner product. The **regular representation** $r = r_H$ of H is defined by $[r(h)f](x) = f(xh)$ for every $h \in H$.

We now introduce "irreducible" representations which can be considered as building blocks from which all representations can be constructed. To this end we first introduce the notion of invariant subspaces of a given representation, which are preserved under the action of all matrices in it. Our first step in this direction is to define invariant vectors. These will feature later on as well in our analysis of expansion in Cayley and Schreier graphs.

Definition 11.10. An **invariant vector** of a representation (V, ρ) is a nonzero vector $v \in V$ such that $\rho(h)v = v$ for all $h \in H$. For example, in the regular representation the invariant vectors are the constant functions on the group H.

An **invariant subspace** is a linear subspace $W \subset V$ such that $\rho(h)W = W$ for all $h \in H$. (This means that $\rho(h)(w) \in W$ for every $w \in W$, but not necessarily that $\rho(h)(w) = w$.)

For example, in the regular representation, constant functions on H are invariant, and there are no other invariant vectors. The orthogonal complement of the space of

[21] For our purposes V can simply be taken to be $\mathbb{C}^d$ with the usual Hermitian inner product of complex vectors, $< x, y >= \sum_{i=1}^{d} x_i \overline{y_i}$.

constant functions, the functions whose coordinate sum is zero, forms an invariant subspace.

Definition 11.11. A representation is called **irreducible** if it has no nontrivial invariant subspaces.

Representations which are identical up to a change of basis in the linear space are considered identical for the purpose of representation theory.

Definition 11.12. Two representations $(V,\rho),(W,\eta)$ are called **isomorphic** if there is an isomorphism of vector spaces $f : V \to W$ such that $f(\rho(g)v) = \eta(f(v))$. Equivalently, there is a common change of basis which moves the matrices in ρ to the corresponding matrices in η.

Therefore, if (V,ρ) is not irreducible and has a nontrivial invariant subspace V', we can carry out a change of basis to make the action of ρ on V' and on its orthogonal complement V'' take place in disjoint sets of coordinates. This yields a representation isomorphic to (V,ρ) which now decomposes into two representations of smaller dimension, (V',ρ') and (V'',ρ''). This process can be iterated on the resulting representations until no more invariant subspaces are found. Thus every representation can be decomposed to irreducible components. We will soon see that the outcome of this process is unique.

The following lemma states that every representation decomposes (uniquely) into irreducible ones.

Proposition 11.13. *Let (V,ρ) be a unitary representation of a group H. Then there is an orthogonal decomposition of V into a direct sum of invariant subspaces $V = \oplus V_i$ with $\rho(h)V_i = V_i$ for all i and $h \in H$, and (V_i,ρ) are irreducible representations. Moreover, this decomposition is unique up to isomorphism.*

We next see that the **regular representation** contains all possible irreducible representations in its decomposition.

Proposition 11.14. *Let (V,r) be the regular representation of a group H, and let (U,ρ) be any irreducible representation of H. Then (U,ρ) appears in the decomposition of (V,r) into irreducible representations.*

We finally get to use this apparatus to analyze Cayley graphs. We proceed in a way analogous to the method used in Proposition 11.7. Fix a generating set S for H and a representation (V,ρ) of H, and consider the matrix

$$A_\rho = (1/|S|) \cdot \sum_{s \in S} \rho(s).$$

The following lemma is simple but useful. It shows how the matrices above, when we range over all **irreducible** representations, capture **all** eigenvalues of the Cayley graph with generators S.

Lemma 11.15. *Let H be a finite group and let S be a symmetric subset of H. Then:*

- *The normalized adjacency matrix of $C(H,S)$ is A_r for r the regular representation of H.*
- *Every eigenvalue of A_r is an eigenvalue of A_ρ for some irreducible representation ρ.*

- *The converse also holds: If ρ is a representation of H, then every eigenvalue of A_ρ is also an eigenvalue of A_r.*

In principle, at least, this is a recipe by which we can calculate all the eigenvalues of $\mathrm{C}(H,S)$. In practice, however, it is usually quite hard to analyze all the matrices A_ρ. It is more manageable, though, for generating sets S of special structure. Also, as mentioned above, it is often the case that results from various areas of mathematics can be used to this end.

11.1.2. *Schreier graphs.* Much of what we have just done applies in a context more general than that of Cayley graphs of groups, namely for graphs corresponding to the action of a group on a set. Let H be a group and X be some set. An **action** of H on X is a group homomorphism $\pi : H \to \mathrm{Sym}(X)$ that sends each element h to a permutation of the elements of X. For any subset S of H we define the **Schreier graph** $\mathrm{Sch}(H,X,S)$ whose vertex set is X and whose edges are $(x, \pi(s)x)$ for every $s \in S$ and every $x \in X$. Here are a few examples.

- Schreier graphs are indeed more general than Cayley graphs, since any group H acts on itself by sending an element h to the permutation $g \to gh$. In this case $\mathrm{Sch}(H,X,S)$ is simply the Cayley graph $\mathrm{C}(H,S)$.
- The symmetric group S_n acts on the set $[n]$ by $\pi(\sigma)(i) = \sigma(i)$.
- The group $\mathrm{GL}_n(p)$ of invertible matrices over the field $\mathbb{F}_p$ acts on the set $\mathbb{F}_p^n$ of n-dimensional vectors by $\pi(M)v = M \cdot v$.

While these examples still look special and "algebraic", the following basic fact shows that essentially every regular graph is a Schreier graph of some group.

Theorem 11.16 (Gross [Gro77])**.** *Every finite regular graph of even degree is a Schreier graph corresponding to some finite group acting on some finite set.*

The idea of the proof is simple: if the degree is $2d$, then the edges can be partitioned to d cycle covers. To each we assign a permutation (and its inverse) on the vertex set. The group generated by all these permutations naturally acts on the graph and provides a natural labeling of the edges by generators.

We now explain how the eigenvalues of the adjacency matrix of a Schreier graph are obtained from those of the naturally associated Cayley graph of the acting group.

An action π of H on a set X naturally defines a **permutation representation** (V,ρ) of H. Here V is the vector space of all complex-valued functions on X, and $(\rho(h)f)(x) = f(\pi(h)x)$. The normalized adjacency matrix of $\mathrm{Sch}(H,X,S)$ is equal to $A_\rho = 1/|S| \cdot \sum_{s\in S} \rho(s)$. By Proposition 11.15 every eigenvalue of A_ρ is an eigenvalue of A_r, where r is the regular representation. Consequently, every eigenvalue of the Schreier graph is an eigenvalue of $\mathrm{C}(H,S)$. In particular

Proposition 11.17. *Let H be a finite group acting on the set X. Let S be a subset of H and let Z be a connected component of $Sch(H,S,X)$. Then $\lambda(Z) \le \lambda(H,S)$.*

The proposition implies that if a Cayley graph of H is an expander, then so are all the corresponding connected Schreier graphs. We apply this to show that two families of graphs we previously met in subsection 2.2 are expanding.

- Let $\mathrm{SL}_2(m)$ be the group of 2×2 matrices with determinant 1 over the ring of integers modulo m. It can be proved (see [Lub94]) that the set
$$S = \left(\begin{array}{cc} 1 & \pm 1 \\ 0 & 1 \end{array}\right), \left(\begin{array}{cc} 1 & 0 \\ \pm 1 & 1 \end{array}\right)$$
generates a Cayley graph with second eigenvalue at most $1 - 1/2000$. The group $\mathrm{SL}_2(m)$ acts on the set $(\mathbb{Z}_m)^2$ (multiply the matrix by the vector). Consider the corresponding Schreier graph $\mathrm{Sch}(\mathrm{SL}_2(m), (\mathbb{Z}_m)^2, S)$. Here a vertex (x, y) is adjacent to $(x, y \pm x)$ and $(x \pm y, y)$. This is a 4-regular subgraph of the 8-regular graph defined in subsection 2.2 and a close relative of the graphs discussed in Section 8. By Proposition 11.17, every connected component of this Schreier graph has second eigenvalue at most $1 - 1/2000$, and it is not hard to see that the connected components are the vector $(0, 0) \in (\mathbb{Z}_m)^2$ and its complement.
- Let p be a prime. The **projective line** $\mathbf{P}^1(\mathbb{Z}_p)$ is the punctured plane $(\mathbb{Z}_p)^2 \setminus \{(0, 0)\}$ modulo the equivalence relation $(x, y) \equiv (ax, ay)$ for all $0 \neq a \in \mathbb{Z}_p$. The set $\mathbf{P}^1(\mathbb{Z}_p)$ consists of $p + 1$ points, which are the equivalence classes of $(x, 1)$ for all $x \in \mathbb{Z}_p$, plus the equivalence class of $(1, 0)$.

  The group $\mathrm{SL}_2(p)$ acts on $\mathbf{P}^1(\mathbb{Z}_p)$, since it acts on $(\mathbb{Z}_p)^2 \setminus \{(0, 0)\}$ and respects the equivalence relation. Identify the equivalence class $(x, 1)$ of the projective line with the element $x \in \mathbb{Z}_p$ and define the equivalence class $(1, 0)$ to be "the point at infinity". Then the action of $\mathrm{SL}_2(p)$ on the projective line can be defined as follows (we extend the arithmetic operations to include infinity in the obvious way):
$$\left(\begin{array}{cc} a & b \\ c & d \end{array}\right) x = \frac{ax + b}{cx + d}.$$
  We say that $\mathrm{SL}_2(p)$ acts on the projective line by **fractional linear transformations** (also known as **Möbius transformations**). Now consider the following generating set S of $\mathrm{SL}_2(p)$:
$$S = \left(\begin{array}{cc} 1 & \pm 1 \\ 0 & 1 \end{array}\right), \left(\begin{array}{cc} 0 & 1 \\ -1 & 0 \end{array}\right).$$
  In the corresponding Cayley graph on the projective line $\mathbf{P}^1(\mathbb{Z}_p)$, the neighbors of x are the points $(-1/x)$ and $(x \pm 1)$. It can be shown that $\lambda(\mathrm{SL}_2(p), S) < 1 - 1/10^4$, so the same holds for the connected graph $\mathrm{Sch}(\mathrm{SL}_2(p), \mathbf{P}^1(\mathbb{Z}_p), S)$. If one removes the point at infinity and adds a self-loop to the point 0 (to compensate for the missing edge), one obtains the 3-regular expander graph described in subsection 2.2.

11.1.3. *Kazhdan constant and expansion of Cayley graphs.* Let (V, ρ) be an irreducible unitary representation of the finite group H. Consider some $v \in V$ of unit norm and let us investigate the distance from v to its images $\{\rho(h)v | h \in H\}$. Since $\rho(h)$ is always a unitary matrix, all the vectors $\rho(h)v$ have unit norm, and in particular $||\rho(h)v - v|| \leq 2$. On the other hand, it is easy to verify that the vector $\sum_{h \in H} \rho(h)v$ is invariant, so by irreducibility it must be zero. But the term corresponding to $h = \mathrm{id}$, the identity in H, gives $\rho(h)v = v$, so there must be some $s \in H$ for which $\rho(s)v$ has negative inner product with v, and thus $||\rho(s)v - v|| > \sqrt{2}$. The **Kazhdan constant** of a subset $S \subseteq H$ quantifies the extent to which a similar conclusion still holds when s must be selected only from S.

This discussion illustrates the significance of smallest displacement of any vector by any nontrivial representation acting on it. In this section we point out that the spectral gap of a Cayley graph as well as the Kazhdan constant are closely related to this minimum displacement and are therefore closely related to each other. Kazhdan [Kaz67] originally introduced this constant as well as the closely related Kazhdan's property (T) for **infinite** groups. These concepts have played a key role in the solution of several important open problems. It turns out that these concepts are very useful for finite groups as well. Although the Kazhdan constant and the spectral gap in Cayley graphs are closely related, there are different situations where one or the other is easier to work with. This dual perspective was a key feature in Kassabov's recent breakthrough result [Kas05b] that the symmetric groups can be made into a family of expanding graphs.

We have already often used the variational definition of eigenvalues to estimate the spectral gap of graphs. In the case of a Cayley graph $\mathrm{C}(H,S)$, recalling that A_r is its normalized adjacency matrix, this can be stated as follows:

$$\lambda_2(H,S) = \max_{v\perp 1} \frac{v^T A_r v}{||v||^2} = \max_{v\perp 1} \frac{1}{|S|} \sum_{h\in S} \frac{v^T r(h) v}{||v||^2}$$

where r is the regular representation. We use the expression for the normalized adjacency matrix of $\mathrm{C}(H,S)$ from Lemma 11.15. Since r is unitary, it is easily verified that

$$\frac{||r(h)v - v||^2}{||v||^2} = 2\left(1 - \frac{v^T r(h) v}{||v||^2}\right).$$

Consequently,

$$g(H,S) = 1 - \lambda_2(H,S) = \min_{v\perp 1} \frac{1}{2|S|} \sum_{h\in S} \frac{||r(h)v - v||^2}{||v||^2}.$$

If the degree is a constant and we do not seek the optimal expansion, the largest term in this sum gives a constant approximation. This leads us to the following definition.

Definition 11.18. Let S be a subset of a group H. The **Kazhdan constant** of S is defined by

$$\mathrm{K}(H,S) = \min_{v\perp 1} \max_{h\in S} \frac{||r(h)v - v||^2}{||v||^2}.$$

We already know by Lemma 11.13 that the regular representation decomposes into an orthogonal sum of irreducible representations and that all the irreducible representations of H appear in this decomposition. The vector with minimal displacement will therefore appear in some irreducible representation, so the Kazhdan constant may also be defined as

$$\mathrm{K}(H,S) = \min_{\rho} \min_{v\perp 1} \max_{h\in S} \frac{||\rho(h)v - v||^2}{||v||^2}$$

where (ρ, V) ranges over all nontrivial irreducible representations of H.

The spectral gap of $\mathrm{C}(H,S)$ and the Kazhdan constant are therefore related as follows:

$$\mathrm{K}(H,S)/(2|S|) < \mathrm{g}(H,S) < \mathrm{K}(H,S)/2.$$

Thus, as mentioned, when we consider generator sets of bounded cardinality and if we do not strive for optimal parameters, the spectral gap and the Kazhdan constant can be used interchangeably. The Kazhdan constant is sometimes more convenient to handle than the spectral gap. For example, enlarging the generating set S may decrease $\mathrm{g}(H,S)$, while $\mathrm{K}(H,S)$ can only increase. Here is a useful observation in the same spirit.

Claim 11.19. *Let H be a group, and let $S,\tilde{S}$ be subsets of H such that each element of $\tilde{S}$ can be written as a product of at most m elements of S. Then $K(H,S) \geq K(H,\tilde{S})/m$.*

Note that this property is not shared by the spectral gap. If the set $\tilde{S}$ is obtained from S by adding many copies of the identity element, then the assumption of the claim holds with $m=1$, but $g(H,S)$ may be much smaller than $\mathrm{g}(H,\tilde{S})$.

11.2. **The replacement product and semidirect product.** In Section 9 we defined the zig-zag product of graphs and used it to construct expander families. In this section we construct families of expander **Cayley** graphs using the zig-zag product. In general, a zig-zag product of two Cayley graphs is not a Cayley graph. Nevertheless, under certain conditions on the underlying groups and on the generating sets, the zig-zag product of two Cayley graphs $\mathrm{C}(A,S_A) ⓩ \mathrm{C}(B,S_B)$ is indeed a Cayley graph $\mathrm{C}(C,S_C)$ where C is the **semidirect** product of A and B. This connection and its utility in establishing expansion were first described in [ALW01]. We first present the connection in this section and then the applications of this and other papers in the subsequent sections.

We start by defining semidirect product and then describe its relation to the zig-zag product. The notion of a group acting on a set was already mentioned above. When the set has a group structure, this is specialized as follows:

Definition 11.20. An **action** of a group B on a group A is a group homomorphism $\phi : B \to Aut(A)$. In other words, each element $b \in B$ corresponds to an automorphism ϕ_b of A, where $\phi_{b_1 \cdot b_2} = \phi_{b_1}\phi_{b_2}$.

For example, the cyclic group $\mathbb{Z}_d$ acts on the group $\mathbb{Z}_2^d$, the additive group of the d-dimensional vector space over $\mathbb{Z}_2$, by cyclically shifting the coordinates of the d-dimensional vectors.

When a group B acts on A we can define a group structure on the set $A \times B$.

Definition 11.21. Suppose a group B acts on a group A. The **semidirect product** $A \rtimes B$ is a group whose elements are pairs (a,b) where $a \in A$ and $b \in B$. We define

$$(a_1,b_1)\cdot(a_2,b_2) = (a_1 \cdot \phi_{b_1}(a_2), b_1 \cdot b_2).$$

The simplest example of the semidirect product is the special case of the direct product $A \times B$. In this case ϕ_b is the identity automorphism of A for all $b \in B$, and so $(a_1,b_1)\cdot(a_2,b_2) = (a_1 \cdot a_2, b_1 \cdot b_2)$. Let us move to a more interesting example. As we saw the group $\mathbb{Z}_d$ acts on $\mathbb{Z}_2^d$. The element set of the resulting semidirect product is $(\bar{v},x) \in \mathbb{Z}_2^d \times \mathbb{Z}_d$. The multiplication rule is

$$((v_i)_{i=0}^{d-1}, x)\cdot((w_i)_{i=0}^{d-1}, y) = ((v_i + w_{i+x})_{i=0}^{d-1}, x+y).$$

Notice that this group is not commutative although the constituent groups are.

We now return to Cayley graphs. Continuing with this example, pick $\{+1,-1\}$ as a generating set of $\mathbb{Z}_d$, and $e_0,\ldots,e_{d-1}$ for $\mathbb{Z}_2^d$. The corresponding Cayley graphs are the d-cycle and the d-dimensional binary cube respectively. Since the degree of the d-dimensional cube is equal to the number of vertices in the d-cycle, we can form a zig-zag product of the two. Let us consider the simpler **replacement product**, defined in subsection 9.3. In this product we replace every vertex of $\mathbb{Z}_2^d$ by a cloud of d vertices representing the set $\mathbb{Z}_d$. On each cloud we preserve the edges of the original d-cycle. We also connect each vertex in the cloud to one of the d neighbors of the cloud in the binary cube. In our discussion in Section 9 it did not matter much how the clouds are connected to each other. Now we choose to connect the vertex (v, h) to $(v + e_h, h)$; the usefulness of this specific choice will become clear later. Like the zig-zag construction, the replacement product is an expander if the original two graphs are expanders. The resulting graph for $d = 3$ is depicted in Figure 14.

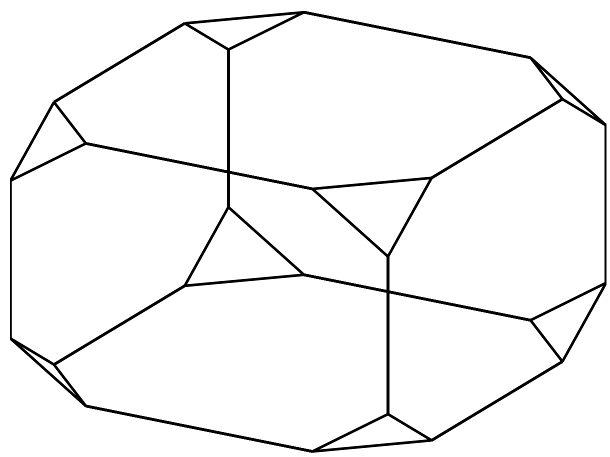


FIGURE 14. A Cayley graph of $(\mathbb{Z}_2)^3 \rtimes Z_3$.

We have described two operations above: The semidirect product takes groups A, B and puts a group structure on the set $A \times B$. The replacement product (or zig-zag product) takes graphs on vertex sets A, B and constructs a graph on vertex set $A \times B$. The two constructions are closely related. Under certain assumptions, the replacement product of the Cayley graphs of the groups A, B is also a Cayley graph of the semidirect product of $A \rtimes B$. In our running example, consider the replacement product of the d-cycle and the d-dimensional binary cube we defined above. It is a Cayley graph of the semidirect product $\mathbb{Z}_2^d \rtimes \mathbb{Z}_d$, with the three generators $\{(\bar{0}, \pm 1) \cup (e_0, 0)\}$.

We now turn to the general statement relating replacement products and semidirect products. Suppose a group B acts on a group A. The **orbit** of an element $a \in A$ under the action of B is the set $\{\phi_b(a) | b \in B\}$. For example, the orbit of $v \in \mathbb{Z}_2^d$ under the action of $\mathbb{Z}_d$ is the set of all cyclic shifts of v. Suppose that some (possibly very large) generating set S_A of A is such an orbit of some element $x \in A$. We prove in the following lemma that we can find a generating set for $A \rtimes B$ consisting of some generating set S_B for B (embedded in $A \rtimes B$), plus the single element x, instead of the whole set S_A. Furthermore, the resulting Cayley graph is the replacement product of $\mathrm{C}(A, S_A)$ and $\mathrm{C}(B, S_B)$. The replacement product is an expander if the constituent graphs are expanders, and as long as S_A is a single

B orbit, the degree of the replacement product is dominated by the size of S_B, even though S_A may be as large as B. We state the conclusion formally in the following theorem.

Theorem 11.22 (Alon-Lubotzky-Wigderson [ALW01]). *Let A, B be two groups with generating sets S_A, S_B such that $|B| = |S_A|$. Furthermore, suppose that B acts on A in such a way that S_A is the orbit of one of its elements $x \in S_A$ under this action. Then $S := \{(1,s)|s \in S_B\} \cup \{(x,1)\}$ generates $A \rtimes B$, and $C(A \rtimes B, S)$ is a replacement product of $C(A, S_A)$ and $C(B, S_B)$. More generally, if S_A is a union (taken with multiplicities) $S_A^1 \cup \ldots \cup S_A^k$ of k orbits under the action of B, and if $x_1, \ldots, x_k$ are representatives of these orbits, then $S := \{(1,s)|s \in S_B\} \cup \{(x_1,1), \ldots, (x_k,1)\}$ is a generating set for $A \rtimes B$, and $C(A \rtimes B, S)$ is the union of the k replacement products of $C(A, S_A^i)$ and $C(B, S_B)$ when $1 \leq i \leq k$.*

Proof. We will prove the single-orbit case and leave the general case to the reader. Suppose then that S_A is a single orbit under the action of B, with representative x. To see that S generates $A \rtimes B$, first note that each element $(a,b) \in A \rtimes B$ can be written as the product $(a,1) \cdot (1,b)$ and can therefore be written as a product of elements from $(1, S_B) \cup (S_A, 1)$. By assumption, any $s_a \in S_A$ is equal to $\phi_{s_b}(x)$ for some $s_b \in S_B$, and therefore $(1, s_b) \cdot (x, 1) \cdot (1, s_b^{-1}) = (\phi_b(x), 1) = (s_a, 1)$. This implies that indeed S generates the group $A \rtimes B$.

Now consider the graph $\mathrm{C}(A \rtimes B, S)$. This graph consists of clouds of the elements of B, where each cloud is interconnected by the edges of $\mathrm{C}(B, S_B)$, since $(a,b) \cdot (1, s_b) = (a, b \cdot s_b)$. Between the clouds we have edges of the form $(a,b) \cdot (x,1) = (a \cdot \phi_b(x), b)$. So the cloud of the element a is connected by one edge to each of the clouds corresponding to the neighbors of a in the graph $\mathrm{C}(A, S_A)$, as required by the replacement product definition. $\square$

Remark 11.23. Under the assumptions of the lemma, we can describe also the zig-zag product (instead of the replacement product) of the same two Cayley graphs as a Cayley graph on $A \rtimes B$ with the generating set $\{(1, s_1) \cdot (x, 1) \cdot (1, s_2) \ : \ s_1, s_2 \in S_B\}$.

11.3. **Constructing expander families by iterated semidirect products.** As we saw, it is possible to construct a family of constant degree expanders using the zig-zag product. We have just seen that the zig-zag product of two Cayley graphs is sometimes a Cayley graph. Can these two ideas be combined to construct a family of Cayley expander graphs? This is indeed possible, and we will present two constructions. One, by Meshulam and Wigderson [MW02], constructs a family of expanders with nonconstant but very slowly growing degree. The other construction, by Rozenman, Shalev, and Wigderson [RSW04], yields a family of constant-degree expanders.

11.3.1. *Cayley expanders from group rings.* [MW02] For a finite group H let $F_p[H]$ be the **group ring** over the finite field F_p, namely the set of all formal sums $\{\sum_{h \in H} \alpha_h h \ : \ \alpha_h \in F_p\}$, with the obvious addition and multiplication. We think of $F_p[H]$ as an additive group, so the group H naturally acts on $F_p[H]$. Therefore, we can iteratively construct a sequence of groups by $H_{i+1} = F_{p_i}[H_i] \rtimes H_i$. We later argue that one can find a constant number of orbits under the action of H_i that make $F_{p_i}[H_i]$ an expander. Using Theorem 11.22 we obtain a family of expanders.

Theorem 11.24 (Meshulam-Wigderson [MW02]). *There exists a group H_1, a sequence of primes p_i, and a sequence of generating sets U_i for H_i such that*

$\lambda(H_n, U_n) \le 1/2$ and $|U_n| \le \log^{(n/2)} |H_n|$, where log^r is the r times iterated logarithm function.

This construction is nearly optimal, since the groups H_n are solvable with huge Abelian subgroups. More precisely, H_n is a solvable group with solvability index at most n (as H_{n-1} is a normal subgroup with Abelian quotient $F_{p_{n-1}}[H_{n-1}]$). In this case it is known that any generating set achieving $\lambda \le 1/2$ must have cardinality at least $\log^{(n)} |H_n|$ [LW93].

To achieve the above construction, Meshulam and Wigderson prove a sufficient condition to make $F_p[H]$ an expander with a generating set consisting of a constant number of H orbits. Interestingly, this condition relies on the distribution of dimensions of the irreducible representations of H (we note that similar but simpler statements appear already in section 3 of [ALW01]).

Theorem 11.25. *Let d_r be the number of irreducible representations of H of dimension smaller than r. If $d_r < c^r$ for some constant c and for all r, then there is a constant number of orbits of $F_p[H]$ that make it an expander.*

One family of groups that have this property are the **monomial groups.**[22] Such groups have the nice property that if H_i is monomial, then H_{i+1} is monomial as well, and one can explicitly find generating orbits for $F_p[H_i]$. This gives a sequence of explicit expanding Cayley graphs X_i on the groups H_i, where the degree of X_n is exponential in n and the size of X_{n+1} is a tower function in n. The degree of X_n is therefore indeed an iterated logarithm in the size of X_n, yielding a family of expanders with degrees that grow extremely slowly.

11.3.2. *Cayley expanders from iterated wreath products.* For every group H, the symmetric group S_d acts naturally on the group H^d, the direct product of d copies of H, by permuting the coordinates. The resulting semidirect product $H^d \rtimes S_d$ is called the **wreath product** of H and S_d. The wreath product may also be defined for any subgroup of S_d, for example the alternating group A_d. Take $H_1 = A_d$, and define $H_{i+1} = H_i^d \rtimes A_d$. Rozenman, Shalev and Wigderson [RSW04] prove that this family of groups can be made into an expander family. The main lemma in the proof is a construction of a single A_d orbit that generates H^d as an expander (under a certain condition on H).

Here is the idea of this main lemma. Let U be some expanding generating set for H. It is not hard to check that U^d, the set of all U-valued vectors of length d, is an expanding generating set for H^d (with the same second eigenvalue). However, the set U^d is far from being a single A_d orbit. Consider the subset $U^{(d)} \subset U^d$, consisting of U-valued vectors of length d in which every element of U appears the same number of times (assume that $|U|$ divides d). This subset is indeed a single S_d orbit, but is it still expanding? The next theorem answers this question positively, if $|U|$ is small relative to d and if every element of H can be written as a commutator. We note that this property carries itself inductively if $H_1 = A_d$ has it, which indeed it does.

Theorem 11.26 (Rozenman-Shalev-Wigderson [RSW04])**.** *Let H be a group in which every element h can be written as a commutator $h = aba^{-1}b^{-1}$. Let U be an*

[22] I.e., all their irreducible representations are induced from one-dimensional representations of subgroups of H.

expanding generating set for H with second eigenvalue at most $1/4$, and suppose $|U| < d/10^6$. Then $\lambda(H^d, U^{(d)}) < 1/2$.

Theorem 11.26 carries out part of the induction step in the construction of the expander family. We want to show that when H_i has a small enough expanding generating set, the same holds for $H_{i+1} = H_i^d \rtimes A_d$. We found a generating set for H^d which is a single A_d orbit. We now need to find an expanding generating set for A_d. Since the size of the generating set of H_{i+1} is dominated by the size of the generating set of A_d, we need A_d to have a small expanding generating set. This is also required for the first group H_1 of the construction (which is also A_d). We can now use:

Theorem 11.27 (Kassabov [Kas05b]). *The alternating group A_d has an explicit generating set U of size independent of d such that $\lambda(A_d, U) < 1/100$.*

These two theorems imply that the construction above works for large enough d.

Corollary 11.28 ([RSW04]). *Consider the sequence of groups defined by $H_1 = A_d$ and $H_{i+1} = H_i \rtimes A_d$. If d is large enough, then there exists a sequence of generating sets U_i such that $\lambda(H_i, U_i) \leq 1/2$ and $|U_i|$ is independent of i.*

11.4. **Expansion is not a group property.** Consider the following question raised by Lubotzky and Weiss [LW93]:

Question 11.29. Let H_i be a family of groups with generating sets U_i, of bounded size independent of i, so that $C(H_i, U_i)$ is a family of expanders. Is it true that for every other bounded size generating sets V_i, the Cayley graphs $C(H_i, V_i)$ form an expanding family as well?

In simple words this asks whether expansion is a property of the groups H_i that is independent of the choice of a specific generating set. The answer is negative, as was first demonstrated via the connection of semidirect product to the zig-zag product of [ALW01] discussed above. We explain this example and then the far simpler example which follows from Kassabov's recent work [Kas05b].

First counterexample. Consider the vector space $A_p = F_2^{p+1}$ for some prime p. This is an Abelian group and of course has no bounded-size expanding generating set. We saw in subsection 11.1.2 that $\mathrm{SL}_2(p)$ acts on the projective line $\mathbf{P}^1(\mathbb{Z}_p)(F_p)$ which has $p+1$ points. Therefore, $\mathrm{SL}_2(p)$ acts on F_2^{p+1} by permuting the coordinates of the vectors. Here we exhibit two generating sets for $F_2^{p+1} \rtimes \mathrm{SL}_2(p)$ yielding a negative answer to the above question. It can be shown that (i) the standard basis of F_2^{p+1} is a single $\mathrm{SL}_2(p)$ orbit and (ii) the union of two **random** orbits is an expanding generating set for F_2^{p+1}. As we mentioned in subsection 11.1.2, there exists a set S of four matrices in $\mathrm{SL}_2(p)$ which are expanding.

By Theorem 11.22, combining our two generating sets on F_2^{p+1} with the generating set of $\mathrm{SL}_2(p)$ creates two Cayley graphs on $F_2^{p+1} \rtimes \mathrm{SL}_2(p)$. The first Cayley graph comes from a set of six generators: the four generators of $\mathrm{SL}_2(p)$ and representatives of two random orbits. By Theorem 11.22 this graph is a replacement product of two expander graphs, which is therefore an expander. The second Cayley graph comes from a set of five generators for $F_2^{p+1} \rtimes \mathrm{SL}_2(p)$: the four generators of $\mathrm{SL}_2(p)$ and a representative of the orbit of standard basis vectors. This Cayley graph is a replacement product, where one of the constituent graphs is the

Cayley graph of F_2^{p+1} with the standard basis vectors. This graph is simply the discrete $p+1$-dimensional cube, which has second eigenvalue at $1-1/(p+1)$. A replacement product can never have a smaller second eigenvalue than any of the constituent graphs, which proves the nonexpansion result.

Second counterexample. By Theorem 11.27 the alternating group A_d has a bounded generating set. Consequently, for every d, the symmetric group S_d has a bounded generating set U_d such that $C(S_d, U_d)$ are an expanding family. On the other hand, S_d is generated by the permutation (12), the cycle $(12\ldots d)$ and its inverse. It is not hard to check that the resulting 3-regular Cayley graphs are not expanders. (It is not hard to show that second eigenvalue which tends to zero when d grows.)

11.5. **Hypercontractive inequalities in groups?** Isoperimetric problems on Cayley graphs suggest numerous interesting questions that make sense only in this more specialized domain. Consider the Cayley graph $\mathrm{C}(H, S)$ of a group H with a (symmetric) generating set S. When we consider an edge cut $E(T, V\setminus T)$, we can do more than just count the number of edges in this set. Rather, for every $s\in S$ we can consider the number $\epsilon_{s,T}$ of edges in $E(T, V\setminus T)$ that correspond to the generator s or s^{-1}. The usual isoperimetric inequality concerns only $|E(T, V\setminus T)| = \sum_{s\in S}\epsilon_{s,T}$. There are, however, several other natural quantities to consider, the most natural of which is

$$\max_{s\in S}\epsilon_{s,T}.$$

A theorem of Kahn Kalai and Linial [KKL88] reveals an interesting phenomenon in the graph of the discrete cube (which for the present discussion is the Cayley graph of the group F_2^n with the standard set of generators). For simplicity we quote only a special case of this theorem.

Theorem 11.30 (Kahn-Kalai-Linial [KKL88]). *Let $T\subseteq\{0,1\}^n$ be a set of cardinality $|T| = 2^{n-1}$. Then there is an index $1\le j\le n$ such that*

$$\epsilon_{j,T}\ge\Omega(\frac{\log n}{n}\cdot 2^n).$$

The bound is tight.

Note that the isoperimetric inequality on the cube (see subsection 4.2.1) says that $\sum_j \epsilon_{j,T}\ge 2^{n-1}$, and the crux of the matter is the additional logarithmic factor in the theorem. The proof of this theorem is based on a hypercontractive inequality due to Bonami [Bon70] and Beckner [Bec75]. We should also note that Theorem 11.30 has found numerous applications in different fields. This makes it very interesting to seek similar phenomena in other groups.

12. **Error correcting codes**

Connections between error correcting codes, a central area in communication and engineering, and expander graphs have seen enormous growth in the past two decades, even though their roots go back to the 1960's. Here we describe in detail only one connection: the application of explicit lossless expanders of Section 10 to the construction of the simplest known efficient asymptotically good codes. We also

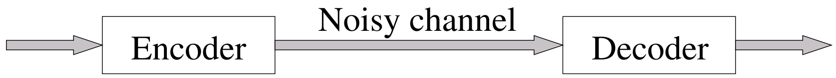


FIGURE 15. Transmitting a message over a noisy channel.

briefly review other variants and connections and give some references. But first we give some background to this important field.

Consider the problem of sending information through a noisy channel. Some of the bits we transmit may be flipped due to noise in the channel. One way to overcome channel errors is to use error correcting codes. Let us restrict our attention to schemes that transmit the information in n-bit blocks. To make sure the receiver can recover from bit-flip errors, we agree in advance that only a subset of the 2^n possible n-bit strings may be transmitted. We call such possible strings *codewords* and would like our set of codewords to have two conflicting properties. The first is that many bit flips are needed to transform one codeword into another, and the second is that there are many codewords. These properties guarantee (respectively) the robustness of our transmission scheme to channel errors and the efficiency of the channel utilization. One further property that we crucially need is algorithmic in nature. We would like an efficient implementation of the encoder and the decoder in such a scheme.

To proceed, we need a better formulation of the concepts involved. We have to be brief here, and the interested reader can refer to one of the standard texts in this area, e.g., [MS77a, MS77b] and [vL99]. Sudan's notes [Sud00] cover some of the more recent developments and the algorithmic aspects of the field.

12.1. **Definition of error correcting codes.** We should mention that our discussion is restricted only to binary codes and to adversarial errors. We do not consider codes over other alphabets and do not even mention any other interesting error models that appear in the literature.

A code is a set of n-bit binary strings $C \subseteq \{0,1\}^n$. The **distance** of C, denoted $\mathrm{dist}(C)$, is the minimum Hamming distance between a pair of distinct codewords $x, y \in C$. That is $\min_{x \neq y \in C} d_H(x,y)$, where the Hamming distance $d_H(x,y)$ is the number of coordinates where x and y differ. The **rate** of C is $\mathrm{rate}(C) = \log|C|/n$.

The most natural way to decode received messages which are possibly noisy is to decide for each received message y that the message sent was the codeword $x \in C$ that is closest (in Hamming distance) to y. It is easily verified that this method is guaranteed to be correct as long as the number of bit-flips is bounded by $\lfloor(\mathrm{dist}(C)-1)/2\rfloor$. Also, since we transmit $\log|C|$ bits of information for every n channel bits, the efficiency of the code is $\log|C|/n$, which is the code rate.

As mentioned above, our goal in designing a code is to maximize the rate as well as the distance. Practical considerations make it also necessary that the code be efficiently encodable and decodable. Ideally, both tasks should take only linear time. The following definition formulates these basic requirements from a code.

Definition 12.1. A family $C_n \subset \{0,1\}^n$ of codes is **asymptotically good** if there are some fixed constants $r > 0$ and $\delta > 0$ such that for all n both $\mathrm{dist}(C) > \delta n$ and $\mathrm{rate}(C) > r$. The family is called **efficient** if encoding and decoding (with $\leq \delta n/2$ errors) can be performed in polynomial time in n.

The existence of asymptotically good codes can be easily proved by a probabilistic argument. The quest for good *efficient* codes took over two decades. Here we will see a simple such construction, based on expander graphs. But first we need to introduce linear codes.

12.2. **Linear codes.** It is often convenient to use **linear codes**. That is, C is to be a linear subspace of F_2^n. Such codes can be described concisely by specifying a basis and therefore can be efficiently encoded in $O(n^2)$ time. On the other hand, the decoding problem (that is, finding the closest codeword to a given word in a given code) is already **NP**-hard. However, as we shall see, efficient algorithms exist for words that are "close enough" to the code.

It is a simple but useful fact that the distance of a linear code equals the smallest weight (number of ones) of a nonzero codeword, since the Hamming distance between two vectors is the weight of their bit-wise sum.

12.3. **Asymptotic bounds.** Here are the basic upper and lower bounds on the trade-off between the distance of a code and its size (and hence, also its rate). Denote by $v(n, r)$ the volume of the radius r ball in the Hamming cube, namely $v(n, r) = \sum_{i=0}^{r} \binom{n}{i}$.

12.3.1. *Lower bounds on size: The Gilbert-Varshamov bound.*

Theorem 12.2. *There exists a length n code with distance $\geq d$ and size $\geq 2^n/v(n, d)$. Moreover, this statement holds even for linear codes.*

Proof. This follows from the following (exponential time) greedy algorithm that constructs a distance d code. We initialize $S = \{0,1\}^n, C = \emptyset$. At each step we pick any point $x \in S$ and add x to C. We then remove from S all the points that are within distance $\leq d$ from x. The bound follows since the initial size of S is 2^n, and at each iteration the size of S is reduced by at most $v(n, r)$.

As mentioned above, a linear code can be specified by giving a basis for the subspace $C \subseteq F_2^n$. Alternatively, we can also view C as the right kernel $C = \{x | Ax = 0\}$ of some $m \times n$ matrix A. (These equations are over the field F_2.) The matrix A is called a **parity check** matrix for C. By a previous comment, the distance of C is the smallest number of columns in A whose sum is the zero vector. We construct A a column at a time, always making sure that no dependent set of fewer than d columns is created. We are able to construct the j-th column under this condition provided that

$$\sum_{r<d} \binom{j-1}{r} < 2^m$$

because the next columns must not coincide with the sum of any set of $d-1$ or fewer columns already in A. The most stringent case of this inequality is when $j = n$ and the resulting code C has dimension $\geq n - m$ and so $|C| \geq 2^{n-m}$ (with equality iff A has rank m). The conclusion follows. $\square$

Let $\delta = d/n$ be the normalized distance of the code, for some $\delta \leq 1/2$. Then the sum $v(n, d) = \sum_{i=0}^{\delta n} \binom{n}{i}$ is dominated by the last term $\binom{n}{\delta n}$. Therefore, the rate of the resulting code is at least $\log(2^n/\binom{n}{\delta n}))/n$. Since $\log \binom{n}{\delta n}/n$ is approximately the binary entropy function $H(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$, we obtain the following asymptotic version of the Gilbert-Varshamov bound.

Corollary 12.3. *For every $\delta \leq 1/2$ and for large n, there exist (linear) codes with normalized distance δ and rate $r \geq 1 - H(\delta)$.*

It is not known how to explicitly construct such codes. Neither is it known how close this bound is to the optimum. As we'll see momentarily, the best known upper bound is exponentially far from this lower bound on code size.

12.3.2. *Upper bounds: Sphere packing and linear programming.*

Theorem 12.4. *Any code C of length n and distance d satisfies $|C| \leq 2^n / v(n, d/2)$.*

Corollary 12.5. *Every code of relative distance δ has rate $r \leq 1 - H(\delta/2)$.*

Proof of theorem. For a code C of distance d, all radius $d/2$ balls around the points of C must be disjoint. The bound follows by dividing the size of the whole space 2^n by the size of each such ball. □

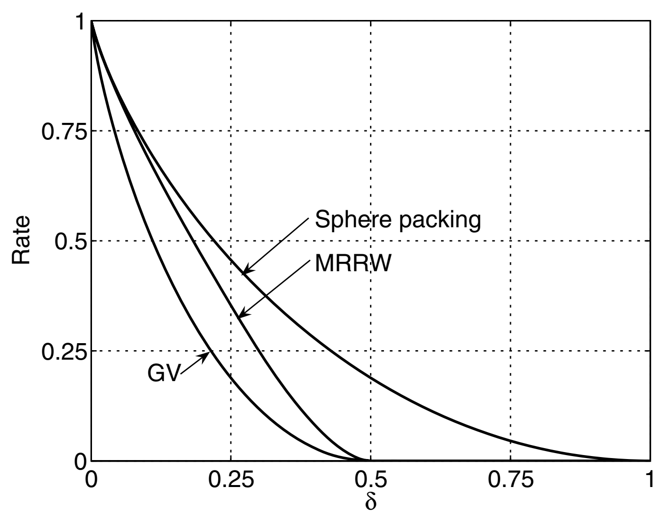


FIGURE 16. Asymptotic upper and lower bounds on rate vs. the relative distance.

As Figure 16 shows, the sphere-packing bound (Theorem 12.4) is rather weak. Indeed a much better upper bound and the "record holder" for nearly three decades was found by McEliece, Rodemich, Rumsey, and Welch [MRRW77].[23]

Theorem 12.6. *Every code of relative distance δ has rate $r \leq H(1/2 - \sqrt{\delta(1-\delta)})$.*

The relations between the lower bound and two upper bounds is illustrated in Figure 16.

[23] See [MRRW77] or [vL99], Section 5.4, for a slight improvement due to the same authors.

12.4. **Codes from graphs.** We now explain the basic connection between graphs and linear codes, which was one of the initial motivations for the very definition of expanders and the quest for their explicit construction.

Let A be a parity check matrix for C, i.e., $C = \{x | Ax = 0\}$ where A is an $m \times n$ matrix. Needless to say, a code C has many such representations, which depend on the choice of a basis for the orthogonal complement of C. Remarkably, properties of the underlying bipartite graph, most notably expansion properties, are crucial for establishing lower bounds on its distance and designing efficient decoding of the code. The connection is created by considering the bipartite graph G associated with A. The m rows (n columns) of A correspond to m vertices on the right and n on the left in G. There is an edge between the i-vertex on the right and the j-th on the left in G iff $a_{ij} = 1$. We occasionally denote the resulting code by $C(G)$. It will be convenient to switch between the graph-theoretic and linear-algebraic terminologies in our discussion.

As an example, the graph in Figure 17 represents a code of length 10, defined by 5 equations, the first of which is $v_1 + v_2 + v_3 + v_6 + v_7 + v_8 = 0 \mod 2$. It can be verified that all equations are independent in this case, and therefore this is a code of rate 1/2.

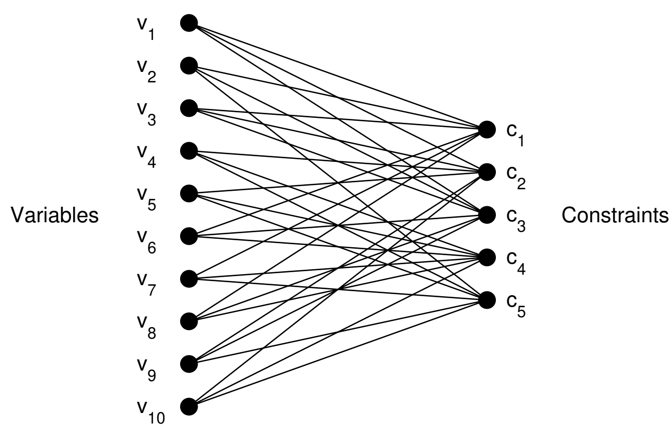


FIGURE 17. The constraints graph of a code.

Below we briefly (and incompletely...) sketch some of the history of this branch of error correction.

The idea of constructing codes this way was first suggested by Gallager [Gal63], who used sparse bipartite graphs, giving rise to the term **LDPC** codes, for Low Density Parity Check codes (as each parity check equation has a few nonzero coefficients). Gallager used random graphs, as explicit expanders did not exist at the time. This idea was picked up in Russia by Pinsker and Bassalygo [BP73], who sought out explicit constructions. They realized that expansion was at the heart of Gallager's arguments, formally defined expanders, and observed their utility in other basic applications, most notably to robust communication networks. It sparked the first two basic results in expander theory: Pinsker's probabilistic proof that most sparse graphs are expanders [Pin73] mentioned in the first section, and Margulis' first explicit construction [Mar73], which is discussed several times in this article. (See also his early paper on LDPC codes and girth, in [Mar82].)

The area fell dormant for quite a few years, and coding theorists concentrated on algebraic techniques, until graph codes stormed again into consciousness in the 80's and 90's, with important works of Tanner [Tan81], Alon et al. [ABN92], Sipser and Spielman [SS96], and more. This particular direction culminated in the paper of Spielman [Spi96], which achieved linear time encoding and decoding for asymptotically good codes.

All these constructions use regular graphs. An important idea due to Luby, Mitzenmacher, Shokrollahi and Spielman [LMSS01] was to use bipartite graphs with **irregular** degree distributions. These ideas were pushed further by Richardson and Urbanke [RSU01, RU01], resulting in nearly linear time decoding of codes with rates approaching channel capacity. (Our basic definition of rate applies only to a very simple model of noise. These new bounds apply in a variety of models for noisy communication channels.) We note that the underlying graphs here are typically chosen at random; the required properties that guarantee the superior behavior of the resulting codes cannot at present be obtained explicitly. Indeed the explicit construction of such expanders is an excellent challenge for those interested in research in this area. Work of Thorpe [Tho03] indicates that high lifts of carefully chosen base graphs (see Section 6) may be the place to look for good graphs. Consult the book of Richardson and Urbanke [RU] for further information on this fast growing area.

Our next step is to describe a prototype of this family, achieving the goal of efficient asymptotically good codes via the lossless expanders of the previous chapter.

12.5. **Efficient asymptotically good codes from lossless expanders.** Let $G = (V_L; V_R, E)$ be a bipartite graph that is k-regular[24] on the left, such that $|V_L| = n$ and $|V_R| = m$. We will need the following variant of vertex expansion for bipartite graphs.

Definition 12.7. The **left vertex expansion ratio** $L(G, d)$ is the minimum of $|\Gamma(S)|/|S|$ over all nonempty sets S in V_L of size at most d. In other words, every such set satisfies $|\Gamma(S)| \geq L(G, d)|S|$.

Note that $L(G, d)$ cannot exceed k for any k-left-regular graph G and any d. On the other hand, the results of [CRVW02] explained in Section 10 provide us with explicit bipartite graphs G for any $m = \Omega(n)$ (and hence constant rate codes $C(G)$) and expansion $L(G, d) > .99k$ for $d = \Omega(n)$. We first see how these yield asymptotically good codes, and later show that they actually suggest naturally efficient (indeed, linear time) decoding. (Encoding in $O(n^2)$ time comes free with the codes being linear. Linear time encoding algorithms can be achieved by paying more attention to the graph's structure.) The two theorems below, due to Sipser and Spielman, were proved before explicit construction of lossless expanders were available.

The code $C(G)$ has large distance.

Theorem 12.8 (Sipser-Spielman [SS96])**.** *If $L(G, d) > k/2$, then dist$(C(G)) \geq d$.*

[24] We use k for the degree here, as d is reserved for distance of the codes.

Proof. Observe that in a k-left-regular bipartite graph with left expansion $> k/2$, every nonempty set $S \subseteq V_L$ of size at most d has a **Unique Neighbor**, namely, a vertex $c \in V_R$ such that $|\Gamma(c) \cap S| = 1$. To see this, note that $e(S, \Gamma(S)) = k|S|$. But $|\Gamma(S)| > k|S|/2$, since the expansion exceeds $k/2$. Therefore the average right degree is less than 2, implying there must be at least one vertex with exactly one neighbor in S.

We now use the unique neighbor property of sets of size at most d to prove that every nonzero codeword $x \in C(G)$ has weight at least d. Let $S \subset V_L$ be the *support* of x, i.e., the set of coordinates v where $x_v = 1$. We saw above that $\Gamma(S)$ must contain a vertex ν which has only one neighbor in S. But then the ν-th coordinate of Ax is one, so x cannot be in the code $C(G)$. $\square$

The code $C(G)$ can be efficiently decoded. We now turn to the decoding of graph codes. Consider the following naïve iterative decoding algorithm: Upon receiving the input n-bit string y, as long as there exists a variable such that **most** of its neighboring constraints are not satisfied, flip it. In other words, given $x \notin C$, we flip the i-th bit provided that $|A(x + e_i)| < |Ax|$ (the first vector has smaller Hamming weight).

This algorithm (and variants thereof) is often called **belief propagation**. This is a powerful theme not only in error correction but in artificial intelligence and computational learning theory. More often than not these are only heuristics, while here this strategy always works.

Theorem 12.9 (Efficient Decoding, Sipser-Spielman [SS96])**.** *Let G be a k-left-regular bipartite graph in which $L(G, d) > \frac{3}{4}k$. Let y be an n-bit string whose distance from a codeword x is at most $d/2$. Then a repeated application of the naïve decoding algorithm to y will return x after a linear number of iterations.*

Proof. Let $y^{(i)}$ be the vector generated by the algorithm after i iterations, where $y = y^{(0)}$. Let A_i be the set of errors at iteration i, i.e. $A_i = \{v : y_v^{(i)} \neq x_v\}$. Then we have to prove that A_t is empty for $t = O(n)$. Consider the set $A = A_i$ at iteration i (we discard the index i to avoid cumbersome notation), and assume that A is not empty and that $|A| \leq d$. Partition $\Gamma(A)$ into satisfied neighbors S and unsatisfied neighbors U. (That is U is the support of the vector Ax.) Then

$$|S| + |U| = |\Gamma(A)| > \frac{3}{4}k|A|. \tag{21}$$

Now, count the edges between A and $\Gamma(A) = U \cup S$. There are at least $|U|$ edges leaving U and at least $2|S|$ edges leaving S (every vertex in S has an even number of neighbors in A). Therefore,

$$|U| + 2|S| \leq k|A|.$$

A linear combination of these inequalities yields

$$|U| > \frac{1}{2}k|A|. \tag{22}$$

Consequently, there is a variable in A with more than $k/2$ unsatisfied neighbors. This implies that as long as there are errors and $|A| \leq d$, some variable will be flipped by the algorithm. Since we flip a vertex with more unsatisfied neighbors than satisfied ones, $|U|$ decreases with every step. We deduce that if the distance from x does not exceed d throughout the run of the algorithm, then the algorithm will halt with the codeword x after a linear number of iterations.

To show that $|A_i|$ can never exceed d, note that at any step $|A_i|$ changes by ± 1. Therefore, if at any iteration $|A_i|$ exceeds d, there must be an index (say i) such that $|A_i| = d$. Then by (22), $|U_i| > kd/2$. On the other hand, by assumption, in the beginning $|A_0| \leq d/2$, and therefore $|U_0| \leq |\Gamma(A_0)| \leq kd/2$, contradicting the fact that $|U_i|$ is decreasing with i. □

We conclude by mentioning a parallel version of the "belief propagation" algorithm, where at every phase, **all** variables with a majority of violated constraints flip their value. A similar analysis can be carried out to show that under similar assumptions this algorithm converges to the correct codeword in $O(\log n)$ phases, and with total work (=number of flips) $O(n)$. The analysis (which we leave for the reader, who can peek e.g. at [CRVW02]) utilizes lossless expansion to show that the error sets A_i shrink by a constant factor with each iteration.

13. Metric embedding

Any metric space can be embedded into Euclidean space with some distortion of the distances. In this section we prove that the graph metric of expander graphs is the hardest metric to embed, in the sense that of all finite metric spaces with the same number of points, expanders require the largest distortion.

The study of finite metric spaces and their embeddings has undergone tremendous growth in the last decade. These developments are too wide in scope to be surveyed here, and a few pointers are provided in subsection 13.5. Much like expander graphs, discrete metric spaces are studied from at least three different perspectives: geometric, combinatorial and computational. Our main goal here is to briefly mention some of the main problems and results and show how the two theories interact.

13.1. **Basic definitions.** A metric space is the pair (X, d), where X is a set of points and d is the distance function $d : X \times X \to \mathbb{R}^+$. The distance function is a symmetric nonnegative function satisfying the triangle inequality; that is, $d(x, y) \geq 0$ with equality if $x = y$, $d(x, y) = d(y, x)$, and $d(x, y) \leq d(x, z) + d(z, y)$ for every $x, y, z \in X$.

A main question we consider here is how well a finite metric space (X, d) can be approximated by Euclidean metric. That is, our "model space" is $\mathbb{R}^n$ with l_2 norm. Let $f : X \to \mathbb{R}^n$ be an embedding of the metric space (X, d) into $(\mathbb{R}^n, l_2)$, where the l_2 distance between $x, y \in \mathbb{R}^n$ is $||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$. We define:

$$\begin{aligned}
\text{expansion}(f) &= \max_{x_1, x_2 \in X} ||f(x_1) - f(x_2)|| \,/\, d(x_1, x_2),\\
\text{contraction}(f) &= \max_{x_1, x_2 \in X} d(x_1, x_2) \,/\, ||f(x_1) - f(x_2)||,\\
\text{distortion}(f) &= \text{expansion}(f) \cdot \text{contraction}(f).
\end{aligned}$$

It is clear that there are metric spaces that cannot be embedded without distortion. For example, consider the metric $(\{1, 2, 3, 4\}, d)$ of the star graph, depicted in Figure 18, where $d(1, 4) = d(2, 4) = d(3, 4) = 1$, and $d(i, j) = 2$ otherwise.

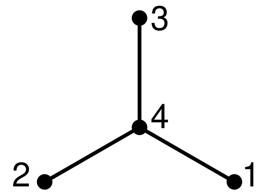


FIGURE 18. A graph that cannot be isometrically embedded into a Euclidean metric.

Embedding this space into $\mathbb{R}^n$ with no distortion implies that each of the triplets $\{1,2,4\}$, $\{1,3,4\}$, and $\{2,3,4\}$ are on a single line. Therefore, all four points must be on the same line, which clearly leads to a contradiction.

Aside from embeddings into $\mathbb{R}^n$ with l_2 norm, there is great interest in embeddings into $\mathbb{R}^n$ with l_1 norm. We will return to this issue below.

13.2. **Finding the minimal l_2 distortion.** Let $c_2(X,d)$ denote the least possible distortion in any embedding of the finite metric space (X,d) into $(\mathbb{R}^n, l_2)$. The dimension n of the host space is insignificant for the moment, since it is easy to see that for finite X, the smallest distortion achievable can always be achieved with dimension $n \leq |X|$. In this section we present some bounds on the dimension and distortion achievable for arbitrary metric spaces and derive an efficient algorithm to calculate $c_2(X,d)$. The algorithm relies on a quadratic semidefinite characterization of the problem. Using semidefinite duality, it leads to a simple way to prove lower bounds on $c_2(X,d)$.

We start with the existential upper bound, the grand ancestor of this area.

Theorem 13.1 (Bourgain [Bou85]). *Any n-point metric space (X,d) can be embedded into Euclidean space with distortion $\leq O(\log n)$.*

We need some additional terminology here. Let $S \subset \mathbb{R}^n$ be a finite set $S = \{z_1, \ldots, z_N\}$. This yields an N-point metric space (X,d) where $X = \{x_1, \ldots, x_N\}$. The metric is defined via $d(x_i, x_j) = ||z_i - z_j||$, where $||\cdot||$ stands for the l_2 norm. Such a metric space (X,d) is called "an l_2 metric". A similar definition is made for l_1 metrics.

We recall the following easy fact which is proved e.g. in [DL97].[25]

Claim 13.2. *That every l_2 metric is also an l_1 metric.*

When we simply speak of an l_1 (or l_2) space, this indicates that for that statement the dimension of the underlying space is immaterial.

The original statement of Theorem 13.1 was about embedding into l_1, but the same proof yields the stronger statement for embedding into l_2. A major reason that dimension plays only a minor role when we consider embeddings into l_2 is the following theorem which shows that in l_2 the dimension can be significantly reduced without significant loss in distortion. (We note that the analogous statement for l_1 does not hold; see [BC03, LN04].)

[25] There is much more to be said about comparisons between norms in general and l_2 in particular. Much of the local theory of Banach Spaces, e.g. [MS86], revolves around such problems. For an accessible proof of Dvoretzky's Theorem, the grandfather of this discipline, see e.g. [Mat02].

Theorem 13.3 (Johnson-Lindenstrauss [JL84]). *Any n-point l_2 metric can be embedded into an $O(\frac{\log n}{\epsilon^2})$-dimensional Euclidean space with distortion $\leq 1+\epsilon$.*

The proof of this extremely useful theorem is simple. It uses random linear projections to a low dimensional subspace and follows directly from the concentration of measure under such mappings. Combining the two theorems, we see that a logarithmic distortion can be achieved even in logarithmic dimension. We now turn to the algorithmic problem of computing distortion.

Theorem 13.4 (Linial-London-Rabinovich [LLR95]). *There is a polynomial time algorithm that given a metric space (X, d) (say by a matrix of distances) computes $c_2(X, d)$.*

Proof. The proof is based on semidefinite programming. Let (X, d) be a metric space with $|X| = n$. Let $f : X \to \mathbb{R}^n$. Since one can always scale f so that $contraction(f) = 1$, we have $distortion(f) \leq \gamma$ if and only if

$$d(x_i, x_j)^2 \leq ||f(x_i) - f(x_j)||^2 \leq \gamma^2 d(x_i, x_j)^2 \quad \text{for all } 1 \leq i < j \leq n. \tag{23}$$

Let us recall some standard facts from linear algebra next. A symmetric n by n matrix Z is said to be positive semidefinite if $v^T Z v \geq 0$ for all $v \in \mathbb{R}^n$. This is equivalent to each of the following two conditions: (i) All eigenvalues of Z are nonnegative, and (ii) $Z = WW^T$ for some matrix W. We denote by $\mathrm{PSD} = \mathrm{PSD}_n$ the collection of all $n \times n$ positive semidefinite matrices. We consider any embedding f of X into $\mathbb{R}^n$ where $f(x_i) = u_i$ and let U be the matrix whose i-th row is u_i and $Z = UU^T$.

It follows that finding $u_i = f(x_i)$ that satisfy (23) is equivalent to the existence of a matrix $Z \in \mathrm{PSD}$ such that

$$d(x_i, x_j)^2 \leq z_{ii} + z_{jj} - 2z_{ij} \leq \gamma^2 d(x_i, x_j)^2 \quad \text{for all } 1 \leq i < j \leq n, \tag{24}$$

since $||u_i - u_j||^2 = z_{ii} + z_{jj} - 2z_{ij}$ for $Z = UU^T$. Thus we conclude that $c_2(X, d) \leq \gamma$ if and only if there is a positive semidefinite matrix Z satisfying (24). Such an optimization problem can be solved in polynomial time by the ellipsoid algorithm. (See the book [GLS93] for general background in discrete optimization and the ellipsoid algorithm.) □

The algorithm above constructs a primal problem and solves it by the ellipsoid algorithm. The dual problem gives us an interesting characterization of $c_2(X, d)$ that is handy in proving lower bounds on the distortion. When we transform a primal problem to its dual we take a nonnegative combination of its constraints. But how do we look at the constraint $Z \in \mathrm{PSD}$? Again we need a simple but useful fact from linear algebra.

Claim 13.5. *A matrix Z is positive semidefinite if and only if $\sum_{i,j} q_{ij} z_{ij} \geq 0$ for all positive semidefinite matrices Q.*

Proof.

- **$\Leftarrow$:** For $v \in \mathbb{R}^n$, let Q be the PSD matrix defined by $(Q)_{ij} = v_i \cdot v_j$. Then, $v^T Z v = \sum_{i,j} q_{ij} z_{ij} \geq 0$, implying that $Z \in \mathrm{PSD}$.
- **$\Rightarrow$:** If Q is a PSD matrix of rank 1, then it has the form $(Q)_{ij} = v_i \cdot v_j$ for some $v \in \mathbb{R}^n$. Therefore, $\sum_{i,j} q_{ij} z_{ij} \geq 0$ for any PSD matrix Z. As mentioned above, any PSD matrix Q can be written as WW^T for some matrix W with orthogonal rows. Therefore any $Q \in \mathrm{PSD}$ is the sum of rank 1 PSD matrices, implying the required claim. □

Theorem 13.6 (Linial-London-Rabinovich [LLR95]). *The least distortion of any finite metric space (X, d) in the Euclidean space is given by*

$$c_2(X,d) = \max_{P \in PSD,\, P\cdot\mathbf{1}=0} \sqrt{\frac{\sum_{p_{ij}>0} p_{ij} d(x_i,x_j)^2}{-\sum_{p_{ij}<0} p_{ij} d(x_i,x_j)^2}}.$$

Proof. As we saw, the primal problem is:

$$\begin{aligned} \sum_{ij} q_{ij} z_{ij} &\geq 0 && \text{for all } Q \in \text{PSD}, \\ z_{ii} + z_{jj} - 2z_{ij} &\geq d(x_i,x_j)^2 && \text{for all } i,j, \\ \gamma^2 d(x_i,x_j)^2 &\geq z_{ii} + z_{jj} - 2z_{ij} && \text{for all } i,j. \end{aligned}$$

The dual program is the statement that for $\gamma < c_2(X,d)$, there must exist a non-negative combination of the constraints of the primal problem that yields a contradiction.

We are looking for a linear combination of the constraints that yields the contradiction $0 \geq 1$. By Claim 13.5 this combination comes down to selecting a specific $Q \in$ PSD and writing $\sum_{ij} q_{ij} z_{ij} \geq 0$. The rest of the inequalities should be combined with coefficients so as to eliminate all z_{ij} from our inequalities. To eliminate the off-diagonal entries, z_{ij} for $i \neq j$, these coefficients are necessarily as follows:

- If $q_{ij} > 0$, then we add the constraint $z_{ii} + z_{jj} - 2z_{ij} \geq d(x_i,x_j)^2$ multiplied by $q_{ij}/2$.
- If $q_{ij} < 0$, then we add the constraint $\gamma^2 d(x_i,x_j)^2 \geq z_{ii} + z_{jj} - 2z_{ij}$ multiplied by $-q_{ij}/2$.

In order that at the end of this process the coefficients of the diagonal entries z_{ii} will be zero, we need that all the row sums of Q will be zero. Therefore, assuming that $\sum_j q_{ij} = 0$ for all i, we obtained the following inequality:

$$0 \geq \sum_{q_{ij}>0} q_{ij} d(x_i,x_j)^2 + \gamma^2 \sum_{q_{ij}<0} q_{ij} d(x_i,x_j)^2.$$

The theorem follows by observing that this is a contradiction if γ^2 is smaller than

$$\sum_{q_{ij}>0} q_{ij} d(x_i,x_j)^2 / (-\sum_{q_{ij}<0} q_{ij} d(x_i,x_j)^2).$$ □

13.3. **Distortion bounds via semidefinite duality.** We now demonstrate the power of the above characterization to obtain optimal bounds on the distortion of two natural metrics both of which are graph metrics. With every graph $G = (V, E)$ we associate the metric space $(V(G), d_G)$ where $d_G(u,v)$ is the distance in G between the two vertices u, v. We use the shorthand $c_2(G)$ for $c_2(V(G), d_G)$. Below we derive optimal bounds on $c_2(G)$ for the discrete cube, as well as for expanders. In particular, we show that the graph metrics of expanders are as far from l_2 metrics as possible.

13.3.1. *Embedding the cube into l_2.* As in subsection 4.2.1 we denote the r-dimensional discrete cube by Q_r. Note that the graph metric of this graph coincides with the Hamming metric. The identity embedding of this graph into $(\mathbb{R}^r, l_2)$ (every vertex is viewed as an r-dimensional vector) can be easily seen to have distortion $\sqrt{r}$. Indeed the identity map has contraction $\sqrt{r}$ and expansion 1. We use Theorem 13.6 to show that this embedding is the best possible. Let us define the $2^r \times 2^r$ matrix P:

$$P(x,y) = \begin{cases} -1 & \text{if } d(i,j) = 1 \\ r-1 & \text{if } i = j \\ 1 & \text{if } d(i,j) = r \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that $P\mathbf{1} = 0$ and that $P \in \mathrm{PSD}$. The latter holds, since P has the same eigenvectors as the r-dimensional cube. Since $\sum_{p_{ij}>0} p_{ij} d(x_i,x_j)^2 = 2^r \cdot r^2$ and $-\sum_{p_{ij}<0} p_{ij} d(x_i,x_j)^2 = 2^r \cdot r$, we get that $c_2(Q_r) \geq \sqrt{r}$.

13.3.2. *Embedding expander graphs into l_2.* Consider some k-regular[26] graph G of size n and $\lambda_2 \leq k - \epsilon$, for fixed $k > 2$ and $\epsilon > 0$. It is again simple to see that G can be embedded with distortion $O(\log n)$ in l_2. In fact, any graph can be embedded with distortion equal to its diameter, which is $O(\log n)$ for an expander, as observed in subsection 2.4. To see this, consider embedding a graph to the vertices of a simplex in $\mathbb{R}^n$: namely, mapping the i-th vertex to $e_i/\sqrt{2}$, where e_i is the i-th vector of the standard basis for $\mathbb{R}^n$. Since every two vertices of the simplex have distance 1, the expansion is one and the contraction is equal to the diameter of the graph. We prove that for constant degree expander graphs, such an embedding is optimal up to a multiplicative constant independent of n.

Lemma 13.7. *Let $G = (V,E)$ be a k-regular graph of even size n. Let $H = (V,E')$ be the graph on the same vertex set as G, where two vertices are adjacent if their distance in G is at least $\lfloor \log_k n \rfloor$. Then H has a perfect matching.*

Proof. Since G is a k-regular graph, then every vertex has at most k^r vertices within distance r. If $r = \lfloor \log_k n \rfloor - 1$, then there at most $n/2$ vertices within distance r, and therefore H has minimal degree at least $n/2$. Therefore, it has a Hamiltonian cycle by Dirac's Theorem (e.g., Theorem 10.1.1 in [Die97]). It follows that H has a perfect matching, as claimed. □

Theorem 13.8 (Linial-London-Rabinovich [LLR95])**.** *Let $k \geq 3$ be an integer and let $\epsilon > 0$. If $G = (V,E)$ is an (n,k)-graph with $\lambda_2(G) \leq k - \epsilon$, then $c_2(G) = \Omega(\log n)$ where the implicit constant depends only on k and ϵ.*

Proof. Let B be the adjacency matrix of a perfect matching in H, whose existence is guaranteed by Lemma 13.7. Let $P = kI - A_G + \frac{\epsilon}{2}(B - I)$. It is easy to verify that $P\mathbf{1} = 0$. To check that P is PSD, it suffices to consider the form $x^T P x$ for some vector $x \perp \mathbf{1}$. Then

$$\begin{aligned} x^T(kI - A_G)x &\geq (k - \lambda_2)||x||^2 \geq \epsilon ||x||^2, \\ x^T(B - I)x &= \sum_{(i,j)\in B} (2x_i x_j - x_i^2 - x_j^2) \geq -2 \sum_{(i,j)\in B} (x_i^2 + x_j^2) = -2||x||^2. \end{aligned}$$

[26] We use k for the degree in this section, as d is reserved for the distance.

Therefore, $x^T P x = x^T(kI - A_G)x + x^T(\epsilon/2)(B - I)x \geq 0$. To obtain the lower bound on $c_2(G)$ we evaluate

$$\begin{aligned} -\sum_{p_{ij}<0} d(i,j)^2 p_{ij} &= kn, \\ \sum_{p_{ij}>0} d(i,j)^2 p_{ij} &\geq \frac{\epsilon}{2} \cdot n \lfloor \log_k n \rfloor^2, \end{aligned}$$

where the inequality follows since distances of edges in B are at least $\lfloor \log_k n \rfloor$. This implies that $c_2(G) = \Omega(\log n)$, as claimed. $\square$

The last two examples are both instances of Poincaré-type inequalities on graphs (see [LMN02] for more on this). Let f be an embedding of the vertices of a graph G into $(\mathbb{R}^n, l_2)$. Poincaré-type inequalities compare between the average of $\|f(u) - f(v)\|^2$ over *all pairs of vertices* u, v and the same average on the edges (u, v). Here is such an inequality that illustrates this idea.

Theorem 13.9. *Let $G = (V, E)$ be a k-regular graph with second eigenvalue λ_2. For every embedding $f : V \to \mathbb{R}^n$*

$$\mathbb{E}_{(u,v)\in V\times V}\|f(u) - f(v)\|^2 \leq \frac{k}{k - \lambda_2}\mathbb{E}_{(u,v)\in E}\|f(u) - f(v)\|^2.$$

Proof Sketch: Observe first that it suffices to prove the inequality for real functions f, since both sides of the inequality are additive over dimensions. We can also assume that f has zero average, since both sides are invariant under shifting f by a constant. At this stage this is just the variational definition of the second eigenvalue. $\square$

The result both for the cube and for expanders can be derived (at least up to a constant factor) from this inequality. In [LMN02] similar questions are considered for graphs of high girth. We still do not have sharp bounds on that question.

13.4. **Algorithms for cut problems via embeddings.** Let us consider the following natural computational problem: Given an n-vertex graph G, compute or estimate its expansion ratio $h(G)$. We mention that there are many variants, like balanced cut (estimating the number of edges to separate the graph into roughly equal pieces), and others. These arise naturally as subroutines in many graph algorithms as part of a "divide-and-conquer" approach, where small cuts guarantee smaller interference between sub-solutions in the separate pieces.

Is has been known for a long time that the exact determination of $h(G)$ is difficult (co-**NP** hard) [BKV81]. There are numerous computational problems about cuts in graphs which are known to be hard, and it is an open and fascinating problem how well they can be approximated in polynomial time. A first breakthrough was achieved by Leighton and Rao [LR99], who presented a polynomial time algorithm that approximates $h(G)$ to within an $O(\log n)$ factor for an n-vertex graph.

A different proof for this result was given in [LLR95]. This paper has established the connection between this circle of problems and low-distortion embeddings of metric spaces. Before we survey this result, we should mention a recent breakthrough [ARV04] which gives a polynomial time approximation algorithm for computing $h(G)$ up to a factor of only $O(\sqrt{\log n})$. It is a beautiful open question of whether an efficient algorithm exists that approximates $h(G)$ up to a constant factor approximation. Some indications that this may be too optimistic were given in [CKK05]. We will not be able to review here any of these exciting recent findings.

As mentioned above, there is a host of important and mostly difficult optimization problems where the input is a graph $G = (V, E)$ and where the objective is to find a subset $S \subseteq V$ so as to optimize some quantity related to the cut $E_G(S, \bar{S})$. Up to a small factor, the question of finding $h(G)$ falls into this category, since it is easily verified that

$$\frac{2h(G)}{n} \geq \min \frac{|E_G(S, \bar{S})|}{|S|(n - |S|)} \geq \frac{h(G)}{n}$$

where the minimum is over all subsets $S \subseteq V$. So let us concentrate on the problem of finding $\min \frac{|E_G(S,\bar{S})|}{|S|(n-|S|)}$. It is closely related to another classical computational problem called the **all-pairs multicommodity flow** problem, which we now define.

The input to this problem is an n-vertex graph $G = (V, E)$. Between every pair of vertices we should "ship" $\delta > 0$ units of a commodity, and we should maximize δ subject to the following constraints:

- There are $\binom{n}{2}$ distinct and unexchangable commodities, one for each pair of vertices in G.
- Edge capacities are 1. Namely, it is possible to ship different commodities through each edge as long as their total amount does not exceed 1.
- The flow of each commodity satisfies conservation of matter at each vertex.

Again, the problem is to find the largest $\delta > 0$ for which this can be done. The all-pairs multicommodity flow problem is a linear programming problem and can, therefore, be solved in polynomial time. We denote the largest attainable δ by $\delta_{\max}$. The following lemma shows that the graph parameter $\delta_{\max}(G)$ provides an $O(\log n)$ approximation for $\min_S \frac{|E_G(S,\bar{S})|}{|S|(n-|S|)}$. Consequently, we achieve a polynomial-time algorithm that approximates $h(G)$ to within an $O(\log n)$ factor.

Theorem 13.10 ([LLR95]). *For every graph $G = (V, E)$ the following inequality holds:*

$$\delta_{\max}(G) \leq \min_S \frac{|E_G(S, \bar{S})|}{|S|(n - |S|)} \leq O(\delta_{\max}(G) \cdot \log n).$$

Proof Sketch: The left inequality is easy. Indeed, for any nonempty vertex set $S \subset V$ one has a total flow of $|S|(n - |S|)$ through the cut. Since each edge has capacity one, the flow cannot exceed $E(S, \bar{S})$, yielding the bound.

To obtain the lower bound on $\delta_{\max}$, we use linear programming duality. It is not hard to show that the optimum of the all-pairs multicommodity flow problem equals

$$\delta_{\max} = \min \frac{\sum_{(i,j)\in E} d_{i,j}}{\sum_{i,j\in V} d_{i,j}}, \tag{25}$$

where the minimum is with respect to all metrics d on the set V.

The basic idea of the proof is that if we restrict d to be an l_1 metric, then two things happen:

- $$\min \frac{\sum_{(i,j)\in E} d_{i,j}}{\sum_{i,j\in V} d_{i,j}}$$

minimized over all l_1 metrics d equals

$$\min_S \frac{|E_G(S,\bar{S})|}{|S|(n-|S|)}.$$

- Restricting d in (25) to be an l_1 metric increases this expression by a factor of at most $O(\log n)$.

Together these claims imply the theorem. Somewhat like the argument in Theorem 13.9, we'd like to reduce the problem to one-dimensional embeddings. Now a simple property of l_1 metrics is that they form a polyhedral convex cone; namely, if d_1 and d_2 are two l_1 metrics on the same finite set X, then so is $a_1d_1 + a_2d_2$ for every two nonnegative constants a_1, a_2. This cone is called the **cut cone** for reasons which we will soon see. A whole book is dedicated to the basic properties of this cone and its relatives [DL97].

The extreme rays of the cut cone are easy to determine: Associated with any proper $S \subseteq [n]$ is a metric d_S on $[n]$ as follows: $d_S(x,y) = 1$ if exactly one of x, y is in S and 0 otherwise. This is the so-called **cut metric associated with** S which is easily seen to be an l_1 metric. We can now approach the proof of Theorem 13.10 as follows: By Bourgain's Theorem only a factor of $O(\log n)$ is lost if in minimizing the right-hand side in Equation 25 we insist that the metric be in l_1. Since the cut cone is convex, the optimum among l_1 metrics is obtained at an extreme ray,[27] i.e., by a cut metric d_S. In this case the right-hand side in Equation 25 becomes $\frac{|E_G(S,\bar{S})|}{|S|(n-|S|)}$, as claimed. □

This type of argument seems very alluring from a computational perspective. Consider any optimization problem where we are given a graph $G = (V, E)$ and we seek a cut $E_G(S, \bar{S})$ which is optimal in some sense. By the same reasoning, such a problem can be turned into a problem of convex optimization, in which we are trying to optimize some objective function on the cut cone. There is a general theory of discrete optimization as laid out in [GLS93], in which we try to optimize some linear function over a convex domain Ω. In order to use this machinery we must be able to solve efficiently two basic questions for Ω: (i) **Membership** - To determine, given a point x, whether it belongs to Ω; and (ii) **Separation** - Same as above, but if $x \notin \Omega$, find a hyperplane that separates x from Ω. Unfortunately, the cut cone is computationally quite bad. Even the membership problem (which is the simpler of the two) for the cut cone is **NP**-hard. In words: Given a metric, it is difficult to decide whether it is an l_1 metric.

This suggests the following natural problem first raised independently by Linial and by Goemans:

Open problem 13.11. Is there a cone that is a good (preferably only constant distortion) approximation to the cut cone and for which membership and separation can be solved in time that is polynomial in the dimension?

They have also suggested a candidate for this job. We say that a metric space (X, d) is of **negative type** if $\sqrt{d}$ is an l_2 metric. It is easy to show that metrics of negative type form a convex cone for which both the membership and separation problems can be solved efficiently. Also, every l_1 metric is of negative type. In this

[27] There is a small technicality which we neglect here: We actually intersect the cone with some hyperplane so as to do the optimization on some bounded domain.

light, Linial and Goemans have posed the following conjecture, a positive answer to which would realize the above-mentioned program.

Conjecture 13.12. Every metric of negative type can be embedded into an l_1 metric with bounded distortion.

In a fascinating and still quite mysterious paper, Khot and Vishnoy [KV05] have recently refuted this conjecture. However the lower bound they derive for the distortion is very small $(\log\log n)^c$ for some $1 > c > 0$ for n-point metrics. There is clearly still a good story here waiting to unfold.

13.5. **A glimpse into the bigger picture.** As mentioned above, the previous sections give only a very brief summary of a fascinating and highly active area of research. For general reviews of this area see [Lin02], Chapter 15 in the book [Mat02], and [IM04]. The quadratic programming method presented here has been used in several additional cases [LMN02, LM00]. Progress in this area has been extremely fast, and already these fairly recent surveys are by now outdated. To keep track of the present state of affairs, the reader is encouraged to look up a list of open problems in this area that is being curated by J. Matousek: `http://kam.mff.cuni.cz/~matousek/`.

We also mention briefly another connection of (finite) expanders to metric embeddings of (infinite) graphs. There are conjectures in topology, most notably the Baum-Connes Conjecture [Val03], which address the embeddability of manifolds into Hilbert space with "uniform" distortion (a term which has to be properly defined). Some extensions of these conjectures were recently disproved by Gromov [Gro03] in an ingenious way whose spirit is the following. The "hard-to-embed" manifolds will be (as in the finite setting) derived from expanders. The manifold is defined via its "fundamental group", an infinite group whose Cayley graph carries the metric structure of the manifold. The definition of this group via generators and relations guarantees that the above Cayley graph contains in it larger and larger finite expanders, which (as we already know) do not embed well into l_2. Carrying out this plan is highly nontrivial, and we refer the reader to [Val03].

About the authors

Shlomo Hoory recently joined the machine learning group at IBM Haifa Research Lab. He held postdoctoral positions at the theory group at the University of Toronto and at the University of British Columbia. He received his Ph.D. in computer science from the Hebrew University in 2002.

Nathan Linial is a professor of computer science at the Hebrew University. He is mostly interested in the interplay between mathematics, computer science, and other fields of science.

Avi Wigderson is a professor at the Institute for Advanced Study in Princeton. Before that he was a professor of computer science at the Hebrew University for fifteen years. He is a winner of the Nevanlinna Prize.

References

[ABN92] N. Alon, J. Bruck, J. Naor, M. Naor, and R. M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992.

[ABSRW04] M. Alekhnovich, E. Ben-Sasson, A. A. Razborov, and A. Wigderson. Pseudorandom generators in propositional proof complexity. *SIAM J. Comput.*, 34(1):67–88 (electronic), 2004. MR2114305 (2006a:03078)

[AC89] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks, *Discrete Math.*, 72:15–19, 1989. MR0975519 (90f:05080)

[AC02] N. Alon and M. Capalbo. Explicit unique-neighbor expanders. In *43nd IEEE Symposium on Foundations of Computer Science (Vancouver, BC)*, pages 73–79. IEEE Computer Society, 2002.

[AC04] N. Alon and M. Capalbo. Smaller explicit superconcentrators. *Internet Math.*, 1(2):151–163, 2004. MR2077222 (2006a:05140)

[AFH] O. Angel, J. Friedman, and S. Hoory. The non-backtracking spectrum of the universal cover of a graph. Manuscript.

[AFWZ95] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Comput. Complexity*, 5(1):60–75, 1995. MR1319493 (95m:68082)

[AKL79] R. Aleliunas, R. M. Karp, R. J. Lipton, L. Lovász, and C. Rackoff. Random walks, universal traversal sequences, and the complexity of maze problems. In *20th Annual Symposium on Foundations of Computer Science (San Juan, Puerto Rico, 1979)*, pages 218–223. IEEE, New York, 1979. MR0598110 (82h:68084)

[AKS83] M. Ajtai, J. Komlós, and E. Szemerédi. An $O(n \log n)$ sorting network. *Combinatorica*, 3(1):1–19, 1983.

[AKS87] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in LOGSPACE. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, pages 132–140, 1987.

[AL06] A. Amit and N. Linial. Random lifts of graphs II: Edge expansion. *Combinatorics Probability and Computing*, 15:317–332, 2006. MR2216470

[AL96] S. Arora and C. Lund. Hardness of approximations. In D. S. Hochbaum, editor, *Approximation Algorithms for NP-hard Problems.* PWS Publishing Company, Boston, MA, 1996.

[AL02] A. Amit and N. Linial. Random graph coverings. I. General theory and graph connectivity. *Combinatorica*, 22(1):1–18, 2002. MR1883559 (2003a:05131)

[ALM96] S. Arora, F. T. Leighton, and B. M. Maggs. On-line algorithms for path selection in a nonblocking network. *SIAM J. Comput.*, 25(3):600–625, 1996. MR1390029 (97b:68014)

[ALM98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. Also in FOCS, 1992. MR1639346 (99d:68077b)

[ALM02] A. Amit, N. Linial, and J. Matoušek. Random lifts of graphs: independence and chromatic number. *Random Structures Algorithms*, 20(1):1–22, 2002. MR1871947 (2003a:05132)

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of Computing (Singer Island, FL, 1984). MR0875835 (88e:05077)

[Alo97] N. Alon. On the edge-expansion of graphs. *Combin. Probab. Comput.*, 6(2):145–152, 1997. MR1447810 (98c:05084)

[ALW01] N. Alon, A. Lubotzky, and A. Wigderson. Semi-direct product in groups and zig-zag product in graphs: connections and applications (extended abstract). In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 630–637. IEEE Computer Society, 2001. MR1948752

[AM85] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985. MR0782626 (87b:05092)

[AR94] N. Alon and Y. Roichman. Random Cayley graphs and expanders. *Random Structures Algorithms*, 5(2):271–284, 1994. MR1262979 (94k:05132)

[AR01] M. Alekhnovich and A. A. Razborov. Lower bounds for polynomial calculus: non-binomial case. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 190–199. IEEE Computer Society, 2001. MR1948707

[ARV04] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 222–231, 2004. MR2121604 (2005j:68081)

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *J. ACM*, 45(1):70–122, 1998. Also in FOCS, 1992. MR1614328 (99d:68077a)

[AS00] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], New York, second edition, 2000. With an appendix on the life and work of Paul Erdős. MR1885388 (2003f:60003)

[BC03] B. Brinkman and M. Charikar. On the impossibility of dimension reduction in l_1. In *44nd IEEE Symposium on Foundations of Computer Science (Cambridge, MA)*, pages 514–523. IEEE Computer Society, 2003.

[BCCF05] K. Burgin, P. Chebolu, C. Cooper, and A.M. Frieze. Hamilton cycles in random lifts of graphs. Preprint at `http://www.math.cmu.edu/~af1p/papers.html`, 2005.

[Bec75] William Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975. MR0385456 (52:6317)

[BFU99] A. Z. Broder, A. M. Frieze, and E. Upfal. Static and dynamic path selection on expander graphs: a random walk approach. *Random Structures Algorithms*, 14(1):87–109, 1999. MR1662203 (99m:05144)

[BH04] Y. Bilu and S. Hoory. On codes from hypergraphs. *European J. Combin.*, 25(3):339–354, 2004. MR2036471 (2005d:94200)

[BKV81] M. Blum, R. M. Karp, O. Vornberger, C. H. Papadimitriou, and M. Yannakakis. The complexity of testing whether a graph is a superconcentrator. *Inform. Process. Lett.*, 13(4-5):164–167, 1981. MR0651465 (84f:68031)

[BL] Y. Bilu and N. Linial. Lifts, discrepancy and nearly optimal spectral gaps. Combinatorica, to appear.

[BMRV02] H. Buhrman, P. B. Miltersen, J. Radhakrishnan, and S. Venkatesh. Are bitvectors optimal? *SIAM J. Comput.*, 31(6):1723–1744 (electronic), 2002. MR1954872 (2004a:68014)

[BOGH03] J. Buresh-Oppenheim, N. Galesi, S. Hoory, A. Magen, and T. Pitassi. Rank bounds and integrality gaps for cutting planes procedures. In *44nd IEEE Symposium on Foundations of Computer Science (Cambridge, MA, 2003)*, pages 318–327. IEEE Computer Society, 2003.

[Bol86] B. Bollobás. *Combinatorics.* Cambridge University Press, Cambridge, 1986. Set systems, hypergraphs, families of vectors and combinatorial probability. MR0866142 (88g:05001)

[Bon70] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)*, 20(fasc. 2):335–402 (1971), 1970. MR0283496 (44:727)

[Bou85] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel J. Math.*, 52(1-2):46–52, 1985. MR0815600 (87b:46017)

[BP73] L. A. Bassalygo and M. S. Pinsker. The complexity of an optimal non-blocking commutation scheme without reorganization. *Problemy Peredači Informacii*, 9(1):84–87, 1973. Translated into English in Problems of Information Transmission, 9 (1974) 64-66. MR0327387 (48:5729)

[BS87] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *The 28th Annual Symposium on Foundations of Computer Science*, pages 286–294, 1987.

[BS92] P. Berman and G. Schnitger. On the complexity of approximating the independent set problem. *Inform. and Comput.*, 96(1):77–94, 1992. MR1142227 (92k:90110)

[BSW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *J. ACM*, 48(2):149–169, 2001. MR1868713 (2002j:03066)

[Bus82] P. Buser. A note on the isoperimetric constant. *Ann. Sci. École Norm. Sup. (4)*, 15(2):213–230, 1982. MR0683635 (84e:58076)

[BZ88] Yu. D. Burago and V. A. Zalgaller. *Geometric inequalities*, volume 285 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1988. Translated from the Russian by A. B. Sosinskiĭ, Springer Series in Soviet Mathematics. MR0936419 (89b:52020)

[Cam] P. J. Cameron. A Markov chain for Steiner triple systems. Manuscript.

[Car72] P. Cartier. Fonctions harmoniques sur un arbre. In *Symposia Mathematica, Vol. IX (Convegno di Calcolo delle Probabilità, INDAM, Rome, 1971)*, pages 203–270. Academic Press, London, 1972. MR0353467 (50:5950)

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pages 195–199. Princeton Univ. Press, Princeton, NJ, 1970. MR0402831 (53:6645)

[Cio06] S. M. Cioabă. On the extreme eigenvalues of regular graphs. *Journal of Combinatorial Theory, Ser. B*, 96(3):367–373, 2006. MR2220665

[CKK05] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *IEEE Conference on Computational Complexity*, 2005.

[CLRS01] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to algorithms.* MIT Press, Cambridge, MA, second edition, 2001. MR1848805 (2002e:68001)

[CRVW02] M. Capalbo, O. Reingold, S. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 659–668, 2002. MR2121193

[CT65] J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965. MR0178586 (31:2843)

[DDPW83] D. Dolev, C. Dwork, N. Pippenger, and A. Wigderson. Superconcentrators, generalizers and generalized connectors with limited depth. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing*, pages 42–51, 1983.

[Die97] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997. Translated from the 1996 German original. MR1411445

[DL] Y. Drier and N. Linial. Minors in lifts of graphs. Random Structures and Algorithms, to appear.

[DL97] M. M. Deza and M. Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997. MR1460488 (98g:52001)

[Dod84] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Trans. Amer. Math. Soc.*, 284(2):787–794, 1984. MR0743744 (85m:58185)

[DSV03] G. Davidoff, P. Sarnak, and A. Valette. *Elementary number theory, group theory, and Ramanujan graphs*, volume 55 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2003. MR1989434 (2004f:11001)

[ER59] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959. MR0120167 (22:10924)

[Fel68] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons Inc., New York, 1968. MR0228020 (37:3604)

[FGL91] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Approximating clique is almost np-complete. In *32nd IEEE Symposium on Foundations of Computer Science*, pages 2–12. IEEE Computer Society, 1991.

[FK81] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981. MR0637828 (83e:15010)

[FKS89] J. Friedman, J. Kahn, and E. Szemeredi. On the second eigenvalue of random regular graphs. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 587–598, 1989.

[Fri] J. Friedman. A proof of Alon's second eigenvalue conjecture. Memoirs of the A.M.S., to appear.

[Fri91] J. Friedman. The spectra of infinite hypertrees. *SIAM J. Comput.*, 20(5):951–961, 1991. MR1115660 (92k:05096)

[Fri93] J. Friedman. Some geometric aspects of graphs and their eigenfunctions. *Duke Math. J.*, 69(3):487–525, 1993. MR1208809 (94b:05134)

[Fri03] J. Friedman. Relative expanders or weakly relatively Ramanujan graphs. *Duke Math. J.*, 118(1):19–35, 2003. MR1978881 (2004m:05165)

[FW95] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995. MR1325271 (96b:05110)

[Gal63] R. G. Gallager. *Low Density Parity Check Codes*. MIT Press, Cambridge, MA, 1963. MR0136009 (24:B2048)

[GG81] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. *J. Comput. System Sci.*, 22(3):407–420, 1981. Special issue dedicated to Michael Machtey. MR0633542 (83d:94029)

[Gil98] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220 (electronic), 1998. MR1621958 (99e:60076)

[GLS93] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, second edition, 1993. MR1261419 (95e:90001)

[Gol97] O. Goldreich. A sample of samplers – a computational perspective on sampling (survey). Technical Report TR97-020, Electronic Colloquium on Computational Complexity (ECCC), 1997. `http://www.eccc.uni-trier.de/eccc/`.

[Gra05] A. Granville. It is easy to determine whether a given integer is prime. *Bull. Amer. Math. Soc. (N.S.)*, 42(1):3–38 (electronic), 2005. MR2115065 (2005k:11011)

[Gre95] Y. Greenberg. *On the Spectrum of Graphs and Their Universal Covering*. PhD thesis, Hebrew University of Jerusalem, 1995 (in Hebrew).

[Gro77] J. L. Gross. Every connected regular graph of even degree is a Schreier coset graph. *J. Combinatorial Theory Ser. B*, 22(3):227–232, 1977. MR0450121 (56:8419)

[Gro83] M. Gromov. Filling Riemannian manifolds. *J. Differential Geom.*, 18(1):1–147, 1983. MR0697984 (85h:53029)

[Gro03] M. Gromov. Random walk in random groups. *Geom. Funct. Anal.*, 13(1):73–146, 2003. MR1978492 (2004j:20088a)

[Hås99] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Math.*, 182(1):105–142, 1999. MR1687331 (2000j:68062)

[H50] R. Hamming, Error detecting and error correcting codes, *Bell System Technical Journal*, 29:147–160, 1950. MR0035935 (12:35c)

[HMP06] S. Hoory, A. Magen, and T. Pitassi. Monotone circuits for the majority function. 10th International Workshop on Randomization and Computation (RANDOM), 2006.

[Hoo02] S. Hoory. *On graphs of high girth*. PhD thesis, Hebrew University of Jerusalem, 2002.

[Hoo05] S. Hoory. A lower bound on the spectral radius of the universal cover of a graph. *J. Combin. Theory Ser. B*, 93(1):33–43, 2005. MR2102266 (2005k:05144)

[IM04] P. Indyk and J. Matoušek. Low-distortion embeddings of finite metric spaces. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of discrete and computational geometry*, Discrete Mathematics and Its Applications (Boca Raton), pages 177–196. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[INW94] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, pages 356–364, 1994.

[Jer03] M. Jerrum. *Counting, sampling and integrating: algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003. MR1960003 (2004a:68044)

[JL84] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, CT, 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984. MR0737400 (86a:46018)

[JŁR00] S. Janson, T. Łuczak, and A. Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. MR1782847 (2001k:05180)

[JM87] S. Jimbo and A. Maruoka. Expanders obtained from affine transformations. *Combinatorica*, 7(4):343–355, 1987. MR0931192 (89d:68071)

[JS89] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing Markov chains. *Inform. and Comput.*, 82(1):93–133, 1989. MR1003059 (91g:68084)

[JS96] M. Jerrum and A. Sinclair. The Markov chain Monte Carlo method: an approach to approximate counting and integration. In D. S. Hochbaum, editor, *Approximation Algorithms for NP-Hard Problems*. PWS Publishing Company, Boston, MA, 1996.

[JSV04] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697 (electronic), 2004. MR2147852 (2006b:15013)

[Kah95] N. Kahale. Eigenvalues and expansion of regular graphs. *J. Assoc. Comput. Mach.*, 42(5):1091–1106, 1995. MR1412045 (97f:68149)

[Kar72] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, NY, 1972)*, pages 85–103. Plenum, New York, 1972. MR0378476 (51:14644)

[Kas05a] M. Kassabov. Kazhdan constants for $SL_n(Z)$. *International Journal of Algebra and Computation*, 15(5-6), 971–995, 2005. MR2197816

[Kas05b] M. Kassabov. Symmetric groups and expander graphs. `http://www.arxiv.org/abs/math.GR/0505624`, 2005.

[Kaz67] D. A. Kazhdan. On a connection between the dual space of a group and the structure of its closed subgroups. *Func. Anal. Appl.*, 1:63–65, 1967.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *29th IEEE Symposium on Foundations of Computer Science (White Planes)*, pages 68–80. IEEE Computer Society, 1988.

[KLN05] M. Kassabov, A. Lubotzky, and N. Nikolov. Finite simple groups as expanders. `http://www.arxiv.org/abs/math.GR/0510562`, 2005.

[Kör89] T. W. Körner. *Fourier analysis*. Cambridge University Press, Cambridge, second edition, 1989. MR1035216 (90j:42001)

[KPS85] R. Karp, N. Pippenger, and M. Sipser. A time-randomness tradeoff. In *AMS Conference on Probabilistic Computational Complexity*, 1985.

[KV05] S. Khot and N. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into l1. In *The 46th Annual Symposium on Foundations of Computer Science*, 2005.

[Lin02] N. Linial. Finite metric-spaces—combinatorics, geometry and algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, pages 573–586, Beijing, 2002. Higher Ed. Press. MR1957562 (2003k:05045)

[LL06] N. Linial and E. London. On the expansion rate of Margulis expanders. *Journal of Combinatorial Theory, Ser. B*, 96(3):436–442, 2006. MR2220670

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. MR1337355 (96e:05158)

[LM00] N. Linial and A. Magen. Least-distortion Euclidean embeddings of graphs: products of cycles and expanders. *J. Combin. Theory Ser. B*, 79(2):157–171, 2000. MR1769197 (2002c:05062)

[LMN02] N. Linial, A. Magen, and A. Naor. Girth and Euclidean distortion. *Geom. Funct. Anal.*, 12(2):380–394, 2002. MR1911665 (2003d:05054)

[LMSS01] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman. Improved low-density parity-check codes using irregular graphs. *IEEE Trans. Inform. Theory*, 47(2):585–598, 2001. MR1820478 (2002h:94115)

[LN98] A. Lubotzky and T. Nagnibeda. Not every uniform tree covers Ramanujan graphs. *J. Combin. Theory Ser. B*, 74(2):202–212, 1998. MR1654133 (2000g:05052)

[LN04] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geom. Funct. Anal.*, 14(4):745–747, 2004. MR2084978 (2005g:46035)

[Lok01] S. V. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. *J. Comput. System Sci.*, 63(3):449–473, 2001. MR1892859 (2003a:68059)

[Lov93] L. Lovász. *Combinatorial problems and exercises*. North-Holland Publishing Co., Amsterdam, second edition, 1993. MR1265492 (94m:05001)

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. MR0963118 (89m:05099)

[LR99] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999. MR1753034 (2001e:05125)

[LR05] N. Linial and E. Rozenman. Random lifts of graphs: perfect matchings. *Combinatorica*, 25(4):407–424, 2005. MR2143247 (2006c:05110)

[LSV05] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Explicit constructions of Ramanujan complexes of type A_d. *European J. Combin.*, 26(6):965–993, 2005. MR2143204 (2006g:20043)

[Lub] A. Lubotzky. Finite simple groups of lie type as expanders. In preparation.

[Lub94] A. Lubotzky. *Discrete groups, expanding graphs and invariant measures*, volume 125 of *Progress in Mathematics*. Birkhäuser Verlag, Basel, 1994. With an appendix by Jonathan D. Rogawski. MR1308046 (96g:22018)

[LW93] A. Lubotzky and B. Weiss. Groups and expanders. In *Expanding graphs (Princeton, NJ, 1992)*, volume 10 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 95–109. Amer. Math. Soc., Providence, RI, 1993. MR1235570 (95b:05097)

[LW98] L. Lovász and P. Winkler. Mixing times. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 85–133. Amer. Math. Soc., Providence, RI, 1998. MR1630411 (99h:60138)

[LZ] A. Lubotzky and A. Zuk. On property (t). In preparation.

[Mar73] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredači Informacii*, 9(4):71–80, 1973. MR0484767 (58:4643)

[Mar82] G. A. Margulis. Explicit constructions of graphs without short cycles and low density codes. *Combinatorica*, 2(1):71–78, 1982. MR0671147 (83j:05053)

[Mar88] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, 1988. MR0939574 (89f:68054)

[Mat02] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. MR1899299 (2003f:52011)

[McK81] B. D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40:203–216, 1981. MR0629617 (84h:05089)

[Mor94] M. Morgenstern. Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q. *J. Combin. Theory Ser. B*, 62(1):44–62, 1994. MR1290630 (95h:05089)

[MR95] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge University Press, Cambridge, 1995. MR1344451 (96i:65003)

[MR00] R. A. Martin and D. Randall. Sampling adsorbing staircase walks using a new Markov chain decomposition method. In *41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000)*, pages 492–502. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000. MR1931846

[MRRW77] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Trans. Information Theory*, IT-23(2):157–166, 1977. MR0439403 (55:12296)

[MS77a] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. I.* North-Holland Publishing Co., Amsterdam, 1977. North-Holland Mathematical Library, Vol. 16. MR0465509 (57:5408a)

[MS77b] F. J. MacWilliams and N. J. A. Sloane. *The theory of error-correcting codes. II.* North-Holland Publishing Co., Amsterdam, 1977. North-Holland Mathematical Library, Vol. 16. MR0465510 (57:5408b)

[MS86] V. D. Milman and G. Schechtman. *Asymptotic theory of finite-dimensional normed spaces*, volume 1200 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1986. With an appendix by M. Gromov. MR0856576 (87m:46038)

[MT] R. Montenegro and P. Tetali. A primer on modern techniques for bounding mixing times. Preprint at `http://www.math.gatech.edu/%7Etetali/RESEARCH/pubs.html`.

[MW02] R. Meshulam and A. Wigderson. Expanders from symmetric codes. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 669–677, 2002. MR2121194

[Nik05] N. Nikolov. A product decomposition for the classical quasisimple groups. `http://www.arxiv.org/abs/math.GR/0510173`, 2005.

[Nil91] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991. MR1124768 (92j:05124)

[Nil04] A. Nilli. Tight estimates for eigenvalues of regular graphs. *Electron. J. Combin.*, 11:N9, 4 pp. (electronic), 2004. MR2056091

[Nov02] T. Novikoff. Asymptotic behavior of the random 3-regular bipartite graph. Preprint, 2002.

[NZ96] Noam Nisan and David Zuckerman. Randomness is linear in space. *J. Comput. System Sci.*, 52(1):43–52, 1996. MR1375803 (97c:68050)

[Pap94] C. H. Papadimitriou. *Computational complexity.* Addison-Wesley Publishing Company, Reading, MA, 1994. MR1251285 (95f:68082)

[Pin73] M. S. Pinsker. On the complexity of a concentrator. In *7th International Telegraffic Conference*, pages 318/1–318/4, 1973.

[PU89] D. Peleg and E. Upfal. Constructing disjoint paths on expander graphs. *Combinatorica*, 9(3):289–313, 1989. MR1030383 (91h:05115)

[Rab80] M. O. Rabin. Probabilistic algorithm for testing primality. *J. Number Theory*, 12(1):128–138, 1980. MR0566880 (81f:10003)

[Rei05] O. Reingold. Undirected st-connectivity in log-space. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 376–385, 2005. MR2181639 (2006g:68118)

[Roi96] Y. Roichman. Upper bound on the characters of the symmetric groups. *Invent. Math.*, 125(3):451–485, 1996. MR1400314 (97e:20014)

[RS03] R. Raz and A. Shpilka. Lower bounds for matrix product in bounded depth circuits with arbitrary gates. *SIAM J. Comput.*, 32(2):488–513 (electronic), 2003. MR1969401 (2004d:68052)

[RSU01] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke. Design of capacity-approaching irregular low-density parity-check codes. *IEEE Trans. Inform. Theory*, 47(2):619–637, 2001. MR1820480 (2002d:94068)

[RSW04] E. Rozenman, A. Shalev, and A. Wigderson. A new family of Cayley expanders (?). In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 445–454, 2004. MR2121630 (2005j:68120)

[RTV05] O. Reingold, L. Trevisan, and S. Vadhan. Pseudorandom walks in biregular graphs and the RL vs. L problem. Technical Report TR05-022, Electronic Colloquium on Computational Complexity (ECCC), 2005. `http://www.eccc.uni-trier.de/eccc/`.

[RU] T. Richardson and R. Urbanke. Modern coding theory. Draft of the book, `http://lthcwww.epfl.ch/mct/index.php`.

[RU01] T. J. Richardson and R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Trans. Inform. Theory*, 47(2):599–618, 2001. MR1820479 (2002c:94048)

[Rud91] W. Rudin. *Functional analysis.* International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition, 1991. MR1157815 (92k:46001)

[RV05] E. Rozenman and S. Vadhan. Derandomized squaring of graphs. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM), volume 3624 of Lecture Notes in Computer Science*, pages 436–447, 2005. MR2193707

[RVW02] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Ann. of Math. (2)*, 155(1):157–187, 2002. MR1888797 (2003c:05145)

[Sar04] P. Sarnak. What is... an expander? *Notices Amer. Math. Soc.*, 51(7):762–763, 2004. MR2072849

[Ser77] J.-P. Serre. *Linear representations of finite groups.* Springer-Verlag, New York, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42. MR0450380 (56:8675)

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948. MR0026286 (10:133e)

[Sha99] Y. Shalom. Bounded generation and Kazhdan's property (T). *Inst. Hautes Études Sci. Publ. Math.*, 90:145–168 (2001), 1999. MR1813225 (2001m:22030)

[Sha04] R. Shaltiel. Recent developments in extractors. In G. Paun, G. Rozenberg, and A. Salomaa, editors, *Current trends in theoretical computer science*, volume 1: Algorithms and Complexity. World Scientific Publishing Co., 2004. MR2193911

[Sie] A. Siegel. A historical review of the isoperimetric theorem in 2-d, and its place in elementary plane geometry. http://www.cs.nyu.edu/faculty/siegel/SCIAM.pdf.

[Sim03] M. Simonovits. How to compute the volume in high dimension? *Math. Program.*, 97(1-2, Ser. B):337–374, 2003. ISMP, 2003 (Copenhagen). MR2004402 (2004j:68194)

[Sip97] M. Sipser. *Introduction to the Theory of Computation*. PWS Publishing Company, 1997.

[Spi96] D. A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1723–1731, 1996. Codes and complexity. MR1465732 (98g:94034)

[SS77] R. Solovay and V. Strassen. A fast Monte-Carlo test for primality. *SIAM J. Comput.*, 6(1):84–85, 1977. MR0429721 (55:2732)

[SS96] M. Sipser and D. A. Spielman. Expander codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1710–1722, 1996. Codes and complexity. MR1465731 (98d:94031)

[ST] H. Stark and A. Terras. Zeta functions of finite graphs and coverings, part iii. *Advances in Mathematics*, to appear.

[Sud00] M. Sudan. A crash course on coding theory. http://theory.lcs.mit.edu/~madhu/coding/ibm/, 2000.

[Sud04] M. Sudan. Probabilistically checkable proofs. In *Computational complexity theory*, volume 10 of *IAS/Park City Math. Ser.*, pages 349–389. Amer. Math. Soc., Providence, RI, 2004. MR2089735

[Tan81] R. M. Tanner. A recursive approach to low complexity codes. *IEEE Trans. Inform. Theory*, 27(5):533–547, 1981. MR0650686 (83i:94017)

[Tan84] R. M. Tanner. Explicit concentrators from generalized N-gons. *SIAM J. Algebraic Discrete Methods*, 5(3):287–293, 1984. MR0752035 (85k:68080)

[Tho03] J. Thorpe. Low-Density Parity-Check codes constructed from protographs. In *The Interplanetary Network Progress Report 42-154*, Jet Propulsion Laboratory, Pasadena, CA, pages 1–7, 2003.

[TW96] C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3):727–754, 1996. MR1385083 (97a:82055)

[UW87] E. Upfal and A. Wigderson. How to share memory in a distributed system. *J. Assoc. Comput. Mach.*, 34(1):116–127, 1987. MR0882664 (88c:68009)

[Val76] L. G. Valiant. Graph-theoretic properties in computational complexity. *J. Comput. System Sci.*, 13(3):278–285, 1976. Working papers presented at the ACM-SIGACT Symposium on the Theory of Computing (Albuquerque, NM, 1975). MR0441008 (55:13874)

[Val03] A. Valette. On the Baum-Connes assembly map for discrete groups. In *Proper group actions and the Baum-Connes conjecture*, Adv. Courses Math. CRM Barcelona, pages 79–124. Birkhäuser, Basel, 2003. With an appendix by Dan Kucerovsky. MR2027170 (2005d:19008)

[vL99] J. H. van Lint. *Introduction to coding theory*, volume 86 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 1999. MR1664228 (2000a:94001)

[vLW01] J. H. van Lint and R. M. Wilson. *A course in combinatorics*. Cambridge University Press, Cambridge, second edition, 2001. MR1871828 (2002i:05001)

[Vu05] V. H. Vu. Spectral norm of random matrices. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 423–430, 2005. MR2181644 (Review)

[Wig58] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958. MR0095527 (20:2029)

[Wig06] A. Wigderson. P, NP and mathematics – a computational complexity perspective. In *Proc. of the 2006 International Congress of Mathematicians*. Madrid, August 2006. Preprint at http://www.math.ias.edu/~avi/PUBLICATIONS/MYPAPERS/W06/W06.pdf.

[Wor99] N. C. Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999. Corrections and addenda at http://www.math.uwaterloo.ca/~nwormald/papers/regcorrs. MR1725006 (2000j:05114)

[WZ93] A. Wigderson and D. Zuckerman, Expanders that beat the eigenvalue bound: explicit contruction and applications, *Combinatorica*, 19(1):125–138, 1999. MR1722214 (2000j:05081)

[Zuc05] D. Zuckerman. Linear degree extractors and inapproximability of MAX-CLIQUE and chromatic number. Manuscript, 2005.

Index

IBM Research Laboratory, Haifa, Israel
E-mail address: shlomoh@il.ibm.com

The Hebrew University, Jerusalem, Israel
E-mail address: nati@cs.huji.ac.il

Institute of Advanced Study, Princeton, New Jersey 08544
E-mail address: avi@ias.edu