On spectral properties of graphs, and their application to clustering

Thesis submitted for the degree
Doctor of Philosophy

by

Yonatan Bilu

Submitted to the Senate of the Hebrew University
August 2004

This work was carried out under the supervision of

Professor Nati Linial

“The truth is out there, Mulder. But so are lies.”
D. K. Scully

Abstract

This work studies the connection between various properties of a graph, and its spectrum - the eigenvalues of its adjacency matrix. Of particular interest in this work are properties related to the *clustering* problem.

In most of the work we are interested in d-regular graphs. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of such a graph. It is easy to see that $\lambda_1 = d$. Denote $\lambda(G) = \max\{\lambda_2, |\lambda_n|\}$.

The first property we discuss is the so called "jumbleness" of a graph. A d-regular graph on n vertices is (n, d, α)-jumbled (Definition 2.2.1) if for every two subsets of vertices, S and T,

$$|e(S,T) - d|S||T|/n| \leq \alpha\sqrt{|S||T|},$$

where $e(S,T) = |\{(i,j) \in E \ : \ i \in S, \ j \in T\}|$. It is well known that $\alpha \leq \lambda(G)$. We show that "jumbleness" is almost the same as $\lambda(G)$. Namely, if G is (n, d, α)-jumbled then

$$\alpha \leq \lambda(G) \leq O(\alpha(\log(d/\alpha) + 1)),$$

and that the second inequality can not, in general, be improved.

The result follows from a general relation between the "discrepancy" of a matrix (Definition 2.1.1), and its spectral radius. In chapter 3 we use this result to give a new, nearly optimal, construction of expander graphs. Namely, we construct d-regular graphs with $\lambda(G) = O(\sqrt{d \log^3 d})$.

The main tool in this construction is an iterative use of the 2-lift operation. A graph $\hat{G}$ is said to be a 2-lift of G is there's a $2:1$ covering map from $\hat{G}$ to G. Such lifts correspond to a *signing* of the edges of G - a symmetric matrix obtained from the adjacency matrix of A by replacing some of the 1 entries with -1. It turns out that the spectrum of $\hat{G}$ is simply the union of that of G, and that of the signing. This characterization implies that to construct expanders, it is enough to find a signing with a small spectral radius. We show that a signing with spectral radius $O(\sqrt{d \log^3 d})$ always exits. Under a natural assumption on the structure of a graph (that it is sparse in the sense of Definition 3.3.1) it can also be found efficiently.

In chapter 4 we study regular graphs where the degree is linear, yet the second eigenvalue is bounded. One example of such graphs is $K_{n,n}$, the complete bipartite graph with n vertices on each side. We show that this is essentially the only possible example when the degree is at most $n/2$. Namely, $n/2$-regular graphs with bounded second eigenvalue are close to $K_{n/2,n/2}$, and for every $0 < \delta < \frac{1}{2}$ and c, there are only finitely many δn-regular graphs with $\lambda_2 < c$.

Our interest in graphs with bounded second eigenvalue and linear degree stems from our study of *sphericity* of graphs. A graph has sphericity k, if its vertices can be embedded in $\mathbb{R}^k$ so that the distance between the images of two vertices is < 1 iff they are adjacent. This is a special case of

what we call *monotone maps* of metric spaces. An embedding from one metric space into another is called a monotone map if it preserves the order of the pairwise distances between points.
In chapter 5 we give lower and upper bounds on the minimal dimension of normed spaces into which every metric space has a monotone map. We also discuss known results about the sphericity of graphs, and give a lower bound for it in terms of its second eigenvalue. This bound is linear for graphs with bounded second eigenvalue and linear degree.
Our study of monotone maps has to do with understanding the clustering problem. Informally, in this problem the input is a graph with distances on the edges, and the objective is to partition the vertices into subsets, "clusters", such that the distances between vertices within the same cluster are small, and those between two vertices in different clusters are big.
It seems that in any plausible formulation of the problem, it is NP-hard. Yet good heuristics exist, especially when the metric space is a low dimensional normed space, and in particular Euclidean. Thus, it is natural to ask when is it possible to find an embedding into such a space, that preserves the relation between distances.
In the same vein, it is interesting to identify non-trivial instances of this problem for which the solution can be found efficiently. We study this question in chapter 6. Specifically, we say that an instance is *stable* if small perturbations of the distances do not change the structure of the optimal clustering. Our case in point is the MaxCut problem, in which the objective is to partition a weighted graph into two clusters. We show that if the input is very stable, then a greedy algorithm finds the optimal partition. We then suggest a spectral algorithm, and show sufficient conditions under which it finds the optimal solution.
Finally, in chapter 7 we extend Hoffman's bound on the chromatic number of a graph to other graph parameters. Namely, we show that it holds for the vector chromatic number, a graph parameter introduced by Karger, Motwani and Sudan, and that a slight variation of it holds for the ψ-covering numbers suggested by Amit, Linial and Matoušek. In addition, we define a new graph parameter, so called the λ-clustering number. This is the least number k such that the graph can be partitioned into k subsets $S_1, \ldots, S_k$, where the spectral radius of $G[S_i]$ is at most λ for each i. We show that this number is bounded from below by $\frac{\lambda_1 - \lambda_n}{\lambda - \lambda_n}$.

Contents

Chapter 1

Introduction

1.1 Discrepancy vs. Spectra

A major effort in modern graph theory focuses on studying the connection between the eigenvalues of the adjacency matrix of a graph, the graph's *spectrum*, and its combinatorial properties. At first glance it might be surprising that such connections exist at all. The ongoing research in this field unravels more and more of them.
One such connection is an equivalence between the *spectral gap* in a regular graph and its *edge expansion*. Let $d = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of a d-regular graph G (we shall only discuss regular graphs here). The *spectral gap* of G, $\lambda(G)$, is $\lambda_1 - \max\{\lambda_2, |\lambda_n|\}$. Let $S \subset V$ be a set of vertices. Denote by E_S the set of edges with exactly one end point in S. The *edge expansion* of S is $|E_S|/|S|$. The *edge expansion of* G, $h(G)$, is the minimal edge expansion over all subsets S with $|S| \leq |V|/2$. Graphs with a big spectral gap, or a big edge expansion, are very useful in computer science (cf. [48]). In fact, it is well known (cf. [9]) that the two properties are equivalent after a fashion:

$$h(G) \geq \frac{\lambda(G)}{2} \tag{1.1}$$

$$\lambda(G) \geq \frac{h(G)^2}{d^2}. \tag{1.2}$$

Spectra allows us to compare how similar a given d-regular graph is to a random d-regular graph. For disjoint $S, T \subset V$ the *discrepancy* between them is:

$$\Delta_{S,T}(A) = \frac{|e(S,T) - d|S||T|/n|}{\sqrt{|S||T|}},$$

where $e(S,T)$ is the number of edges with one end point in S and one in T. Note that $d|S||T|/n$ is the expected number of edges in a random d-regular graph. The discrepancy of a graph, $\Delta(G)$, is the maximal discrepancy over all such pairs. A well known property of d regular graphs (cf. [9])

is:

$$\Delta(G) \leq d - \lambda(G)$$
$$h(G) \geq \frac{d - \Delta(G)}{2}$$

where the first inequality is sometimes called *The Expander Mixing Lemma*.
In Chapter 2 we show that $\Delta(G)$ also gives an upper bound on $d - \lambda(G)$, which is surprisingly tight in comparison with estimates derived from edge expansion. Namely:

$$d - \lambda(G) = O(\Delta(G)(\log(d/\Delta(G)) + 1)). \tag{1.3}$$

The proof relies on a more general result, connecting the spectral radius of a matrix with a similar notion of discrepancy.
It known that inequality 1.2 is actually tight, in the sense that there are graphs for which the ratio between the sides is a small constant. Similarly, we prove that inequality 1.3 is tight up to a constant factor.
Turning our viewpoint, inequality 1.3 says that if $\lambda(G)$ is small, then there are $S, T \subset V$ with a large discrepancy - either a lot of edges, or very few. In fact, from the proof it follows that if the least eigenvalue is close to $-d$ then the former holds, and if the second largest eigenvalue is close to d then the latter. Furthermore, such S and T can be found efficiently. We therefore speculate that this lemma might prove useful in clustering algorithms, as identifying such subsets is a natural problem.

1.2 Constructing expander graphs from 2 lifts

Constructing large regular graphs with a large spectral gap is not an easy task. The first to do so was Margulis, in [51]. In a celebrated result, Margulis and Lubotzky, Phillips and Sarnak [49, 52] construct arbitrarily large d-regular (for d a prime power plus one), with $\lambda(G) \geq d - 2\sqrt{d-1}$. A matching upper bound, known as the Alon-Boppana bound, shows that this is optimal (cf. [57]). Such optimal construction were called *Ramanujan graphs* by Lubotzky, Phillips and Sarnak, as their construction is based on deep results concerning certain number-theoretic conjectures by Ramanujan.
In Chapter 3 we suggest a construction that is nearly optimal, and uses only elementary algebra. The building block of the construction is the 2-*lift* operation on graphs.
A *signing* of a graph $G = (V, E)$ is a function $s : E \to \{-1, 1\}$. A signing defines a graph $\hat{G}$, called a *2-lift of* G with vertex set $V(G) \times \{-1, 1\}$. The vertices (u, x) and (v, y) are adjacent iff $(u, v) \in E(G)$, and $x \cdot y = s(u, v)$. The corresponding *signed adjacency matrix* $A_{G,s}$ is a symmetric $\{-1, 0, 1\}$-matrix, with $(A_{G,s})_{u,v} = s(u, v)$ if $(u, v) \in E$, and 0 otherwise.
It is an easy fact that the spectrum of $\hat{G}$ is simply the union of that of G and that of $A_{G,s}$. Consider the following scheme for constructing Ramanujan graphs. Start with $G_0 = K_{d+1}$, the complete graph on $d+1$ vertices or any other small d-regular Ramanujan graph. (Note that small Ramanujan graphs are easy to come by.) Consider the following conjecture:

Conjecture. *Every Ramanujan d-regular graph G has a signing s with spectral radius $\rho(A_{G,s}) \leq 2\sqrt{d-1}$.*

Assuming this to be true, define successively G_i as a 2-lift of G_{i-1}, obtained from a signing s of G_{i-1} that satisfies the conjecture. These G_i constitute an infinite family of d-regular Ramanujan graphs.
In fact, we make the even more daring:

Conjecture. Every *d-regular graph G has a signing s with spectral radius $\rho(A_{G,s}) \leq 2\sqrt{d-1}$.*

We note that by the Alon-Boppana bound, the conjecture can not, in general, be improved. In Chapter 3 we establish a somewhat weaker result:

Theorem. *(*3.3.1*) Every graph G of maximal degree d has a signing s with spectral radius $\rho(A_{G,s}) = O(\sqrt{d \cdot \log^3 d})$.*

Moreover, we show how to find such a signing efficiently. The proof of the theorem relies on the intimate connection between discrepancy and eigenvalues, mentioned in the previous section.
We also give examples of families of graphs where we know how to find an s such that $\rho(A_{G,s}) \leq 2\sqrt{d-1}$, and show empirically that most randomized algorithms we came up with succeed in finding such a signing on random graphs. Finally, we discuss *group lifts*, of which 2-lifts are a special case, and characterize their spectrum.

1.3 Graphs with bounded second eigenvalue

The Alon-Boppana bound says that the second eigenvalue of a d-regular graph is at least $2\sqrt{d-1} - o(1)$, where the $o(1)$ goes to 0 as n/d grows (n denotes throughout the number of vertices). For graphs of linear degree the bound does not give anything. In Chapter 4 we ask: are there such graphs with bounded second eigenvalue (independent of d and n), and if so, what do they look like?
The second eigenvalue of $K_{n,n}$, the complete bipartite graph with n vertices on each side, is 0. Clearly, it is regular, of linear degree, and the second eigenvalue is bounded. We show that for linear degree at most $n/2$, this is essentially the only example: $n/2$-regular graphs with bounded second eigenvalue are close to $K_{n/2,n/2}$, and for every $0 < \delta < \frac{1}{2}$ and c, there are only finitely many δn-regular graphs with $\lambda_2 < c$.
By close to $K_{n/2,n/2}$ we mean that there is a partition of the vertices $(S, [n]\backslash S)$, such that $|S| = n/2$, the number of edges between S and $[n]\backslash S$ is $n^2/4 - o(n^2)$ and the number of edges with both end points in S or both end points in $[n]\backslash S$ is $o(n^2)$.
In other words, if the two first eigenvalues of an $n/2$-regular graph are similar to those of $K_{n/2,n/2}$, the graph itself is similar to $K_{n/2,n/2}$. If the spectrum is even more similar, namely the second smallest eigenvalue is also bounded, a stronger of notion can be proved. Rather than $o(n^2)$, the difference between the graph and $K_{n/2,n/2}$ is $O(n^{3/2})$ edges.

1.4 Monotone maps, sphericity, and inverting the KL-divergence

The problem of clustering arises is many scientific disciplines. The input is a set of points, with a measure of similarity or distance between them. The objective is to find a structure within these points. *Clustering* is one possible way to define this structure. A clustering of the points is their partition into subsets, or *clusters*, such that close points are within the same cluster, and remote points are in different clusters.
Simple algorithms for doing so rely only on the order of the distances. For example, one version of the neighbor-joining algorithm is as follows. It starts with each point as a cluster to itself, and proceeds by merging the two closest clusters until some terminating criteria is reached (e.g., number of clusters). The distance between two clusters is taken as the minimum over all pairs of points, one from each cluster.
The clustering problem tends to be easier when the points reside in a low dimensional Euclidean space. One approach towards clustering is therefore to try and embed the points into such a space, without losing too much of the structure of the original distances. In particular, if we only care about the order of the distances, this embedding is what we call a *monotone map* - it preserves the order of the distances between points.
In Chapter 5 we give lower and upper bounds on the dimension of Euclidean space into which monotone maps exist. Specifically, we show that $n-1$ dimensions are enough to embed any metric space on n points, and that "almost all" metric spaces do require linear dimension. We also give bounds for other normed spaces.
Looking for a specific example for which linear dimension is required, we discuss metrics arising from graphs and *sphericity* of graphs. An embedding of the vertices of a graph in Euclidean space $f : V \to \mathbb{R}^m$ is called a spherical embedding if $||f(u) - f(v)|| < 1$ iff $(u, v) \in E$. It is easy to see that any properly scaled monotone map of a graph metric is a spherical embedding of the underlying graph. The sphericity of a graph G, $Sph(G)$, is the minimal dimension into which it can be embedded this way. Reiterman, Rödl and Šiňajová ([61]) show that the sphericity of $K_{n,n}$ is n. In particular, this gives an example of a metric for which any monotone map requires linear dimension.
We give a lower bound on the sphericity of a graph in terms of its second eigenvalue. Namely, that for $0 < \delta \leq \frac{1}{2}$, for any n-vertex δn-regular graph G, with bounded diameter, $Sph(G) = \Omega(\frac{n}{\lambda_2+1})$. Unfortunately, this turns out to be too weak to derive new graphs with linear sphericity, by the results of Chapter 4.
We also discuss the related notion of the *margin* of an embedding, and the sphericity of quasi-random graphs. We show that although the sphericity of a graph in $G(n, \frac{1}{2})$ is linear, there are families of quasi-random graphs with sub-linear sphericity.
Finally, we consider a related problem of *soft clustering*. In this variation, rather than partitioning the points into clusters, the objective is to assign to each point a distribution over the clusters. As before, nearby points should have similar distributions, and faraway points should have little correlation. Let f be some distance measure on distributions. Given a metric space on n points with distance function d, the soft clustering problem is to find $D_1, \ldots, D_n$, such that $f(D_i, D_j) = d(i, j)$. A natural question is - for what metric spaces do such distributions exist? At the end of Chapter 5 we show that when f is the KL-divergence between distributions, there is a solution to the soft clustering problem for any metric space.

1.5 Efficient algorithms for stable problem instances

Many computational problems that arise in practice are particular instances of NP-hard problems. We have seen one such example - the clustering problem. As mentioned before, essentially any concrete formulation of the problem is known to be NP-hard. Should this deter us from trying to solve it?
Since the problem is NP-hard, we believe that for any algorithm there exist problem instances for which it will not produce the correct solution in polynomial time. However, do such instances actually arise in practice? Are they actually interesting to solve?
In Chapter 6 we focus on the Max-Cut problem, the case of the clustering problem when two clusters are sought. Let G be a graph on n vertices, W a weight function on its edges and $(S, [n]\backslash S)$ the maximal cut in it. Fix $\gamma \geq 1$. Let W' be a weight function on the edges of G, such that for every edge e, $W(e) \leq W'(e) \leq \gamma \cdot W(e)$. We say that (G, W) is γ-*stable* if for all such W', $(S, [n]\backslash S)$ is also the maximal cut in (G, W'). Thinking of W' as a noisy version of W, stability means that the solution is resilient to noise.
We also study a related property, which we call *distinctness*. Roughly, it measures how much the weight of a cut decreases as a function of its distance from the maximal one.
We ask the following questions: Is there some value of stability γ_0 for which Max-Cut can be solved efficiently? If so, how small can γ_0 be taken? How well do known approximation algorithms and heuristics perform, as a function of stability? What about distinctness?
We show a greedy algorithm that finds the maximal cut when the input is $\sqrt{Dn}$-stable, where D is the maximal degree in the graph. This algorithm is generalized at the end of the chapter to other problems.
We show a spectral algorithm that finds the maximal cut for instances where the least eigenvector of the weighted adjacency matrix is balanced. Namely, let u be an eigenvector corresponding to the least eigenvalue of W, if the input is $\frac{\max_{u_i u_j < 0} -u_i u_j}{\min_{u_i u_j \geq 0} u_i u_j}$-stable, then partitioning the vertices according to the signs of the elements in u gives the maximal cut.
For regular unweighted graphs we define:

Defintion. *(*6.2.3*) Let $C \subset E$ be the set of edges of a cut in a d-regular graph G, and $T \subset V$. The* relative net contribution *of T to the cut is:*

$$\rho_C(T) = \frac{|\{e \in (T \times \bar{T}) \cap C\}| - |\{e \in (T \times \bar{T}) \cap (E\backslash C)\}|}{d \min\{|T|, n - |T|\}}.$$

*Let $C^* \subset E$ be the set of edges in a maximal cut of G. We say that it is ρ-*distinct, *if for all $T \subset V$, $\rho_{C^*}(T) \geq \rho$.*

We show:

Theorem. *(*6.2.1*) Let $C^* \subset E$ be the set of edges in a maximal cut of d-regular graph G. Assume that this cut is ρ - distinct, and, furthermore, that for all $v \in V$, $\rho_{C^*}(\{v\}) \geq \alpha$. If*

$$\alpha + \frac{\rho^2}{10} > 1$$

then the maximal cut can be found efficiently.

Finally, we survey results from computer simulations on random graphs, testing the stability required for various algorithms to find the maximal cut.

1.6 Extensions of Hoffman's bound

The chromatic number of a graph G, χ, is the minimal k such that G can be partitioned into k independent sets. In the context of clustering, thinking of edges as denoting two points which are far apart, we are looking for a partition into as few clusters as possible, such that all points within a cluster a close to each other.
One of the first non-trivial lower bounds on the chromatic number is that of Hoffman [39]: let λ_1 and λ_n be the largest and least eigenvalues of G, then $\chi(G) \geq 1 - \lambda_1/\lambda_n$.
Karger, Motwani and Sudan [44] define a quadratic programming relaxation of the chromatic number, χ_v, called the *vector chromatic number*. They show that $\chi \geq \chi_v$. Amit, Linial and Matoušek [12] define a set of graph parameters, so called ψ*-covering numbers*, and show that (under some assumptions) they take values in $[\sqrt{\chi}, \chi]$. In chapter 7 we extend Hoffman's bound to the vector chromatic number, and to ψ-covering numbers. Specifically, we show that $\chi_v(G) \geq 1 - \lambda_1/\lambda_n$ (Theorem 7.1.2). The bound for ψ-covering numbers is somewhat involved and requires some further definitions (see Theorem 7.1.3).
We also suggest a new relaxation of the chromatic number. Rather than requiring that all subsets in the partition be independent, we ask that they have spectral radius at most λ. We call the minimal k for which there exists a partition into k such clusters the λ*-clustering number* of a graph. In particular, the 0-clustering number in exactly χ. We show that this parameter is at least $\frac{\lambda_1 - \lambda_n}{\lambda - \lambda_n}$ (Theorem 7.1.4).

Chapter 2

Discrepancy vs. Eigenvalues

2.1 Discrepancy vs. spectral radius

2.1.1 Introduction

In this chapter we are interested in understanding the relation between the *discrepancy* of a real symmetric matrix, which we shall define shortly, and its spectral radius. Recall that the spectral radius of such a matrix is the largest absolute value of its eigenvalue. By the Rayleigh-Ritz characterization, if A be a $n \times n$ real symmetric matrix, then its spectral radius is

$$\rho(A) = \max_{x \in \mathbb{R}^n} \frac{xAx}{||x||_2^2} = \max_{x,y \in \mathbb{R}^n} \frac{xAy}{||x||_2||y||_2}.$$

Definition 2.1.1. *Let A be a $n \times n$ real symmetric matrix, and $S, T \in [n]$. Define the discrepancy between S and T:*

$$\Delta_{S,T}(A) = \frac{\sum_{i \in S, j \in T} A_{i,j}}{\sqrt{|S||T|}}.$$

Define the discrepancy of A *as:*

$$\Delta(A) = \max_{S,T \subset [n]} \Delta_{S,T}(A) = \max_{x,y \in \{0,1\}^n} \frac{xAy}{||x||_2||y||_2}.$$

Intuitively, we think of the entries of A as being random variables with expectation 0, and measure how much the sum of the entries in the sub-matrix defined by S and T deviate from the expected value of 0. The normalization by $\sqrt{|S||T|}$ should be thought of as measuring this quantity in units proportional to standard deviation. The discrepancy of a matrix is the largest deviation among all pairs of subsets.

Clearly $\Delta(A) \leq \rho(A)$. But can the gap between the two values by arbitrarily large? A related result appears in the work of Kahn and Szemeredi [32]. In their work on the second eigenvalue of a random d-regular graph they observe that if $E \subset S^{n-1}$ in an ϵ-net on the $n-1$ dimensional

sphere, then a bound on $\max_{x,y \in E} xAy$ yields a bound on $\rho(A)$. However, discrepancy has to do with points in $\{0,1\}^n$ - certainly not an ϵ-net.

Nonetheless, assuming a bound on the l_1 norm of the rows of A, leads to a surprisingly tight bound on its spectral radius, in terms of its discrepancy.

2.1.2 The main lemma

The key lemma for the results we describe in this chapter, and in chapter 3 is the following:

Lemma 2.1.1. *Let A be an $n \times n$ real symmetric matrix such that the l_1 norm of each row in A is at most d. Assume that for any two vectors, $u, v \in \{0,1\}^n$, with $supp(u) \cap supp(v) = \emptyset$:*

$$\frac{|uAv|}{||u||||v||} \leq \alpha,$$

and that all diagonal entries of A are, in absolute value, $O(\alpha(\log(d/\alpha)+1))$. Then the spectral radius of A is $O(\alpha(\log(d/\alpha)+1))$.

Note that $\alpha \leq \Delta(A)$, as we define it as the maximal discrepancy among pairs of *disjoint* subsets.

Proof: For simplicity, let us first assume that all diagonal entries of A are zeros. Note that our assumptions imply that for any $u \in \{0,1\}^n$,

$$\frac{|uAu|}{||u||^2} \leq 2\alpha :$$

For any $u_1, u_2 \in \{0,1\}^n$ such $S(u_1) \cap S(u_2) = \emptyset$ we have that

$$|u_1 A u_2| \leq \alpha ||u_1|| ||u_2||. \tag{2.1}$$

Let $u \in \{0,1\}^n$, and denote $k = |S(u)|$. Set $K = \binom{k}{k/2}$. Summing up inequality (2.1) over all subsets of $S(u)$ of size $k/2$, we have that:

$$\sum_{u_1 : S(u_1) \subset S(u), |S(u_1)| = k/2} u_1 A(u - u_1) \leq K\alpha k/2.$$

For each $i \neq j \in S(u)$, $a_{i,j}$ is added up $\binom{k-2}{k/2-1}$ times in the sum on the LHS, hence (since diagonal entries are zero):

$$\binom{k-2}{k/2-1} uAu \leq K\alpha k/2,$$

or:

$$uAu \leq \alpha 2k.$$

Next, it follows that for any $u, v \in \{-1,0,1\}^n$, such that $S(u) = S(v)$, or $S(u) \cap S(v) = \emptyset$:

$$\frac{|uAv|}{||u||||v||} \leq 4\alpha.$$

Fix $u, v \in \{-1, 0, 1\}^n$. Denote $u = u^+ - u^-$ and $v = v^+ - v^-$, where $u^+, u^-, v^+, v^- \in \{0, 1\}^n$, and $S(u^+) \cap S(u^-) = S(v^+) \cap (v^-) = \emptyset$.

$$\begin{aligned}
&|uAv| = |(u^+ - u^-)A(v^+ - v^-)| \\
&\leq 2\alpha(||u^+||||v^+|| + ||u^+||||v^-|| + ||u^-||||v^+|| + ||u^-||||v^-||) \\
&\leq 4\alpha\sqrt{||u^+||^2||v^+||^2 + ||u^+||^2||v^-||^2 + ||u^-||^2||v^+||^2 + ||u^-||^2||v^-||^2} \\
&= 4\alpha\sqrt{(||u^+||^2 + ||u^-||^2)(||v^+||^2 + ||v^-||^2)} = 4\alpha||u||||v||.
\end{aligned}$$

The first inequality follows from our assumption on vectors in $\{0, 1\}^n$, and the second from the l_2 to l_1 norm ratio.

Fix $x \in \mathbb{R}^n$. We need to show that $\frac{|xAx|}{||x||^2} = O(\alpha \log(d/\alpha))$. By losing only a multiplicative factor of 2, we may assume that the absolute value of every non-zero entry in x is a negative powers of 2: Clearly we may assume that $||x||_\infty < \frac{1}{2}$. To bound the effect of rounding the coordinates, denote $x_i = \pm(1 + \delta_i)2^{t_i}$, with $0 \leq \delta_i \leq 1$ and $t_i < -1$, an integer. Now round x to a vector x' by choosing the value of x'_i to be $\text{sign}(x_i) \cdot 2^{t_i+1}$ with probability δ_i and $\text{sign}(x_i) \cdot 2^{t_i}$ with probability $1 - \delta_i$. The expectation of x'_i is x_i. As the coordinates of x' are chosen independently, and the diagonal entries of A are 0's, the expectation of $x'Ax'$ is xAx. Thus, there's a rounding, x', of x, such that $|xAx| \leq |x'Ax'|$. Clearly $||x'||^2 \leq 2||x||^2$, so $\frac{|xAx|}{||x||^2} \leq 2\frac{|x'Ax'|}{||x'||^2}$.

Denote $S_i = \{j : x_j = \pm 2^{-i}\}$, $s_i = |S_i|$. Denote by k the maximal index i such that $s_i > 0$. Denote by x^i the sign vector of x restricted to S_i, that is, the vector whose j'th coordinate is the sign of x_j if $j \in S_i$, and zero otherwise. By our assumptions, for all $1 \leq i \leq j \leq k$:

$$|x^i A x^j| \leq \alpha\sqrt{s_i s_j}. \tag{2.2}$$

Also, since the l_1 norm of each row is at most d, for all $1 \leq i \leq k$:

$$\sum_j |x^i A x^j| \leq d s_i. \tag{2.3}$$

We wish to bound:

$$\frac{|xAx|}{||x||^2} \leq \frac{\sum_{i,j=1}^k |x^i A x^j| 2^{-(i+j)}}{\sum_i 2^{-2i} s_i}. \tag{2.4}$$

Denote $\gamma = \log(d/\alpha)$, $q_i = s_i 2^{-2i}$ and $Q = \sum_i q_i$. Add up inequalities (2.2) and (2.3) as follows. For $i = j$ multiply inequality (2.2) by 2^{-2i}. When $i < j \leq i + \gamma$ multiply it by $2^{-(i+j)+1}$. Multiply inequality (2.3) by $2^{-(2i+\gamma)}$. (We ignore inequalities (2.2) when $j > i + \gamma$.)

We get that:

$$\begin{aligned}
&\sum_i 2^{-2i}|x^iAx^i| + \sum_i \sum_{i<j\le i+\gamma} 2^{-(i+j)+1}|x^iAx^j| + \sum_i 2^{-(2i+\gamma)} \sum_j |x^iAx^j| \\
&\le \sum_i \alpha q_i + \sum_i \sum_{i<j<i+\gamma} 2\alpha\sqrt{q_i q_j} + \sum_i 2^{-\gamma} d \cdot q_i \\
&\le \alpha \sum_i q_i + \alpha \sum_i \sum_{i<j<i+\gamma} (q_i + q_j) + 2^{-\gamma} d \sum_i \cdot q_i \\
&< (2^{-\gamma} d + 2\gamma\alpha) \sum_i q_i = (\alpha + \alpha \log(d/\alpha))Q.
\end{aligned}$$

Note that the denominator in (2.4) is Q, so to prove the lemma it's enough to show that the numerator,

$$\sum_{i<j} 2^{-(i+j+1)}|x^iAx^j| + \sum_i 2^{-2i}|x^iAx^i|, \tag{2.5}$$

is bounded by

$$\sum_i 2^{-2i}|x^iAx^i| + \sum_i \sum_{i\le j\le i+\gamma} 2^{-(i+j)+1}|x^iAx^j| + \sum_i 2^{-(2i+\gamma)} \sum_j |x^iAx^j|. \tag{2.6}$$

Indeed, let us compare the coefficients of the terms $|x^iAx^j|$ in both expressions (Since $|x^iAx^j| = |x^jAx^i|$, it's enough to consider $i \le j$). For $i = j$ this coefficient is 2^{-2i} in (2.5), and $2^{-2i} + 2^{-(2i+\gamma)}$ in (2.6). For $i < j \le i + \gamma$, it is $2^{-(i+j)+1}$ in (2.5), and in (2.6) it is $2^{-(i+j)+1} + 2^{-(2i+\gamma)} + 2^{-(2j+\gamma)}$. For $j > i + \gamma$, in (2.5) the coefficient is again $2^{-(i+j)+1}$. In (2.6) it is:

$$2^{-(2i+\gamma)} + 2^{-(2j+\gamma)} > 2^{-(2i+\gamma)} \ge 2^{-(i+j)+1}.$$

It remains to show that the lemma holds when the diagonal entries of A are not zeros, but $O(\alpha(log(d/\alpha)+1))$ in absolute value. Denote $B = A - D$, with D the matrix having the entries of A on the diagonal, and zero elsewhere. We have that for any two vectors, $u, v \in \{0,1\}^n$, with $supp(u) \cap supp(v) = \emptyset$:

$$\frac{|uAv|}{||u||||v||} \le \alpha.$$

For such vectors, $uBv = uAv$, so by applying the lemma to B, we get that its spectral radius is $O(\alpha(log(d/\alpha) + 1))$. By the assumption on the diagonal entries of A, the spectral radius of D is $O(\alpha(log(d/\alpha) + 1))$ as well. The spectral radius of B is at most the sum of these bounds - also $O(\alpha(log(d/\alpha) + 1))$. ∎

2.1.3 Tightness

Lemma 2.1.1 is tight up to constant factors. To see this, consider the n-dimensional vector x whose i'th entry is $1/\sqrt{i}$. Let A be the outer product of x with itself, that is, the matrix whose

(i,j)'th entry is $1/\sqrt{i \cdot j}$. Clearly x is an eigenvector of A corresponding to the eigenvalue $||x||^2 = \Theta(\log(n))$. Also, the sum of each row in A is $O(\sqrt{n})$. To prove that the lemma is essentially tight, we need to show that $max_{u,v \in \{0,1\}^n} \frac{uAv}{||u||||v||}$ is constant. Indeed, fix $k, l \in [n]$. Let $u, v \in \{0,1\}^n$ be such that $||u|| = k$ and $||v|| = l$. As the entries of A are decreasing along the rows and the columns, uAv is maximized for such vectors when their support is the first k and l coordinates. For these optimal vectors, $uAv = \Theta(\sqrt{k \cdot l})$. Thus,

$$max_{u,v \in \{0,1\}^n} \frac{uAv}{||u||||v||} = \Theta(1).$$

This example was independently discovered by Bollobás and Nikiforov [14] who prove a result in the same vein as Lemma 2.1.1 (they do not assume a bound on the l_1 norm of the rows, and hence get a bound of $O(\alpha \log n)$, rather than $O(\alpha(log(d/\alpha) + 1))$).

In the applications of the lemma in this work, the matrices will actually have all row sums *equal*. Thus, one might hope that under this stronger assumption, a better bound on the spectral radius might be given. Although this hope is not proscribed by the example above, it would nonetheless be in vain. This is shown by the more elaborate example, given in section 2.2.3.

2.1.4 Finding the proof: LP-duality

As the reader might have guessed, the proof for Lemma 2.1.1 was discovered by formulating the problem as a linear program. Define $\Delta_{i,j} = |x^i A x^j|$. Our assumptions translate to:

$$\forall 1 \leq i \leq j \leq k : |\Delta_{i,j}| \leq \alpha \sqrt{s_i s_j},$$
$$\forall 1 \leq i \leq k : \sum_j |\Delta_{i,j}| \leq d s_i.$$

We want to deduce an upper bound on $|xAx|$. In other words, we are asking, under these constraints, how big

$$\frac{|xAx|}{||x||^2} \leq \frac{\sum_{i,j=1}^k \Delta_{i,j} 2^{-(i+j)}}{\sum_i 2^{-2i} s_i}$$

can be.

The dual program is to minimize:

$$\frac{\alpha \sum_{i \leq j} b_{i,j} \sqrt{s_i s_j} + d \sum_i c_i s_i}{\sum_i 2^{-2i} s_i}$$

under the constraints:

$$\begin{aligned}
\forall 1 \leq i < j \leq k, &\quad b_{i,j} + c_i + c_j \geq 2^{-(i+j)+1} \\
\forall 1 \leq i \leq k, &\quad b_{i,i} + c_i \geq 2^{-2i} \\
\forall 1 \leq i \leq j \leq k, &\quad b_{i,j} \geq 0 \\
\forall 1 \leq i \leq k, &\quad c_i \geq 0
\end{aligned}$$

The following choice of b's and c's satisfies the constraints, and gives the desired bound. These indeed appear in the proof above:

$$\begin{aligned} \forall 1 \leq i < j \leq k, j < i + \gamma, \quad & b_{i,j} = 2^{-(i+j)+1} \\ \forall 1 \leq i < j \leq k, j \geq i + \gamma, \quad & b_{i,j} = 0 \\ \forall 1 \leq i \leq k, \quad & b_i = 2^{-2i} \\ \forall 1 \leq i \leq k, \quad & c_i = 2^{-2i-\gamma+1} \end{aligned}$$

2.2 A converse to the Expander Mixing Lemma

A useful property of expander graphs is the so-called Expander Mixing Lemma. Roughly, this lemma states that the number of edges between two subsets of vertices in an expander graph is what is expected in a random graph, up to an additive error that depends on the second eigenvalue. In this section we conclude from Lemma 2.1.1 a converse to it.

2.2.1 Introduction and definitions

An d-regular graph is called a λ-*expander,* if all its eigenvalues but the first are in $[-\lambda, \lambda]$. Such graphs are interesting when d is fixed, $\lambda < d$, and the number of vertices in the graph tends to infinity. Applications of such graphs in computer science and discrete mathematics are many, see for example [48] for a survey.
A d-regular graph on n vertices is an (n, d, c)-edge expander if every set of vertices, W, of size at most $n/2$, has at least $c|W|$ edges emanating from it.
The two notions are closely related. An (n, d, λ)-expander is also an $(n, d, \frac{d-\lambda}{2})$-edge expander (cf. [9]). Conversely, an (n, d, c)-edge expander is also an $(n, d, d - \frac{c^2}{2d})$-expander[1]. Though the two notions of expansion are qualitatively equivalent, they are far from being quantitatively the same. While algebraic expansion yields good bounds on edge expansion, the reverse implications are very weak. It is also known that this is not just a failure of the proofs - indeed this estimate is nearly tight [2]. Is there, we ask, another combinatorial property that is equivalent to spectral gaps? We next answer this question.
A third notion of expansion is what we call here, following the terminology of Thomason in [66], *jumbleness*.

Definition 2.2.1. *For two subsets of vertices, S and T, denote*

$$e(S, T) = |\{(i, j) : i \in S,\ j \in T,\ (i, j) \in E\}|.$$

[1] A related result, showing that *vertex* expansion implies spectral gap appears in [3]. The implication from edge expansion is easier, and the proof we are aware of is also due to Noga Alon.

A d-regular graph G on n vertices is (n, d, α)-jumbled, *if for every two subsets of vertices, S and T,*

$$|e(S,T) - d|S||T|/n| \leq \alpha\sqrt{|S||T|}.$$

An (n, d, λ)-expander is (n, d, λ)-jumbled. This is a fairly easy observation, known as the *Expander Mixing Lemma*. To see this Let G be an (n, d, λ)-expander, and let A be its adjacency matrix. Consider the matrix $B = A - \frac{d}{n}J$, where J is the all ones matrix. Observe that J is a rank 1 matrix, and that the non-zero eigenvalue is d, which corresponds to the all ones eigenvector. Hence, the eigenvectors of B are the same as those of A, and they correspond to the same eigenvalues, with the exception of the all ones vector, which corresponds to 0. In other words, as A is an (n, d, λ)-expander, $\rho(B) = \lambda$.
By the Rayleigh-Ritz characterization of the spectral radius, we have that

$$\lambda = \max_{x,y\in\mathbb{R}^n} \frac{|xBy|}{||x||_2||y||_2} \geq \max_{x,y\in\{0,1\}^n} \frac{|x(A - \frac{d}{n}J)y|}{||x||_2||y||_2} = \max_{S,T\subset[n]} \frac{|e(S,T) - d|S||T|/n|}{\sqrt{|S||T|}}.$$

An (n, d, λ)-jumbled graph is an $(n, d, \frac{d-\lambda}{2})$-edge expander. Again, this follows easily from the definition. Consider a subset of vertices S, $|S| \leq \frac{n}{2}$. As the graph is (n, d, λ)-jumbled, we have that $|e(S, S) - d|S|^2/n| \leq \lambda|S|$. Hence, $\frac{e(S,V\setminus S)}{|S|} \geq \frac{d-\lambda}{2}$.

As mentioned above, edge expansion implies spectral expansion, but in a weak way. As this brings us a full circle between these three notions, we can also deduce that edge expansion implies jumbleness, and that jumbleness implies spectral expansion. However, we are constrained by the weakest link in this chain. Namely, the implications are that an (n, d, c)-edge expander is $(n, d, d - \frac{c^2}{2d})$-jumbled, and that an (n, d, α)-jumbled graph is an $(n, d, d - \frac{(d-\alpha))^2}{8d})$-expander. The example in [2] shows that the first implication, though weak, is essentially the best we can hope for. However, in the next subsection we show that we can do much better than the latter.

2.2.2 The Lemma Mixing Expander

The following lemma states the promised converse to the Expander Mixing Lemma:

Lemma 2.2.1. *Let G be a d-regular graph on n vertices. Suppose that for any $S, T \subset V(G)$, with $S \cap T = \emptyset$*

$$|e(S,T) - \frac{|S||T|d}{n}| \leq \alpha\sqrt{|S||T|}.$$

Then all but the largest eigenvalue of G are bounded, in absolute value, by $O(\alpha(1 + \log(d/\alpha)))$.

Note 2.2.1. *In particular, this means that for a d-regular graph G, $\lambda(G)$ is a $\log d$ approximation of the "jumbleness" parameter of the graph.*

Proof: Let A be the adjacency matrix of G. Denote $B = A - \frac{d}{n}J$, where J is the all ones $n \times n$ matrix. Clearly B is symmetric, and the sum of the absolute value of the entries in each row is at most $2d$. The first eigenvalue of A is d. The other eigenvalues of A are also eigenvalues of B. Thus, for the corollary to follow from Lemma 2.1.1 it suffices to show that for any two vectors, $u, v \in \{0,1\}^n$:

$$|uBv| = |u\frac{d}{n}Jv - uAv| \leq \alpha||u||||v||.$$

This is exactly the hypothesis for the sets $S(u)$ and $S(v)$. ∎

Note 2.2.2. *For a* bipartite *d-regular graph $G = (L, R; E)$ on n vertices, if for any $S \subset R, T \subset L$, with $S \cap T = \emptyset$*

$$|e(S,T) - \frac{|S||T|d}{2n}| \leq \alpha\sqrt{|S||T|}$$

Then all but the largest eigenvalue of G are bounded, in absolute value, by $O(\alpha(1 + \log(d/\alpha)))$. The proof is essentially identical to the one above, taking $B = A - d/nC$, instead of $B = A - d/nJ$, where $C_{i,j}$ is 0 if i, j are on the same side, and 2 otherwise.

2.2.3 Tightness

Lemma 2.2.1 is actually tight, up to a constant multiplicative factor:

Theorem 2.2.1. *For any large enough d, and $7\sqrt{d} < \alpha < d$, there exist infinitely many (d, α)-jumbled graphs with second eigenvalue $\Omega(\alpha(\log(d/\alpha) + 1))$.*

It will be useful to extend Definition 2.2.1 to unbalanced bipartite graphs:

Definition 2.2.2. *A bipartite graph $G = (U, V, E)$ is* (c, d, α)-jumbled*, if the vertices in U have degree c, those in V have degree d, and for every two subsets of vertices, $A \subset U$ and $B \subset V$,*

$$|e(A,B) - d|A||B|/|U|| \leq \alpha\sqrt{|A||B|}.$$

We note that such bipartite graphs exit:

Lemma 2.2.2. *For $c|d$ and $\alpha = 2\sqrt{d}$, there exist (c, d, α)-jumbled graphs.*

Proof: Let $G' = (U', V', E')$ be a c-regular Ramanujan bipartite graph, such that $|U'| = |V'| = n$. Let $G = (U, V, E)$ be a bipartite graph obtained from G' by partitioning V' into subsets of size d/c, and merging each subset into a vertex, keeping all edges (so this is a multi-graph) .
Let $A \subset U$ and $B \subset V$. Let $A' = A$, and B' the largest set whose merger gives B. Clearly $e(A,B) = e(A',B')$, $|B'| = d/c|B|$. As G' is Ramanujan, by the expander mixing lemma

$$|e(A',B') - c|A'||B'|/n| \leq 2\sqrt{c|A'||B'|},$$

or:

$$|e(A,B) - d|A||B|/|U|| \leq 2\sqrt{d|A||B|}.$$

∎

We shall need the following inequality, that can be easily proven by induction:

Lemma 2.2.3. *For $i = 0, \ldots, t$ let a_i be numbers in $[0, 2^{2i} \cdot N]$, for some $N > 0$. Then*

$$(\sum a_i 2^{-i})^2 \leq 3N \sum a_i.$$

Proof: (Theorem 2.2.1) Fix d, and $\sqrt{d} < \Delta < d$. Set $t = \frac{1}{2}\log(\frac{3d}{4\Delta})$, $\tau = \sum_{i=0}^{t} 4^i = \frac{d}{\Delta} < \Delta$. Let N be some large number and $n = \tau N$ (this will be the number of vertices). For $i, j = 0, \ldots, t$ set $d_{i,j} = \frac{d}{\tau}2^{2j} + \Delta 2^{j-i}$ when $i, j < t$ or $i = j = t$, and $d_{i,j} = \frac{d}{\tau}2^{2j} - \Delta 2^{j-i}$ otherwise. Note that in this case $\frac{d}{\tau}2^{2j} - \Delta 2^{j-i} \geq 2^j(\Delta 2^j - \Delta 2^{-i}) > 0$. Set $\alpha_{i,j} = 2\sqrt{\min\{d_{i,j}, d_{j,i}\}}$.
Let $|V_i|$ be subsets of size $4^i N$, and G a graph on vertices $V = \cup_{i=0}^{t} V_i$ (hence, $|V|$ = n). For $0 < i, j \leq t$ construct a $(d_{j,i}, d_{i,j}, \alpha_{i,j})$-jumbled graph between V_i and V_j (or $(d_{i,i}, \alpha_{i,i})$-jumbled if $i = j$).
The theorem follows from the following two lemmata.

Lemma 2.2.4. *G is $(d, 7\Delta)$-jumbled.*

Proof: It is not hard to verify that G is indeed d regular. Take $A, B \subset V$, and denote their size by a and b. Denote $A_i = A \cap V_i$, $B_i = B \cap V_i$, and their size by a_i and b_i. We want to show that:

$$|e(A, B) - dab/n| \leq 7\Delta\sqrt{ab}.$$

For simplicity we show that $e(A, B) \leq dab/n + 7\Delta\sqrt{ab}$. A similar argument bounds the number of edges from below. From the construction, $|e(A_i, B_j) - d_{i,j}|A_i||B_j|/|V_j|| \leq \alpha_{i,j}\sqrt{|A_i||B_j|}$, or:

$$e(A_i, B_j) \leq d_{i,j}a_i b_j/(4^j N) + \alpha_{i,j}\sqrt{a_i b_j}.$$

Summing up over $i, j = 0, \ldots, t$ we get:

$$\begin{aligned} e(A, B) &\leq \sum d_{i,j}a_i b_j/(4^j N) + \alpha_{i,j}\sum \alpha_{i,j}\sqrt{a_i b_j} \\ &\leq d/n \sum a_i b_j + \Delta/N \sum a_i b_j 2^{-(i+j)} + \sum \alpha_{i,j}\sqrt{a_i b_j}. \end{aligned}$$

$\sum a_i b_j = ab$, so it remains to bound the error term. As $a_i, b_i \in [0, N \cdot 2^{2i}]$, by Lemma 2.2.3:

$$\begin{aligned} \Delta/N \sum a_i b_j 2^{-(i+j)} &= \Delta/N(\sum a_i 2^{-i})(\sum b_i 2^{-i}) \\ &\leq \Delta/N(\sqrt{3N\sum a_i})(\sqrt{3N\sum b_i}) = 3\Delta\sqrt{ab}, \end{aligned}$$

so it remains to show that $\sum \alpha_{i,j}\sqrt{a_i b_j} \leq 4\Delta\sqrt{ab}$. It's enough to show that:

$$\sum 2\sqrt{\frac{d}{\tau}a_i b_j 2^{2\max\{i,j\}}} + \sum 2\sqrt{\Delta 2^{|j-i|}a_i b_j} \leq 4\Delta\sqrt{ab}.$$

Indeed, as $i, j \leq t$, and $2^{2t} < \tau$,

$$\sum \sqrt{\frac{d}{\tau}a_i b_j 2^{2\max\{i,j\}}} < \sqrt{d}\sum \sqrt{a_i b_j} < \Delta\sqrt{ab}.$$

Similarly, $2^{|j-i|} \leq 2^t < \sqrt{\tau} < \sqrt{\Delta}$, and so:

$$\sum \sqrt{\Delta 2^{j-i}a_i b_j} < \Delta \sum \sqrt{a_i b_j} = \Delta\sqrt{ab}.$$

∎

Lemma 2.2.5. $\lambda(G) \geq \Delta(t+1)$

Proof: Take $x \in \mathbb{R}^n$ to be -2^{-t} on vertices in V_t, and 2^{-i} on vertices in V_i, for $i < t$. It is easy to verify that $x \perp \vec{1}$, and that $||x||^2 = N \cdot (t+1)$. Let M be the adjacency matrix if G. Since $\vec{1}$ is an eigenvector of M corresponding to the largest eigenvalue, by the variational characterization of eigenvalues, $\lambda(G) \geq \frac{xMx}{||x||^2}$. Hence, to prove the lemma it suffices to show that $xMx \geq \Delta N(t+1)^2$. Indeed:

$$\begin{aligned} xMx &= \sum_{i,j=0}^{t} d_{i,j} 4^i N 2^{-(i+j)} - 4 \sum_{i=0}^{t-1} d_{i,t} 4^i N 2^{-(i+t)} \\ &= \frac{d}{\tau} N \sum_{i,j=0}^{t} 2^{i+j} + \Delta N \sum_{i,j=0}^{t} 1 - 4 \frac{d}{\tau} N \sum_{i=0}^{t-1} 2^{i+t} + 4\Delta N \sum_{i=0}^{t-1} 1 \\ &= \frac{d}{\tau} N (2^{2(t+1)} - 4 \cdot 2^{2t}) + \Delta N((t+1)^2 + 4t) > \Delta N (t+1)^2. \end{aligned}$$

∎

Chapter 3

Constructing expander graphs by 2-lifts

3.1 Introduction

An d-regular graph is called a λ-*expander*, if all its eigenvalues but the first are in $[-\lambda, \lambda]$. Such graphs are interesting when d is fixed, $\lambda < d$, and the number of vertices in the graph tends to infinity. Applications of such graphs in computer science and discrete mathematics are many, see for example [48] for a survey.

It is known that random d-regular graphs are good expanders ([17], [32], [29]), yet many applications require an explicit construction. Some known construction appear in [51], [35], [8], [49], [6], [52], [1] and [59]). The Alon-Boppana bound says that $\lambda \geq 2\sqrt{d-1} - o(1)$ (cf. [57]). The graphs of [49] and [52] satisfy $\lambda \leq 2\sqrt{d-1}$, for infinitely many values of d, and are constructed very efficiently. However, the analysis of the eigenvalues in these construction relies on deep mathematical results. Thus, it is interesting to look for construction whose analysis is elementary.

The first major step in this direction is a construction based on iterative use of the zig-zag product [59]. This construction is simple to analyze, and is very explicit, yet the eigenvalue bound falls somewhat short of what might be hoped for. The graphs constructed with the zig-zag product have second eigenvalue $O(d^{3/4})$, which can be improved, with some additional effort to $O(d^{2/3})$. Here we introduce an iterative construction based on 2-lifts of graphs, which is close to being optimal and gives $\lambda = O(\sqrt{d \log^3 d})$.

A graph $\hat{G}$ is called a k-lift of a "base graph" G if there is a $k:1$ *covering map* $\pi : V(\hat{G}) \to V(G)$. Namely, if $y_1, \ldots, y_d \in G$ are the neighbors of $x \in G$, then every $x' \in \pi^{-1}(x)$ has exactly one vertex in each of the subsets $\pi^{-1}(y_i)$. See [10] for a general introduction to graph lifts.

The study of lifts of graphs has focused so far mainly on random lifts [10, 11, 12, 47, 31]. In particular, Amit and Linial show in [11] that w.h.p. a random k-lift has a strictly positive edge expansion. It is not hard to see that the eigenvalues of the base graph are also eigenvalues of the lifted graph. These are called by Joel Friedman the "old" eigenvalues of the lifted graph. In [31] he shows that w.h.p. a random k-lift of a d-regular graph on n vertices is "weakly Ramanujan". Namely, that all eigenvalues but, perhaps, those of the base graph, are, in absolute value, $O(d^{3/4})$. In both cases the probability tends to 1 as k tends to infinity.

Here we study 2-lifts of graphs. We conjecture that every d regular graph has a 2-lift with all new

eigenvalues at most $2\sqrt{d-1}$ in absolute value. It is not hard to show (e.g., using the Alon-Boppana bound [57]) that if this conjecture is true, it is tight. We prove (in Theorem 3.3.1) a slightly weaker result; every graph of maximal degree d has a 2-lift with all new eigenvalues $O(\sqrt{d\log^3 d})$ in absolute value. Under some natural assumptions on the base graph, such a 2-lift can be found efficiently. This leads to a polynomial time algorithm for constructing families of d-regular expander graphs, with second eigenvalue $O(\sqrt{d\log^3 d})$.

3.2 The construction scheme

3.2.1 Definitions

Let $G = (V, E)$ be a graph on n vertices, and let A be its adjacency matrix. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of A. We denote by $\lambda(G) = \max_{i=2,\ldots,n} |\lambda_i|$. We say that G is an $(n, d, \mu) - expander$ if G is d-regular, and $\lambda(G) \leq \mu$. If $\lambda(G) \leq 2\sqrt{d-1}$ we say that G is *Ramanujan*.

A *signing* of the edges of G is a function $s : E(G) \to \{-1, 1\}$. The *signed adjacency matrix* of a graph G with a signing s has rows and columns indexed by the vertices of G. The (x, y) entry is $s(x, y)$ if $(x, y) \in E$ and 0 otherwise.
A 2-*lift* of G, associated with a signing s, is a graph $\hat{G}$ defined as follows. Associated with every vertex $x \in V$ are two vertices, x_0 and x_1, called the *fiber* of x. If $(x, y) \in E$, and $s(x, y) = 1$ then the corresponding edges in $\hat{G}$ are (x_0, y_0) and (x_1, y_1). If $s(x, y) = -1$, then the corresponding edges in $\hat{G}$ are (x_0, y_1) and (x_1, y_0). The graph G is called the *base graph*, and $\hat{G}$ a 2-lift of G. By the *spectral radius of a signing* we refer to the spectral radius of the corresponding signed adjacency matrix. When the spectral radius of a signing of a d-regular graph is $\tilde{O}(\sqrt{d})$ we say that the signing (or the lift) is *Quasi-Ramanujan*.

For $v, u \in \{-1, 0, 1\}^n$, denote $S(u) = supp(u)$, and $S(u, v) = supp(u) \cup supp(v)$.
It will be convenient to assume throughout that $V(G) = \{1, \ldots, n\}$.

3.2.2 The spectrum of 2-lifts

The eigenvalues of a 2-lift of G can be easily characterized in terms of the adjacency matrix and the signed adjacency matrix:

Lemma 3.2.1. *Let A be the adjacency matrix of a graph G, and A_s the signed adjacency matrix associated with a 2-lift $\hat{G}$. Then every eigenvalue of A and every eigenvalue of A_s are eigenvalues of $\hat{G}$. Furthermore, the multiplicity of each eigenvalue of $\hat{G}$ is the sum of its multiplicities in A and A_s.*

Proof: It is not hard to see that the adjacency matrix of $\hat{G}$ is:

$$\hat{A} = \begin{pmatrix} A_1 & A_2 \\ A_2 & A_1 \end{pmatrix}$$

Where A_1 is the adjacency matrix of $(V, s^{-1}(1))$ and A_2 the adjacency matrix of $(V, s^{-1}(-1))$. (So $A = A_1 + A_2$, $A_s = A_1 - A_2$). Let v be an eigenvector of A with eigenvalue μ. It is easy to check that $\hat{v} = (v\ v)$ is an eigenvector of $\hat{A}$ with eigenvalue μ.
Similarly, if u is an eigenvector of A_s with eigenvalue λ, then $\hat{u} = (u\ -u)$ is an eigenvector of $\hat{A}$ with eigenvalue λ.
As the $\hat{v}$'s and $\hat{u}$'s are perpendicular and $2n$ in number, they are all the eigenvectors of $\hat{A}$. ∎

We follow Friedman's ([31]) nomenclature, and call the eigenvalues of A the *old* eigenvalues of $\hat{G}$, and those of A_s the *new* ones.

3.2.3 The iterative construction

Consider the following scheme for constructing (n, d, λ)-expanders. Start with $G_0 = K_{d+1}$, the complete graph on $d + 1$ vertices [1] . Its eigenvalues are d, with multiplicity 1, and -1, with multiplicity d. We want to define G_i as a 2-lift of G_{i-1}, such that all new eigenvalues are in the range $[-\lambda, \lambda]$. Assuming such a 2-lifts always exist, the G_i constitute an infinite family of (n, d, λ)-expanders.
It is therefore natural to look for the smallest $\lambda = \lambda(d)$ such that every graph of degree at most d has a 2-lift, with new eigenvalues in the range $[-\lambda, \lambda]$. In other words, a signing with spectral radius $\leq \lambda$.

We note that $\lambda(d) \geq 2\sqrt{d-1}$ follows from the Alon-Boppana bound. More concretely:

Proposition 3.2.1. *Let G be a d-regular graph which contains a vertex that does not belong to any cycle of bounded length, then no signing of G has spectral radius below $2\sqrt{d-1} - o(1)$.*

To see this, note first that all signing of a tree have the same spectral radius. This follows e.g., from the easy fact that any 2-lift of a tree is a union of two disjoint trees, isomorphic to the base graph. The assumption implies that G contains an induced subgraph that is a full d-ary tree T of unbounded radius. The spectral radius of T is $2\sqrt{d-1} - o(1)$. The conclusion follows now from the interlacing principle of eigenvalues.

We conjecture that this lower bound is tight:

Conjecture 3.2.1. *Every d-regular graph has a signing with spectral radius at most $2\sqrt{d-1}$.*

We have numerically tested this conjecture quite extensively. A close upper bound is proved in section 3.3.1.

3.2.4 Examples

We give a few examples of specific graphs for which we know how to construct a 2-lift where all new eigenvalues are small.

[1] We could start with any small d-regular graph with a large spectral gap. Such graphs are easy to find.

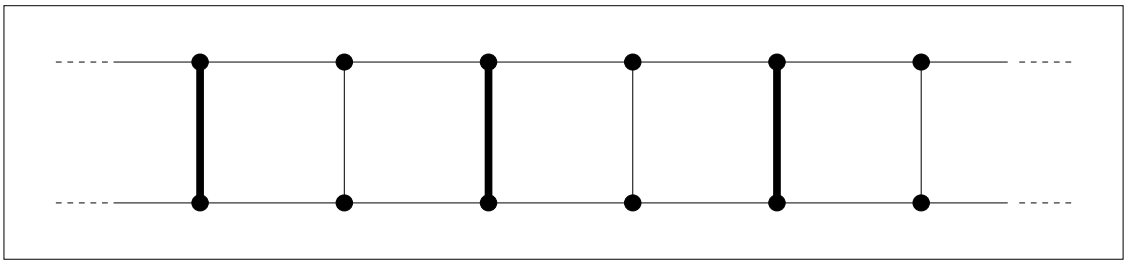

Figure 3.1: The Railway Graph. Edges where the signing is -1 are bold.

1. **The complete graph on $d+1$ vertices, and the complete bipartite graph with d vertices on each side**:
   We show a good signing of K_{d+1}, the complete graph on $d+1$ vertices. Suppose there exists an Hadamard matrix, H, of size $d+1$. Note that $H-I$ is a signed adjacency matrix for K_{d+1}. It is well known that its spectral radius is $\sqrt{d+1}$. Hence, the spectral radius of $H-I$, is at most $\sqrt{d+1}+1$.
   For a general d, it is known that there exists an Hadamard matrix H of size n, with $d < n \leq 2d$. Take H as the submatrix defined by the first $d+1$ rows and columns. By the Interlacing Theorem its spectral radius is at most $\sqrt{2d}$, and the same argument as above gives a 2-lift of K_{d+1} with all new eigenvalues at most $\sqrt{2d}+1$ in absolute value.
   Note that a similar argument works for a complete bipartite graph with the same number of vertices on each side.

2. **The grid**:
   Let G be the Cayley graph of the group $\mathbb{Z}_{2n} \times \mathbb{Z}_{2n}$, with generators $(1,0)$, $(-1,0)$, $(0,1)$ and $(0,-1)$. Let s be -1 on edges corresponding to $(0,1)$ (or $(0,-1)$), and connecting vertices where the first coordinate is even. Let s be 1 elsewhere.
   Consider A_s^2, and think of it as counting length 2 paths. Clearly it has 4 on the diagonal. Now consider two vertices, x and y, at distance 2 in the graph. If x and y share the same value in some coordinate, then the corresponding entry in A_s^2 is either 1 or -1, altogether 4 such entries in each row. Otherwise, there are 2 length 2 paths that connect x and y - either moving first in the first coordinate, and then in the second, or vice versa. One of these paths has both edges signed as 1, while the other has the horizontal edge marked as 1 and the vertical edge marked as -1. Thus, the (x,y) entry in A_s^2 is 0. Clearly the spectral radius of A_s^2 is at most that of $|A_s^2|$, the matrix obtained from A_s^2 by taking the absolute value of each entry. This, in turn, is 8-regular. Hence the spectral radius of A_s is at most $\sqrt{8} \leq 2\sqrt{3} = \sqrt{12}$.

3. **The railway graph**:
   Let R be the 3-regular graph defined as follows. $V(R) = \{0,\ldots,2k-1\} \times \{0,1\}$. For $i \in [2k], j \in \{0,1\}$, the neighbors of $(i,j) \in R$ are $((i-1) \bmod 2k, j)$, $((i+1) \bmod 2k, j)$ and $(i, 1-j)$. Define s, a signing of R, to be -1 on the edges $((2i,0),(2i,1))$, for $i \in \{0,\ldots,k-1\}$, and 1 elsewhere (see Figure 3.1). Let A_s be the signed adjacency matrix. It is easy to see that A_s^2 is a matrix with 3 on the diagonal, and two 1's in each row and column. Thus, its spectral radius is 5, and that of A_s is $\sqrt{5} < 2\sqrt{2}$.

4. **The r-dimensional railway**:
   Let R^r be the Cayley graph of the group $\mathbb{Z}_{2n} \times \mathbb{Z}_2^r$, with the generator set of all elements which are either 1 or -1 in one coordinate, and 0 in r coordinates (so it is $(r+2)$-regular). Think of R as being constructed by starting from the railway graph, and iteratively multiplying the vertex set by $\mathbb{Z}_2$. We define a signing s of the graph by induction on r. For $r = 1$ use the signing depicted in Figure 3.1. We shall define a signing s, such that in A_s^2 the only non-zero entries are $r+2$ on the diagonal, and ± 1 in entries corresponding to vertices $(x_0, x_1, \ldots, x_r)$ and $(y_0, y_1, \ldots, y_r)$ such that $|x_0 - y_0| = 2$, and for $i > 0$, $x_i = y_i$. Clearly this holds for $r = 1$.
   Think of the $(r+1)$-dimensional railway as two copies of the r-dimensional railway, $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{0\}$ and $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{1\}$. Let s' be the signing for R^r. Define a signing s of R^{r+1} to be the same as s' on the edges spanned by $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{0\}$, and $-s'$ on the edges spanned by $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{1\}$. Define s to be 1 on the remaining edges (those crossing from $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{0\}$ to $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{1\}$).
   Consider two vertices, x and y, at distance 2 from each other in R^{r+1}. If both are in $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{0\}$, or both are in $\mathbb{Z}_n \times \mathbb{Z}_2^r \times \{1\}$, then the claim holds by the induction hypothesis. Otherwise, there are two paths from x to y, one that first crosses, then takes one step within the copy of R^r, and one then first takes one step within the copy of R^r, and then crosses. The crossing edges have the same sign (1). The edges within the two copies of R^r are copies of the same edge, hence, by the construction of s, they have opposite signs. Summing the signs of these two paths, we get that the (x, y) entry in A_s^2 is 0.
   Using the same argument as in example 2, the spectral radius of A_s^2 is at most $r+4$, and thus that of A_s at most $\sqrt{r+4} \leq 2\sqrt{r+1}$.

3.3 Quasi-ramanujan 2-lifts

3.3.1 Quasi-ramanujan 2-lifts for every graph

In this subsection we prove a weak version of Conjecture 3.2.1:

Theorem 3.3.1. *Every graph of maximal degree d has a signing with spectral radius $O(\sqrt{d \cdot \log^3 d})$.*

By the relation between discrepancy and eigenvalues shown in Lemma 2.1.1, it is enough to show that with positive probability the discrepancy of randomly signed graph is small. Hence, the theorem will follow from the following lemma:

Lemma 3.3.1. *For every graph of maximal degree d, there exists a signing s such that for all $v, u \in \{-1, 0, 1\}^n$ the following holds:*

$$\frac{|v^t A_s u|}{||v|| ||u||} \leq 10\sqrt{d \log d}, \tag{3.1}$$

where A_s is the signed adjacency matrix.

Proof: First note that it's enough to prove this for u's and v's such that the set $S(u,v)$ spans a connected subgraph. Indeed, assume that the claim holds for all connected subgraphs and suppose that $S(u,v)$ is not connected. Split u and v according to the connected components of $S(u,v)$, and apply the claim to each component separately. Summing these up and using the Cauchy-Schwartz inequality, we conclude that the claim for u and v as well. So henceforth we assume that $S(u,v)$ is a connected.

Consider some $u, v \in \{-1,0,1\}^n$. Suppose we choose the sign of each edge uniformly at random. Denote the resulting signed adjacency matrix by A_s, and by $E_{u,v}$ the "bad" event that $\frac{|v^t A_s u|}{||v||||u||} > 10\sqrt{d \log d}$. Assume w.l.o.g. that $|S(u)| \geq \frac{1}{2}|S(u,v)|$. By the Chernoff inequality ($v^t A_s u$ is the sum of independent variables, attaining values of either ± 1 or ± 2):

$$\begin{aligned} Pr[E_{u,v}] \leq \quad & 2exp(-\frac{100d \log d |S(u)||S(v)|}{8|e(S(u),S(v))|}) \\ & \leq 2exp(-\frac{100d \log d |S(u)||S(v)|}{8d|S(v)|}) \\ & < d^{(-10|S(u,v)|)} \end{aligned}$$

We want to use the Lovász Local Lemma [26], with the following dependency graph on the $E_{u,v}$: There is an edge between $E_{u,v}$ and $E_{u',v'}$ iff $S(u,v) \cap S(u',v') \neq \emptyset$. Denote $k = |S(u,v)|$. How many neighbors, $E_{u',v'}$, does $E_{u,v}$ have, with $|S(u',v')| = l$?

Since we are interested only in connected subsets, this is clearly bounded by the number of rooted directed subtrees on l vertices, with a root in $S(u,v)$. It is known (cf. [45]) that there are at most $k\binom{d(l-1)}{l-1} \approx kd^{l-1}$ such trees (a similar argument appears in [33]. The bound on the number of trees is essentially tight by [4]).

In order to apply the Local Lemma, we need to define for such u and v numbers $0 \leq X_{u,v} < 1$ such that:

$$X_{u,v} \prod_{(u',v'):E_{u,v} \sim E_{u',v'}} (1 - X_{u',v'}) \geq d^{-10|S(u,v)|}. \tag{3.2}$$

Observe that for $S \subseteq [n]$ there are at most $2^{4|S|}$ distinct pairs $v, u \in \{-1,0,1\}^n$ with $S(u,v) = S$. For all u, v set $X_{u,v} = d^{-3k}$, where $k = |S(u,v)|$. Then in (3.2) we get:

$$\begin{aligned} & X_{u,v} \cdot \prod_{(u',v'):E_{u,v} \sim E_{u',v'}} (1 - X_{u',v'}) = d^{-3k} \prod_{l=1}^{n} (1 - d^{-3l})^{kd^l 2^{4l}} \approx \\ & d^{-3k} exp(-k \sum_{l=1}^{n} d^{-3l} d^l 2^{4l}) \geq d^{-3k} e^{-2k} > d^{-10k} \end{aligned}$$

as required. ∎

3.3.2 An explicit construction of quasi-ramanujan graphs

For the purpose of constructing expanders, it is enough to prove a weaker version of Theorem 3.3.1. Roughly, that every *expander* graph has a 2-lift with small spectral radius. In this subsection we show that when the base graph is a good expander (in the sense of the definition below), then w.h.p. a random 2-lift has a small spectral radius. We then derandomize the construction to get a deterministic polynomial time algorithm for constructing arbitrarily large expander graphs.

Definition 3.3.1. *We say that a graph G on n vertices is (β, t)-sparse if for every $u, v \in \{0,1\}^n$, with $|S(u,v)| \leq t$,*

$$uAv \leq \beta ||u|| ||v||.$$

Lemma 3.3.2. *Let A be the adjacency matrix of a d-regular $(\gamma(d), \log n)$-sparse G graph on n vertices, where $\gamma(d) = 10\sqrt{d \log d}$. Then for a random signing of G (where the sign of each edge is chosen uniformly at random) the following holds w.h.p.:*

1. $\forall u, v \in \{-1, 0, 1\}^n : |uA_s v| \leq \gamma(d)||u||||v||$.
2. *$\hat{G}$ is $(\gamma(d), 1 + \log n)$-sparse.*

A_s denotes the random signed adjacency matrix, and $\hat{G}$ the corresponding 2-lift.

Proof: Following the same arguments and notations as in the proof of Lemma 3.3.1, we have that there are at most $n \cdot d^k$ connected subsets of size k. For a given pair u, v such that $|S(u,v)| = k$, the probability that requirement (1) is violated is at most d^{-10k}. Since for each S there are at most $2^{4|S|}$ pairs u, v such that $S(u,v) = S$, by the union bound, w.h.p. no pair u, v such that $|S(u,v)| > \log n$ violates (1). If $|S(u,v)| \leq \log n$ then by (2) there are simply not enough edges between $S(u)$ and $S(v)$ for (1) to be violated.

Next, we show that w.h.p. (2) holds as well. Let s be a signing, and define A_1, A_2 and $\hat{A}$ as in Lemma 3.2.1. Given $u = (u_1 \ u_2), v = (v_1 \ v_2) \in \{0,1\}^n \times \{0,1\}^n$, we wish to prove that $uAv \leq \gamma(d)||u||||v||$. As in the proof of Lemma 3.3.1 we may assume that $S(u,v)$ is connected - in fact, that it is connected via the edges between $S(u)$ and $S(v)$. Hence, we may assume that the ratio of the sizes of these subsets is at most d. Define $x = u_1 \vee u_2$, $y = v_1 \vee v_2$, $x' = u_1 \wedge u_2$, and $y' = v_1 \wedge v_2$ (the characteristic vectors of $S(u_1, u_2)$, $S(v_1, v_2)$, $S(u_1) \cap S(u_2)$ and $S(v_1) \cap S(v_2)$). It is not hard to verify that:

$$u\hat{A}v = u_1 A_1 v_1 + u_1 A_2 v_2 + u_2 A_2 v_1 + u_2 A_1 v_2 \leq xAy + x'Ay'. \tag{3.3}$$

If $|S(x,y)| \leq \log n$, then clearly $|S(x',y')| \leq \log n$ and from the assumption that G is $(\gamma(d), \log n)$-sparse

$$xAy + x'Ay' \leq \gamma(d)(\sqrt{|S(x)||S(y)|} + \sqrt{|S(x')||S(y')|}).$$

Observe that $|S(u)| = |S(x)| + |S(x')|$ and $|S(v)| = |S(y)| + |S(y')|$, so in particular $u\hat{A}v \leq \gamma(d)\sqrt{|S(u)||S(v)|}$, and requirement (2) holds.
So assume $|S(x,y)| = |S(u,v)| = \log n + 1$. It is not hard to see that this entails $S(u_1, v_1) \cap S(u_2, v_2) = \emptyset$. In other words, $S(u,v)$ contains at most one vertex from each fiber. Hence,

$x' = y' = \vec{0}$ and $|S(u)| = |S(x)|$, $|S(v)| = |S(y)|$.
Denote $S = S(x, y)$, and assume w.l.o.g. that $|S(y)| > \frac{1}{2}\log n$. From (3.3) $u\hat{A}v \leq xAy$, so it's enough to show that $xAy \leq \gamma(d)\sqrt{|S(x)||S(y)|}$. If this is not the case, we can bound the ratio between $|S(x)|$ and $|S(y)|$: Since the graph is of maximal degree d we have $xAy \leq d|S(X)|$. Hence, $\frac{|S(x)|}{|S(y)|} > \frac{\gamma(d)^2}{d^2} = \frac{100\log d}{d}$.
Observe that the edges between $S(u)$ and $S(v)$ in $\hat{G}$ originate from edges between $S(x)$ and $S(y)$ in G in the following way - for each edge between $S(x)$ and $S(y)$ in G there is, with probability $\frac{1}{2}$, an edge between $S(u)$ and $S(v)$ in $\hat{G}$.
Next we bound xAy. Averaging over all $S\backslash\{i\}$, for $i \in S(y)$ we have that:

$$(|S(y)| - 2)xAy \leq |S(y)|\gamma(d)\sqrt{(|S(y)| - 1)|S(x)|}.$$

Hence the expectation of $u\hat{A}v$ is at most $\frac{1}{2}c\gamma(d)||u||||v||$, where

$$c = \frac{\sqrt{|S(y)|(|S(y)| - 1)}}{|S(y)| - 2} \leq 1.1$$

(assuming n is not very small). By the Chernoff bound, the probability that $u\hat{A}v > \gamma(d)||u||||v||$ is at most:

$$\begin{aligned} &2exp(-\frac{0.9}{2.2}\gamma(d)||u||||v||) \leq \\ &exp(-0.2\gamma(d)\frac{(\log n + 1)(10\sqrt{\log d})}{\sqrt{d}}) = \\ &exp(-20\log d(\log n + 1)), \end{aligned}$$

Since $\frac{|S(x)|}{|S(y)|} > \frac{100\log d}{d}$, and $|S(v)| > \frac{1}{2}\log n$.
There are at most $d^{\log n+1}4^{\log n+1}$ pairs u, v with $S(u, v)$ connected and of size $\log n + 1$, so by the union bound, w.h.p., requirement (2) holds. ∎

Corollary 3.3.1. *Let A be the adjacency matrix of a d-regular $(\gamma(d), \log n)$-sparse G graph on n vertices, where $\gamma(d) = 10\sqrt{d\log d}$. Then there is a deterministic polynomial time algorithm for finding a signing s of G such that the following hold:*

1. *The spectral radius of A_s is $O(\sqrt{d\log^3 d})$.*
2. *$\hat{G}$ is $(\gamma(d), 1 + \log n)$-sparse,*

where A_s is the signed adjacency matrix, and $\hat{G}$ is the corresponding 2-lift.

Proof: Consider a random signing s. For each closed path p in G of length $l = 2\lceil \log n \rceil$ define a random variable Y_p equal to the product of the signs of its edges. From lemmas 3.3.2 and 2.1.1, the expected value of the trace of A_s^l, which is the expected value of the sum of these variables, is $O(\sqrt{d \log^3 d})^l$ (since l is even the sum is always positive). For each $u, v \in \{0,1\}^n$, with $|S(u,v)| = \log n + 1$, and $S(u,v)$ connected, define $Z_{u,v}$ to be d^l if $u\hat{A}v \geq \gamma(d)||u||||v||$, and 0 otherwise. In the proof of Lemma 3.3.2 we've seen that the probability that $Z_{u,v}$ is not 0 is at most $d^{-10 \log n}$, thus the expected value of $Z_{u,v}$ is at most $d^{-8 \log n}$. Let Z be the sum of the $Z_{u,v}$'s. Recall that there are at most $n(4d)^{\log n + 1}$ pairs (u,v) such that $|S(u,v)| = \log n + 1$, and $S(u,v)$ connected. Hence, the expected value of Z is less than $d^{-6 \log n}$.

Let $X = Y + Z$. Note that the expected value of X is approximately that of Y, $O(\sqrt{d \log^3 d})^l$. The expectation of Y_p and $Z_{u,v}$ can be easily computed even when the sign of some of the edges is fixed, and that of the other is chosen at random. As there is only a polynomial number of variables, using the method of conditional probabilities (cf. [9]) one can find a signing s such that the value of X is at most its expectation. For this value of X, $tr(A_s^l) = Y = O(\sqrt{d \log^3 d})^l$, and $Z = 0$ since if $Z \neq 0$ then $Z \geq d^l$. Clearly, the spectral radius of A_s is $O(\sqrt{d \log^3 d})$. In the proof of Lemma 3.3.2 we've seen that if G is $(\gamma(d), \log n)$-sparse then so is $\hat{G}$, for any signing of G. For our choice of s all $Z_{u,v} = 0$, hence $\hat{G}$ is actually $(\gamma(d), \log n + 1)$-sparse. ∎

An alternative method for derandomization, using an almost k-wise independent sample space, is given in the appendix.

Recall the construction from the beginning of this section. Start with a d-regular graph G_0 which is an (n_0, d, μ)-expander, for $\mu = 10\sqrt{d \log d}$ and $n_0 > d \log^2 n_0$. From the Expander Mixing Lemma (cf. [9]), G_0 is $(\mu, \log n_0)$-sparse. Iteratively chose G_{i+1} to be a 2-lift of G_i according to Corollary 3.3.1, for $i = 1, \ldots, \log(n/n_0)$. Clearly this is a polynomial time algorithm that yields a $(n, d, O(\sqrt{d \log^3 d}))$-expander graph.

3.3.3 Random 2-lifts

Theorem 3.3.1 states that for every graph there exists a signing such that the spectral radius of the signed matrix is small. The proof shows that for a random signing, this happens with positive, yet exponentially small, probability. The following example shows the limitations of this argument and in particular, that there exist graphs for which a random signing almost surely fails to give a small spectral radius.

Consider a graph composed of $n/(d+1)$ disjoint copies of K_{d+1} (the complete graph on $d+1$ vertices). If all edges in one of the components are equally signed, then A_s has spectral radius d. For d fixed and n large, this event will occur with high probability. Note that connectivity is not the issue here - it is easy to modify this example and get a connected graph for which, w.h.p., the spectral radius of A_s is $\Omega(d)$.

However, for a random d-regular graph, it is true that a random 2-lift will, w.h.p., yield a signed matrix with small spectral radius. This follows from the fact that, w.h.p., a random d-regular graph is an $(n, d, O(\sqrt{d}))$-expander ([32, 30, 29]). In particular, by the Expander Mixing Lemma, it is $(O(\sqrt{d}), \log n)$-sparse. By Lemma 3.3.2, w.h.p., a random 2-lift yields a signed matrix with small

spectral radius.

3.4 A stronger notion of explicitness

In this section we suggest an alternative derandomization scheme of section 3.3.2. We use the construction of Naor and Naor [56] of a small, almost k-wise independent sample space. This derandomization scheme leads to a construction which, in a sense, is more explicit than that in section 3.3.2.

3.4.1 Derandomization using an almost k-wise independent sample space

Recall that in the proof of Corollary 3.3.1 we defined two types of random variables: Let $l = 2\lceil \log n \rceil$. For each closed path p of length l, Y_p is the product of the signs of the edges of p. For each $u, v \in \{0,1\}^n$, with $|S(u,v)| = \log n + 1$, and $S(u,v)$ connected, let $Z_{u,v}$ be d^l if $u\hat{A}v \geq \gamma(d)||u||||v||$, and 0 otherwise. Define X to be the sum of all these random variables.
For brevity it will be convenient to make the following ad-hoc definitions:

Definition 3.4.1. *A signing s of a d-regular graph G is* (n,d)-good*, if the spectral radius of A_s is $O(\sqrt{d \log^3 d})$ and $\hat{G}$ is $(\gamma(d), 1+\log n)$-sparse.*
A d-regular graph G is an (n,d)-good expander*, if it's an $(n, d, O(\sqrt{d \log^3 d})$-expander, and is $(\gamma(d), 1+\log n)$-sparse.*

The proof showed that finding a good signing is equivalent to finding a signing such that the value of X is at most its expected value. We now show that this conclusion is also true when rather than choosing the sign of each edge uniformly and independently, we choose the signing from an (ϵ, k)-wise independent sample space, with $k = d \log n$ and $\epsilon = d^{-2d \log n}$.

Definition 3.4.2. *([56]) Let Ω_m be a sample space of m-bit strings, and let $S = s_1 \ldots s_m$ be chosen uniformly at random from Ω_m. We say that Ω_m is an (ϵ, k)-wise independent sample space if for any $k' \leq k$ positions $i_1 < i_2 \ldots < i_{k'}$,*

$$\sum_{\alpha \in \{-1,1\}^{k'}} | \, Pr[s_{i_1} \ldots s_{i_{k'}} = \alpha] - 2^{-k'} | < \epsilon.$$

Naor and Naor [56] suggest an explicit construction of such sample spaces. When $k = O(\log m)$ and $1/\epsilon = poly(m)$, the size of the sample space is polynomial in m (simpler constructions are also given in [7]).
We shall immediately see that the expected value of X does not change significantly when the signing is chosen from such a sample space. Hence, an alternative way of efficiently finding a good signing is to go over the entire sample space. For at least one point in it, the value of X is at most its expected value, and thus the signing is good.

Lemma 3.4.1. *Let $m = dn/2$, $k = d \log n$ and $\epsilon = d^{-2d \log n}$. Let Ω_m be an (ϵ, k)-wise independent sample space. Let X be as in the proof of Corollary 3.3.1. Let U_m be the uniform distribution on m bits. Then*

$$|\mathbb{E}_{\Omega_m}[X] - \mathbb{E}_{U_m}[X]| = o(1).$$

Proof: Recall that $X = \sum_p Y_p + \sum_{u,v} Z_{u,v}$, where Y_p and $Z_{u,v}$ are as above, the first sum is over all closed paths of length $l = 2\lceil \log n \rceil$, and the second sum is over all $u, v \in \{0,1\}^n$, with $|S(u,v)| = \log n + 1$ and $S(u,v)$ connected. Hence

$$|\mathbb{E}_{\Omega_m}[X] - \mathbb{E}_{U_m}[X]| \leq \sum_p |\mathbb{E}_{\Omega_m}[Y_p] - \mathbb{E}_{U_m}[Y_p]| + \sum_{u,v} |\mathbb{E}_{\Omega_m}[Z_{u,v}] - \mathbb{E}_{U_m}[Z_{u,v}]|.$$

Let p be a path of length l, and denote the edges that appear in it an odd number of times by $i_1, \ldots i_{l'}$, for some $l' < l$. Let $s_{i_1}, \ldots s_{i_{l'}}$ be the signs of these edges. Then the value of Y_p is $\prod_{j=1}^{l'} s_{i_j}$, and (for every distribution)

$$\mathbb{E}[Y_p] = \sum_{\alpha \in \{-1,1\}^{l'}} \Pr[s_{i_1}, \ldots s_{i_{l'}} = \alpha] \cdot \prod_{j=1}^{l'} \alpha_j.$$

Thus,

$$\begin{aligned}
&|\mathbb{E}_{\Omega_m}[Y_p] - \mathbb{E}_{U_m}[Y_p]| = \\
&|\sum_{\alpha \in \{-1,1\}^{l'}} (\prod_{j=1}^{l'} \alpha_j)(\Pr_{\Omega_m}[s_{i_1}, \ldots s_{i_{l'}} = \alpha] - \Pr_{U_m}[s_{i_1}, \ldots s_{i_{l'}} = \alpha])| \leq \\
&|\sum_{\alpha \in \{-1,1\}^{l'}} (\Pr_{\Omega_m}[s_{i_1}, \ldots s_{i_{l'}} = \alpha] - 2^{-l'})| < \epsilon.
\end{aligned}$$

As there are less than d^l closed paths p of length l, $\sum_p |\mathbb{E}_{\Omega_m}[Y_p] - \mathbb{E}_{U_m}[Y_p]| < \epsilon d^l = o(1)$. A similar argument shows that $|\sum_{u,v} |\mathbb{E}_{\Omega_m}[Z_{u,v}] - \mathbb{E}_{U_m}[Z_{u,v}]| = o(1)$ as well. ∎

In fact, it follows that w.h.p. (say, $1 - \frac{1}{n^2}$ for an appropriate choice of ϵ), choosing an element uniformly at random from Ω_m leads to a good signing.

3.4.2 A probabilistic strongly explicit construction

The constructions of section 3.3.2 and of the previous subsection are explicit in the sense that given n and d they suggest a polynomial (in n) time algorithm for constructing an (n,d)-good expander. However, in some applications a stronger notion of explicitness is required. Namely, an algorithm that given n, d and $i, j \in [n]$, decides in time polylog(n) whether i and j are adjacent. We do not know how to achieve such explicitness using the 2-lifts schema, but we can do so probabilistically. Consider the following:

Definition 3.4.3. *Let $f_n : \{0,1\}^t \times [n] \times [n] \to \{0,1\}$, with $t = O(\log n)$. Given $r \in \{0,1\}^t$, f_n defines a graph $G_{f_n}(r)$, on n vertices, where i and j are adjacent iff $f_n(r,i,j) = 1$. We say that f_n is a* δ-probabilistic strongly explicit *description of an (n,d)-good expander graph, if given n, f_n can be computed in time polylog(n), and, with probability at least $1-\delta$ (over a uniform choice of r), $G_{f_n}(r)$ is an (n,d)-good expander graph.*

We show that this notion of explicitness can be achieved by our construction. It will be convenient to give a similar definition for a signing of a graph, and for a composition of such functions:

Definition 3.4.4. *Let $h_n : \{0,1\}^t \times [n] \times [n] \to \{-1,1\}$, with $t = O(\log n)$. Given $r \in \{0,1\}^t$, and a graph G on n vertices, h_n defines a signing s_{h_n} of G by $s_{h_n}(r)(i,j) = h_n(r,i,j)$. We say that h_n is a* δ-probabilistic strongly explicit *description of an (n,d)-good signing, if given n, h_n can be computed in time polylog(n), and, for any $(\log n, \gamma(d))$-sparse d-regular graph G on n vertices, with probability at least $1-\delta$ (over a uniform choice of r), h_n defines an (n,d)-good signing.*

Definition 3.4.5. *Let $f_n : \{0,1\}^{t_1} \times [n] \times [n] \to \{0,1\}$, and $h_n : \{0,1\}^{t_2} \times [n] \times [n] \to \{-1,1\}$ be as above. Their* composition, *$f_{2n} : \{0,1\}^t \times [2n] \times [2n] \to \{0,1\}$, with $t = max\{t_1, t_2\}$ is as follows. For $r \in \{0,1\}^t$, let r_1 be the first t_1 bits of r, and r_2 the first t_2 bits in r. f_{2n} is such that the graph $G_{f_{2n}}(r)$ is the 2-lift of $G_{f_n}(r_1)$ described by the signing $s_{h_n}(r_2)$.*

The following lemma is easy, and we omit the proof:

Lemma 3.4.2. *Let f_n be a δ_1-probabilistic strongly explicit description of an (n,d)-good expander, and h_n a δ_2-probabilistic strongly explicit description of an (n,d)-good signing. Then their composition is a $(\delta_1+\delta_2)$-probabilistic strongly explicit description of an $(2n,d)$-good expander.*

Think of an (ϵ,k)-wise independent space Ω_m as a function $\omega : \{0,1\}^t \to \{-1,1\}^m$, where $|\Omega_m| = 2^t$. It follows from the work of Naor and Naor ([56]), that not only can ω be computed efficiently, but that given $r \in \{0,1\}^t, p \in [m]$ $\omega(r)_p$ (the p'th coordinate of $\omega(r)$) can be computed efficiently (i.e. in time polylog(m)). Take $m = \binom{n}{2}$, and think of the elements of $\{-1,1\}^m$ as being indexed by unordered pairs $(i,j) \in \binom{[n]}{2}$. Define $h_n(r,i,j) = \omega(r)_{i,j}$. It follows from the above discussion than h_n is a $\frac{1}{n^2}$-probabilistic strongly explicit description of an (n,d)-good signing, for k and ϵ as above.

We now describe how to construct a δ-probabilistic strongly explicit description of an (N,d)-good expander graph. Let G be an (n,d)-good expander, with $n \geq \frac{1}{\delta}$. For $i = 0, \ldots, l = log(N/n)$, define $n_i = n \cdot 2^i$ and $m_i = \binom{n_i}{2}$. Define $k_i = d \log n_i$. Let $\omega_i : \{0,1\}^{t_i} \to \{-1,1\}^{m_i}$ be a description of an (ϵ_i, k_i)-wise independent space of bit strings of length m_i, where ϵ_i is such that an element chosen uniformly at random from this space yields an (n_i,d)-good signing with probability at least $1 - \frac{1}{n_i^2}$.

The functions $h_{n_i}(r,p,q) = \omega_i(r)_{p,q}$ are $\frac{1}{n_i^2}$-probabilistic strongly explicit descriptions of (n_i,d)-good signings. Let f_n be a description of G. For simplicity, assume that adjacency in G can be decided in time polylog(n). Thus, f_n is, trivially, a 0-probabilistic strongly explicit description of an (n,d)-good expander. Define f_{n_i} as the composition of $f_{n_{i-1}}$ and $h_{n_{i-1}}$. It follows from this construction and Lemma 3.4.2 that:

Lemma 3.4.3. *f_{n_l} is an $\frac{1}{n}$-probabilistic strongly explicit description of an (N,d)-good expander graph.*

3.4.3 Algorithmic aspect

Lemma 2.1.1 is algorithmic, in the sense that given a matrix with a large eigenvalue, we can efficiently construct, from its eigenvector, a pair $u, v \in \{0,1\}^n$ such that $S(u) \cap S(v) = \emptyset$, and $|uAv| \geq \alpha ||u|| ||v||$ (There is a small caveat - in the proof we used a probabilistic argument for rounding the coordinates. This can be easily derandomized using the conditional probabilities method). Taking into consideration Note 2.2.1, given a d-regular graph G where $\lambda(G)$ is large, one can efficiently find disjoint subsets S and T, such that $e(S,T) - \frac{d}{n}|S||T| \geq c \cdot \lambda(G)/\log d \sqrt{|S||T|}$ (for some constant c). We speculate that this might be useful in designing graph partitioning algorithms.

In Lemma 3.3.1 we showed the with positive probability a random signing has a small spectral radius. It is interesting to find a sample space where this happens with high probability. We conjecture that the following algorithm works: Choose a signing at random. If some eigenvalue of the signed matrix is big, use Lemma 2.1.1 to find a pair $u, v \in \{0,1\}^n$ such that $|uAv| \geq \alpha ||u|| ||v||$. Choose a signing at random for the edges between $S(u)$ and $S(v)$. Repeat until all eigenvalues are small.

3.5 Group Lifts

The stepping-stone in the construction of expander graphs from 2-lifts was the simple characterization of the new eigenvalues of. Here we generalize this result, and characterize the new eigenvalues of general *group-lifts*.

Let $G = (V, E)$ be a graph on n vertices, and $(\Gamma, +)$ a finite group of size r (and $0 \notin \Gamma$). A $\Gamma - signing$ of G is function $s : V \times V \to \Gamma \cup \{0\}$ such that $s(x,y) = 0$ iff $(x,y) \notin E$, and for $(x,y) \in E$, $s(x,y) = -s(x,y)$. The Γ-*signed adjacency matrix* of a graph G (with respect to s) is the following matrix over $\Gamma \cup \{0\}$. Its rows and columns are indexed by the vertices of G. The (x,y)'th entry is 0 if (x,y) is not an edge in the graph, and $s((x,y))$ if it is.
A Γ-signing, s, of G defines a graph $\hat{G}$, a Γ-*lift* of G, as follows. For every vertex $x \in V$ and every element $g \in \Gamma$ define a vertex x_g in $\hat{G}$. For every edge $(x,y) \in E$ define in $\hat{G}$ the r edges $\{(x_g, y_{g+s((x,y))})\}_{g \in \Gamma}$.
Let ρ be an irreducible representation of Γ of dimension d. In particular, we think of ρ as associating with each group element a $d \times d$ unitary matrix. Fix s to be a $\Gamma - signing$ of G. Define the $nd \times nd$ matrix $A_{s,\rho}$ as follows. It is composed of $n \times n$ blocks of $d \times d$ submatrices. Each such submatrix is thus indexed by a pair of vertices. The (x,y)'th submatrix is the all zeros matrix if x and y are not adjacent in G. Otherwise, it is $\rho(s(x,y))$. Note that $A_{s,\rho}$ is Hermitian.

Lemma 3.5.1. *Let $G = (V, E)$ be a graph on n vertices and Γ a finite group of size r. Let $\rho_1, \ldots, \rho_t$ be the irreducible representations of Γ, of dimensions $d_1, \ldots, d_t$. Let $\hat{G}$ be a $\Gamma - lift$ of G defined by a Γ-signing s. Then the eigenvalues of $\hat{G}$ are the union of the eigenvalues of the matrices A_{s,ρ_i} for $i = 1, \ldots, t$.*
Furthermore, let μ be an eigenvalue of $\hat{G}$, and denote by m_i its multiplicity in A_{s,ρ_i}. Then its multiplicity in $\hat{G}$ is $\sum_{i=1}^{t} d_i \cdot m_i$.

Proof: We start by showing how the eigenvectors of $\hat{G}$ relate to those of A_{s,ρ_i}. Denote by k the number of irreducible representation of Γ. Fix some representation ρ, and let d be its dimension. Let $v \in \mathbb{C}^{dn}$ be an eigenvector of $A_{s,\rho}$ with eigenvalue μ. Think of v as being the concatenation of n vectors, $v_1, \ldots, v_n \in \mathbb{C}^d$. For every $g \in \Gamma$ define $w_{g,x} \in \mathbb{C}^d$ by $w_{g,x} = \rho(g)v_x$. Next we define d vectors, $u^1, \ldots, u^d \in \mathbb{C}^{rn}$ as follows. Think of the coordinates of the u^l's as being indexed by pairs (g, x) where $g \in \Gamma$ and $x \in V$. Define the (g, x) coordinate of u^l to be the l'th coordinate of $w_{g,x}$. We now show that the vectors u^l are eigenvectors of $\hat{G}$, and that all eigenvectors of $\hat{G}$ can be obtained in this way. As $\mu v = A_{s,\rho}v$, by the construction of $A_{s,\rho}$, we have that for all $x \in [n]$:

$$\mu v_x = \sum_{y:(x,y)\in E} \rho(s(x,y))v_y.$$

Fix some $g \in \Gamma$. Applying $\rho(g)$ to both sides we get:

$$\begin{aligned} \mu w_{g,x} &= \mu\rho(g)v_x = \sum_{y:(x,y)\in E} \rho(g)\rho(s(x,y))v_y \\ &= \sum_{y:(x,y)\in E} \rho(g + s(x,y))v_y = \sum_{y:(x,y)\in E} w_{g+s(x,y),y}. \end{aligned}$$

In particular, this holds point-wise, so for every $l \in [d]$, we have:

$$\mu(w_{g,x})_l = \sum_{y:(x,y)\in E} (w_{g+s(x,y),y})_l.$$

The neighbors of the vertex (g, x) in $\hat{G}$ are exactly all the $(g + s(x, y), y)$, such that $(x, y) \in E$. Hence, for every $l \in [d]$, a concatenation of these values is an eigenvector of $\hat{G}$ corresponding to eigenvalue μ. These are exactly the vectors u^l defined above.
It remains to show that all eigenvectors are obtained in this way. Every irreducible representation ρ_i of dimension d_i defines a matrix A_{s,ρ_i} that has $n \cdot d_i$ eigenvectors. Each of these defines d_i eigenvectors of $\hat{G}$, for a total of $n \sum d_i^2 = |G||\Gamma| = |\hat{G}|$. Hence, it suffices to show that these vectors are linearly independent.
We shall need some ungainly indexing. Denote by $u^{i,p,l}$, for $i \in [k]$, $p \in [d_i n]$ and $l \in [d_i]$, the eigenvector obtained in the following way. Let $v^{i,p}$ be the p'th eigenvector of A_{s,ρ_i}. Let $w_{g,x}$ be as above. $u^{i,p,l}$ is the vector obtained from concatenating the l'th coordinates of the $w_{g,x}$'s (as above). Consider a linear combination of these vectors that equals zero: $\sum_{i,p,l} c_{i,p,l}u^{i,p,l} = 0$. In particular, for every $g \in \Gamma$ and $x \in V$, we have that:

$$\begin{aligned} 0 &= \sum_{i,p,l} c_{i,p,l}u^{i,p,l}_{g,x} \\ &= \sum_{i,p,l} c_{i,p,l} < \rho_i(g)_{\vec{l}}, v^{i,p}_x > \\ &= \sum_{i,p,l,h} c_{i,p,l}\rho_i(g)_{l,h}(v^{i,p}_x)_h \\ &= \sum_{i,l,h} \rho_i(g)_{l,h} \sum_p c_{i,p,l}(v^{i,p}_x)_h, \end{aligned}$$

where $\rho_i(g)_{l'}$ is the l'th row of $\rho_i(g)$.
Define $r^{i,l,h} \in \mathbb{C}^{|\Gamma|}$ by $r_g^{i,l,h} = \rho_i(g)_{l,h}$. A well known corollary from Shur's Lemma is that these vectors are linearly independent. Hence, we have that for every x, i, l and h: $\sum_p c_{i,p,l}(v_x^{i,p})_h = 0$. Fixing i and l, and writing this as an equation in vector form, we get:

$$\sum_p c_{i,p,l} v^{i,p} = 0.$$

But the $v^{i,p}$ are linearly independent, so for all p, $c_{i,p,l} = 0$. As this is true for any choice of i and l, we get that all coefficients are zero, and hence the vectors $\{u^{i,p,l}\}$ are linearly independent. ∎

3.6 Experimental results

In this section we describe various heuristics for finding a good signing. We've tested these heuristics on two datasets - a list of all 41301 3-regular graphs on 18 vertices and a list of 1000 randomly generated 3-regular graphs. We assume that the graph is connected - otherwise the heuristic is applied to each connected component. The 12 heuristics we've tested are the following:

1. Choose a signing for each edge independently and uniformly at random.
2. Choose uniformly at random a signing that assigns half the edges 1 and half the edges -1.
3. Pick an arbitrary spanning tree of the graph. Choose a random signing for edges outside the tree. For each edge marked with -1, flip the value of the edges along the path in the spanning tree that connects its two vertices.
4. Pick an arbitrary spanning tree of the graph. Sign its edges to -1, and the other edges to 1.
5. Choose a spanning tree at random, via a random walk. Sign its edges -1, and the rest 1.
6. Choose a spanning tree at random, via a random walk. Sign its edges -1, and the rest 1. Repeat the following step until the spectral radius no longer decreases: Let v be an eigenvector associated with the spectral radius. Let (i, j) be an edge outside the tree such that $v_i \cdot v_j$ is maximal. Change the sign of (i, j) to -1. Choose an edge (i', j') from among the edge in the cycle that was just created such that $v_{i'} \cdot v_{j'}$ is minimal, and change its sign to 1.
7. Pick a random maximal matching in the graph. Sign its edges to -1, and the other edges to 1.
8. Choose some subset of vertices S. Go over all sign assignments to the vertices in S. For each edge e let v be the vertex in S closest to it (if well defined). Let d be the distance from e to v (if well defined). If both v and d are well defined, the sign of e is the parity of d times the sign of v. Otherwise, it is 1.
   Here we choose $|S| = 4$.

9. Start with all edges signed 1. Go over the edges in an arbitrary order, and choose the preferable value with respect to the spectral radius of the current signed matrix.

10. First partition the edges into pairs of adjacent edges and "loner" edges in the following way. To begin with, all edges are "free". Each edge chooses one of its "free" neighbors uniformly at random. If two edges choose each other they are added to the partition as a pair, and are no longer free. If an edge has no free neighbors, it is added to the partition as a "loner". Pairs are signed arbitrarily with opposite signs. Loners are signed at random.

11. Start with all signs being 1. Look at the eigenvectors associated with eigenvalues, which, in absolute value, are at least 0.999 the spectral radius. Choose an edge such that the product of the entries associated with its vertices contribute most to the expression vA_sv (among all such eigenvectors v, normalized to 1). Flip the sign of this edge. Repeat while this decreases the spectral radius.

12. As 11, but start from a random signing.

Performance on dataset of all 41301 3-regular graphs on 18 vertices.						
Algorithm	Iter	Good	Always	% Rmnjn	Avg. Rmnjn	Avg. Rad.
1	100	41301	2	0.82	41297	2.75
2	100	41301	1	0.82	41300	2.75
3	100	41297	3	0.77	38114	2.76
4	1	30999	30999	0.75	30999	2.76
5	100	41301	28	0.85	41285	2.74
6	100	41301	32216	1.00	41301	2.63
7	100	41301	162	0.80	39806	2.76
8	10	37884	2730	0.60	27743	2.81
9	1	0	0	0.00	0	2.93
10	100	41301	5	0.79	41237	2.76
11	1	4078	4078	0.10	4078	2.89
12	100	41301	710	0.89	41301	2.73

Figure 3.2: Iter - Number of iterations per graph.
Good - Number of graphs such that in some iteration the signing was Ramanujan.
Always - Number of graphs such that in all iterations the signing was Ramanujan.
% Rmnjn - Fraction of signings that were Ramanujan.
Avg. Rmnjn - Number of graphs with average spectral radius of signings $\leq 2\sqrt{2}$.
Avg. Rad. - Average spectral radius ($2\sqrt{2} \approx 2.8284$).

Performance on dataset of all 41301 3-regular graphs on 18 vertices. Distribution of graphs by number of iterations giving a Ramanujan signing.										
Alg.	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
1	0	0	0	17	164	847	3358	11031	20768	5116
2	0	0	0	24	179	908	3356	11048	20782	5004
3	16	111	315	907	1907	2643	3707	7989	18581	5125
5	0	0	2	18	135	603	2305	6697	18078	13463
6	0	0	0	0	0	0	3	45	194	41059
7	15	84	274	755	1405	2138	3370	7561	16942	8757
10	0	0	18	118	565	2217	5063	11851	17253	4216
12	0	0	0	4	17	139	962	3959	13722	22498

Performance on 999 3-regular graphs on 100 vertices.						
Algorithm	Iter	Good	Always	% Rmnjn	Avg. Rmnjn	Avg. Rad.
1	100	999	0	0.77	998	2.81
2	100	999	0	0.77	996	2.81
3	100	999	0	0.77	982	2.81
4	1	817	817	0.82	817	2.81
5	100	999	0	0.81	999	2.81
6	100	999	626	0.99	999	2.77
7	100	999	0	0.77	990	2.81
9	1	0	0	0.00	0	2.99
10	100	999	0	0.76	996	2.81
11	1	0	0	0.00	0	2.94
12	100	999	0	0.83	999	2.81

Figure 3.3: Iter - Number of iterations per graph.
Good - Number of graphs such that in some iteration the signing was Ramanujan.
Always - Number of graphs such that in all iterations the signing was Ramanujan.
% Rmnjn - Fraction of signings that were Ramanujan.
Avg. Rmnjn - Number of graphs with average spectral radius of signings $\leq 2\sqrt{2}$.
Avg. Rad. - Average spectral radius ($2\sqrt{2} \approx 2.8284$).

Performance on 999 3-regular graphs on 100 vertices.

Distribution of graphs by number of iterations giving a Ramanujan signing.

Alg.	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
1	0	0	0	0	1	20	108	546	319	5
2	0	0	0	0	3	16	117	541	314	8
3	0	0	1	13	3	13	111	511	338	9
5	0	0	0	0	1	8	51	358	549	32
6	0	0	0	0	0	0	0	0	0	999
7	1	0	1	3	6	30	146	465	341	6
10	0	0	0	1	4	26	171	487	306	4
12	0	0	0	0	0	8	33	261	604	93

Chapter 4

Graphs with bounded λ_2

In chapters 2 and 3 we were interested in regular graphs, where the degree is bounded, and the second eigenvalue is small. In this chapter we investigate what the graph looks like when the degree is linear in the number of vertices, yet the second eigenvalue remains bounded by some constant. Note that here we shall discuss the second largest eigenvalue, rather than the second largest in absolute value.
We show that if the graph is $n/2$-regular, then it is close to being a complete bipartite graph, and that for $0 < \delta < \frac{1}{2}$, and λ_2 there are only finitely many δn-regular graphs with second eigenvalue at most λ_2.
It is worth noting that graphs with bounded second eigenvalue have been previously studied. The apex of these works is probably that of Cameron, Goethals, Seidel and Shult, who characterize in [19] graphs with second eigenvalue at most 2.

4.1 $n/2$-regular graphs

In this section we consider the family $\mathbb{G}$ of $n/2$-regular graphs, and second largest eigenvalue λ_2 bounded by a constant. We prove that, asymptotically, they are nearly complete bipartite.

Definition 4.1.1. *Let G and H be two graphs on n vertices. We say that G and H are* close*, if there is a labeling of their vertices such that $|E(G) \triangle E(H)| = o(n^2)$.*

Theorem 4.1.1. *Every $G \in \mathbb{G}$ is close to $K_{n/2,n/2}$, where n is the number of vertices in G.*

Note 4.1.1. *By passing to the complement graph, if $\lambda_n = O(1)$, then G is close to the disjoint union of two cliques, $K_{n/2} \dot{\cup} K_{n/2}$.*

It is worthwhile noting that the only example we know of $n/2$-regular graphs with λ_2 bounded are constructed by taking a complete bipartite graph, and changing a constant number of edges for each vertex. These graphs are $O(n)$-close to being complete bipartite. Thus, we conjecture that perhaps a stronger notion of closeness be proved.

We need several lemmas before getting to the proof of the theorem. The first is the well-known expander mixing lemma (cf. [9]). The second is a special case of Simonovitz's stability theorem ([65]), for which we give a simple proof here. The third is a commonly used corollary of Szemeredi's Regularity Lemma. We shall also make use of the Regularity Lemma itself (see e.g. [24]).

Lemma 4.1.1. *Let G be an $\frac{n}{2}$-regular graph on n vertices with second largest eigenvalue λ_2. Then every subset of vertices with k vertices has at most $\frac{1}{4}k^2 + \frac{1}{2}\lambda_2 k$ internal edges.*

Lemma 4.1.2. *Let R be a triangle-free graph on n vertices, with $n^2/4 - o(n^2)$ edges. Then R is close to $K_{n/2,n/2}$. Furthermore, all but $o(n)$ of the vertices have degree $\frac{n}{2} \pm o(n)$.*

Proof: Denote by d_i the degree of the ith vertex in R, and by m the number of edges. Then:

$$\sum_{(i,j)\in E(R)} (d_i + d_j) = \sum_{i\in V(R)} d_i^2 \geq \frac{1}{n}(\sum_{i\in V(R)} d_i)^2 = \frac{4m^2}{n}.$$

Thus, there is some edge $(i,j) \in E(R)$ such that $d_i + d_j \geq \frac{4m}{n} = n - o(n)$. Let Γ_i and Γ_j be the neighbor sets of i and j. Since i and j are adjacent, and R has no triangles, the sets Γ_i and Γ_j are disjoint and independent. If we delete the $o(n)$ of vertices in $V\backslash(\Gamma_i \cup \Gamma_j)$ we obtain a bipartite graph. We have deleted only $o(n^2)$ edges, so the remaining graph still has $n^2/4 - o(n^2)$ edges. But this means that $|\Gamma_i|, |\Gamma_j| = \frac{n}{2} - o(n)$, and that the degree of each vertex in these sets is $\frac{n}{2} \pm o(n)$. ∎

Recall that the Regularity Lemma states that for every $\epsilon > 0$ and $m \in \mathbb{N}$ there's an M, such that the vertex set of every large enough graph can be partitioned into k subsets, for some $m \leq k \leq M$ with the following properties: All subsets except one, the "exceptional" subset, are of the same size. The exceptional subset contains less than an ϵ-fraction of the vertices. All but an ϵ-fraction of the pairs of subsets are ϵ-regular.
The regularity graph with respect to such a partition and a threshold d, has the k subsets as vertices. Two subsets, U_1 and U_2 are adjacent, if they are ϵ-regular, and $e(U_1, U_2) > dn^2$.

Lemma 4.1.3 ([24], Lemma 7.3.2). *Let G be a graph on n vertices, $d \in (0,1]$, $\epsilon = d^{-4}$. Let R be an ϵ-regularity graph of G, with (non exceptional) sets of size at least $\frac{s}{\epsilon}$, and threshold d. If R contains a triangle, then G contains a complete tripartite subgraph, with each side of size s.*

Corollary 4.1.1. *If $G \in \mathbb{G}$, and R is as in the lemma, with $s = 10\lambda_2$, then R is triangle free. In this case, if R has $\frac{k^2}{4} - o(k^2)$ edges, then R is close to complete bipartite.*

Proof: If R contains a triangle, then G contains a complete tripartite subgraph, with s vertices on each side. Let U be the set vertices in this subgraph. Then $e(U) = 3s^2 = 300\lambda_2^2$, but by lemma 4.1.1 $e(U) \leq 50\lambda_2^2$ - a contradiction. The second part now follows from Lemma 4.1.2. ∎

Proof: (Theorem) We would like to apply the Regularity Lemma to graphs in $\mathbb{G}$, and have $\epsilon = o(1)$, and $k = \omega(1)$ as well as $k = o(n)$. Indeed, this can be done. Since M depends only on m and ϵ, choose $d = o(1)$, and $m = \omega(1)$, such that the M given by the lemma satisfies $\frac{n}{(M+1)} \geq \frac{s}{\epsilon}$. As M depends only on m and ϵ, $\frac{M}{\epsilon}$ can be made small enough, even with the requirements on d and m.

Let R be the regularity graph for the partition given by the Regularity Lemma, with threshold d as above. Denote by k the number of sets in the partition, and their size by l (so $k \cdot l = n(1-\eta)$, for some $\eta \leq \epsilon$). We shall show that R is close to complete bipartite, and that G is close the graph obtained by replacing each vertex in R with l vertices, and replacing each edge in R by a $K_{l,l}$.

Call an edge in G (i) "irregular" if it belongs to an irregular pair; (ii) "internal" if it connects two vertices within the same part; (iii) "redundant" if it belongs to a pair of edge density smaller than d, or touches a vertex in the exceptional set. Otherwise (iv), call it "good".
Recall that $\epsilon = o(1)$, so only $o(k^2)$ pairs of sets are not ϵ-regular. Thus, G can have only $o(l^2k^2) = o(n^2)$ irregular edges. Also, $d = o(1)$, so the number of redundant edges is $k^2 \cdot o(l^2) + o(l)\frac{n}{2} = o(n^2)$. Finally, the number of internal edges is at most $\frac{1}{2}l^2k$, hence there are $\frac{n^2}{4} - o(n^2)$ good edges. The number of edges between two sets is at most l^2, so R must have at least

$$\frac{n^2 - o(n^2)}{4l^2} = \frac{k^2}{4} - o(k^2)$$

edges. The corollary implies that it is close to complete bipartite. By lemma 4.1.2, the valency of all but $o(k)$ of the vertices in R is indeed $\frac{k}{2} \pm o(k)$. This means that every edge in R corresponds to $l^2 - o(l^2)$ good edges in G (as the number of edges in R is also no more than $\frac{k^2}{4} + o(k^2)$).
To see that G is close to complete bipartite, let's count how many edges need to be modified. First, delete $o(n^2)$ edges that are not "good". Next add all possible $o(n^2)$ new edges between pairs of sets that have "good" edges between them. As R is close to complete bipartite, we need to delete or add all edges between $o(k^2)$ pairs. Each such step modifies l^2 edges, altogether $o(l^2k^2) = o(n^2)$ modifications. Finally, divide the $o(n)$ vertices of the exceptional set evenly between the two sides of the bipartite graph, and add all the required edges, and the tally remains $o(n^2)$. ∎

Note 4.1.2. *In essence, the proof shows that a graph with no dense induced subgraphs is close to complete bipartite. This claim is similar in flavor to Bruce Reed's* Mangoes and Blueberries *theorem [58]. Namely, that if every induced subgraph G' of G has an independent set of size $\frac{1}{2}|G'| - O(1)$, then G is close to being bipartite. The conclusion in Reed's theorem is stronger in that only a* linear *number of edges need to be deleted to get a bipartite graph.*

Note 4.1.3. *In fact, the proof gives something a bit stronger. Let $t_r(n)$ be the number of edges in an n-vertex complete r-partite graph, with parts of equal size. Using the general Stability Theorem ([65]) instead of Lemma 4.1.2, the same proof shows that if a graph has $t_n - o(n^2)$ edges and no dense induced subgraphs, then it is close to being complete r-partite.*

4.2 δn-regular graphs

In Theorem 4.1.1 we required that the degree is $n/2$. We can deduce from the theorem that this requirement can be relaxed:

Corollary 4.2.1. *Let $\mathbb{G}$ be a family of d-regular graphs, with $d \leq \frac{n}{2}$, (n being the number of vertices in the graph) and bounded second eigenvalue, then every $G \in \mathbb{G}$ is close to a complete bipartite graph.*

Proof: Let $M \in \mathbb{M}_n$ be the adjacency matrix of such a d-regular graph, and denote $\bar{M} = J - M$, where J is the all ones matrix. Consider the graph H corresponding to the following matrix:

$$N = \begin{pmatrix} M & \bar{M} \\ \bar{M}^t & M \end{pmatrix}$$

Clearly H is an n-regular graph on $2n$ vertices. Denote by (x, y) the concatenation of two n-dimensional vectors, x, y, into a $2n$ dimensional vector. Let v be an eigenvector of M corresponding to eigenvalue λ. It is easy to see that v is also an eigenvalue of $\bar{M}$: If $v = \vec{1}$ (and thus $\lambda = d$) it corresponds to eigenvalue $n - \lambda$, otherwise to $(-\lambda)$.
Thus, (v, v) and $(v, -v)$ are both eigenvectors of N. If $v = \vec{1}$ they correspond to eigenvalues n, $2d - n$, respectively, otherwise to 0, 2λ. Since the v's are linearly independent, so are the $2n$ vectors of the form (v, v) and $(v, -v)$: Consider a linear combination of these vectors that gives 0. Both the sum and the difference of the coefficients of each pair have to be 0, and thus both are 0. So we know the entire spectrum of N, and see, since $d \leq \frac{n}{2}$, that theorem 4.1.1 holds for it.
Let H' be a complete bipartite graph that is close to H. Since H differs from H' by $o(n^2)$ edges, the same holds for subgraphs over the same set of vertices. In particular, G is close to the subgraph of H' spanned by the first n vertices. Obviously, every such subgraph is itself complete bipartite. ∎

Corollary 4.2.2. *For every $0 < \delta < \frac{1}{2}$ and c, there are only finitely many δn-regular graphs with $\lambda_2 < c$.*

Proof: Consider such a graph with n large. By the previous corollary it is close to complete bipartite. Since it is also regular, it must be close to $K_{\frac{n}{2},\frac{n}{2}}$, which contradicts the constraint $\delta < \frac{1}{2}$. ∎

4.3 Graphs with both λ_2 and λ_{n-1} bounded by a constant

Theorem 4.1.1 can loosely be stated as follows: A regular graph with spectrum similar to that of a bipartite graph (λ_1 being close to $n/2$ and λ_2 being close to 0) is close to being complete bipartite. We conclude this section by noting that if we strengthen the assumption on how close the spectrum of a graph is to that of a bipartite graph, we get a stronger result as to how close it is to a complete bipartite graph.

Theorem 4.3.1. *Let $\mathbb{G}$ be a family of $\frac{n}{2}$-regular graphs on n vertices, with both λ_2 and λ_{n-1} bounded by a constant. Then every $G \in \mathbb{G}$ is close to a $K_{\frac{n}{2},\frac{n}{2}}$, in the sense that such a graph can be obtained from G by modifying a linear number of edges for $O(\sqrt{n})$ vertices of G, and $O(\sqrt{n})$ edges for the rest.*

Proof: First note that it follows that $\lambda_n(G) = -\frac{n}{2} + O(1)$. Take $G \in \mathbb{G}$, and let A be its adjacency matrix. Clearly $tr(A^2) = \frac{n^2}{2}$. If $\lambda_{n-1}(G) = -O(1)$, then

$$\frac{n^2}{2} = tr(A^2) = \lambda_1^2 + \lambda_n^2 + \sum_{i=2,\ldots,n-1} \lambda_i^2$$

Since $\lambda_1 = \frac{n}{2}$

$$\lambda_n^2 = \frac{n^2}{2} - \left(\frac{n}{2}\right)^2 - \sum_{i=2,\ldots,n-1} \lambda_i^2$$

As $\lambda_2, \ldots, \lambda_{n-1} = O(1)$ we have

$$\lambda_n^2 = \frac{n^2}{4} + O(n)$$

And since λ_n is negative, and is smaller than λ_1 in absolute value:

$$\lambda_n = -\frac{n}{2} + O(1).$$

Let x be an eigenvector corresponding to λ_n. Suppose, w.l.o.g that $||x||_\infty = 1$ and that $x_v = 1$. Denote $A = \{u : x_u \leq -(1 - \frac{1}{\sqrt{n}})\}$, and $B = \{w : x_w \geq (1 - \frac{1}{\sqrt{n}})\}$. The eigenvalue condition on v entails:

$$\frac{n}{2} - O(1) = - \sum_{u:(u,v)\in E} x_u.$$

Thus, there is a vertex u such that $x_u \leq -(1 - O(\frac{1}{n}))$. It is not hard to verify that v must have $\frac{n}{2} - O(\sqrt{n})$ neighbors in A, and that u must have $\frac{n}{2} - O(\sqrt{n})$ neighbors in B.

Now denote $A' = \{u : x_u \leq -\frac{1}{2}\}$, and $B' = \{w : x_w \geq \frac{1}{2}\}$. Again, it is not hard to check that each vertex in A must have $\frac{n}{2} - O(\sqrt{n})$ neighbors in B', and vice versa. Thus, delete the $O(\sqrt{n})$ vertices that are neither in A nor in B. For each remaining vertex in A (similarly in B), its degree is at most $\frac{n}{2}$, and at least $\frac{n}{2} - O(\sqrt{n})$. It has $\frac{n}{2} - O(\sqrt{n})$ neighbors in B, so the number of its neighbors in A, and the number of its non-neighbors in B is $O(\sqrt{n})$. By deleting and adding $O(\sqrt{n})$ edges to each vertex, we get a complete bipartite graph. ∎

Chapter 5

Monotone maps, sphericity and the KL-divergence

5.1 Introduction

The motivation for the work in this chapter arised from our involvement in the ProtoNet project ([63]). The aim of this project is to generate a (meaningful) hierarchical clustering of all known protein sequences in an unsupervised manner. Roughly speaking, the collection of all these sequences can be thought of as a finite metric space, where the points are the sequences, and the distances are (roughly) the edit distances between the sequences.
A hierarchical clustering of a metric space, at least for our purpose, is a binary tree with leaves corresponding to the points in the metric space. Each internal node in the tree defines a cluster of points in a natural way - those which correspond to leaves that are descendents of this internal node. A cluster is considered "good" if the points in it are close to each other. The goal of the ProtoNet project is to create a "good" hierarchical clustering, where clusters reveal connections between proteins, that are not readily evident otherwise.
A hierarchical clustering algorithm usually builds the tree structure in a greedy manner. It starts with the leaves of the tree. Think of them as a forest of trees of depth 0. It then iteratively chooses two "close" trees in the forest, and merges them into a single tree, by adding a new node as the parent of the roots of both these trees. For the algorithm to be well defined, one needs to define a distance between trees. Let C_1 and C_2 be the clusters defined by the leaves of two trees in the forest. In the ProtoNet project several ways defining the distance between two clusters as the average over all pairwise distances are used. Another plausible way of doing so (though not implemented in ProtoNet) is to define the distance as $\max_{i\in C_1, j\in C_2} dist(i,j)$ or $\min_{i\in C_1, j\in C_2} dist(i,j)$.
A different approach to clustering of a metric space is to first embed the space into, say, Euclidean space, and then do the the clustering in that space. One could hope that if the structure of the metric space is not lost in the embedding, the additional structure induced by having the points in Euclidean space will give rise to better clustering. Indeed, many clustering algorithms are designed to address only the Euclidean problem.
Euclidean embeddings of finite metric spaces have been extensively studied, with the aim of find-

ing an embedding that doesn't distort the metric too much. We refer the reader to the survey papers of Indyk ([41]) and Linial ([46]), as well as chapter 15 of Matoušek's Discrete Geometry book [53]. Here we focus on a different type of embeddings. Namely, those that preserve the order relation of the distances. We call such embeddings *monotone*. There are quite a few applications that make this concept natural and interesting, since there are numerous algorithmic problems whose solution depends only on the order among the distances. Specifically, questions that concern nearest neighbors. The notion of monotone embeddings suggests the following general strategy toward the resolution of such problems. Namely, embed the metric space at hand monotonically into a "nice" space, for which good algorithms are known to solve the problem. Solve the problem in the "nice" space - the same solution applies as well for the original space. "Nice" often means a low dimensional normed space. Thus, we focus on the minimal dimension which permits a monotone embedding.

In section 5.2 we observe that any metric on n points can be monotonically embedded into an n-dimensional Euclidean space, and that the bound on the dimension is asymptotically tight. The embedding clearly depends only on the order of the distances (Lemma 5.2.1). We show that for almost every ordering of the $\binom{n}{2}$ distances among n points, the host space of a monotone embedding must be $\Omega(n)$-dimensional. Similar bounds are given for embeddings into l_∞, and some bounds are also deduced for other norms.

Next we consider embeddings that are even less constrained. Given a metric space (X, δ) and some threshold t, we seek a mapping f that only respects this threshold. Namely, $||f(x) - f(y)|| < 1$ iff $\delta(x, y) < t$. The input to this problem can thus be thought of as a graph (adjacency indicating distances below the threshold t). The minimal dimension d, such that a graph G can be mapped this way into l_2^d is known as the *sphericity* of G, and denoted $Sph(G)$. Reiterman, Rödl and Šiňajová ([61]) show that the sphericity of $K_{n,n}$ is n. This is, then, an explicit example of a metric space which requires linear dimension to be monotonically embedded into l_2. Other than that, the best lower bounds previously known to us are logarithmic. In section 5.3 we prove a better lower bound, namely that for $0 < \delta \leq \frac{1}{2}$, $Sph(G) = \Omega(\frac{n}{\lambda_2+1})$, for any n-vertex δn-regular graph, with bounded diameter. Here λ_2 is the second largest eigenvalue of the graph. We also show examples of quasi-random graphs of logarithmic sphericity. This is somewhat surprising since quasi-random graphs tend to behave like random graphs, yet the latter have linear sphericity.

Finally, we consider "soft clustering". Here the objective is not to partition the metric space into clusters, but rather define a distribution over clusters. Intuitively, this distribution measures how likely it is that the point belongs to each cluster.

Thus, suppose that each point i in the metric space has a distribution p_i (over k clusters) associated with it. Consider an operator $\phi : \Delta^{k-1} \times \Delta^{k-1} \to \mathbb{R}_+$, where Δ^{k-1} is the $(k-1)$-dimensional simplex. We can think of our distance function as arising from this operator, that is, $dist(i, j) = \phi(p_i, p_j)$. Given such an operator, a natural question is which metric spaces can arise form it.

The case when ϕ is the inner product operator has been studied extensively ([13]), yet is still largely open. Here we study the KL-divergence operator, and show that all metric spaces may arise from it.

5.2 Monotone Maps

5.2.1 Definitions

Let $X = ([n], \delta)$ be a metric space on n points, such that all pairwise distances are distinct. Let $||\ ||$ be a norm on $\mathbb{R}^d$. We say that $\phi : X \to (\mathbb{R}^d, ||\ ||)$ is a *monotone map* if for every $w, x, y, z \in X$, $\delta(x, y) < \delta(w, z) \Leftrightarrow ||\phi(x) - \phi(y)|| < ||\phi(w) - \phi(z)||$.

We denote by $d(X, ||\ ||)$ the minimal t such that there exists a monotone map from X to $(\mathbb{R}^t, ||\ ||)$. We denote by $d(n, ||\ ||) = \max_X d(X, ||\ ||)$, the smallest dimension to which every n point metric can be mapped monotonically.

The first thing to notice is that we are actually concerned only with the *order* among the distances between the points in the metric space, and not with the actual distances. Let (X, δ) be a finite metric space, and let ρ be a linear order on $\binom{X}{2}$. We say that ρ and (X, δ) are *consistent* if for every $w, x, y, z \in X$, $\delta(x, y) < \delta(w, z) \Leftrightarrow (x, y) <_\rho (w, z)$.

We start with an easy, but useful observation.

Lemma 5.2.1. *Let X be a finite set. For every strict order relation ρ on $\binom{X}{2}$, there exists a distance function δ on X, that is consistent with ρ.*

Proof: Let $\{\epsilon_{ij}\}_{(i,j)\in\binom{X}{2}}$ be small, non-negative numbers, ordered as per ρ. Define $\delta(i, j) = 1+\epsilon_{ij}$. It is obvious that δ induces the desired order on the distances of X, and, that if the ϵ's are small, the triangle inequality holds. ∎

When we later (Section 5.2.3) use this observation, we refer to it as a *standard ϵ-construction*, where $\epsilon = \max \epsilon_{ij}$. It is not hard to see that this metric is Euclidean, that is, the resulting metric can be isometrically embedded into l_2, see Lemma 5.2.3 below.

We say that an order relation ρ on $\binom{[n]}{2}$ is *realizable* in $(\mathbb{R}^d, ||\ ||)$ if there exists a metric space (X, δ) on n points which is consistent with ρ, and a monotone map $\phi : X \to \mathbb{R}^d$. We say that ϕ is a realization of ρ. (In other words, $d(n, ||\ ||)$ is the minimal d such that any linear order on $\binom{[n]}{2}$ is realizable in $(\mathbb{R}^d, ||\ ||)$.)

We denote by $J = J_n$ the $n \times n$ all ones matrix, and by PSD_n the cone of real symmetric $n \times n$ positive semidefinite matrices. We omit the subscript n when it is clear from the context.

Finally, for a graph G, and U, V subsets of its vertices, we denote by $e(U, V) = |\{(u, v) \in E(G) : u \in U, v \in V\}|$, and $e(U) = |\{(u, u') \in E(G) : u, u' \in U\}|$.

5.2.2 Monotone Maps into l_∞.

Lemma 5.2.2. $\frac{n}{2} - 1 \leq d(n, l_\infty) \leq n$

Proof: It is well known that any metric X on n points can be embedded into l_∞^n isometrically, hence $d(n, l_\infty) \leq n$.

For the lower bound, we define a metric space (X, δ) with $2n+2$ points that cannot be realized in l_∞^n. By lemma 5.2.1, it suffices to define an ordering on the distances. In fact, we define only a partial order, any linear extension of which will do. The $2n+2$ points come in $n+1$ pairs, $\{x_i, y_i\}_{i=1,...,n+1}$. If $z \notin \{x_i, y_i\}$, we let $\delta(x_i, y_i) > \delta(x_i, z), \delta(y_i, z)$. Assume for contradiction that a monotone map ϕ into l_∞^n does exist. For each pair (x, y) define $j(x, y)$ to be some index i for which $|\phi(x)_i - \phi(y)_i|$ is maximized, that is, an index i for which $|\phi(x)_i - \phi(y)_i| = \|\phi(x) - \phi(y)\|_\infty$.

By the pigeonhole principle there exist two pairs, say (x_1, y_1) and (x_2, y_2), for which $j(x_1, y_1) = j(x_2, y_2) = j$. It is easy to verify that our assumptions on the four real numbers $\phi(x_1)_j$, $\phi(x_2)_j$, $\phi(y_1)_j$, $\phi(y_2)_j$, are contradictory. Thus $d(n, l_\infty) \geq \frac{n}{2} - 1$. ∎

5.2.3 Monotone Maps into l_2.

Lemma 5.2.3. *$\frac{n}{2} \leq d(n, l_2) \leq n$. Furthermore, for every $\delta > 0$, and every large enough n, almost no linear orders ρ on $\binom{[n]}{2}$ can be realized in dimension less than $\frac{n}{2+\delta}$.*

Note 5.2.1. *The upper bound is apparently folklore. As we could not find a reference for it, we give a proof here.*

Proof: Let ρ be a linear order on $\binom{[n]}{2}$. Let ϵ be a real symmetric matrix with the following properties:

- $\epsilon_{ii} = 0$ for all i.
- $\frac{1}{n} > \epsilon_{ij} > 0$, for all $i \neq j$.
- The numbers $\epsilon_{i,j}$ are consistent with the order ρ.

Since the sum of each row is strictly less than one, all eigenvalues of ϵ are in the open interval $(-1, 1)$. It follows that the matrix $I - \epsilon$ is positive definite. Therefore, there exists a matrix V such that $VV^t = I - \epsilon$. Denote the i'th row of V by v_i. Clearly, the v_i's are unit vectors, and $< v_i, v_j > = -\epsilon_{i,j}$ for $i \neq j$. Therefore, $||v_i - v_j||_2^2 = < v_i, v_i > + < v_j, v_j > -2 < v_i, v_j > = 2 + 2\epsilon_{i,j}$. It follows that the map $\phi(i) = v_i$ is a realization of ρ, and the upper bound is proved. In fact, one can add another point without increasing the dimension, by mapping it to 0, and perturbing the diagonal.
For the lower bound, it is essentially known that if X is the metric induced by $K_{n,n}$, then $d(X, l_2) \geq n$. We discuss this in more detail in the next section.

For the second part of the lemma we need a bound on the number of *sign-patterns* of a sequence of real polynomials. Let $p_1, ..., p_m$ be real polynomials in l variables of (total) degree d, and let $x \in \mathbb{R}^l$ be a point where none of them vanish. The sign-pattern at x is $(\text{ sign}(p_1(x)), ..., \text{ sign}(p_m(x)))$. Denote the total number of different sign-patterns that can be obtained from $p_1, ..., p_m$ by $s(p_1, ..., p_m)$. A variation of the Milnor-Thom theorem [55] due to Alon, Frankl and Rödl [5] shows:

Theorem 5.2.1. *[5] Let $p_1, ..., p_m$ be real polynomials as above. Then for any integer k between 1 and m:*

$$s(p_1, ..., p_m) \leq 2kd \cdot (4kd - 1)^{l + \frac{m}{k} - 1}$$

Set $n = c \cdot d$, for some constant c, and $l = n \cdot d$. Consider a point $x \in R^l$, and think of it as an $n \times d$ matrix. Denote the ith row of this matrix by x_i. As before, x *realizes* an order ρ on $\binom{[n]}{2}$ if the distances $||x_i - x_j||$ are consistent with ρ.

For two different pairs, (i_1, j_1) and (i_2, j_2), define the polynomial

$$p_{(i_1,j_1),(i_2,j_2)}(x) = ||x_{i_1} - x_{j_1}||^2 - ||x_{i_2} - x_{j_2}||^2.$$

The list contains $m = \binom{\binom{n}{2}}{2}$ polynomials of degree 2. Note that there is a $1:1$ correspondence between orders on $\binom{[n]}{2}$ and sign-patterns of $p_1, ..., p_m$, thus no more than $s = s(p_1, ..., p_m)$ orders may be realized in l_2^d.

Take $k = \mu n^2$, for some large constant μ. Then $\log s$ is approximately $2cd^2 \log d$. By contrast, that total number of orders is $\binom{n}{2}!$, so its log is about $c^2 d^2 \log d$. If c is bigger than 2, almost all order relations can not be realized. ∎

5.2.4 Other Norms

We conclude this section with three easy observations about monotone maps into other normed spaces. The first gives an upper bounds on the dimension required for embedding into l_p:

Lemma 5.2.4. $d(n, l_p) \leq \binom{n}{2}$.

Proof: By Lemma 5.2.3, any metric space on n points can be mapped monotonically into l_2. It is known (see [23] and also chapter 15 of [53]) that any l_2 metric on n points can be isometrically embedded into $\binom{n}{2}$-dimensional l_p. The composition of these mappings is a monotone mapping of the metric space into $\binom{n}{2}$-dimensional l_p. ∎

The second observation is that the second part of Lemma 5.2.3 holds for any l_p, where p is even and independent of n. Namely:

Lemma 5.2.5. $d(n, l_{2p}) = \Omega(n)$
(the constant of proportionality depends on p)

The proof is essentially the same as that in Lemma 5.2.3. Simply repeat the argument with polynomials of degree $2p$ rather than quadratic polynomials.

The third observation gives a lower bound for arbitrary norms. We first note the following:

Lemma 5.2.6. *Let $||\ ||$ be an arbitrary n-dimensional norm and let $x_1, ..., x_{5^n}$ be points in $\mathbb{R}^n$, such that $||x_i - x_j|| > 1$ for all $i \neq j$. Then there exits a pair (x_i, x_j) such that $||x_i - x_j|| \geq 2$*

Proof: Denote by v the volume of B, the unit ball in $(\mathbb{R}^n, ||\ ||)$. The translates $x_i + \frac{1}{2}B$ are obviously non intersecting, so the volume of their union is $(\frac{5}{2})^n v$. Assume for contradiction that all pairwise distances are less than 2, then all these balls are contained in a single ball of radius less than $\frac{5}{2}$. But this is impossible, since the volume of this ball is less than $(\frac{5}{2})^n v$. ∎

Note that the l_∞ norm shows that indeed an exponential number of points is required for the lemma to follow. We do not know, however, the smallest base of the exponent for which the claim holds. The determination of this number seems to be of some interest.

Corollary 5.2.1. *There exists an n-point metric spaces (X, δ) such that for any norm $||\ ||$, $d(n, ||\ ||) = \Omega(\log n)$.*

Proof: We construct a distance function on $5^n + 1$ points which can not be realized in any n-dimensional norm. By lemma 5.2.1 it suffices to define a partial order on the distances. Denote the points in the metric space $0, \ldots, 5^n$. Let the distance between 0 and any other point be smaller than any distance between any two points $i \neq j > 0$. Consider a monotone map ϕ of the metric space into n-dimensional normed space. Assume, w.l.o.g., that $\min_{i,j=1,\ldots,5^n} ||\phi(i) - \phi(j)|| = 1$. By the previous lemma there exists a pair of points, $i, j \neq 0$, such that $||\phi(i) - \phi(j)|| > 2$. But for ϕ to be monotone it must satisfy $||\phi(0) - \phi(i)|| < 1$ and $||\phi(0) - \phi(j)|| < 1$, contradicting the triangle inequality. ∎

5.3 Sphericity

So far we have concentrated on embeddings of a metric space into a normed space, that preserve the order relations between distances. However, in the examples that gave us the lower bounds for l_∞ and for arbitrary norms, we actually only needed to distinguish between "long" and "short" distances. This motivates the introduction of a broader class of maps, that need only respect the distinction between short and long distances. More formally, let $X = ([n], \delta)$ be a metric space. Its *proximity graph* with respect to some threshold τ, is a graph on n vertices, with an edge between i and j iff $\delta(i, j) \leq \tau$. An embedding of a proximity graph, is a mapping ϕ of its vertices into normed space, such that $||\phi(i) - \phi(j)|| < 1$ iff (i, j) is an edge in the proximity graph (We assume that no distance is exactly 1). The minimal dimension in which a graph can be so embedded (in Euclidean space) was first studied by Maehara in [50] under the name *sphericity*, and denoted $Sph(G)$. Following this terminology, we call such an embedding *spherical*.
The sphericity of graphs was further studied by Maehara and Frankl in [28], and then by Reiterman, Rödl and Šiňajová in [61], [60], [62]. Breu and Kirkpatrick have shown in [16] that it is NP-hard to recognize graphs of sphericity 2 (also known as *unit disk graphs*) and graphs of sphericity 3. We refer the reader to [60] for a survey of most known results regarding this parameter, and mention only a few of them here.

Theorem 5.3.1. *Let G be graph on n vertices with minimal degree δ. Let λ_n the least eigenvalue of its adjacency matrix.*

1. *$Sph(K_{m,n}) \leq m + \frac{n}{2} - 1$ [50].*

2. $Sph(G) = O(\lambda_n^2 \log n)$ *[28].*
3. $Sph(G) = O((n-\delta)\log(n-\delta))$ *[60].*
4. $Sph(K_{n,n}) \geq n$ *[61].*
5. *All but a* $\frac{1}{n}$ *fraction of graphs on* $n > 37$ *vertices have sphericity at least* $\frac{n}{15} - 1$ *[60].*
6. $Sph(G) \geq \frac{\log \alpha(G)}{\log(2r(G)+1)}$*, where* $\alpha(G)$ *is the independence number of* G*, and* $r(G)$ *is its radius [61].*

The first thing to notice is that any lower bower on the sphericity of some graph on n vertices is also a lower bound on $d(n, l_2)$. In particular, the fact that $Sph(K_{n,n}) \geq n$ proves the lower bound in Lemma 5.2.3. (Similarly, any upper bound on the former also applies to the latter.)
In this section we are interested in graphs of large sphericity. The above results tell us that they exist in abundance, yet that graphs of very small or very large degree have small sphericity (the maximal degree is an upper bound on $|\lambda_n|$, hence by (2) the sphericity is small if all degrees are small). Other than the complete bipartite graph, the above results do not point out an explicit graph with super-logarithmic sphericity.

5.3.1 Upper bound on margin

Following Frankl and Maehara [28], consider an embedding of a proximity graph where there is a large margin between short and long distances. In such a situation, the Johnson-Lindenstrauss Lemma ([43]) would yield a spherical embedding into lower dimension. Specifically suppose all short distances are at most $1 - m$, and all long distances are at least $1 + m$. The Johnson-Lindenstrauss Lemma says that the points can be embedded in dimension $\frac{1}{\epsilon^2}\log n$ with distortion at most $1 + \epsilon$. In particular, if we can find an initial spherical embedding where $(1+m)/(1+\epsilon) > (1-m)(1+\epsilon)$, then the Lemma yields a new spherical embedding, into dimension $\frac{1}{\epsilon^2}\log n$.
In other words, an embedding with margin m allows ϵ to be taken (when m is small) roughly as big as m. For this to give sub-linear sphericity, we need that $\epsilon^2 = \omega(\frac{\log n}{n})$, or

$$m = \omega(\sqrt{\frac{\log n}{n}}).$$

Alas, we show that for most regular graphs the margin can not be so large:

Theorem 5.3.2. *Let* G *be a* δn*-regular graph, with second eigenvalue* $\lambda_2 > \frac{2}{n}$*. Let* ϕ *be a spherical embedding of* G*. Denote* $a = \max_{u \sim v} ||\phi(u) - \phi(v)||_2^2$*, and* $b = \min_{u \not\sim v} ||\phi(u) - \phi(v)||_2^2$*. Then* $b - a = O(\frac{\lambda_2 + \delta}{\delta n})$.

Think of δ as fixed. For a random δn-regular graph, w.h.p. $\lambda_2 = O(\sqrt{n})$. Hence, w.h.p. in any spherical embedding the margin is $O(\frac{1}{\sqrt{n}})$, and in particular does not allow for dimension reduction.

Proof: Denote $m = \min\{1-a, b-1\}$, and for a vertex i, denote $v_i = \phi(i)$. The largest value m can attain, over all embeddings ϕ, is given by the following quadratic semidefinite program:

$$\begin{array}{ll} \max m & \\ s.t. \forall (i,j) \in E(G) & ||v_i - v_j||^2 \leq 1-m \\ \forall (i,j) \notin E(G) & ||v_i - v_j||^2 \geq 1+m \end{array}$$

Its dual turns out to be:

$$\begin{array}{ll} \min \frac{1}{2} trA & \\ s.t. & A \in PSD \\ \forall (i,j) \in E(G) & A_{ij} \leq 0 \\ \forall (i,j) \notin E(G), i \neq j & A_{ij} \geq 0 \\ \forall i & \sum_{j=1,\ldots,n} A_{ij} = 0 \\ & \sum_{i \neq j} |A_{ij}| = 1 \end{array}$$

Equivalently, we can drop the last constraint, and change the objective function to $\min \frac{trA}{\sum_{i \neq j} |A_{ij}|}$. Next we construct an explicit feasible solution for the dual program, and conclude from it a bound on m.

Let M be the adjacency matrix of G. Define $A = I + \alpha J - \beta M$. To satisfy the constraints we need:

$$\begin{array}{l} A \in PSD \\ \beta \geq \alpha \geq 0 \\ 1 + \alpha n - \beta \delta n = 0 \end{array}$$

The last condition implies $\alpha = \beta\delta - \frac{1}{n}$, so it follows that $\beta \geq \alpha$, and the constraint on β is $\beta \geq \frac{1}{\delta n}$.

Now, since we assume that the graph is δn-regular, its Perron eigenvector is $\vec{1}$, corresponding to eigenvalue δn. Therefore, we can consider the eigenvectors of M to be eigenvectors of J and I as well, and hence also eigenvectors of A. If $\lambda \neq \delta n$ is an eigenvalue of M, then $1 - \beta\lambda$ is an eigenvalue of A, corresponding to the same eigenvector. Denote by λ_2 the second largest eigenvalue of M, then in order to satisfy the condition $A \in PSD$ it is enough to set $\beta = \frac{1}{\lambda_2}$, in which case all the constraints are fulfilled.

We conclude that:

$$\begin{aligned} m &\leq \frac{trA}{\sum_{i \neq j} |A_{ij}|} = \frac{n(1+\alpha)}{\delta n^2(\beta - \alpha) + ((1-\delta)n^2 - n)\alpha} \\ &= \frac{n + \frac{\delta n}{\lambda_2} - 1}{\delta n(\frac{n+\delta n}{\lambda_2} - 1) + ((1-\delta)n - 1)(\frac{\delta n}{\lambda_2} - 1)} < 4\frac{1 + \frac{\delta}{\lambda_2}}{\frac{\delta n}{\lambda_2}} = 4\frac{\lambda_2 + \delta}{\delta n}. \end{aligned}$$

In particular, $b - a = O(\frac{\lambda_2+\delta}{\delta n})$. ∎

Think of δ as fixed. The theorem shows we can't get sub-linear sphericity via the Johnson-Lindenstrauss lemma, if $\lambda_2 = O(\sqrt{n \log n})$.
Frankl and Maehara show that the J-L lemma does yield sub-linear sphericity when $\lambda_n = o(\sqrt{\frac{n}{\log n}})$. Consequently, we get that a δn-regular graph can't have both $\lambda_2 = O(\sqrt{n \log n})$ and $\lambda_n = o(\sqrt{\frac{n}{\log n}})$. This is a bit more subtle than what one gets from the second moment argument, namely, that the graph can't have both $\lambda_2 = o(\sqrt{n})$ and $\lambda_n = o(\sqrt{n})$.

5.3.2 Lower bound on sphericity

Theorem 5.3.3. *Let G be a d-regular graph with diameter D and λ_2, the second largest eigenvalue of G's adjacency matrix, at least $d - \frac{1}{2}n$. Then $Sph(G) = \Omega(\frac{d-\lambda_2}{D^2(\lambda_2+O(1))})$.*

In the interesting range where $d \leq \frac{n}{2}$, and $\lambda_2 \geq 1$ the bound is $Sph(G) = \Omega(\frac{d-\lambda_2}{D^2\lambda_2})$. Note that for this bound to be linear, we would like λ_2 to be bounded (independently of d). However, as we saw in chapter 4, this holds only for graphs which are close to being complete bipartite.

Proof: It will be useful to consider the following operation on matrices. Let A be an $n \times n$ symmetric matrix, and denote by $\vec{a}$ the vector whose i-th coordinate is A_{ii}. Define $R(A)$ to be the $n \times n$ matrix with all rows equal to $\vec{a}$, and $C(A) = R(A)^t$. Define:

$$\breve{A} = 2A - C(A) - R(A) + J$$

First note that the rank of $\breve{A}$ and that of A can differ by at most 3. Now, consider the case where A is the Gram matrix of some vectors $v_1, ..., v_n \in R^d$. Then all diagonal entries of $\breve{A}$ equal one, and the (i, j) entry is $2 < v_i, v_j > - < v_i, v_i > - < v_j, v_j > +1 = 1 - ||v_i - v_j||^2$.

We will need the following lemma (see [40], p.175):

Lemma 5.3.1. *Let X be a real symmetric matrix, then $rank(X) \geq \frac{(trX)^2}{\sum_{i,j} X_{i,j}^2}$*

Applying this to $\breve{A}$, we conclude that:

$$rank(\breve{A}) \geq \frac{n^2}{n + \sum_{i \neq j}(1 - ||v_i - v_j||^2)^2} \tag{5.1}$$

Let $v_1, ..., v_n \in \mathbb{R}^d$ be an embedding of G. By the discussion above it is enough to show that

$$\sum_{i \neq j}(1 - ||v_i - v_j||^2)^2 = O(D^2 n^2 \frac{\lambda_2}{d - \lambda_2}). \tag{5.2}$$

By the triangle inequality $||v_i - v_j|| \leq D$ for any two vertices. So the LHS of (5.2) is bigger by at most a factor of D^2 than:

$$\sum_{(i,j) \notin E}(||v_i - v_j||^2 - 1) + \sum_{(i,j) \in E}(1 - ||v_i - v_j||^2) =$$

$$\sum_{(i,j)\notin E} ||v_i - v_j||^2 - \sum_{(i,j)\in E} ||v_i - v_j||^2 - \binom{n}{2} + nd \tag{5.3}$$

We can bound this sum from above, by solving the following SDP:

$$\begin{array}{ll} \max & \sum_{(i,j)\notin E} (V_{ii} + V_{jj} - 2V_{ij}) + \sum_{(i,j)\in E} (-V_{ii} - V_{jj} + 2V_{ij}) - \binom{n}{2} + nd \\ s.t. & V \in PSD \\ \forall (i,j) \in E & V_{ii} + V_{jj} - 2V_{ij} \leq 1 \\ \forall (i,j) \notin E & V_{ii} + V_{jj} - 2V_{ij} \geq 1 \end{array}$$

The dual problem is:

$$\begin{array}{ll} \min & \frac{1}{2} trA \\ s.t. & A \in PSD \\ \forall (i,j) \in E & A_{ij} \leq -1 \\ \forall (i,j) \notin E, i \neq j & A_{ij} \geq 1 \\ \forall i \in [n] & \sum_{j=1,\ldots,n} A_{ij} = 0 \end{array}$$

Let M by the adjacency matrix of the graph, and set $A = (\alpha d - n)I + J - \alpha M$, where $\alpha \geq 2$ will be determined shortly. This takes care of the all constraints except for $A \in PSD$. Note that since M is regular, its eigenvectors are also eigenvectors of A. Moreover, if $Mu = \lambda u$ for a non constant u, then $Au = \alpha d - n - \alpha\lambda$ (and $A\vec{1} = 0$). So take $\alpha = \frac{n}{d-\lambda_2}$, and by our assumption on λ_2, $\alpha \geq 2$.

Now A gives an upper bound on (5.3):

$$\frac{1}{2} trA = \frac{1}{2} n(\alpha d - n + 1) = \frac{1}{2} n^2 \frac{d}{d-\lambda_2} - \frac{1}{2} n^2 + \frac{1}{2} n = \frac{1}{2} n^2 \frac{\lambda_2}{d-\lambda_2} + \frac{1}{2} n.$$

This, by (5.1), shows that the dimension of the embedding is $\Omega\left(\frac{d-\lambda_2}{D^2(\lambda_2 + O(1))}\right)$. ∎

5.3.3 A Quasi-random graph of logarithmic sphericity

It is an interesting problem to construct new examples of graphs of linear sphericity. Since random graphs have this property, it is natural to search among quasi-random graphs. There are several equivalent definitions for such graphs (see [9]). The one we adopt here is:

Definition 5.3.1. *A family of graphs is called* quasi-random *if the graphs in the family are* $(1 + o(1))\frac{n}{2}$*-regular, and all their eigenvalues except the largest one are (in absolute value)* $o(n)$.

Counter-intuitively, perhaps, quasi-random graphs may have very small sphericity.

Lemma 5.3.2. *Let* $\mathbb{G}$ *be the family of graphs with vertex set* $\{0,1\}^k$*, and edges connecting vertices that are at Hamming distance at most* $\frac{k}{2}$*. Then* $\mathbb{G}$ *is a family of quasi-random graphs of logarithmic sphericity.*

Proof: The fact that the sphericity is logarithmic is obvious - simply map each vertex to the vector in $\{0,1\}^n$ associated with it. The Krawtchouk polynomial of order s over $\mathbb{Z}_2^k$ is defined by $K_s^{(k)}(i) = \sum_{j=0}^{s}(-1)^j\binom{i}{j}\binom{k-i}{s-j}$. To show that all eigenvalues except the largest one are $o(2^k)$ we need the following facts about Krawtchouk polynomials (see [67]. For simplicity we assume that k is odd):

1. For any $x \in \mathbb{Z}_2^k$ with $|x| = i$, $\sum_{z\in\mathbb{Z}_2^k|z|=s}(-1)^{<x,z>} = K_s^{(k)}(i)$.
2. $\sum_{s=0}^{l} K_s^{(k)}(i) = K_l^{(k-1)}(i-1)$.
3. For any s and k, $\max_{i=0,\ldots,n}|K_s^{(k)}(i)| = K_s^{(k)}(0) = \binom{k}{s}$.

Observe that G is a Cayely graph for the group $\mathbb{Z}_2^k$ with generator set $\{g \in \mathbb{Z}_2^k : |g| \leq \frac{k}{2}\}$. Since $\mathbb{Z}_2^k$ is abelian, the eigenvectors of the graphs are independent of the generators, and are simply the characters of the group written as the vector of their values. Namely, corresponding to each $y \in \mathbb{Z}_2^k$ we have an eigenvector v^y, such that $v_x^y = (-1)^{<x,y>}$. For every y, $v_0^y = 1$, so to figure out the eigenvalue corresponding to v^y, we simply need to sum the value of v^y on the neighbors of 0. Note that for $y = 0$ we get the all 1s vector, which corresponds to the largest eigenvalue. So we are interested in y's such that $|y| > 0$. By the first two facts above we have:

$$\lambda_y = \sum_{g\in\mathbb{Z}_2^k,|g|\leq\frac{k}{2}}(-1)^{<y,g>} = \sum_{s=0}^{\frac{k-1}{2}} K_s^{(k)}(|y|) = K_{\frac{k-1}{2}}^{(k-1)}(|y|-1).$$

By the third fact, this is at most $\binom{k-1}{\frac{k-1}{2}} \approx \frac{2^{k-1}}{\sqrt{k-1}} = o(2^{k-1})$. ∎

5.4 Inverting the KL-divergence

5.4.1 Introduction

Consider the following clustering problem. Let $P_1, \ldots, P_n$ be unknown distributions on $\{1, \ldots, k\}$. Let ϕ be a (known) operator that measures distance or similarity between distributions. The input to the problem is a matrix A defined by $A_{i,j} = \phi(P_i, P_j)$, and the task is to reconstruct (or apporximate) the distributions $P_1, \ldots, P_n$.
For example, one of the main questions of sequence analysis in computational biology is the following. Given a set of n proteins and their pairwise sequence similarity, compute a set of k clusters, and a measure of how likely each protein is to belong to each cluster. These clusters might correspond to protein functions, folds or domains. In other words, given a similarity matrix, we want to compute, for each protein, a probability distribution over the clusters.
Given ϕ, the first question is for what A's a solution exists. One possible choice for ϕ is the inner-product operator. Let C_0 by the cone of all matrices which are both non-negative and positive semi-definite. Let C_1 by the cone of all matrices A such that there exist a non-negative matrix B with $BB^t = A$ (such matrices are called *completely positive*). It is not hard to verify that both are

indeed cones. Note that C_1 is not exactly the set we're interested in, as we did not require that B be stochastic. However, the entries in each row of B can be normalized to sum up to 1.
It is not hard to see that $C_1 \subset C_0$. It is known that for $n > 4$, $C_1 \neq C_0$. The question of characterizing C_1 by its supporting hyperplanes (or deciding membership in C_1) has recieved much attention, yet much is still unknown (cf. [13]).
Here we study a different operator - the so-called *KL-divergence* or *cross-entropy*. Let $P = (p_1, \ldots, p_k)$ and $Q = (q_1, \ldots, q_k)$ be two distributions on $\{1, \ldots, k\}$, we define and denote the KL-divergence of P w.r.t. Q as follows (cf. [21]):

$$D(P||Q) = \sum_{i=1}^{k} p_i \log \frac{p_i}{q_i}.$$

We say that a $n \times n$ matrix is *realizable* (w.r.p.t. KL-divergence) if there are distinct distributions $P_1, \ldots, P_n$ such that $A_{i,j} = D(P_i||P_j)$. Our goal is to prove that following theorem:

Theorem 5.4.1. *An $n \times n$ matrix A is realizable if and only if it has zeros on the diagonal and positive entries elsewhere.*

The fact that A must have zeros on the diagonal follows from the fact that $D(P||P) = 0$. Hence, for any distribution P, $D(P||P) = 0$. The fact that for $P \neq Q$ $D(P||Q) > 0$ is not too hard to see and is well known (cf. [21]).

5.4.2 The KL-divergence of composed distributions

The key lemma in proving Theorem 5.4.1 expresses the KL-divergence of distributions obtained by composing some base distributions.

Lemma 5.4.1. *Let $\Omega = \dot{\cup}_{i=1}^{m} \Omega_i$ be a set with a partition $\{\Omega_i\}$. Let X and Y be distributions on $[m]$, and, for $i = 1, \ldots, m$, let U_i and V_i be distributions on Ω_i. Let p be the distributon on Ω defined by choosing an $i \in [m]$ according to X, and then an element in Ω_i according to U_i. Let q be the distributon on Ω defined by choosing an $i \in [m]$ according to Y, and then an element in Ω_i according to V_i. Then,*

$$D(p||q) = D(X||Y) + \sum_{i=1}^{m} Pr[X = i] \cdot D(U_i||V_i).$$

Proof: Denote $x_i = Pr[X = i]$, $y_i = Pr[Y = i]$. For $i = 1, \ldots, m$ and $j \in \Omega_i$, denote

$u_{i,j} = Pr[U_i = j]$ and $v_{i,j} = Pr[V_i = j]$.

$$\begin{aligned} D(p||q) &= \sum_{i=1}^{m} \sum_{j \in \Omega_i} x_i u_{i,j} \log \frac{x_i u_{i,j}}{y_i v_{i,j}} \\ &= \sum_{i=1}^{m} \sum_{j \in \Omega_i} x_i u_{i,j} (\log \frac{x_i}{y_i} + \log \frac{u_{i,j}}{v_{i,j}}) \\ &= \sum_{i=1}^{m} x_i \log \frac{x_i}{y_i} \sum_{j \in \Omega_i} u_{i,j} + \sum_{i=1}^{m} x_i \sum_{j \in \Omega_i} u_{i,j} \log \frac{u_{i,j}}{v_{i,j}} \\ &= D(X||Y) + \sum_{i=1}^{m} Pr[X = i] \cdot D(U_i||V_i), \end{aligned}$$

Where the last equality follows from $\sum_{j \in \Omega_i} u_{i,j} = 1$ for all $i = 1, \ldots, m$. ∎

Corollary 5.4.1. *For $X = Y$ we get $D(p||q) = \sum x_i D(U_i||V_i)$, hence any convex combination of realizable matrices can be realized.*

Corollary 5.4.2. *Let U_1 and V_1 be two distributions on Ω_1. Let $U_2 = V_2$ be some distribution on Ω_2. For some $\epsilon \leq 1$, let $X = Y$ be the distribution $(\epsilon, 1-\epsilon)$. This gives $D(p||q) = \epsilon D(U_1||V_1)$. Hence, any linear down scaling of a realizable matrix can be realized.*

5.4.3 Proof of the Theorem

We now give two proofs for Theorem 5.4.1. Both are algorithmic, in the sense that given a matrix A one can efficiently compute distributions the realize it. However, typically one would like the distribution space (the number of clusters) to be small. Unfortunatley, in both proofs this is not the case.

Proof: (I) Let A be an $n \times n$ matrix with zero on the diagonal, and positive entries elsewhere. We construct realizable matrices $B^{i,j}$ for $1 \leq i \neq j \leq n$, such that A is in the convex hull of the $B^{i,j}$'s. By Corollay 5.4.1, this will prove the theorem.
Let N be a large number such that for all $1 \leq i \neq j \leq n$, $1/N < A_{i,j} < N$. Fix $i \neq j$. Let $U_1, \ldots, U_n$ be the following distributions on 2 points. For $k \neq i, j$,
$U_k = (\frac{1}{N \cdot 2^N}, 1 - \frac{1}{N \cdot 2^N})$ (to make them distinct, one can take small perturbations thereof).
$U_i = (\frac{1}{N^2}, 1 - \frac{1}{N^2})$.
$U_j = (\frac{1}{2^{2^N}}, 1 - \frac{1}{2^{2^N}})$.
It is not hard to verify that $D(U_i||U_j) > N$, and that for any other pair of distributions, the KL-divergence is at most $\frac{1}{N}$. Let $B^{i,j}$ be the matrix obtained from these distributions, then A is in the convex hull of all the $B^{i,j}$. By Caratheodory's theorem, A can be written as a convex combination of at most $n+1$ $B^{i,j}$'s, hence the realization of A in this way requires distributions on $2(n+1)$ points. ∎

Proof: (II) Fix $0 < \epsilon < \frac{1}{2}$ such that $2^{-\frac{1}{\epsilon}} < \frac{1}{n}\epsilon^2$. Let $\epsilon_{i,j}$ be such that $2^{-\frac{1}{\epsilon}} < \epsilon_{i,j} < \frac{1}{n}\epsilon^2$. For $i = 1, \ldots, n$ let D_i be the following distribution on n points. For $j \neq i$, $Pr[D_i = j] = \epsilon_{i,j}$. Denote $s_i = \sum_j \epsilon_{i,j}$, so $Pr[D_i = i] = 1 - s_i$. We have that:

$$D(D_i||D_j) = \sum_{k \neq i,j} \epsilon_{i,k} \frac{\epsilon_{i,k}}{\epsilon_{j,k}} + \epsilon_{i,j} \log \frac{\epsilon_{i,j}}{1 - s_j} + (1 - s_i) \log \frac{1 - s_i}{\epsilon_{j,i}}$$

If ϵ is small, this is dominated by the last summand:

$$\sum_{k \neq i,j} \epsilon_{i,k} \frac{\epsilon_{i,k}}{\epsilon_{j,k}} \leq \sum_{k \neq i,j} (\frac{1}{n}\epsilon^2 \frac{1}{\epsilon}) = \frac{n-2}{n}\epsilon.$$

$$-\epsilon_{i,j} \log \frac{\epsilon_{i,j}}{1 - s_j} < \frac{1}{n}\epsilon^2 \frac{1}{\epsilon} \log 2 = \frac{\log 2}{n}\epsilon.$$

$$(1 - s_i) \log \frac{1 - s_i}{\epsilon_{j,i}} \approx \log \frac{1}{\epsilon_{j,i}}.$$

This allows us to approximate a realization of any matrix A whose entries satisfy $\frac{1}{\epsilon} > A_{i,j} > \log \frac{n}{\epsilon^2}$. However, by Corollary 5.4.2, we can also realize any linear scaling down of A. Hence, as ϵ tends to 0, we can approximate as well as we want any matrix A. Taking the limit distributions gives us a realization of A. Note that this realization requires $n + 1$ points. ∎

5.5 Open Problems

The only explicit examples known so far for graphs that have linear sphericity are $K_{n,n}$ and small modifications of it. We conjecture that more complicated graphs, such as the Paley graph, also have linear sphericity. Note that the lower bound presented here only shows a bound of $\Omega(\sqrt{n})$ for the Paley graph. It is also interesting to know if the bound can be improved, either as a pure spectral bound, or with some further assumptions on the structure of the graph.

What is the largest sphericity, $d = d(n)$, of an n-vertex graph? We know that $\frac{n}{2} \leq d \leq n - 1$. Can this gap be closed? For a seemingly related question, the smallest dimension required to realize a sign matrix (see [5]) the answer is known to be $\frac{n}{2} \pm o(n)$. We have also seen a similar gap for $d(n, l_2)$ and $d(n, l_\infty)$. Can this be closed? Can some kind of interpolation arguments generalize the bounds we know for these two numbers to bounds on $d(n, l_p)$ for $p > 2$?

Our interest in sphericity arose from a search for a lower bound on $d(n, l_2)$. But why limit the discussion to Euclidean space? What can be said of spherical embeddings into l_1 or l_∞? The former may be particularly interesting, as it will give a non-trivial lower bound on $d(n, l_1)$.

In inverting the KL-divergence, we showed that this can be done with distributions on $n+1$ clusters, where the distribution for each point in the metric space is concentrated on its own cluster. Clearly this is useless for clustering. It is interesting what can be said about this problem when the number of clusters is bounded.

Chapter 6

Algorithms for stable problem instances

6.1 Introduction and definitions

In this chapter we try to develop efficient algorithms for optimization problem instances under the assumption that the solution is "stable" or "distinct". Our case in point in studying this notion is the problem of data clustering. This problem is NP-hard in essentially any plausible formulations. Nonetheless, many heuristics for clustering are successful in practice. When such heuristics fail to produce a good solution, it is often because the data has no firm underlying structure that clearly defines the "correct solution".
Consider the points in the plane shown in Figure 6.1. The points in the left plot clearly define three clusters, while those on the right do not. The definition of what is the correct solution in the latter plot is likely to depend of the exact formalization of the problem. Similarly, small perturbation of the points will probably lead to different "optimal" solutions, and there will tend to be many solutions which are near-optimal.
Our motto in this chapter is: "Clustering is either easy or boring". Put another way, we would like to show that either an optimal solution can be found efficiently, or there's a good excuse for not finding it, namely, that in an informal sense the optimal solution is not well defined. We start by giving a formal definition of the clustering problem, and what instances of it are considered "stable".

Definition 6.1.1. *The input to the* k-clustering problem *is a graph* $G = (V, E)$*, and a weight function* $w : E \to \mathbb{R}$*. The objective is to partition* V *into* k *subsets,* $V_1, \ldots, V_k$*, such that* $\sum_{i<j} \sum_{u \in V_i, v \in V_j} w(u, v)$ *is maximized.*
We say that a weight function w' *is a* γ-perturbation *of* w *(for* $\gamma \geq 1$*) if* $\forall e \in E$*,* $w(e) \leq w'(e) \leq \gamma \cdot w(e)$*.*
An instance (G, w) *to the* k*-clustering problem is* γ-stable *if for all* γ*-perturbations* w' *of* w*, the optimal partition of* (G, w') *is unique, and these partitions are all identical.*

Note that γ-stability for every $\gamma > 1$ is equivalent to G being k-partite.
In fact, there is only one perturbation, w' that needs to be considered. Let $V_1, \ldots, V_k$ be an optimal partition of (G, w). It is not hard to see that in order to establish γ-stability it is enough to consider

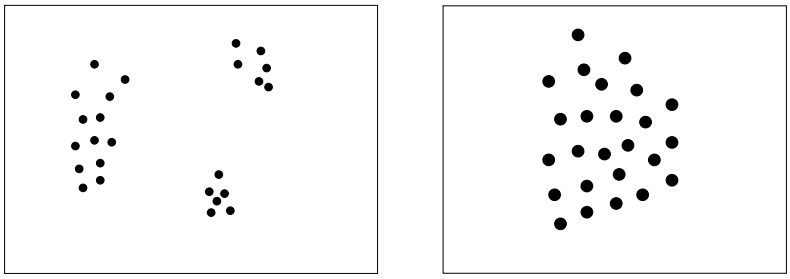

Figure 6.1: Good (left) and bad (right) instances of the clustering problem.

the perturbation w' that multiplies by γ all edges which are internal to the V_i's.
We could have also allowed dividing some values of w by γ, but for simplicity we do not. It is an easy fact that γ-stability under this alternative definition is equivalent γ^2-stability under the previous one.

We are interested in the following questions:

1. Is there a γ such that there exists a polynomial time algorithm that solves correctly γ-stable instances of the k-clustering problem (possibly, $\gamma = \gamma(n)$)? If so, how small can γ be taken? Since this problem in NP-hard, the promise of stability is essential (if P $\neq$ NP).

2. For values of γ below this bound, how well can polynomial time algorithms approximate the solution to γ-stable instances, as a function of γ?

3. For known approximation algorithms, can better approximation ratios be guaranteed when the input is γ-stable?

4. Given an instance to the k-clustering problem, is it possible to efficiently compute, or approximate, its stability? Is it possible to do so when the optimal partition is supplied as well?

Another interesting question is how to efficiently sample γ-stable instances, in a non-trivial way. This is necessary should we want to consider the above questions for random instances. It is possible to generate a γ-stable instance from an given instance (G, w) in the following way. Let $V_1, \ldots, V_k$ be an optimal partition. Let $w'(i,j) = w(i,j)$ if i, j belong to the same V_l, and $w'(i,j) = \gamma \cdot w(i,j)$ otherwise. It is not hard to see that (G, w') is γ-stable, and that all γ-stable instances are obtainable in this way. We do not know, though, how to generate "typical" γ-stable instances.

The concept can be applied to other combinatorial optimization problems. We focus on the 2-clustering problem, more commonly known as the Max-Cut problem, in the section 6.2, and discuss the Vertex Cover problem and the Multi-way Cut problem in the section 6.3.

6.2 Stable Max-Cut

6.2.1 Notation

The *partition induced by* v on $[n]$, for $v \in \mathbb{R}^n$, is the partition $(\{i : v_i > 0\}, \{i : v_i \leq 0\})$. If G is a graph on n vertices, this partition is also called the *cut induced by* v in G.
The *indicator vector* of a partition $(S, \bar{S})$ of $[n]$ (or a cut in graph on $[n]$), is the vector $v \in \{-1, 1\}^n$, with $v_i = 1$ iff $i \in S$.
The *weighted adjacency matrix* corresponding to an input (G, w), is a matrix W indexed by the vertices of the graph with $W_{i,j} = w(i, j)$. We generally assume throughout that w is a non-negative function.
For a weighted graph G, we denote the indicator vector of its maximal cut by mc^*. We generally assume that this cut is unique, otherwise mc^* is an indicator of some maximal cut.
For two disjoint subsets of vertices in the graph, A, B, $E(A, B)$ denotes the set of edges going between them. For a set of edges $F \subset E$, denote $w(F) = \sum_{e \in F} w(e)$. For a subset $A \subset [n]$, we denote $\bar{A} = [n] \backslash A$.
Let (G, w) be an instance of the Max-Cut problem. Let $(S, \bar{S})$ be the optimal partition of V. We say that (G, w) is γ-*locally stable* if for all $v \in S$

$$\gamma \cdot \sum_{u \in S} w(u, v) < \sum_{u \in \bar{S}} w(u, v),$$

and for all $v \in \bar{S}$

$$\gamma \cdot \sum_{u \in \bar{S}} w(u, v) < \sum_{u \in S} w(u, v),$$

It is known that Max-Cut is NP-hard even when it is γ-locally stable (for γ at most exponential in the size of the input). In fact, one can easily "force" local stability:

Definition 6.2.1. *Let G be a graph with weighted adjacency matrix W, and ω the sum of the weights of its edges. $G^\times$ is a graph on $V \times \{0, 1\}$, with weighted adjacency matrix*

$$G^\times = \begin{pmatrix} W & \omega \cdot I \\ \omega \cdot I & W \end{pmatrix}$$

It is not hard to see that the maximal cut in $G^\times$ is essentially two copies of that in G. Specifically, $(S, \bar{S})$ is a maximal cut in G iff $(S \times \{0\} \cup \bar{S} \times \{1\}, S \times \{1\} \cup \bar{S} \times \{0\})$ is a maximal cut in $G^\times$. It is also not hard to see that G is γ-stable, iff $G^\times$ is.

6.2.2 The Goemans-Williamson algorithm

The best currently known approximation algorithm for the Max-Cut problem is that of Goemans and Williamson ([36]). It is based on a semi-definite relaxation of the problem, which we now

describe. Ideally, we would like to solve the following problem:

$$\text{Maximize } \frac{1}{2}\sum_{(i,j)\in E} w(i,j)(1-y_iy_j)$$
$$\text{over } y \in \{-1,1\}^n.$$

Clearly, this is equivalent to minimizing $\sum_{(i,j)\in E} w(i,j)y_iy_j$, over $y \in \{-1,1\}^n$, or minimizing $\sum_{(i,j)\in E} W_{i,j}Y_{i,j}$ over all $\{-1,1\}$-matrices Y that are positive semi-definite and of rank 1. Goemans and Williamson's algorithm relaxes the rank constraint, yielding a semi-definite programming problem which can be solved efficiently. The approximation guarantee of their algorithm is ~ 0.8786. However, they show that when the weight of the maximal cut is relatively big, this guarantee can be improved. Namely, let R be the ratio between the weight of the maximal cut and the total weight of the edges. Let $h(t) = arccos(1-2t)/\pi$. Then the approximation ratio is at least $h(R)/R$.
It is not hard to see that when the input is very stable, the maximal cut is indeed relatively big. First observe that γ-stability implies local γ-stability. Otherwise, multiplying the weights of all edges (u,v), for $u \in S$, by γ, would make the cut $(S\backslash\{v\}, \bar{S}\cup\{v\})$ preferable to $(S,\bar{S})$.
It follows that for γ-stable instances the maximal cut weighs at least $\frac{\gamma}{\gamma+1}$ of the total weight. Hence, the performance guarantee of the G-W algorithm on γ-stable instances is at least $(1-O(\frac{1}{\sqrt{\gamma}}))$.
The semi-definite program used in the G-W algorithm can be strengthened when the input is γ-stable, by inequalities that express this stability. It is interesting whether these additional constraints can improve the approximation ratio further.

6.2.3 A greedy algorithm

We begin with an algorithm that gives an exact solution when the input is very stable. Since the maximal cut is unique, its edges form a connected spanning bipartite graph. The algorithm iteratively identifies sets of edges which are in the maximal cut, until they form such a graph.

FindMaxCut(G, w)

1. Initialize $T = (V(G), \emptyset)$ Throughout the algorithm T will be a bipartite subgraph of G.
2. While T is not connected, do:
   (a) Let $C_1, \ldots, C_t$ be the connected components of T. Each of them is a bipartite graph, so denote $V(C_i) = (L_i, R_i)$.
   (b) Let C_{i^*} be a component with the least number of vertices. For each $j = 1, \ldots, t$, $j \neq i$, let $E_j^0 = E(L_i, L_j) \cup E(R_i, R_j)$ and $E_j^1 = E(L_i, R_j) \cup E(R_i, L_j)$. Let j^* and c^* be such that the weight of $E_{j^*}^{c^*}$ is maximal among all E_j^c.
   (c) Add the edges of $E_{j^*}^{c^*}$ to T
3. Output the cut defi ned by the two sides of T.

Lemma 6.2.1. *The above algorithm is well defined, and outputs the correct solution on $\sqrt{n\Delta}$-stable instances of Max-Cut (where n is the number of vertices in the graph, and Δ the maximal degree),*

Proof: Let $(S, \bar{S})$ be the maximal cut. We maintain that throughout the algorithm, for each connected component $C_i = (L_i, R_i)$ either $L_i \subset S$ or $R_i \subset S$.
This clearly holds at the outset. If it holds at termination, the algorithm work correctly. So consider the first iteration when this does not hold. Let C_{i^*} be a smallest connected component at this stage, and denote $k = |C_{i^*}|$. Up to this point our assumption holds, so either $L_{i^*} \subset S$ or $R_{i^*} \subset S$. W.l.o.g. assume the former. Let j^* and c^* be those chosen as in step 2b. Since this is the point where the algorithm errs, $E_{j^*}^{c^*}$ is added to T, yet $E_{j^*}^{c^*} \cap E(S, \bar{S}) = \emptyset$.
Now consider the γ-perturbation of the graph obtained by multiplying the edges in $E_{j^*}^{c^*}$ by γ. If the original graph is γ-stable, the maximal cut of the perturbed graph is $(S, \bar{S})$ as well. Consider the cut obtained by flipping the sides of L_{i^*} and R_{i^*}. That is, denote $Z = S \backslash L_{i^*} \cup R_{i^*}$, and consider the cut $(Z, \bar{Z})$.
The cut $(Z, \bar{Z})$ contains the edges $E_{j^*}^{c^*}$, which $(S, \bar{S})$ does not. For each $j \neq j^*$, let c_j be such that $E_j^{c_j}$ is in the cut $(S, \bar{S})$ (we'll be interested only in non-empty subsets). In the extreme case, all these edges are not in the cut $(Z, \bar{Z})$. Observe that all other edges in $E(S, \bar{S})$ are also in $E(Z, \bar{Z})$. Define $J = \{j \neq i : E_j^{c_j} \neq \emptyset\}$. Since the weight of $(Z, \bar{Z})$, even in the perturbed graph, is smaller than that of $(S, \bar{S})$, we have that:

$$\gamma \cdot w(E_{j^*}^{c^*}) < \sum_{j \in J} w(E_j^{c_j}).$$

Recall that $E_{j^*}^{c^*}$ was chosen to be the set of edges with the largest total weight. Hence, $\sum_{j \in J} w(E_j^{c_j}) \leq |J| w(E_{j^*}^{c^*})$, and so $\gamma < |J|$. Clearly, $|J| \leq \min\{\frac{n}{k}, k\Delta\}$, and so:

$$\gamma^2 < \frac{n}{k} k\Delta = n\Delta.$$

This is a contradiction to the assumption that the input is $\sqrt{n\Delta}$-stable. ∎

Note that we have actually proven that the algorithm works as long as it can find a connected component C_{i^*}, such that $|\{j : E_j^c \neq \emptyset\}| < \gamma$, for $c = 0, 1$.

6.2.4 A spectral algorithm

Recall that the semi-definite algorithm of Goemans and Williamson relaxes the problem:

$$\min_{y \in \{-1,1\}^n} \sum_{(i,j) \in E} w(i,j) y_i y_j$$

to

$$\min_{Y \in PSD, Y_{i,i}=1} \sum_{(i,j \in E} W_{i,j} Y_{i,j}.$$

Rather than relaxing the rank constraint, consider the relaxation of the condition $y \in \{-1, 1\}^n$, to $y \in \mathbb{R}^n, ||y||^2 = n$. By the variational characterization of eigenvalues, this relaxation amounts to finding the eigenvector corresponding to the least eigenvalue of W.
In the most general setting, a cut is obtained from such a vector by choosing a threshhold τ, and partitioning the vertices according to whether the corresponding entries in y are greater or lesser than τ. This technique is sometimes called spectral partitioning. Here we always take $\tau = 0$:

Definition 6.2.2. *Let W be an $n \times n$ matrix. Let λ be its least eigenvalue, and u the corresponding eigenvector. The* spectral partitioning induced by W *is the partition of $[n]$ induced by u.*

Lemma 6.2.2. *Let W be the weighted adjacency matrix of a γ-stable instance of Max-Cut. Let D be a diagonal matrix, and u an eigenvector corresponding to the least eigenvalue of $W + D$. If $\gamma \geq \frac{\max_{(i,j)\in E} |u_i u_j|}{\min_{(i,j)\in E} |u_i u_j|}$, then the spectral partitioning induced by $W + D$ yields the maximal cut.*

Proof: For any diagonal matrix D, think of $W + D$ as the weighted adjacency matrix of a graph, with loops added. These loops do not contribute to the weight of any cut, so for any D, a cut is maximal in W iff it is maximal in $W + D$. Furthermore, it is not hard to see that W is γ-stable, iff $W + D$ is.
Let D be a diagonal matrix, and u an eigenvector corresponding to the least eigenvalue of $W + D$. For the lemma to be meaningful, assume that u has no 0 entries. Normalize u so that $\min_{(i,j)\in E} u_i \cdot u_j = 1$. Let D' be the diagonal matrix $D'_{i,i} = D_{i,i} + u_i^2$. Let W' be the matrix $W'_{i,j} = W_{i,j} \cdot |u_i u_j|$. Observe that W' is a γ-perturbation of W, hence the maximal cut in W', and in $W' + D'$, is the same as in W. In other words, mc^* is a vector that minimizes the expression:

$$\min_{x\in\{-1,1\}^n} xAx.$$

Note that since W is stable, the maximal cut is unique. Let y be the indicator vector for the cut induced by u. Assume for contradiction that $mc^* \neq y, -y$. Let v be the vector $v_i = |u_i| mc_i^*$. Observe that:

$$\begin{aligned} mc^*(W' + D')mc^* &< y(W' + D')y \\ mc^*(W' + D')mc^* &= v(W + D)v \\ y(W' + D')y &= u(W + D)u. \end{aligned}$$

This implies $u(W + D)u > v(W + D)v$, which is a contradiction, since $||u|| = ||v||$, and u minimizes the expression $\min_{x\in\mathbb{R}^n} \frac{x(W+D)x}{||x||^2}$. ∎

Note 6.2.1. *In fact, a somewhat stronger result can be proven. It suffices that*

$$\gamma \geq \frac{\max_{(i,j)\in E\,:\,u_i u_j<0} -u_i u_j}{\min_{(i,j)\in E\,:\,u_i u_j\geq 0} u_i u_j}.$$

It remains an intriguing question whether a large enough γ entails the existence of such a diagonal matrix D, and how to find it efficiently. We present some experimental results in section 6.2.7, and conclude this section with one special case where D can be found efficiently.

Lemma 6.2.3. *Let W be the weighted adjacency matrix of a Max-Cut instance. Define a diagonal matrix D by $D_{i,i} = mc_i^* \sum_j W_{i,j} mc_j^*$. If $W + D$ is positive semi-definite, then the maximal cut in W can be found efficiently.*

Proof: It is easy to see that mc^* is an eigenvector of $W + D$ corresponding to eigenvalue 0. As $W + D$ is positive semidefinite, 0 is the least eigenvalue of $W + D$, so this is indeed a special case of the assertion in Lemma 6.2.2. Let c be such that $\text{tr}(D + cI) = 0$ (in other words, $c = -\text{tr}(D)/n$). The eigenvectors of $W + D$ and $W + D + cI$ are the same, and, in particular, mc^* is the eigenvector of $W + D + cI$ associated with its least eigenvalue. Denote $m = mc^*(W + D + cI)mc^* = mc^*Wmc^*$. This is the value we wish to compute.
Consider the following optimization problem: Let $\mathbb{A}$ be the set of all positive definite matrices A such that $A_{i,j} = W_{i,j}$ for $i \neq j$, and $\text{tr}(A) = 0$. Find an $A \in \mathbb{A}$ such that its least eigenvalue of is maximal. Denote this value by m'.
For any $A \in \mathbb{A}$, $mc^*Amc^* = m$, so the least eigenvalue of any such A its least at most $\frac{m}{n}$. Thus, $m' \leq \frac{m}{n}$. However, $W + D + cI \in \mathbb{A}$, and its least eigenvalue is exactly $\frac{m}{n}$, so $m' = \frac{m}{n}$.
Grötschel, Lovász and Schrijver [37, 38] show that the ellipsoid algorithm solves convex optimization problems under very general conditions. Delorme and Poljak ([22] observe that this applies to the above optimization problem (this essentially already appears in [15] as well). In particular, we can compute m, and from it, the value of the maximal cut. If the solution to the optimization problem is unique, we can also get D and mc^* from it. This uniqueness can be achieved by making minute perturbations in W. ∎

6.2.5 Distinctness and spectral stability

So far our way to make the motto "Clustering is either easy or boring" explicit was by saying that if the optimal solution changes under γ-perturbations, then it is not interesting.
Next we study the notion of a *distinctness* of a solution. Intuitively, we want the optimal solution to be significantly better than all other solutions. However, one can modify the optimal cut by switching the position of only few vertices, and the value of the cut will not change much. Thus, for the definition to be non-vacuous, we should consider the drop in the weight of the cut as a function of the (Hamming) distance from the maximal cut. This distinction is reminiscent of the difference between graph connectivity and expansion.
For simplicity, we restrict the definition to d-regular, unweighted graphs:

Definition 6.2.3. *Let $C \subset E$ be the set of edges of a cut in a d-regular graph G, and $T \subset V$. The* relative net contribution *of T to the cut is:*

$$\rho_C(T) = \frac{|\{e \in (T \times \bar{T}) \cap C\}| - |\{e \in (T \times \bar{T}) \cap (E \backslash C)\}|}{d \min\{|T|, n - |T|\}}.$$

*Let $C^* \subset E$ be the set of edges in a maximal cut of G. We say that it is ρ-*distinct, *if for all $T \subset V$, $\rho_{C^*}(T) \geq \rho$.*

In this section we show that if a graph is highly distinct in this sense, spectral partitioning works:

Theorem 6.2.1. *Let $C^* \subset E$ be the set of edges in a maximal cut of d-regular graph G. Assume that this cut is ρ - distinct, and, furthermore, that for all $v \in V$, $\rho_{C^*}(\{v\}) \geq \alpha$. If*

$$\alpha + \frac{\rho^2}{10} > 1$$

then the maximal cut can be found efficiently.

In proving the theorem, it is worthwhile to consider yet another definition of stability.

Definition 6.2.4. *Let (G, w) be an instance of Max-Cut with a unique optimum, and W its weighted adjacency matrix. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of W. (G, w) is called (ϵ, γ)-*spectrally stable *(for $0 \leq \epsilon < 1$, $\gamma \geq 1$) if:*
1. (G, w) is γ-locally stable.
2. $\lambda_{n-1} = \epsilon\lambda_n$.

An Intuitive reason to require a gap at the lower end of the spectrum is that by Lemma 6.2.2 the eigenvector corresponding to the least eigenvalue might suggest a good cut. If the two smallest eigenvalues are close, we get two "good suggestions". Spectral stability rules out this case.

Lemma 6.2.4. *Let (G, w) be an (ϵ, γ)-spectrally stable instance of Max-Cut, and W its weighted adjacency matrix. Let δ be the minimal row sum in W, and λ_n be W's least eigenvalue. If*

$$\delta \cdot \frac{\gamma - 1}{\gamma + 1} > \frac{1}{2}(1 + \epsilon)|\lambda_n| = \frac{1}{2}|\lambda_n + \lambda_{n-1}|$$

then there is a diagonal matrix D such that the spectral partitioning induced by $W + D$ coincides with the maximal cut. Furthermore, such a matrix D can be found efficiently.

For example, if all the row sums in W are the same, then $\beta \geq 1$. In this case, the requirement in the lemma regarding the tradeoff between the spectral gap and the local stability is:

$$\epsilon < \frac{\gamma - 3}{\gamma + 1}.$$

In proving the lemma, we'll need the following observation:

Claim 6.2.1. *Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of a real symmetric matrix W, such that $\lambda_{n-1} = \epsilon\lambda_n$, for some $0 \leq \epsilon < 1$. Let $x, y \in \mathbb{R}^n$ be two orthogonal unit vectors. Then $xWx + yWy \geq (1 + \epsilon)\lambda_n$.*

Proof: Let u be a unit eigenvector of W corresponding to eigenvalue λ_n. Denote $x^{\|} =< x, u > u$, $x^{\perp} = x - x^{\|}$, and similarly $y^{\|}$ and $y^{\perp}$.
Note that $||x + y||^2 = ||x||^2 + ||y||^2 = 2$. Thus,

$$||x^{\|}|| + ||y^{\|}|| =< x, u > + < y, u >=< x + y, u >\leq ||x + y|| \cdot ||u|| = \sqrt{2}.$$

It follows that:

$$\begin{aligned} xWx + yWy &= (x^{\|} + x^{\perp})W(x^{\|} + x^{\perp}) + (y^{\|} + y^{\perp})W(y^{\|} + y^{\perp}) \\ &\geq \lambda_n(||x^{\|}||^2 + ||y^{\|}||^2) + \lambda_{n-1}(||x^{\perp}||^2 + ||y^{\perp}||^2) \\ &= \lambda_n(||x^{\|}||^2 + ||y^{\|}||^2 + \epsilon||x^{\perp}||^2 + \epsilon||y^{\perp}||^2). \end{aligned}$$

Denote $\alpha = ||x^{\|}||^2$, and so $||y^{\|}||^2 \leq (\sqrt{2} - \sqrt{\alpha})^2$. As $\epsilon < 1$, we have that:

$$\begin{aligned} xWx + yWy &\geq \lambda_n(\alpha + (\sqrt{2} - \sqrt{\alpha})^2 + \epsilon(1-\alpha) + \epsilon(1 - (\sqrt{2} - \sqrt{\alpha})^2) \\ &= 2\lambda_n(1 + \alpha - \sqrt{2\alpha} + \epsilon(\sqrt{2\alpha} - \alpha)). \end{aligned}$$

Taking the derivative and comparing to zero, we get that this expression is maximized for $\alpha = \frac{1}{2}$, giving the claim. ∎

Proof: (Lemma 6.2.4) Let $\beta = -\delta/\lambda_n$. Let D be a diagonal matrix with $D_{i,i} = mc_i^* \sum_j W_{i,j} mc_j^*$. From local stability, this is at least $\frac{\gamma-1}{\gamma+1}\delta$, hence this is also a lower bound on all of D's eigenvalues. Note that mc^* is an eigenvector of $W + D$, corresponding to eigenvalue 0. Let x and y be two (orthogonal) unit eigenvectors of $W + D$ corresponding to the two smallest eigenvalues of $W + D$. Since 0 is an eigenvalue of $W + D$, at least one of these two eigenvalues is not positive, say, the one corresponding to x. In particular, $x(W + D)x \leq 0$, or

$$xWx \leq -xDx \leq -\frac{\gamma-1}{\gamma+1}\delta = \frac{\gamma-1}{\gamma+1}\beta\lambda_n.$$

By Claim 6.2.1, $yWy \geq (1 + \epsilon - \beta\frac{\gamma-1}{\gamma+1})\lambda_n$. But the eigenvalue of $W + D$ corresponding to y is given by:

$$\begin{aligned} y(W+D)y &\geq (1 + \epsilon - \beta\frac{\gamma-1}{\gamma+1})\lambda_n + \frac{\gamma-1}{\gamma+1}\delta \\ &= (1 + \epsilon - 2\beta\frac{\gamma-1}{\gamma+1})\lambda_n > 0. \end{aligned}$$

Since 0 is an eigenvalue of $W + D$, it must be the least one. Recall that mc^* is the eigenvector of $W + D$ corresponding to this value, as claimed in the first part of the lemma. The second part now follows from Lemma 6.2.3. ∎

As noted above, the concept of "distinctness" is close in spirit to *edge expansion* (cf. [9]) in a graph. Indeed, the connection between edge expansion and spectral gap shows that distinctness implies spectral stability.

Lemma 6.2.5. *Let $C^* \subset E$ be the set of edges in a maximal cut of a d-regular graph G. Assume that this cut is ρ - distinct, and, furthermore, that for all $v \in V$, $\rho_{C^*}(\{v\}) \geq \alpha$. Suppose $\frac{\rho^2}{2} + 3\alpha - 3 > 0$. Then G is $(4 - \frac{\rho^2}{2} - 3\alpha, \frac{1+\alpha}{1-\alpha})$-spectrally stable.*

Proof: The fact that $\forall v \in V$, $\rho_{C^*}(\{v\}) \geq \alpha$, is equivalent to the local stability being at least $\frac{1+\alpha}{1-\alpha}$. It remains to show a spectral gap.
The Cheeger constant of a d regular graph is defined as:

$$h_G = \min_{T \subset V \,:\, |T| \leq \frac{n}{2}} \frac{|\{e \in (T \times \bar{T}) \cap E\}|}{d|T|}.$$

Clearly, $h_G \geq \rho$. The Cheeger constant is related to the spectral gap between the first and second eigenvalues of G. Namely, let $d = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of G, then (see e.g. [48]):

$$\lambda_2 \leq (1 - \frac{h_G^2}{2})d. \tag{6.1}$$

Let A be the adjacency matrix of G, and $w' \in \mathbb{R}^n$ a unit vector such that $< w', \vec{1} >= 0$, $< w', mc^* >= 0$, and $w'Aw' \leq \lambda_{n-1}$. Such a vector can be constructed by taking a linear combination of the eigenvectors of A which correspond to eigenvalues λ_n and λ_{n-1}. Define $w \in \mathbb{R}^n$ to be the vector obtained by point-wise multiplication of w' and mc^* (i.e. $w_i = w'_i \cdot mc^*_i$). Note that $||w|| = ||w'|| = 1$. Since $< w', \vec{1} >= 0$ and $< w', mc^* >= 0$, it follows that $< w, \vec{1} >= 0$ as well. Consider the graph H induced by the edges $E \backslash C^*$. It has maximal degree $\frac{1-\alpha}{2}d$, hence its least eigenvalue is at least $-\frac{1-\alpha}{2}d$. Let B be the adjacency matrix of H. In particular, $w'Bw' \geq -\frac{1-\alpha}{2}d$. By the definition of w,

$$wAw = w'Bw' - w'(A-B)w' = 2w'Bw' - w'Aw' \geq -\lambda_{n-1} - 2(1-\alpha)d.$$

By the variational characterization of the second eigenvalue,

$$\lambda_2 = \max_{x\,:\,<x,\vec{1}>=0} \frac{xAx}{||x||^2} \geq wAw,$$

and so:

$$\lambda_{n-1} \geq -\lambda_2 - 2(1-\alpha)d.$$

By the Cheeger inequality (6.1),

$$\lambda_{n-1} \geq (\frac{h_G^2}{2} + 2\alpha - 3)d.$$

Using the variational characterization again, we have that

$$\lambda_n \leq \frac{mc^* A mc^*}{||mc^*||^2} \leq -\alpha d.$$

Thus, as $\rho \leq h_G$ and $\lambda_n \geq -d$,

$$\lambda_{n-1} - \lambda_n \geq (\frac{h_G^2}{2} + 2\alpha - 3 + \alpha)d \geq (\frac{\rho^2}{2} + 3\alpha - 3)|\lambda_n|.$$

Hence,

$$\frac{\lambda_{n-1}}{\lambda_n} \leq 4 - \frac{\rho^2}{2} - 3\alpha$$

∎

A straightforward calculation shows that Lemmatta 6.2.4 and 6.2.5 imply Theorem 6.2.1.

6.2.6 Random graphs

Consider the following model for random weighted graphs. Let P be some probability measure on $[0, \infty)$. Generate a matrix W' (a weighted adjacency matrix), by choosing each entry $W'_{i,j}$, $i < j$, independently from P. Set $W'_{i,j} = W'_{j,i}$ for $i > j$, and $W'_{i,i} = 0$. Let C be the set of edges in the maximal cut of W (for "reasonable" P's, this will be unique w.h.p.). Set $W_{i,j} = \gamma \cdot W'_{i,j}$ for

$(i,j) \in C$.
It is easy to see that W is indeed γ-stable, yet for certain probability measures the problem becomes trivial. For example, if P is a distribution on $\{0,1\}$, the maximal cut in W simply consists of all the edges with weight γ.
An even simpler random model is the following. Take n even. Generate an $n \times n$ matrix W' as above. Choose $S \subset [n]$, $|S| = n/2$ uniformly at random. Let C be the set of edges in the cut $(S, \bar{S})$. Set $W_{i,j} = \gamma \cdot W'_{i,j}$ for $(i,j) \in C$. Denote this distribution $\mathbb{G}(n, P, \gamma)$. For an appropriate γ, w.h.p. $(S, \bar{S})$ will be the maximal cut in W. This random model is close to what is sometimes known as "the planted partition model" ([18, 15, 25, 42, 20, 27, 54, 64]).
In this section we show that w.h.p. the maximal cut of graphs from this distribution can be found efficiently. Our technique follows that of Boppana in [15].

Theorem 6.2.2. *Let P be a distribution with bounded support, expectation μ and variance σ^2. There exists a polynomial time algorithm that w.h.p. solves Max-Cut for $G \in \mathbb{G}(n, P, \gamma)$, for $\gamma = 1 + \Omega(\sqrt{\frac{\log n}{n}})$.*

The theorem follows from Lemma 6.2.3 and the following one:

Lemma 6.2.6. *Let P be a distribution with bounded support, expectation μ and variance σ^2. Let $G \in \mathbb{G}(n, P, \gamma)$, and S the subset chosen in the generating G. Let $mc \in \{-1, 1\}^n$ be the indicator vector of the cut $(S, \bar{S})$. Let D be the diagonal matrix defined by $D_{i,i} = mc_i \sum_j W_{i,j} mc_j$. If $\gamma \geq 1 + \Omega(\sqrt{\frac{\log n}{n}})$, then w.h.p.:*
1. mc is the indicator vector of the maximal cut in G.
2. $W + D$ is positive semi-definite.

The key element in the proof is the following result of Füredi and Komlós:

Theorem 6.2.3. ***([34])*** *Let $a_{i,j}$, $i \leq j$ be independent random variables with values in $[-K, K]$, expectation 0 and variance $\leq \sigma^2$. Let A be a matrix with $A_{i,j} = A_{j,i} = a_{i,j}$, for $i \leq j$. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be its eigenvalues. Then w.h.p. $\max |\lambda_i| \leq 2\sigma\sqrt{n} + O(Kn^{1/3} \log n)$.*

Proof: (Lemma 6.2.6) Let W', W and S be as in the definition of $\mathbb{G}(n, P, \gamma)$. Define matrices A, B, C_A and C_B by

$$A_{i,j} = \begin{cases} W'_{i,j} - \mu & \text{if } i \in S,\ j \in \bar{S} \text{ or } j \in S,\ i \in \bar{S} \\ 0 & \text{otherwise} \end{cases}$$

$$B_{i,j} = \begin{cases} W'_{i,j} - \mu & \text{if } i \neq j \in S \text{ or } i \neq j \in \bar{S} \\ 0 & \text{otherwise} \end{cases}$$

$$(C_A)_{i,j} = \begin{cases} \mu & \text{if } i \in S,\ j \in \bar{S} \text{ or } j \in S,\ i \in \bar{S} \\ 0 & \text{otherwise} \end{cases}$$

$$(C_B)_{i,j} = \begin{cases} \mu & \text{if } i \neq j \in S \text{ or } i \neq j \in \bar{S} \\ 0 & \text{otherwise} \end{cases}$$

Let D_A and D_B diagonal matrices defined by $(D_A)_{i,i} = \frac{1}{2}\mu n \sum_j A_{i,j}$ and $(D_B)_{i,i} = \frac{1}{2}\mu n \sum_j B_{i,j}$. To clarify these notations, observe that $W' = A + C_A + B + C_B$, $W = \gamma(A + C_A) + B + C_B$ and

$D = \gamma D_A - D_B$. We need to prove that $W + D$ is, w.h.p., positive semi-definite.
It is not hard to verify that the eigenvalues of C_A are $-\frac{1}{2}\mu n$ and $\frac{1}{2}\mu n$, each with multiplicity 1, and 0 with multiplicity $n-2$. The eigenvalues of C_B are $-\mu$ with multiplicity $n-2$, and $\frac{1}{2}\mu n$ with multiplicity 2.
As D_A and D_B are diagonal, their eigenvalues are simply their diagonal entries. By the Chernoff bound, with probability at least $1-\frac{1}{n}$, every such entry deviates from $\frac{1}{2}\mu n$ by at most $O(\sigma\sqrt{n\log n})$. By the Füredi-Komlós theorem, the absolute value of all eigenvalues of A and B is, w.h.p., bounded by $2\sigma\sqrt{n} + O(n^{1/3}\log n)$.
Let $x \perp mc$ be a unit vector. We shall show that w.h.p. $0 < x(W+D)x = \gamma \cdot x(A+C_A+D_A)x + x(B+C_B-D_B)x$.
Observe that mc is an eigenvector of C_A corresponding to eigenvalue $-\frac{1}{2}\mu n$. Hence $xC_Ax \geq 0$, and w.h.p.

$$x\gamma(A+C_A+D_A)x \geq \gamma(\frac{1}{2}\mu n - O(\sigma\sqrt{n\log n})),$$

since xAx and xD_Ax are no less than the least eigenvalue of A and D_A, respectively.
Similarly, w.h.p. $x(B+C_B-D_B)x \geq -\frac{1}{2}\mu n - O(\sigma\sqrt{n\log n})$. As $\gamma = 1 + O(\sqrt{\frac{\log n}{n}})$, we get that w.h.p. $x(W+D)x > 0$.

It remains to show that w.h.p. mc is indeed an indicator vector for the maximal cut in W. Assume that indeed $W + D$ is positive semi-definite. Let $u \in \{-1,1\}^n$ be the indicator vector for the maximal cut in W, and assume for contradiction that $u \neq mc$. In particular, since the maximal cut is unique (W is γ-stable) $uWu < mcWmc$. Since $uDu = mcDmc = \text{tr}(D)$, $u(W+D)u < mc(W+D)mc$. However, mc as an eigenvector of $W+D$ corresponding to the least eigenvalue. Since $||u|| = ||mc||$, $u(W+D)u \geq mc(W+D)mc$, a contradiction. ∎

Note 6.2.2. *More generally, the distribution P may depend on n. Let P be a distribution on $[0, K+\mu]$, with expectation $\mu \leq K$ and variance σ^2. Then Lemma 6.2.6 holds if*

$$\gamma = 1 + \Omega(\mu^{-1}(\sigma\sqrt{\frac{\log n}{n}} + K\frac{\log n}{n^{-2/3}}).$$

Even more generally, the lemma holds for the following model. Choose S and define C as before. Choose the edge weights independently at random from distributions with variance at most σ^2, and expectation $\gamma\mu$ for $e \in C$, and μ for $e \notin C$. With γ as above, the lemma holds.

6.2.7 Experimental results

In this section we present results from computer simulations. We test variations on the spectral partitioning method to seek the maximal cut, and see how well they perform as a function of stability.
The weighted graphs in the simulations were generated in the following way. For each edge, a value x was sampled independently from a standard normal distribution, and the weight was taken as $|x|$. We chose to use normal, rather than uniform, distribution, to allow for some very big weights. In practice, we performed these experiments using Matlab (version 6).

The results from the simulation are depicted in Figure 6.2. Six variations on spectral partitioning were used, as described below. The basic idea behind all of them is inspired from Lemma 6.2.2. We choose random values for the diagonal (from a standard normal distribution), and look at the eigenvector corresponding to the least eigenvalue of the resulting matrix. We repeat this 1000 times, obtaining 1000 vectors, which we denote $v_1, \ldots, v_{1000}$. We denote by $u_1, \ldots, u_{1000}$ the corresponding sign vectors (i.e. $(u_i)_j = \text{sign}(v_i)_j$).
The last 3 methods described here rely on edge contraction in a weighted graph, where the weights may be negative. Contracting an edge of a weighted graph W on n vertices yields a weighted graph W' on $n-1$ vertices in one of two ways.
A *separating contraction* encodes keeping the two vertices on different sides of the cut. In this case, for each $k \neq i, j$ we define $W'_{i,k} = W'_{k,i} = W_{i,k} - W_{j,k}$. A *juxtaposing contraction* encodes keeping the two vertices on the same side of the cut. In this case, for each $k \neq i, j$ we define $W'_{i,k} = W'_{k,i} = W_{i,k} + W_{j,k}$. In both cases we also assign weights to the vertices (the diagonal in W), which encodes the weight of the contracted edges, and thus define $W'_{i,i} = W_{i,i} + W_{i,j} + W_{j,j}$. and $W'_{i,i} = W_{i,i} + W_{i,j} - W_{j,j}$. Finally, in both cases we remove the j'th row and column from the matrix, and keep all other entries as in W.
The six algorithms tested are the following:

1. **All, best:** For each i, consider the cut induced by v_i. Take the best cut. As can be expected, this method outperforms the others. Even for stability 1, it found the optimal cut in 2558 of the 2600 graphs. For stability 1.12 and higher, the optimal cut was missed in only one or two graphs.

2. **All, val; All, sign:** Both these methods use an average of the generated vectors to construct a cut. The first outputs the cut induced by $\text{sign} \sum v_i$, and the second the one induced by $\text{sign} \sum u_i$. As can be seen in Figure 6.2, both methods performed equally well, and their average success ratio is essentially identical. In fact, it is essentially impossible to discern between the two plots in the figure for stability > 1.

3. **Edge, val:** This method iteratively contracts one edge. At *each iteration* 1000 random diagonals are generated, and from them 1000 vectors v_i as above. The algorithm then looks at the vector $z = \sum v_i$, and contracts the edge (i, j) such that $|z_i z_j|$ is maximized. If $z_i z_j < 0$ a separating contraction is performed, otherwise a juxtaposing one.

4. **Edge, sign:** This method iteratively contracts one edge as above. rather than looking at $\sum v_i$, it looks at $\sum u_i$.

5. **Edge, big:** This method is a greedy algorithm. It does not look at the v_i's, but, rather, serves as a baseline to measure how well the spectral methods perform in comparison with this "trivial" algorithm. The algorithm contracts edges one by one, each time choosing the one with the largest absolute value. If the weight is positive, a separating contraction is performed, otherwise a juxtaposing one. Perhaps surprisingly, it outperforms the algorithms described in (2) and (4) above.

This was carried out for 2600 random weighted graphs on 10 vertices, and repeated 11 times, for values of stability growing from 1 to 1.2 by increments of 0.02. A graph was made γ-stable by

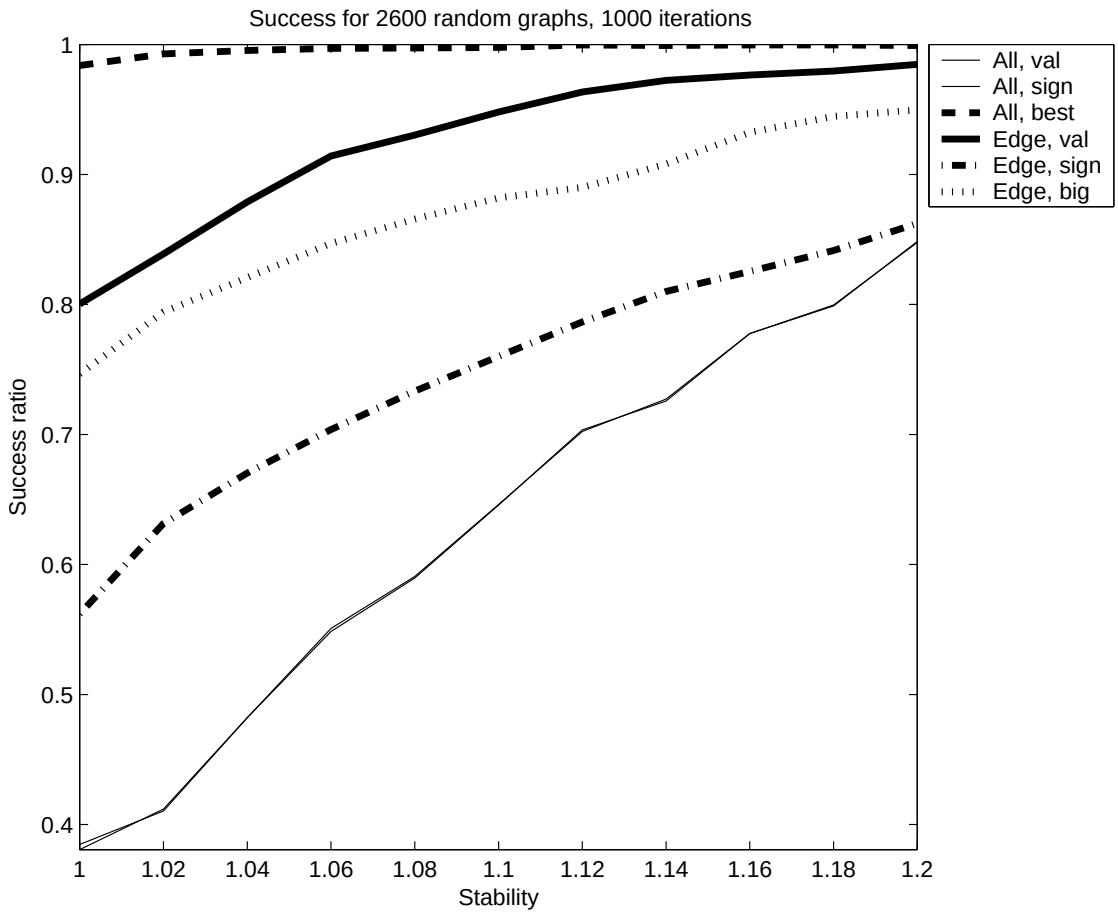

Figure 6.2: Success ratio in identifying the maximal cut as a function of the cut's stability.

multiplying the edges in the maximal cut by γ.

6.3 Other problems

In this section we exemplify how to generalize the greedy algorithm in section 6.2.3 to two other combinatorial optimization problems, Vertex Cover and Multi-way Cut. As before, the input is a graph and a weight function over the edges or the vertices. A solution is either a subset of the vertices or of the edges. An instance is γ-stable, if for all γ-perturbations of the weight function, the solution is identical.

6.3.1 Vertex Cover

The input to the *Vertex Cover* problem is a graph with weights over the vertices. The goal is to find a subset of vertices that intersects all edges, and has minimal total weight. Denote by $VC(G, w)$,

the optimal solution to this problem for input (G, w) (for simplicity assume it is unique). The following algorithm finds the optimal cover if the instance is stable enough:

FindVertexCover(G, w)

1. Initialize $U_0, U_1 = \emptyset$.
2. While U_1 does not intersect all edges in G:
   (a) Choose v a vertex of maximal weight in $V \backslash (U_0 \cup U_1)$.
   (b) Add v to U_0.
   (c) Add all of v's neighbors which are in $V \backslash (U_0 \cup U_1)$ to U_1.
3. Output U_1.

Lemma 6.3.1. *Let (G, w) be a γ-stable instance to the Vertex Cover problem, with γ at least the maximal degree in G. Then the above algorithm outputs the optimal solution.*

Proof: Let S be the optimal solution. We show by induction that throughout the run of the algorithm, $U_1 \subset S \subset V \backslash U_0$.
When $U_0, U_1 = \emptyset$ this is trivial. Suppose the claim holds until a vertex v is chosen in step 2a. Let H be the subgraph of G spanned by $V \backslash (U_0 \cup U_1)$. Denote by $w|_H$ the restriction of w to these vertices. Clearly the maximal degree in H is at most γ. Note that all edges in $G \backslash H$ intersect some vertex in U_1, so by the induction hypothesis, $VC(H, w|_H) = S \backslash U_1$.
It is not hard to see that $(H, w|_H)$ is also γ-stable: If w' is a γ-perturbation of $w|_H$ that leads to a different solution, then its extension to G by identifying it with w over the vertices in U is a γ-perturbation of w that contradicts the stability assumption.
Denote the neighbors of v in H by N. If $N = \emptyset$, then clearly $v \notin VC(H, w|_H)$, and we are done. So assume for contradiction that $|N| > 0$ and $v \in S$. By our assumption on γ, $|N| \leq \gamma$. Consider w', the γ-perturbation of $w|_H$ that multiplies the weight of v by γ, and is identical to $w|_H$ otherwise. Clearly, $S \cup N \backslash \{v\}$ is a cover of the edges. As v has maximal weight, $\gamma \cdot w(v) \geq |N| \cdot w(v) \geq w(N)$. Thus, $w'(S \cup N \backslash \{v\}) \leq w'(S)$, in contradiction with $(H, w|_H)$ being γ-stable.
We conclude that $v \notin S$. But as all its edges need to be covered by some vertex in S, it must be the case that $N \subset S$. Hence, the induction hypothesis continues to hold after steps 2b and 2c. ∎

6.3.2 Multi-way Cut

The input to the *Multi-way Cut* problem is a graph G with weights over the edges, and an independent set of vertices, T. The objective is to find a set of edges with minimal weight, such that their deletion disconnects every vertex in T from all the others. A greedy algorithm gives an exact solution for stable instances:

FindMultiwayCut(G, w)

1. Initialize F to be a graph with no edges over the vertex set of G.
2. While $E\backslash F$ is not a minimal (i.e. there's an $E' \subsetneq E\backslash F$, which is also a multi-way cut) do:
   (a) Let $C_1, \ldots, C_t$ be the connected components of F which are not vertices in T. Let $E_{i,j}$ be the subset of edges between C_i and C_j.
   (b) Let i^* be such that C_{i^*} contains the least number of vertices. Let j^* be such that the weight of E_{i^*,j^*} is maximal among all $E_{i^*,j}$.
   (c) Add the edges of E_{i^*,j^*} to F
3. Output $E\backslash F$.

Lemma 6.3.2. *For $\sqrt{n\Delta}$-stable instances of Multi-way Cut (where n is the number of vertices in the graph, and Δ the maximal degree), the above algorithm outputs the correct solution.*

Proof: The structure of the proof is very similar to that of Lemma 6.2.1. Let M be the set of edges composing the optimal solution. We show by induction that throughout the run of the algorithm, $F \cap M = \emptyset$. When F contains no edges, this holds trivially. Assume that the claim holds up to a point where there are t connected components in T, the smallest of which contains k vertices. Let i^* and j^*, be as in step 2b of the algorithm. So $|C_{i^*}| = k$. Let $J = \{j \neq j^* : E_{i^*,j} \neq \emptyset\}$.
Consider an edge $e = (u, v) \in E_{i^*,j^*}$ ($u \in C_{i^*}, v \in C_{j^*}$). Assume for contradiction that $e \in M$. We first show that if this is the case, then $E_{i^*,j^*} \subseteq M$. From the minimality of M, there are $t_1, t_2 \in T$, such that in $(E\backslash M) \cup \{e\}$ there's a path p from t_1 to t_2. W.l.o.g. assume that it crosses e once, and from u to v. Let (u', v') be some other edge in E_{i^*,j^*}. Consider the path p' from t_1 to t_2, defined as follows. It proceeds from t_1 to u in the same way p does (using no edges of M). It then proceeds from u to u' using edges in F. This can be done since both are in the connected component C_{i^*} of F. By the induction hypothesis, this part of p' is also not in M. Then p' goes from u' to v', from v' to v using edges of F, and from v to t_2 in the same way that p did. Note that the latter two parts do not use edges in M. As t_1 and t_2 are disconnected in $E\backslash M$, it must be the case that $(u', v') \in M$. Hence, $E_{i^*,j^*} \subseteq M$.
Define w' as the γ-perturbed weight function obtained from w by multiplying the weight of the edges in E_{i^*,j^*} by γ. Define $M' = M\backslash E_{i^*,j^*} \cup (\cup_{j\in J} E_{i^*,j})$. Since $T \cap C_{i^*} = \emptyset$, M' is also a multi-way cut. Hence, if the input is γ-stable,

$$\gamma \cdot w(E_{i^*,j^*}) < w(\cup_{j\in J} E_{i^*,j}).$$

As E_{i^*,j^*} was chosen to be of maximal weight, this implies $\gamma < |J|$. Clearly, $|J| \leq t - 2, k\Delta$. In particular, as $tk < n$, $\gamma^2 < tk\Delta \leq n\Delta$. This is a contradiction to the assumption that the input is $\sqrt{n\Delta}$-stable. ∎

Chapter 7

Extensions of Hoffman's bound

7.1 Introduction

7.1.1 Hoffman's bound

Let G be a graph on n vertices, χ its chromatic number, and A its adjacency matrix. Let λ_1 and λ_n the largest and least eigenvalues of A. A theorem of Hoffman [39] states that:

$$\chi \geq 1 - \frac{\lambda_1}{\lambda_n}.$$

The first result in this chapter is a new proof for this bound. We prove, in fact, the following strengthening of it (which also follows from Hofmman's original proof):

Theorem 7.1.1. *[39] Let G be a graph. Let $W \neq 0$ be a symmetric matrix such that $W_{i,j} = 0$ whenever $(i,j) \notin E$. Let λ_1 and λ_n be the largest and least eigenvalues of W. Then $\chi(G) \geq 1 - \frac{\lambda_1}{\lambda_n}$.*

7.1.2 The vector chromatic number

Karger, Motwani and Sudan [44] define a quadratic programming relaxation of the chromatic number, called the *vector chromatic number*. This is the minimal k such that there exist unit vectors $u_1, \ldots, u_n \in \mathbb{R}^n$ with:

$$< u_i, u_j > \leq -\frac{1}{k-1},$$

whenever (i,j) is an edge in the graph.

Let χ_v denote the vector chromatic number of G. Karger, Motwani and Sudan observe that $\chi_v \leq \chi$. In this chapter we show that Hoffman's bound holds for this parameter as well:

Theorem 7.1.2. *Let G be a graph. Let $W \neq 0$ be a symmetric matrix such that $W_{i,j} = 0$ whenever $(i,j) \notin E$. Let λ_1 and λ_n be the largest and least eigenvalues of W. Then*

$$\chi_v(G) \geq 1 - \frac{\lambda_1}{\lambda_n}.$$

7.1.3 The ψ-covering number

Let ψ be a graph parameter, such that $\psi = 1$ on graphs with no edges, and for every graph G, $\psi(G) \geq \chi(G)$. The ψ-*covering* number of a graph G was defined by Amit, Linial and Matoušek [12] to be the minimal k such that there exist k subsets of G, $S_1, \ldots, S_k$ so that for every $v \in G$, $\sum_{i\,:\,v\in S_i} \frac{1}{\psi(G[S_i])} \geq 1$.
They show that this value is bounded between $\sqrt{\chi(G)}$ and $\chi(G)$, and ask whether better lower bounds can be proven when $\psi(G) = dgn(G)+1$ (the degeneracy of G + 1), and $\psi(G) = \Delta(G)+1$ (the maximal degree in G + 1).
To state our result, we'll need a couple of ad-hoc definitions:

Definition 7.1.1. *A graph G has c-*vertex cover *if there exists a cover $E(G) = \cup_{i\in V(G)} E_i$ such that for all $i \in V(G)$, $E_i \subset \{e \in E \,:\, i \in e\}$, and $|E_i| \leq c$.*
Denote by $L_{\psi,\alpha}(G)$ the minimal k such that there exist k subsets of G, $S_1, \ldots, S_k$, so that for all $i = 1, \ldots, k$, $G[S_i]$ has a $\alpha \cdot \psi(G)$-vertex cover, and for every $v \in G$, $\sum_{i\,:\,v\in S_i} \frac{1}{\psi(G[S_i])} \geq 1$.

Observe that all graphs have a $\psi(G)$-vertex cover for $\psi(G) = dgn(G) + 1$, and a $\frac{1}{2}\psi(G)$-vertex cover when $\psi(G) = \Delta(G) + 1$. So in these cases, $L_{\psi,1}$ and $L_{\psi,\frac{1}{2}}$, respectively, are exactly the ψ-covering numbers.
Note also that $\alpha = 0$ means that the S_i are independent sets. Thus, since $\psi = 1$ on such sets, $L_{\psi,0} = \chi$.

Theorem 7.1.3.

$$L_{\psi,\alpha}(G) \geq \frac{d - \lambda_n}{2\alpha - \lambda_n},$$

where λ_n is the least eigenvalue of G, and d the average degree.

Note that when the graph is regular and $\alpha = 0$, this is the same as Hoffman's bound.
For random d-regular graphs, $|\lambda_n| = O(\sqrt{d})$ and $\chi = \Theta(\frac{d}{\log d})$. So in this case (if α is taken small) the bound is slightly better than $\sqrt{\chi}$ mentioned above.

7.1.4 The λ-clustering number

Finally, we are interested in a graph parameter that has to do with how well a graph can be partitioned into sparse clusters:

Definition 7.1.2. *Let W be a weighted adjacency matrix of a graph G. A partition $V = \dot\cup_{i=1}^{k} C_i$ is a* λ-clustering *of G into k clusters if*

$$\max_{i\in[k]} \lambda_1(C_i) \leq \lambda,$$

where $\lambda_1(C_i)$ is the largest eigenvalue of the sub-graph spanned by the vertices in C_i.
The λ-clustering number of G is the minimal k such that there exists a λ-clustering of G into k clusters.

It is not hard to see that the 0-clustering number is identical to the chromatic number.
For example, let G be a weighted graph with $\sum_j W_{i,j} = d$ for all $i = 1, \ldots, n$, which is γ-stable for the k-clustering problem (see Chapter 6). Let $C_1, \ldots, C_k$ denote the optimal solution. From γ-stability it follows that for each i, and each $v \in C_i$, the maximal weight of edges emanating from v to neighbors in C_i is at most $\frac{\gamma-1}{\gamma+1} \cdot d$. The maximal row sum is an upper bound on the spectral radius, and so $\lambda_1(C_i) \leq \frac{\gamma-1}{\gamma+1} \cdot d$. Hence, for such instances, for $\lambda \geq \frac{\gamma-1}{\gamma+1} \cdot d$, the λ-clustering number is at most k.
We show that Hoffman's bound can also be extended to this graph parameter:

Theorem 7.1.4. *Let W be a weighted adjacency matrix. Let λ_1 and λ_n the largest and least eigenvalues of W. The λ-clustering number of the graph is at least:*

$$\frac{\lambda_1 - \lambda_n}{\lambda - \lambda_n}.$$

7.2 Vectorial characterization of the least eigenvalue

The proofs of the four theorems mentioned in the previous section rely on the following observation:

Lemma 7.2.1. *Let A be a real symmetric matrix and λ_n its least eigenvalue.*

$$\lambda_n = \min \frac{\sum_{i,j=1}^n A_{i,j} < v_i, v_j >}{\sum_{i=1}^n ||v_i||_2^2}. \tag{7.1}$$

where the minimum is taken over all $v_1, \ldots, v_n \in \mathbb{R}^n$.

Proof: By the Rayleigh-Ritz characterization, λ_n equals

$$\begin{aligned} \min \quad & \sum_{i,j} A_{i,j} x_i x_j \\ \text{s.t.} \quad & x \in \mathbb{R}^n \\ & ||x||_2 = 1. \end{aligned}$$

Denote by PSD_n the cone of $n \times n$ positive semi-definite matrices. For each unit vector $x \in \mathbb{R}^n$, let X be the matrix $X_{i,j} = x_i x_j$. This is a positive semi-definite matrix of rank 1 and trace 1, and all such matrices are obtained in this way. Hence, λ_n equals:

$$\begin{aligned} \min \quad & \sum_{i,j} A_{i,j} X_{i,j} \\ \text{s.t.} \quad & X \in PSD_n \\ & rank(X) = 1 \\ & tr(X) = 1. \end{aligned}$$

However, the rank restriction is superfluous. It restricts the solution to an extreme ray of the cone PSD_n, but, by convexity, the optimum is attained on an extreme ray anyway. Hence, λ_n equals:

$$\begin{aligned} \min \quad & \sum_{i,j} A_{i,j} X_{i,j} \\ \text{s.t.} \quad & X \in PSD_n \\ & tr(X) = 1. \end{aligned}$$

Now, think of each $X \in PSD_n$ as a Gram matrix of n vectors, $v_1, \ldots, v_n$ (i.e. $X_{i,j} =< v_i, v_j >$). An equivalent formulation of the above is thus:

$$\begin{aligned} \min \quad & \sum_{i,j} A_{i,j} < v_i, v_j > \\ \text{s.t.} \quad & v_i \in \mathbb{R}^n \qquad \text{for } i = 1, \ldots, n \\ & \sum_{i=1}^{n} ||v_i||_2^2 = 1. \end{aligned}$$

Clearly, this is equivalent to 7.1. ∎

7.3 Proofs of the theorems

7.3.1 A new proof of Hoffman's bound

We start by proving the strengthening of Hoffman's bound [39] stated in Theorem 7.1.1. Let G be a graph on n vertices with chromatic number χ. Let $c : V \to [\chi]$ be a χ-coloring of G. Let $W \neq 0$ be a symmetric matrix such that $W_{i,j} = 0$ whenever $(i,j) \notin E$. Let λ_1 and λ_n be the largest and least eigenvalues of W.
By Lemma 7.2.1, any choice of n vectors gives an upper bound on λ_n. We now choose particular vectors, $v_1, \ldots, v_n$. Let $u_1, \ldots, u_\chi \in S^n$ be the vertices of a regular $(\chi - 1)$-dimensional simplex centered at the origin. In other words, $< u_i, u_j >= -\frac{1}{\chi - 1}$ for $i \neq j$, and $||u_i||_2 = 1$. Let $\alpha \in \mathbb{R}^n$ be an eigenvector of W, corresponding to λ_1. Set $v_i = \alpha_i \cdot u_{c(i)}$.
For i and j such that $c(i) = c(j)$ (i.e. v_i and v_j are on the same line), $W_{i,j} = 0$. Hence, by Lemma 7.2.1,

$$\begin{aligned} \lambda_n \quad &\leq \quad \frac{\sum_{i,j} W_{i,j}\alpha_i\alpha_j < u_i, u_j >}{\sum_{i=1}^{n} \alpha_i^2 \cdot ||u_i||_2^2} = -\frac{1}{\chi - 1} \cdot \frac{\sum_{i,j} W_{i,j}\alpha_i\alpha_j}{\sum_i \alpha_i^2} = \\ &= \quad -\frac{1}{\chi - 1} \cdot \frac{\alpha^t W \alpha}{||\alpha||^2} = -\frac{1}{\chi - 1} \cdot \lambda_1. \end{aligned}$$

Equivalently, $\chi \geq 1 - \frac{\lambda_1}{\lambda_n}$, as claimed. ∎

7.3.2 Extending Hoffman's bound to vector coloring

We now prove Theorem 7.1.2. Let G be a graph on n vertices with vector chromatic number χ_v. Let $W \neq 0$ be a symmetric matrix such that $W_{i,j} = 0$ whenever $(i,j) \notin E$. Let λ_1 and λ_n be the largest and least eigenvalues of W.
Again we choose vectors $v_1, \ldots, v_n$, and look at the bound they give on λ_n. Let $u_1, \ldots, u_n \in S^n$ be vectors on which the vector chromatic number is attained. That is, $< u_i, u_j > \leq -\frac{1}{\chi_v - 1}$ for $(i,j) \in E$, and $||u_i||_2 = 1$. Let $\alpha \in \mathbb{R}^n$ be an eigenvector of W corresponding to λ_1. Set $v_i = \alpha_i \cdot u_i$.
Since $W_{i,j} = 0$ whenever $< u_i, u_j > > -\frac{1}{\chi_v - 1}$, by Lemma 7.2.1,

$$\begin{aligned} \lambda_n &\leq \frac{\sum_{i,j} W_{i,j}\alpha_i\alpha_j < u_i, u_j >}{\sum_{i=1}^n \alpha_i^2 \cdot ||u_i||_2^2} \leq -\frac{1}{\chi_v - 1} \cdot \frac{\sum_{i,j} W_{i,j}\alpha_i\alpha_j}{\sum_i \alpha_i^2} = \\ &= -\frac{1}{\chi_v - 1} \cdot \frac{\alpha^t W \alpha}{||\alpha||^2} = -\frac{1}{\chi_v - 1} \cdot \lambda_1. \end{aligned}$$

Equivalently, $\chi_v \geq 1 - \frac{\lambda_1}{\lambda_n}$, as claimed. ∎

7.3.3 Extending the bound to ψ-covering numbers

In this section we prove Theorem 7.1.3. Denote $k = L_{\psi,\alpha}(G)$, and let $u_1, \ldots, u_k$ be the vertices of the regular $k-1$-dimensional simplex centered at 0 - i.e. $< u_i, u_j >= 1$ when $i = j$ and $\frac{-1}{k-1}$ otherwise. Again we choose vectors $v_1, \ldots, v_n$. We do so probabilistically. Let $S_1, \ldots, S_k$ be the subsets attaining the value k. For each i, v_i will be chosen from among the u_j's such that $i \in S_j$. Specifically, let $h_i = \sum_{j\,:\,i \in S_j} \frac{1}{\psi(S_j)}$. The probability that v_i is chosen to be u_j is $p_{i,j} = h_i^{-1}\frac{1}{\psi(S_j)}$. Note that $h_i \geq 1$, and so $p_{i,j} \leq \frac{1}{\psi(S_j)}$. (there is a slight abuse of notation here - by $\psi(S_j)$ we refer to $\psi(G[S_j])$.)
Say that an edge is "bad" if both its endpoints are assigned the same vector. For a given j, the probability that an edge $(i,i') \in E(S_j)$ is "bad" because both endpoints were assigned to u_j is $p_{i,j}p_{i',j}$. Thus, the expected number of "bad" edges is at most:

$$\sum_{j=1}^{k} \sum_{(i,i') \in E(S_j)} p_{i,j}p_{i',j}.$$

Each S_j has a $\alpha \cdot \psi(G)$-vertex cover $E(S_j) = \cup E_j^i$. Summing the expression above according to this cover (some edges might be counted more than once) we get that the expected number of "bad" edges is at most:

$$\sum_{j=1}^{k} \sum_{i \in S_j} \sum_{i'\,:\,(i,i') \in E_j^i} p_{i,j}p_{i',j} \leq \sum_{j=1}^{k} \sum_{i \in S_j} p_{i,j}|E_j^i| \frac{1}{\psi(S_j)} \leq \sum_{j=1}^{k} \sum_{i \in S_j} p_{i,j}\alpha = \alpha \sum_{i \in G} \sum_{j\,:\,i \in S_j} p_{i,j} = \alpha n.$$

In particular, there is a choice of v_i's such that the number of "bad" edges is at most this value. Assume this is the case. If $(i,j) \in E(G)$ is a "bad" edge then $< v_i, v_j >= 1$. Otherwise $< v_i, v_j >= -\frac{1}{k-1}$.

Lemma 7.2.1 now gives:

$$\lambda_n \leq \left(2\alpha n - \frac{1}{k-1}(dn - 2\alpha n)\right)/n = \frac{2k\alpha}{k-1} - \frac{d}{k-1},$$

or $(k-1)\lambda_n \leq 2k\alpha - d$. Equivalently, $k \geq \frac{d-\lambda_n}{2\alpha-\lambda_n}$. ∎

Note 7.3.1. *In the definition of the ψ-covering number, and of $L_{\psi,\alpha}$, it is required that for every $v \in G$, $\sum_{i\,:\,v\in S_i} \frac{1}{\psi(S_i)} \geq 1$. The bounds given in [12] hold also if we demand that all sums* equal 1. *In this case, the theorem holds also if we relax the condition that all S_i have an $\alpha \cdot \psi(G)$-vertex cover, and require only that for each S_i, $|E(S_i)| \leq \alpha \cdot \psi(S_i) \cdot |V(S_i)|$.*

7.3.4 Extending the bound to the λ-clustering number

Here we prove Theorem 7.1.4. Denote the λ-clustering number of W by k. Let $u_1, \ldots, u_k \in R^n$ be the vertices of a regular simplex centered at the origin, as above. Let $\alpha \in \mathbb{R}^n$ be an eigenvector of W, corresponding to λ_1. Let $C_1, \ldots, C_k$ be a λ-clustering of G. Define $\phi : V \to [k]$ to be the index of the cluster containing a vertex. That is, $i \in C_{\phi(i)}$. Define W_t to be the weighted adjacency matrix of the sub-graph spanned by C_t (So $\lambda_1(W_t) \leq \lambda$). Set $v_i = \alpha_i \cdot u_{\phi(i)}$.
By Lemma 7.2.1,

$$\begin{aligned}
\lambda_n &\leq \frac{\sum_{i,j} \alpha_i\alpha_j < u_{\phi(j)}, u_{\phi(j)} > W_{i,j}}{\sum_{i=1}^n \alpha_i^2 \cdot ||u_{\phi(i)}||_2^2} \\
&= -\frac{1}{k-1} \cdot \frac{\sum_{i,j\,:\,\phi(i)\neq\phi(j)} \alpha_i\alpha_j W_{i,j}}{\sum_i \alpha_i^2} + \frac{\sum_{i,j\,:\,\phi(i)=\phi(j)} \alpha_i\alpha_j W_{i,j}}{\sum_i \alpha_i^2} \\
&= -\frac{1}{k-1} \cdot \frac{\alpha^t W \alpha}{||\alpha||^2} + \frac{k}{k-1} \frac{\sum_{t=1}^k \sum_{i,j\in C_t} \alpha^t W_t \alpha}{\sum_i \alpha_i^2} \\
&\leq -\frac{1}{k-1}\lambda_1 + \frac{k}{k-1}\lambda
\end{aligned}$$

Equivalently, $k \geq \frac{\lambda_n-\lambda_1}{\lambda_n-\lambda}$ ∎

Personal comments

Acknowledgments

I was fortunate to have Nati Linial as an advisor. I am grateful for his endless patience and enthusiasm, and his sure-handed guidance.
I was doubly fortunate to have Michal Linial as a de-facto second advisor, who was ever generous with her time and knowledge, and ever patient with my fickle path between computational biology and combinatorics.
I would also like to thank the following people for their encouragement and inspiration, for answering my questions and for listening to my ideas, over the many years of my studies: Alex Samorodnitsky, Avi Wigderson, Avner Magen, Buffy Summers, my love Dana Porrat, Danny Gutfreund, Eli Ben-Sasson, Elon Potrugaly, Eyal Rozenman, Miki Ben-Or, Noam Slonim, Noga Alon, Rachel Kolodny, Ronnen Shaltiel, Shlomo Hoory, Tamir Hazan, Yael Dekel-Vinner and Yoav Yaffe.

Research not described here

I have included in this thesis the parts of my graduate research which do not already appear elsewhere (at least at the time of writing). Nonetheless, I would like the thesis to reflect a full picture of my research, and in particular my work in the field of computational biology. Thus, I somewhat vainly include here a list of papers and technical reports that describe this work:

- *The Advantage of Functional Prediction Based on Clustering of Yeast Genes and Its Correlation with Non-Sequence Based Classifications*. With Michal Linial, in JCB 9 (2002) (Earlier version appeared in RECOMB 2001).
- *Functional Consequences in Metabolic Pathways from Phylogenetic Profiles*. With Michal Linial, in WABI 2002.
- *Locating Transcription Factors Binding Sites Using a Variable Memory Markov Model*. With Michal Linial, Noam Slonim and Naftali Tishby. Leibintz Center TR 2002-57.
- *On codes from hypergraphs*. With Shlomo Hoory, in Eur. J. Comb. 25(3) (2004).
- *A Gap in Average Proof Complexity*. With Eli Ben-Sasson, ECCC report 003 (2002).

I was also involved in the ProtoNet project, described in the following paper:

- *ProtoNet: hierarchical classification of the protein space*. Ori Sasson, Avishay Vaaknin, Hillel Fleischer, Elon Portugaly, Yonatan Bilu, Nathan Linial and Michal Linial. In NAR 31(1) (2003).

Bibliography

[1] M. Ajtai. Recursive construction for 3-regular expanders. *Combinatorica*, 14(4):379–416, 1994.

[2] N. Alon. Personal communication.

[3] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of computing (Singer Island, Fla., 1984).

[4] N. Alon. A parallel algorithmic version of the local lemma. *Random Structures Algorithms*, 2(4):367–378, 1991.

[5] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. *Symposium on Foundations of Computer Science*, 26:277–280, 1985.

[6] N. Alon, Z. Galil, and V. D. Milman. Better expanders and superconcentrators. *J. Algorithms*, 8(3):337–347, 1987.

[7] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. In *31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990)*, pages 544–553. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990.

[8] N. Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Combin. Theory Ser. B*, 38(1):73–88, 1985.

[9] 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.

[10] A. Amit and N. Linial. Random graph coverings. I. General theory and graph connectivity. *Combinatorica*, 22(1):1–18, 2002.

[11] A. Amit and N. Linial. Random lifts of graphs II: Edge expansion. To appear in Combinatorics Probability and Computing, 2003.

[12] A. Amit, N. Linial, and J. Matoušek. Random lifts of graphs: independence and chromatic number. *Random Structures Algorithms*, 20(1):1–22, 2002.

[13] A. Berman and N. Shaked-Monderer. *Completely positive matrices*. World Scientific Publishing Co. Inc., River Edge, NJ, 2003.

[14] B. Bollobás and V. Nikiforov. Graphs and hermitian matrices: discrepancy and singular values. To appear in Discrete Math.

[15] R. Boppana. Eigenvalues and graph bisection: An average case analysis. In *28th Annual Symposium on Foundations of Computer Science, October 12–14, 1987, Los Angeles, California*, pages 280–285. IEEE Comput. Soc. Press, 1987.

[16] H. Breu and D. G. Kirkpatrick. Unit disk graph recognition is NP-hard. *Comput. Geom.*, 9(1-2):3–24, 1998.

[17] A. Broder and E. Shamir. On the second eigenvalue of random regular graphs. In *28th Annual Symposium on Foundations of Computer Science, October 12–14, 1987, Los Angeles, California*, pages 286–294. IEEE Comput. Soc. Press, 1987.

[18] T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, 1987.

[19] P. J. Cameron, J.-M. Goethals, J. J. Seidel, and E. E. Shult. Line graphs, root systems, and elliptic geometry. *J. Algebra*, 43(1):305–327, 1976.

[20] A. Condon and R. M. Karp. Algorithms for graph partitioning on the planted partition model. *Random Structures Algorithms*, 18(2):116–140, 2001.

[21] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley Series in Telecommunications. John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.

[22] C. Delorme and S. Poljak. Laplacian eigenvalues and the maximum cut problem. *Math. Programming*, 62(3, Ser. A):557–574, 1993.

[23] M. Deza and M. Laurent. Applications of cut polyhedra. I, II. *J. Comput. Appl. Math.*, 55(2):191–216, 217–247, 1994.

[24] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2000.

[25] M. E. Dyer and A. Frieze. Fast solution of some random np-hard problems. In *27th Annual Symposium on Foundations of Computer Science, October 27–29, 1986, Toronto, Ontario, Canada*, pages 313–321. IEEE Comput. Soc. Press, 1986.

[26] P. Erdős and L. Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II*, pages 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.

[27] U. Feige and J. Kilian. Heuristics for semirandom graph problems. *J. Comput. System Sci.*, 63(4):639–671, 2001. Special issue on FOCS 98 (Palo Alto, CA).

[28] P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.

[29] J. Friedman. A proof of alon's second eigenvalue conjecture. manuscript.

[30] J. Friedman. On the second eigenvalue and random walks in random d-regular graphs. *Combinatorica*, 11(4):331–362, 1991.

[31] J. Friedman. Relative expanders or weakly relatively Ramanujan graphs. *Duke Math. J.*, 118(1):19–35, 2003.

[32] J. Friedman, J. Kahn, and E. Szemeredi. On the second eigenvalue of random regular graphs. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 587–598. ACM Press, 1989.

[33] A. M. Frieze and M. Molloy. Splitting an expander graph. *J. Algorithms*, 33(1):166–172, 1999.

[34] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

[35] O. Gabber and Z. Galil. Explicit constructions of linear-sized superconcentrators. *J. Comput. System Sci.*, 22(3):407–420, 1981. Special issued dedicated to Michael Machtey.

[36] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42(6):1115–1145, 1995.

[37] M. Grötschel, L. Lovász, and A. Schrijver. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica*, 1(2):169–197, 1981.

[38] M. Grötschel, L. Lovász, and A. Schrijver. Corrigendum to our paper: "The ellipsoid method and its consequences in combinatorial optimization" [Combinatorica **1** (1981), no. 2, 169–197. *Combinatorica*, 4(4):291–295, 1984.

[39] A. J. Hoffman. On eigenvalues and colorings of graphs. In *Graph Theory and its Applications (Proc. Advanced Sem., Math. Research Center, Univ. of Wisconsin, Madison, Wis., 1969)*, pages 79–91. Academic Press, New York, 1970.

[40] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.

[41] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 10–33. IEEE Computer Soc., Los Alamitos, CA, 2001.

[42] M. Jerrum and G. B. Sorkin. The Metropolis algorithm for graph bisection. *Discrete Appl. Math.*, 82(1-3):155–175, 1998.

[43] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in modern analysis and probability (New Haven, Conn., 1982)*, volume 26 of *Contemp. Math.*, pages 189–206. Amer. Math. Soc., Providence, RI, 1984.

[44] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *J. ACM*, 45(2):246–265, 1998.

[45] D. E. Knuth. *The art of computer programming. Vol. 1: Fundamental algorithms*. Second printing. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1969.

[46] 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.

[47] N. Linial and E. Rozenman. Random lifts of graphs: Perfect matchings. To appear in Combinatorica.

[48] N. Linial and A. Wigderson. Expander graphs and their applications. Lecture notes available at http://www.math.ias.edu/∼boaz/ExpanderCourse/.

[49] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[50] H. Maehara. Space graphs and sphericity. *Discrete Appl. Math.*, 7(1):55–64, 1984.

[51] G. A. Margulis. Explicit constructions of expanders. *Problemy Peredači Informacii*, 9(4):71–80, 1973.

[52] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988.

[53] J. Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[54] F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science (Las Vegas, NV, 2001)*, pages 529–537. IEEE Computer Soc., Los Alamitos, CA, 2001.

[55] J. W. Milnor. On the Betti numbers of real algebraic varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[56] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993.

[57] A. Nilli. On the second eigenvalue of a graph. *Discrete Math.*, 91(2):207–210, 1991.

[58] B. Reed. Mangoes and blueberries. *Combinatorica*, 19(2):267–296, 1999.

[59] 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.

[60] J. Reiterman, V. Rödl, and E. Šiňajová. Embeddings of graphs in Euclidean spaces. *Discrete Comput. Geom.*, 4(4):349–364, 1989.

[61] J. Reiterman, V. Rödl, and E. Šiňajová. Geometrical embeddings of graphs. *Discrete Math.*, 74(3):291–319, 1989.

[62] J. Reiterman, V. Rödl, and E. Šiňajová. On embedding of graphs into Euclidean spaces of small dimension. *J. Combin. Theory Ser. B*, 56(1):1–8, 1992.

[63] O. Sasson et al. Protonet: hierarchical classification of the protein space. *Nuc. Acids Res.*, 31(1):348–352, 2003.

[64] R. Shamir and D. Tsur. Improved algorithms for the random cluster graph model. In *Proceedings of the 8th Scandinavian Workshop on Algorithm Theory*, pages 230–239. 2002.

[65] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 279–319. Academic Press, New York, 1968.

[66] A. Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

[67] J. H. van Lint. *Introduction to coding theory*, volume 86 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 1999.

על תכונות ספקטרליות של גרפים ושימושיהן בהצברה

חיבור לשם קבלת תואר דוקטור לפילוסופיה

מאת: יונתן בילו

הוגש לסינט האוניברסיטה העברית, בירושלים
אוגוסט תשס"ד

עבודה זו נעשתה בהדרכתו של
פרופסור נתי ליניאל

תקציר

מטרת העבודה הינה חקר תכונות ספקטרליות של גרפים, תוך דגש על שימושים לשאלת ההצברה. הספקטרום של גרף הוא אוסף הערכים העצמיים של מטריצת השכנויות שלו. עניין מרכזי בתורת הגרפים הוא הבנת הקשר בין ערכים אלו, ובין תכונות קומבינטוריות של הגרף. התכונות בהן אנו מתעניינים בעבודה זו קשורות לשאלת ההצברה – בהנתן גרף ממושקל, מצא חלוקה של קדקודי הגרף לצברים, כך שמשקל הצלעות בין קדקודים באותו הצבר קטן, ומשקל הצלעות בין קדקודים בצברים שונים - גדול.

התכונה הראשונה בה העבודה עוסקת היא מידת הדמיון של התפלגות מספר הצלעות בגרף להתפלגות בגרף מקרי. לגרף d-רגולרי G על n קדקודים מסמנים את ה"ערבוביות" של הגרף:

$\alpha(G) = \max_{S,T\subseteq V} |e(S,T) - \frac{d}{n}|S||T|| / \sqrt{|S||T|}$, כאשר e(S,T) הוא מספר הצלעות עם קדקוד בכ"א משתי הקבוצות.

תכונה זו קשורה הן לערך העצמי השני של הגרף, והן למידת ההרחבה הצלעית שלו. הווה ידוע כי לערך העצמי השני בערכו המוחלט של גרף, λ, מתקיים $\alpha(G) \le \lambda$. בפרק 2 אנו מראים כי ה"ערבוביות" של הגרף חוסמת את הערך העצמי השני גם מלעיל - $\lambda = O(\alpha(G)(\log(d/\alpha(G))+1))$.

יתר על כן, ההוכחה היא קונסטרוקטיבית במובן הבא. בהנתן גרף עם ערך עצמי שני גדול, ניתן למצוא ביעילות זוג קבוצות S ו-T כך ש- $|e(S,T) - \frac{d}{n}|S||T||$ גדול. הדעת נותנת כי עובדה זו יכולה להיות שימושית לפתרון בעיות הצברה, אם כי בשלב זה איננו יודעים כיצד לבנות ממנה אלגוריתם.

החסם לעיל על הערך העצמי השני של גרף נובע מקשר כללי יותר בין האנלוג לתכונת ה"ערבוביות" במטריצה סימטרית ממשית, ובין הרדיוס הספקטרלי שלה. בעזרת קשר זה אנו מראים בפרק 3 בנייה פשוטה של גרפים עם פער ספקטרלי כמעט אופטימלי. חסם אלון-בופאנה אומר כי בגרף d-רגולרי, $\lambda \ge 2\sqrt{d-1} - o(1)$. בבנייה שלנו, $\lambda = O(\sqrt{d\log^3 d})$.

הבנייה מסתמכת על 2-הרמות של גרפים. גרף G' נקרא 2-הרמה של G אם קיימת העתקת כיסוי 2:1 מ-G' ל-G. במילים אחרות, כנגד כל קדקוד x ב-G יש ב-G' שני קדקודים, x1 ו-x2. כנגד כל צלע (x,y) ב-G יש ב-G' זוג צלעות: או זוג הצלעות (x1,y1) ו-(x2,y2) - הרמה "שטוחה" של הצלע, או זוג הצלעות (x1,y2) ו-(x2,y1) - הרמה "מוצלבת" של הצלע.

באופן שקול, ניתן לתאר 2-הרמה של G ע"י סימון הצלעות ב-1 ו-(-1), כאשר 1 מסמן הרמה "שטוחה" ו-(-1) מסמן הרמה "מוצלבת". אבחנה לא קשה היא שהספקטרום של הגרף

המורם הוא סך הערכים העצמיים של מטריצת השכנויות של גרף הבסיס ואלה של המטריצה המסומנת באופן המתואר לעיל. בעקבות יואל פרידמן אנו קוראים לראשונים הערכים העצמיים ה"ישנים" של הגרף, ולאחרונים ה"חדשים".
אנו מציעים לבנות גרפים מרחיבים באופן הבא. התחל מגרף בסיס (קטן) בעל ערך עצמי שני קטן – למשל הגרף השלם. באופן איטרטיבי מצא סימון כך שהרדיוס הספקטרלי של המטריצה המסומנת קטן. ע"פ האבחנה לעיל לאורך כל התהליך הפער הספקטרלי של הגרף גדול: הוא כזה מלכתחילה, ובעקבות כל איטרציה נוספים לגרף ערכים עצמיים "חדשים" שכולם קטנים.
במילים אחרות, כדי לבנות גרפים עם פער ספקטרלי גדול, די למצוא סימון כך שלמטריצה המתארת אותו יש רדיוס ספקטרלי קטן. אנו מראים שלכל גרף קיים סימון עם רדיוס ספקטרלי $O(\sqrt{d\log^3 d})$, וכן כי בתנאים מסוימים (המספיקים לצורך הבנייה) ניתן למצוא את הסימון ביעילות.

פרק 4 דן בגרפים עם ערך עצמי שני קבוע ודרגה לינארית. אנו מראים כי במובן מסוים, הדוגמא היחידה לגרפים כאלה היא גרפים דו-צדדיים מלאים. ליתר דיוק, בהנתן $c \le 1/2$, וקבוע L, עבור n מספיק גדול, אם G גרף cn-רגולרי עם ערך עצמי שני לכל היותר L, אז G "קרוב" לגרף דו"צ מלא. שני גרפים נקראים "קרובים" אם ניתן לקבל מהאחד את השני ע"י מחיקה והוספה של $o(n^2)$ צלעות.

פרק 5 דן בהעתקות מונוטוניות וספריות של גרפים. העתקה ממרחב מטרי אחד למשנהו נקראת מונוטונית אם היא שומרת את יחס הסדר של המרחקים. במובן זה, העתקה כזו שומרת חלק מהמבנה של המרחב המקורי. טכניקה שימושית להתמודדות עם בעיות הצברה היא שיכון של הבעיה המקורית במרחב "טוב" – מרחב בו קיימים אלגוריתמים או היוריסטיקות לפתרון הבעיה – תוך שמירה על המבנה הרלוונטי בבעייה המקורית. דוגמא למרחב שנחשב "טוב" הוא מרחב אוקלידי ממימד נמוך.
אנו מראים חסמים עליונים ותחתונים על המימד המינימלי של מרחבים נורמיים, אליו ניתן להעתיק מונוטונית כל מרחב מטרי. בפרט אנו מראים כי ניתן להעתיק מונוטונית כל מרחב מטרי על n נקודות לתוך מרחב אוקלידי ממימד (n-1), וכי נדרש מימד לינארי לכמעט כל מרחב מטרי (במובן המוגדר שם).
מקרה פרטי של העתקה מונוטונית היא העתקה "ספרית" של גרף. שיכון של קדקודי גרף, $f:V\to\Re^k$ נקרא "ספרי" אם מתקיים: $\|f(x)-f(y)\|<1$ אם ורק אם x ו-y שכנים בגרף.
ה"ספריות" של גרף היא המימד המינימלי בו ניתן לשכן אותו באופן זה. אנו מראים חסם תחתון על ספריות של גרפים, במונחי הערך העצמי השני. חסם זה הוא המוטיבציה העיקרית

לאפיון גרפים עם דרגה לינארית וערך עצמי שני חסום בפרק 4. עבור גרפים כאלה החסם הוא לינארי.

בחלק האחרון של הפרק אנו דנים בשאלה הקשורה להצברה רכה. כאן המטרה אינה לחלק את הנקודות לצברים, אלא להתאים לכל נקודה התפלגות על הצברים. כמו בהצברה "קשה", המטרה היא שלנקודות קרובות תהיינה התפלגויות קרובות, ולהפך.

מנקודת מבט זו, אפשר לחשוב על הקלט לבעיית ההצברה - מטריצת מרחקים בין נקודות - כמטריצת מרחקים בין התפלגויות. בהנתן אופרטור מרחק בין התפלגויות, שאלה טבעית היא איפה אילו מטריצות מרחקים יכולות להתקבל באופן זה. אנו מראים כי כאשר האופרטור הוא אופרטור האנטרופיה היחסית (KL-Divergence), ניתן לקבל כך כל מטריצת מרחק.

פרק 6 דן בקלטים "יציבים" לבעיית ההצברה, ומתמקד במקרה הפרטי של בעיית החתך המקסימלי. בבעיית החתך המקסימלי הקלט הוא גרף ממושקל, והמטרה היא לחלק את הקדקודים לשתי קבוצות, כך שמשקל הצלעות שקדקוד שלהן נמצא בכ"א מהקבוצות הוא מקסימלי. קלט לבעייה נקרא γ-יציב, אם הכפלה של הצלעות בערכים בין 1 ו-γ לא משנה את החלוקה האופטימלית. האם יציבות הופכת הבעיה לפתירה באופן יעיל? אנו מראים כי אם הקלט h-יציב, אז אלגוריתם חמדן מוצא את הפתרון האופטימלי. יחד עם זאת, אנו משערים כי די ביציבות קבועה כדי להפוך את הבעייה לפתירה באופן יעיל.

האם יציבות הופכת את הבעייה לקלה יותר לקירוב, עבור אלגוריתמים מוכרים? האלגוריתם של Goemans-Williamson מבטיח 0.879-קירוב לבעייה. למעשה, GW מבחינים כי יחס הקירוב משתפר כאשר היחס בין משקל החתך המקסימלי וסך משקל הצלעות מתקרב ל-1. על-סמך תוצאה זו, אנו מראים כי אם הקלט γ-יציב, יחס הקירוב המובטח הוא $1-O(1/\sqrt{\gamma})$.

האם יציבות מאפשרת לנתח היוריסטיקות לפתרון הבעייה, ולהראות שהן תמצאנה את הפתרון האופטימלי? היוריסטיקה אחת לפתרון הבעייה נקראת "חלוקה ספקטרלית" (Spectral Partitioning). ע"פ שיטה זו, מתבוננים בוקטור העצמי של מטריצת הגרף המתאים לערך העצמי המינימלי, ומחלקים את הקדקודים ע"פ סימני הקואורדינטות המתאימות. אנו מראים תנאים מספיקים המבטיחים שהיוריסטיקה זו (למעשה, וריאציה קלה שלה) תמצא את החתך המקסימלי.

ב-1970 הופמן הראה חסם ספקטרלי למספר הצביעה של גרף: $\chi \geq 1-\frac{\lambda_1}{\lambda_n}$, באשר χ מספר הצביעה של הגרף, λ_1 הערך העצמי העליון של מטריצת השכנויות שלו ו-λ_n הערך העצמי התחתון. בהקשר של הצברה, ניתן לחשוב על צלעות הגרף כמתארות זוגות של

קדקודים רחוקים. מספר הצביעה הוא המספר המינימלי של צברים להם ניתן לחלק את הגרף, כך שכל הנקודות בתוך צבר קרובות זו לזו.

בפרק 7 אנו מכלילים את חסם הופמן לפרמטרים נוספים, הקשורים במספר הצביעה. ההרחבה הראשונה היא למספר הצביעה הווקטורי, המופיע בעבודתם של מוטוואני, סודאן וקרגר, ומהווה חסם תחתון למספר הצביעה. אנו מראים כי חסם הופמן תקף גם לגביו. ההרחבה השנייה היא ל-"מספרי כיסוי" של גרף, המופיעים בעבודתם של ליניאל, מטושק ועמית, וחוסמים אף הם את מספר הצביעה מלמטה. אנו מראים כי וריאציה של חסם הופמן תופסת גם לגביהם.

ההרחבה השלישית היא לפרמטר אותו אנו מציעים בעבודה זו, המכונה "מספר ההצברה" של גרף. מספר ה-λ-הצברה של גרף הוא המספר המינימלי של צברים להם ניתן לחלק את הגרף, כך שהרדיוס הספקטרלי של כל צבר הוא לכל היותר λ. (בפרט, מספר ה-0-הצברה הוא בדיוק מספר הצביעה). אנו מראים כי מספר זה חסם מלרע ע"י $\frac{\lambda_1 - \lambda_n}{\lambda - \lambda_n}$.

תוכן העניינים