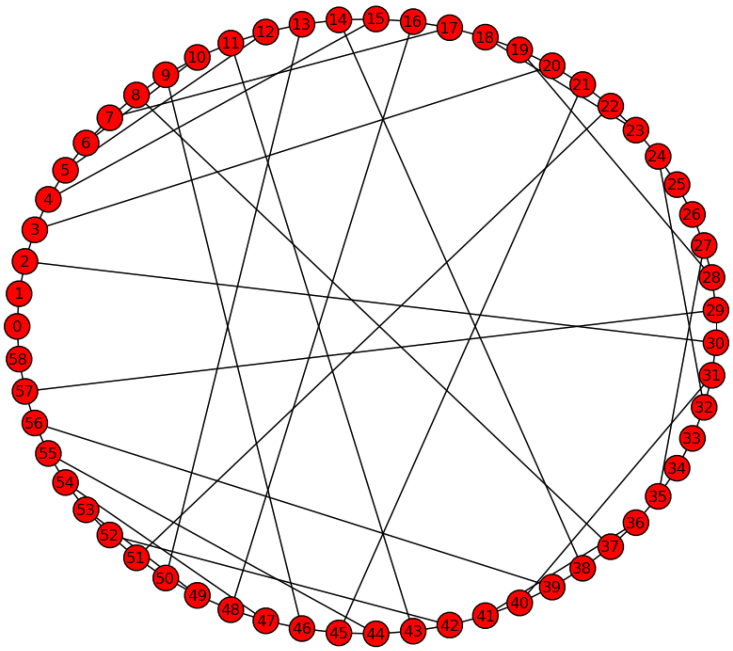
# Lecture Notes on Expansion, Sparsest Cut, and Spectral Graph Theory



Luca Trevisan

University of California, Berkeley

# Foreword

These notes are a lightly edited revision of notes written for the course "Graph Partitioning and Expanders" offered at Stanford in Winter 2011 and Winter 2013.

I wish to thank the students who attended this course for their enthusiasm and hard work.

This material is based upon work supported by the National Science Foundation under Grant No. 1216642 and by the US-Israel BSF Grant No. 2010451. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. As far as I know, the opinions of National Science Foundation concerning expander graphs are completely different from mine.

San Francisco, August 23, 2014.

Luca Trevisan

# Contents

Chapter 1

# The Basics

## 1.1 Expander Graphs and Sparse Cuts

Before giving the definition of expander graph, it is helpful to consider examples of graphs that are not expanders, in order to gain intuition about the type of "bad examples" that the definition is designed to avoid.

Suppose that a communication network is shaped as a path, with the vertices representing the communicating devices and the edges representing the available links. The clearly undesirable feature of such a configuration is that the failure of a single edge can cause the network to be disconnected, and, in particular, the failure of the middle edge will disconnect half of the vertices from the other half.

This is a situation that can occur in reality. Most of Italian highway traffic is along the highway that connect Milan to Naples via Bologna, Florence and Rome. The section between Bologna and Florence goes through relatively high mountain passes, and snow and ice can cause road closures. When this happens, it is almost impossible to drive between Northern and Southern Italy. Closer to California, I was once driving from Banff, a mountain resort town in Alberta which hosts a mathematical institute, back to the US. Suddenly, traffic on Canada's highway 1 came to a stop. People from the other cars, after a while, got out of their cars and started hanging out and chatting on the side of the road. We asked if there was any other way to go in case whatever accident was ahead of us would cause a long road closure. They said no, this is the only highway here. Thankfully we started moving again in half an hour or so.

Now, consider a two-dimensional $\sqrt{n} \times \sqrt{n}$ grid. The removal of an edge cannot disconnect the graph, and the removal of a constant number of edges can only disconnected a constant number of vertices from the rest of the graph, but it is possible to remove just $\sqrt{n}$ edges, a $1/O(\sqrt{n})$ fraction of the total, and have half of the vertices be disconnected from the other half.

A $k$-dimensional hypercube with $n = 2^k$ is considerably better connected than a grid,

although it is still possible to remove a vanishingly small fraction of edges (the edges of a "dimension cut," which are a $1/k = 1/\log_2 n$ fraction of the total number of edges) and disconnect half of the vertices from the other half.

Clearly, the most reliable network layout is the clique; in a clique, if an adversary wants to disconnect a $p$ fraction of vertices from the rest of the graph, he has to remove at least a $p \cdot (1-p)$ fraction of edges from the graph.

This property of the clique will be our "gold standard" for reliability. The expansion and the sparsest cut parameters of a graph measure how worse a graph is compared with a clique from this point.

**Definition 1.1 (Sparsest Cut)** *Let $G = (V, E)$ be a graph and let $(S, V-S)$ be a partition of the vertices (a cut). Then the (normalized) sparsity of the cut is*

$$\sigma(S) := \frac{\mathop{\mathbb{E}}_{(u,v)\sim E} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\mathop{\mathbb{E}}_{(u,v)\sim V^2} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}$$

*the fraction of edges cut by the partition $(S, V-S)$ divided by the fraction of pairs of edges separated by the partition $(S, V-S)$.*

*The sparsest cut problem is, given a graph, to find the set of minimal sparsity. The sparsity of a graph $G = (V, E)$ is*

$$\sigma(G) := \min_{S\subseteq V : S\neq\emptyset, S\neq V} \sigma(S)$$

Note that if $G$ is a $d$-regular graph, then

$$\sigma(S) := \frac{E(S, V-S)}{\frac{d}{|V|} \cdot |S| \cdot |V-S|}$$

In a $d$-regular graph, the *edge expansion* of a set of vertices $S \subseteq V$ is the related quantity

$$\phi(S) := \frac{E(S, V-S)}{d \cdot |S|}$$

in which we look at the ratio between the number of edges between $S$ and $V - S$ and the obvious upper bound given by the total number of edges incident $S$.

The edge expansion $\phi(G)$ of a graph is

$$\phi(G) := \min_{S : |S| \leq \frac{|V|}{2}} \phi(S)$$

the minimum of $\phi(S)$ over all partitions $(S, V-S)$, where $|S| \leq |V-S|$.

(It is common to alternatively define the edge expansion without the normalizing factor of $d$ in the denominator.)

We note that for every regular graph $G$ we have that, for every set $S$ such that $|S| \leq |V|/2$,

$$\frac{1}{2}\sigma(S) \leq \phi(S) \leq \sigma(S)$$

and we have $\sigma(S) = \sigma(V - S)$, hence

$$\frac{1}{2}\sigma(G) \leq \phi(G) \leq \cdot\sigma(G)$$

A family of constant degree expanders is a family of (multi-)graphs $\{G_n\}_{n \geq d}$ such that each graph $G_n$ is a $d$-regular graph with $n$ vertices and such that there is an absolute constant $\phi > 0$ such that $\phi(G_n) \geq \phi$ for every $n$.

Constant-degree graphs of constant expansion are sparse graphs with exceptionally good connectivity properties. For example, we have the following observation.

**Lemma 1.2** *Let $G = (V, E)$ be a regular graph of expansion $\phi$. Then, after an $\epsilon < \phi$ fraction of the edges are adversarially removed, the graph has a connected component that spans at least $1 - \epsilon/2\phi$ fraction of the vertices.*

PROOF: Let $d$ be the degree of $G$, and let $E' \subseteq E$ be an arbitrary subset of $\leq \epsilon|E| = \epsilon \cdot d \cdot |V|/2$ edges. Let $C_1, \ldots, C_m$ be the connected components of the graph $(V, E - E')$, ordered so that $|C_1| \geq |C_2| \geq \cdots \geq |C_m|$. We want to prove that $|C_1| \geq |V| \cdot (1 - 2\epsilon/\phi)$. We have

$$|E'| \geq \frac{1}{2}\sum_{i \neq j} E(C_i, C_j) = \frac{1}{2}\sum_i E(C_i, V - C_i)$$

If $|C_1| \leq |V|/2$, then we have

$$|E'| \geq \frac{1}{2}\sum_i d \cdot \phi \cdot |C_i| = \frac{1}{2} \cdot d \cdot \phi \cdot |V|$$

but this is impossible if $\phi > \epsilon$.

If $|C_1| \geq |V|/2$, then define $S := C_2 \cup \cdots \cup C_m$. We have

$$|E'| \geq E(C_1, S) \geq d \cdot \phi \cdot |S|$$

which implies that $|S| \leq \frac{\epsilon}{2\phi} \cdot |V|$ and so $C_1 \geq \left(1 - \frac{\epsilon}{2\phi}\right) \cdot |V|$. □

In words, this means that, in a $d$-regular expander, the removal of $k$ edges can cause at most $O(k/d)$ vertices to be disconnected from the remaining "giant component." Clearly, it is always possible to disconnect $k/d$ vertices after removing $k$ edges, so the reliability of an expander is essentially best possible.

## 1.2 Eigenvalues and Eigenvectors

*Spectral graph theory* studies how the eigenvalues of the adjacency matrix of a graph, which are purely algebraic quantities, relate to combinatorial properties of the graph.

We begin with a brief review of linear algebra.

If $x = a + ib$ is a complex number, then we let $\bar{x} = a - ib$ denote its *conjugate*. If $M \in \mathbb{C}^{m \times n}$ is a matrix, then $M^*$ denotes the conjugate transpose of $M$, that is, $(M^*)_{i,j} = \overline{M_{j,i}}$. If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ are two vectors, then their inner product is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^* \mathbf{y} = \sum_i \overline{x_i} \cdot y_i \tag{1.1}$$

Notice that, by definition, we have $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$ and $\langle \mathbf{x}, \mathbf{x} \rangle = ||\mathbf{x}||^2$.

If $M \in \mathbb{C}^{n \times n}$ is a square matrix, $\lambda \in \mathbb{C}$ is a scalar, $\mathbf{x} \in \mathbb{C}^n - \{\mathbf{0}\}$ is a non-zero vector and we have

$$M\mathbf{x} = \lambda \mathbf{x} \tag{1.2}$$

then we say that $\lambda$ is an *eigenvalue* of $M$ and that $\mathbf{x}$ is *eigenvector* of $M$ corresponding to the eigenvalue $\lambda$.

When (1.2) is satisfied, then we equivalently have

$$(M - \lambda I) \cdot \mathbf{x} = \mathbf{0}$$

for a non-zero vector $\mathbf{x}$, which is equivalent to

$$\det(M - \lambda I) = 0 \tag{1.3}$$

For a fixed matrix $M$, the function $\lambda \to \det(M - \lambda I)$ is a univariate polynomial of degree $n$ in $\lambda$ and so, over the complex numbers, the equation (1.3) has exactly $n$ solutions, counting multiplicities.

If $G = (V, E)$ is a graph, then we will be interested in the adjacency matrix $A$ of $G$, that is the matrix such that $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise. If $G$ is a multigraph or a weighted graph, then $A_{ij}$ is equal to the number of edges between $(i, j)$, or the weight of the edge $(i, j)$, respectively.

The adjacency matrix of an undirected graph is symmetric, and this implies that its eigenvalues are all real.

**Definition 1.3** *A matrix $M \in \mathbb{C}^{n \times n}$ is Hermitian if $M = M^*$.*

Note that a real symmetric matrix is always Hermitian.

**Lemma 1.4** *If $M$ is Hermitian, then all the eigenvalues of $M$ are real.*

PROOF: Let $M$ be an Hermitian matrix and let $\lambda$ be a scalar and $\mathbf{x}$ be a non-zero vector such that $M\mathbf{x} = \lambda\mathbf{x}$. We will show that $\lambda = \lambda^*$, which implies that $\lambda$ is a real number.

We first see that

$$\begin{aligned}
&\langle M\mathbf{x}, \mathbf{x}\rangle \\
&= (M\mathbf{x})^*\mathbf{x} \\
&= \mathbf{x}^* M^* \mathbf{x} \\
&= \mathbf{x}^* M \mathbf{x} \\
&= \langle \mathbf{x}, M\mathbf{x}\rangle
\end{aligned}$$

where we use the fact that $M$ is Hermitian. Then we note that

$$\langle M\mathbf{x}, \mathbf{x}\rangle = \langle \lambda\mathbf{x}, \mathbf{x}\rangle = \bar{\lambda}||x||^2$$

and

$$\langle \mathbf{x}, M\mathbf{x}\rangle = \langle \mathbf{x}, \lambda\mathbf{x}\rangle = \lambda||x||^2$$

so that $\lambda = \bar{\lambda}$. □

We also note the following useful fact.

**Fact 1.5** *If $M$ is an Hermitian matrix, and $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors of different eigenvalues, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.*

PROOF: Let $\mathbf{x}$ be an eigenvector $\lambda$ and $\mathbf{y}$ be an eigenvector of $\lambda'$, then, from the fact that $M$ is Hermitian, we get

$$\langle M\mathbf{x}, \mathbf{y}\rangle = (M\mathbf{x})^*\mathbf{y} = \mathbf{x}^* M^* \mathbf{y} = \mathbf{x}^* M \mathbf{y} = \langle \mathbf{x}, M\mathbf{y}\rangle$$

but

$$\langle M\mathbf{x}, \mathbf{y}\rangle = \lambda \cdot \langle \mathbf{x}, \mathbf{y}\rangle$$

and

$$\langle \mathbf{x}, M\mathbf{y}\rangle = \lambda' \cdot \langle \mathbf{x}, \mathbf{y}\rangle$$

so that

$$(\lambda - \lambda') \cdot \langle \mathbf{x}, \mathbf{y}\rangle = 0$$

which implies that $\langle \mathbf{x}, \mathbf{y}\rangle = 0$, that is, that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. □

We will be interested in relating combinatorial properties of a graph $G$, such as connectivity and bipartiteness, with the values of the eigenvalues of the adjacency matrix of $G$.

A step in this direction is to see the problem of computing the eigenvalues of a real symmetric matrix as the solution to an *optimization* problem.

**Theorem 1.6 (Variational Characterization of Eigenvalues)** *Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and $\lambda_1 \leq \ldots \leq \lambda_n$ be its real eigenvalues, counted with multiplicities and sorted in nondecreasing order. Let $\mathbf{x}_1, \cdots, \mathbf{x}_k$, $k < n$, be orthonormal vectors such that $M\mathbf{x}_i = \lambda_i \mathbf{x}_i$ for $i = 1, \ldots, k$. Then*

$$\lambda_{k+1} = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}: \mathbf{x}\perp\mathbf{x}_1,\ldots,\mathbf{x}\perp\mathbf{x}_k} \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

*and any minimizer is an eigenvector of $\lambda_{k+1}$.*

In particular, Theorem 1.6 implies that

$$\lambda_1 = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}} \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

and, if we call $\mathbf{x}_1$ a minimizer of the above expression, then

$$\lambda_2 = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}: \mathbf{x}\perp\mathbf{x}_1} \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

and a minimizer $\mathbf{x}_2$ of the above expression is an eigenvector of $\mathbf{x}_1$, and so on.

In order to prove Theorem 1.6, we first prove the following result.

**Lemma 1.7** *Let Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and let $\mathbf{x}_1, \ldots, \mathbf{x}_k$, $k < n$ be orthogonal eigenvectors of $M$. Then there is an eigenvector $\mathbf{x}_{k+1}$ of $M$ that is orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$.*

Proof: Let $V$ be the $(n-k)$-dimensional subspace of $\mathbb{R}^n$ that contains all the vectors orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$. We claim that for every vector $\mathbf{x} \in V$ we also have $M\mathbf{x} \in V$. Indeed, for every $i$, the inner product of $M\mathbf{x}$ and $\mathbf{x}_i$ is

$$\langle \mathbf{x}_i, M\mathbf{x}\rangle = \mathbf{x}_i^T M\mathbf{x} = (M^T\mathbf{x}_i)^T\mathbf{x} = (M\mathbf{x}_i)^T\mathbf{x} = \lambda_i\mathbf{x}_i^T\mathbf{x} = \lambda \cdot \langle \mathbf{x}_1, \mathbf{x}\rangle = 0$$

Let $B \in \mathbb{R}^{n\times(n-k)}$ be the matrix that computes a bijective map from $\mathbb{R}^{n-k}$ to $V$. (If $\mathbf{b}_1, \ldots, \mathbf{b}_{n-k}$ are an orthonormal for basis for $V$, then $B$ is just the matrix whose columns are the vectors $\mathbf{b}_i$.) Let also $B' \in \mathbb{R}^{(n-1)\times n}$ be the matrix such that, for every $\mathbf{y} \in V$, $B'\mathbf{y}$ is the $(n-k)$-dimensional vector such that $BB'\mathbf{y} = \mathbf{y}$. (Let $B' = B^T$ where $B$ is as described above.) Let $\lambda$ be a real eigenvalue of the real symmetric matrix

$$M' := B'MB \in \mathbb{R}^{(n-k)\times(n-k)}$$

and $\mathbf{y}$ be a real eigenvector of $M'$.

Then we have the equation

$$B'MB\mathbf{x} = \lambda\mathbf{y}$$

and so

$$BB'MB\mathbf{y} = \lambda B\mathbf{y}$$

Since $B\mathbf{y}$ is orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$, it follows that $MB\mathbf{y}$ is also orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$, and so

$$BB'MB\mathbf{y} = MB\mathbf{y} \ ,$$

which means that we have

$$MB\mathbf{y} = \lambda B\mathbf{y}$$

and, defining $\mathbf{x}_{k+1} := B\mathbf{y}$, we have

$$M\mathbf{x}_{k+1} = \lambda B\mathbf{x}_{k+1}$$

□

We note that Lemma 1.7 has the following important consequence.

**Corollary 1.8 (Spectral Theorem)** *Let $M \in \mathbb{R}^{n\times n}$ be an real symmetric matrix, and $\lambda_1, \ldots, \lambda_n$ be its real eigenvalues, with multiplicities; then there are orthonormal vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^n$ such that $\mathbf{x}_i$ is an eigenvector of $\lambda_i$.*

We are now ready to prove Theorem 1.6.

PROOF:[Of Theorem 1.6] By repeated applications of Lemma 1.7, we find $n-k$ orthogonal eigenvectors which are also orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The eigenvalues of this system of $n$ orthogonal eigenvectors must include all the eigenvalues of $M$, because if there was any other eigenvalue, its eigenvector would be orthogonal to our $n$ vectors, which is impossible. Let us call the additional $n-k$ vectors $\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$, where $\mathbf{x}_i$ is an eigenvector of $\lambda_i$. Now consider the minimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}}:\mathbf{x}\perp\mathbf{x}_1,\ldots,\mathbf{x}\perp\mathbf{x}_k} \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

The solution $\mathbf{x} := \mathbf{x}_{k+1}$ is feasible, and it has cost $\lambda_{k+1}$, so the minimum is at most $\lambda_{k+1}$.

Consider now an arbitrary feasible solution $\mathbf{x}$. We can write

$$\mathbf{x} = a_{k+1}\mathbf{x}_{k+1} + \cdots + a_n\mathbf{x}_n$$

and we see that the cost of such a solution is

$$\frac{\sum_{i=k+1}^n \lambda_i a_i^2}{\sum_{i=k+1}^n a_i^2} \geq \lambda_{k+1} \cdot \frac{\sum_{i=k+1}^n a_i^2}{\sum_{i=k+1}^n a_i^2}$$

and so the minimum is also at least $\lambda_{k+1}$. Notice also that if $\mathbf{x}$ is a minimizer, that is, if the cost of $\mathbf{x}$ is $\lambda_{k+1}$, then we must $a_i = 0$ for every $i$ such that $\lambda_i > \lambda_{k+1}$, which means that $\mathbf{x}$ is a linear combination of eigenvectors of $\lambda_{k+1}$, and so it is itself an eigenvector of $\lambda_{k+1}$. $\square$

Sometimes it will be helpful to use the following variant of the variational characterization of eigenvalues.

**Corollary 1.9** *Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric matrix, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ its eigenvalues, counted with multiplicities and sorted in nondecreasing order. Then*

$$\lambda_k = \min_{V \text{ } k-\text{dimensional subspace of } \mathbb{R}^n} \max_{\mathbf{x} \in V - \{\mathbf{0}\}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

## 1.3 The Basics of Spectral Graph Theory

From the discussion so far, we have that if $A$ is the adjacency matrix of an undirected graph then it has $n$ real eigenvalues, counting multiplicities of the number of solutions to $\det(A - \lambda I) = 0$.

If $G$ is a $d$-regular graph, then instead of working with the adjacency matrix of $G$ it is somewhat more convenient to work with the *normalized Laplacian matrix* of $G$, which is defined as $L := I - \frac{1}{d} \cdot A$.

In the rest of this section we shall prove the following relations between the eigenvalues of $L$ and certain purely combinatorial properties of $G$.

**Theorem 1.10** *Let $G$ be a $d$-regular undirected graph, and $L = I - \frac{1}{d} \cdot A$ be its normalized Laplacian matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the real eigenvalues of $L$ with multiplicities. Then*

1. *$\lambda_1 = 0$ and $\lambda_n \leq 2$.*
2. *$\lambda_k = 0$ if and only if $G$ has at least $k$ connected components.*
3. *$\lambda_n = 2$ if and only if at least one of the connected components of $G$ is bipartite.*

Note that the first two properties imply that the multiplicity of 0 as an eigenvalue is precisely the number of connected components of $G$.

## 1.4 Proof of Theorem 1.10

We will make repeated use of the following identity, whose proof is immediate: if $L$ is the normalized Laplacian matrix of a $d$-regular graph $G$, and $\mathbf{x}$ is any vector, then

$$\mathbf{x}^T L \mathbf{x} = \frac{1}{d} \cdot \sum_{\{u,v\} \in E} (x_u - x_v)^2 \tag{1.4}$$

and so

$$\lambda_1 = \min_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}\}:} \frac{\mathbf{x}^T L\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \geq 0$$

If we take $\mathbf{1} = (1, \ldots, 1)$ to be the all-one vector, we see that $\mathbf{1}^T L\mathbf{1} = 0$, and so 0 is the smallest eigenvalue of $L$, with $\mathbf{1}$ being one of the vectors in the eigenspace of 1.

We also have the following formula for $\lambda_k$:

$$\lambda_k = \min_{S \text{ } k-\text{dimensional subspace of } \mathbb{R}^n} \max_{\mathbf{x}\in S-\{\mathbf{0}\}} \frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{d\sum_v x_v^2}$$

So, if $\lambda_k = 0$, there must exist a $k$-dimensional space $S$ such that for every $\mathbf{x} \in S$ we have

$$\sum_{\{u,v\}\in E} (x_u - x_v)^2 = 0 \ ,$$

but this means that, for every $\mathbf{x}$, and for every edge $(u, v) \in E$ of positive weight, we have $x_u = x_v$, and so $x_u = x_v$ for every $u, v$ which are in the same connected component. This means that each $\mathbf{x} \in V$ must be constant within each connected component of $G$, and so the dimension of $V$ can be at most the number of connected components of $G$, meaning that $G$ has at least $k$ connected components.

Conversely, if $G$ has at least $k$ connected components, we can let $S$ be the space of vectors that are constant within each component, and $S$ is a space of dimension at least $k$ such that for every element $\mathbf{x}$ of $S$ we have

$$\sum_{\{u,v\}\in E} (x_u - x_v)^2 = 0$$

meaning that $S$ is a witness of the fact that $\lambda_k = 0$.

Finally, to study $\lambda_n$, we first note that we have the formula

$$\lambda_n = \max_{\mathbf{x}\in\mathbb{R}^n-\{\mathbf{0}\}} \frac{\mathbf{x}^T L\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

which we can prove by using the variational characterization of the eigenvalues of $-L$ and noting that $-\lambda_n$ is the smallest eigenvalue of $-L$.

We also observe that for every vector $\mathbf{x} \in \mathbb{R}^n$ we have

$$2 - \mathbf{x}^T L\mathbf{x} = \frac{1}{d} \sum_{\{u,v\}\in E} (x_u + x_v)^2$$

and so

$$\lambda_n \leq 2$$

and if $\lambda_n = 2$ then there must be a non-zero vector $\mathbf{x}$ such that

$$\sum_{\{u,v\}\in E} (x_u + x_v)^2 = 0$$

which means that $x_u = -x_v$ for every edge $(u, v) \in E$.

Let $v$ be a vertex such that $x_v = a \neq 0$, and define the sets $A := \{v : x_v = a\}$, $B := \{j : x_v = -a\}$ and $R = \{v : x_v \neq \pm a\}$. The set $A \cup B$ is disconnected from the rest of the graph, because otherwise an edge with an endpoint in $A \cup B$ and an endpoint in $R$ would give a positive contribution to $\sum_{u,v} A_{u,v}(x_u + x_v)^2$; furthermore, every edge incident on a vertex on $A$ must have the other endpoint in $B$, and vice versa. Thus, $A \cup B$ is a connected component, or a collection of connected components, of $G$ which is bipartite, with the bipartition $A, B$.

Chapter 2

# Eigenvalues and Expansion

## 2.1 Expansion and The Second Eigenvalue

Let $G = (V, E)$ be an undirected $d$-regular graph, $A$ its adjacency matrix, $L = 1 - \frac{1}{d} \cdot A$ its Laplacian matrix, and $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $L$.

We proved that $\lambda_2 = 0$ if and only if $G$ is disconnected, that is, $\lambda_2 = 0$ if and only if $\phi(G) = 0$. In this lecture we will see that this statement admits an *approximate version* that, qualitatively, says that $\lambda_2$ is small if and only if $\phi(G)$ is small. Quantitatively, we have

**Theorem 2.1 (Cheeger's Inequalities)**

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2 \cdot \lambda_2} \tag{2.1}$$

## 2.2 The Easy Direction

In this section we prove

**Lemma 2.2** $\lambda_2 \leq \sigma(G) \leq 2\phi(G)$

From which we have one direction of Cheeger's inequality.

Let us find an equivalent restatement of the sparsest cut problem. We can write

$$\sigma(G) = \min_{\mathbf{x} \in \{0,1\}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|} \tag{2.2}$$

Note that, when $x_u, x_v$ take boolean values, then so does $|x_u - x_v|$, so that we may also equivalently write

$$\sigma(G) = \min_{\mathbf{x} \in \{0,1\}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|^2} \tag{2.3}$$

In a previous section, we gave the following characterization of $\lambda_2$:

$$\lambda_2 = \min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{d \cdot \sum_v x_v^2}$$

Now we claim that the following characterization is also true

$$\lambda_2 = \min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|^2} \tag{2.4}$$

This is because

$$\begin{aligned} &\sum_{u,v} |x_u - x_v|^2 \\ &= \sum_{u,v} x_i^2 + \sum_{u,v} x_v^2 - 2 \sum_{u,v} x_u x_v \\ &= 2n \sum_v x_v^2 - 2 \left( \sum_v x_v \right)^2 \end{aligned}$$

so for every $\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}$ such that $\mathbf{x} \perp \mathbf{1}$ we have that

$$\sum_v x_v^2 = \frac{1}{2n} \sum_{u,v} |x_u - x_v|^2 = \frac{1}{n} \sum_{\{u,v\}} |x_u - x_v|^2 \ ,$$

and so

$$\min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{d \cdot \sum_v x_v^2} = \min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{\{u,v\}} |x_u - x_v|^2}$$

To conclude the argument, take an $\mathbf{x}$ that maximizes the right-hand side of (2.4), and observe that if we shift every coordinate by the same constant then we obtain another optimal solution, because the shift will cancel in all the expressions both in the numerator and the denominator. In particular, we can define $\mathbf{x}'$ such that $x'_v = x_v - \frac{1}{n} \sum_u x_u$ and note that the entries of $\mathbf{x}'$ sum to zero, and so $\mathbf{x}' \perp \mathbf{1}$. This proves that

$$\min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{u,v} |x_u - x_v|^2} = \min_{\mathbf{x} \in \mathbb{R}^V - \{\mathbf{0},\mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\frac{d}{n} \sum_{u,v} |x_u - x_v|^2}$$

and so we have established (2.4).

Comparing (2.4) and (2.3), it is clear that the quantity $\lambda_2$ is a *continuous relaxation* of $\sigma(G)$, and hence $\lambda_2 \leq \sigma(G)$.

## 2.3 Other Relaxations of $\sigma(G)$

Having established that we can view $\lambda_2$ as a *relaxation* of $\sigma(G)$, the proof that $\phi(G) \leq \sqrt{2 \cdot \lambda_2}$ can be seen as a *rounding* algorithm, that given a real-valued solution to (2.4) finds a comparably good solution for (2.3).

Later in the course we will see two more approximation algorithms for sparsest cut and edge expansion. Both are based on continuous relaxations of $\sigma$ starting from (2.2).

The algorithm of Leighton and Rao is based on a relaxation that is defined by observing that every bit-vector $\mathbf{x} \in \{0,1\}^V$ defines the semi-metric $dist(u,v) := |x_u - x_v|$ over the vertices; the Leighton-Rao relaxation is obtained by allowing arbitrary semi-metrics:

$$LR(G) := \min_{\substack{dist : V \times V \to \mathbb{R} \\ dist \text{ semimetric}}} \frac{\sum_{\{u,v\} \in E} dist(u,v)}{\frac{d}{n} \sum_{\{u,v\}} dist(u,v)}$$

It is not difficult to express $LR(G)$ as a linear programming problem.

The algorithm of Arora-Rao-Vazirani is obtained by noting that, for a bit-vector $x \in \{0,1\}^V$, the distances $dist(u,v) := |x_u - x_v|$ define a metric which can also be seen as the Euclidean distance between the $x_v$, because $|x_u - x_v| = \sqrt{(x_u - x_v)^2}$, and such that $dist^2(u,v)$ is also a semi-metric, trivially so because $dist^2(u,v) = dist(u,v)$. If a distance function $dist(\cdot,\cdot)$ is a semi-metric such that $\sqrt{dist(\cdot,\cdot)}$ is a Euclidean semi-metric, then $dist(\cdot,\cdot)$ is called a *negative type* semi-metric. The Arora-Rao-Vazirani relaxation is

$$ARV(G) := \min_{\substack{dist : V \times V \to \mathbb{R} \\ dist \text{ negative type semimetric}}} \frac{\sum_{\{u,v\} \in E} dist(u,v)}{\frac{d}{n} \sum_{\{u,v\}} dist(u,v)}$$

The Arora-Rao-Vazirani relaxation can be expressed as a semi-definite programming problem.

From this discussion it is clear that the Arora-Rao-Vazirani relaxation is a tightening of the Leigthon-Rao relaxation and that we have

$$\sigma(G) \geq ARV(G) \geq LR(G)$$

It is less obvious in this treatment, and we will see it later, that the Arora-Rao-Vazirani is also a tightening of the relaxation of $\sigma$ given by $\lambda_2$, that is

$$\sigma(G) \geq ARV(G) \geq \lambda_2$$

The relaxations $\lambda_2$ and $LR(G)$ are incomparable.

## 2.4 Spectral partitioning and the proof of the difficult direction

The proof of the more difficult direction of Theorem 2.1 will be constructive and algorithmic. The proof can be seen as an analysis of the following algorithm.

Algorithm: *SpectralPartitioning*

- Input: graph $G = (V, E)$ and vector $\mathbf{x} \in \mathbb{R}^V$
- Sort the vertices of $V$ in non-decreasing order of values of entries in $\mathbf{x}$, that is let $V = \{v_1, \ldots, v_n\}$ where $x_{v_1} \leq x_{v_2} \leq \ldots x_{v_n}$
- Let $i \in \{1, \ldots, n-1\}$ be such that $\max\{\phi(\{v_1, \ldots, v_i\}), \phi(\{v_{i+1}, \ldots, v_n\})\}$ is minimal
- Output $S = \{v_1, \ldots, v_i\}$ and $\bar{S} = \{v_{i+1}, \ldots, v_n\}$

We note that the algorithm can be implemented to run in time $O(|V| \log |V| + |E|)$, assuming arithmetic operations and comparisons take constant time, because once we have computed $E(\{v_1, \ldots, v_i\}, \{v_{i+1}, \ldots, v_n\})$ it only takes time $O(degree(v_{i+1}))$ to compute $E(\{v_1, \ldots, v_{i+1}\}, \{v_{i+2}, \ldots, v_n\})$.

We have the following analysis of the quality of the solution:

**Lemma 2.3 (Analysis of Spectral Partitioning)** *Let $G = (V, E)$ be a $d$-regular graph, $\mathbf{x} \in \mathbb{R}^V$ be a vector such that $\mathbf{x} \perp \mathbf{1}$, define*

$$R(\mathbf{x}) := \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{d \cdot \sum_v x_v^2}$$

*and let $S$ be the output of algorithm SpectralPartitioning on input $G$ and $\mathbf{x}$. Then*

$$\phi(S) \leq \sqrt{2R(x)}$$

**Remark 2.4** *If we apply the lemma to the case in which $\mathbf{x}$ is an eigenvector of $\lambda_2$, then $R(\mathbf{x}) = \lambda_2$, and so we have*

$$\phi(S) \leq \sqrt{2 \cdot \lambda_2}$$

*which is the difficult direction of Cheeger's inequalities.*

**Remark 2.5** *If we run the SpectralPartitioning algorithm with the eigenvector* $\mathbf{x}$ *of the second eigenvalue* $\lambda_2$*, we find a set* $S$ *whose expansion is*

$$\phi(S) \leq \sqrt{2 \cdot \lambda_2} \leq 2\sqrt{\phi(G)}$$

*Even though this doesn't give a constant-factor approximation to the edge expansion, it gives a very efficient, and non-trivial, approximation.*

*As we will see in a later lecture, there is a nearly linear time algorithm that finds a vector* $\mathbf{x}$ *for which the Rayleigh quotient* $R(\mathbf{x})$ *is very close to* $\lambda_2$*, so, overall, for any graph* $G$ *we can find a cut of expansion* $O(\sqrt{\phi(G)})$ *in nearly linear time.*

## 2.5 Proof of Lemma 2.3

We saw that $\lambda_2$ can be seen as a relaxation of $\sigma(G)$, and Lemma 2.3 provides a *rounding algorithm* for the real vectors which are solutions of the relaxation. In this section we will think of it as a form of *randomized rounding*. Later, when we talk about the Leighton-Rao sparsest cut algorithm, we will revisit this proof and think of it in terms of *metric embeddings*.

To simplify notation, we will assume that $V = \{1, \ldots, n\}$ and that $x_1 \leq x_2 \leq \cdots x_n$. Thus our goal is to prove that there is an $i$ such that $\phi(\{1, \ldots, i\}) \leq \sqrt{2R(\mathbf{x})}$ and $\phi(\{i+1, \ldots, n\}) \leq \sqrt{2R(\mathbf{x})}$

We will derive Lemma 2.3 by showing that there is a distribution $D$ over sets $S$ of the form $\{1, \ldots, i\}$ such that

$$\frac{\mathbb{E}_{S \sim D} E(S, V - S)}{\mathbb{E}_{S \sim D} d \cdot \min\{|S|, |V - S|\}} \leq \sqrt{2R(\mathbf{x})} \tag{2.5}$$

We need to be a bit careful in deriving the Lemma from (2.5). In general, it is not true that a ratio of averages is equal to the average of the ratios, so (2.5) does not imply that $\mathbb{E}\,\phi(S) \leq \sqrt{2R(\mathbf{x})}$. We can, however, apply linearity of expectation and derive from (2.5) the inequality

$$\mathop{\mathbb{E}}_{S \sim D} \frac{1}{d} E(S, V - S) - \sqrt{2R(\mathbf{x})} \min\{|S|, |V - S|\} \leq 0$$

So there must exist a set $S$ in the sample space such that

$$\frac{1}{d} E(S, V - S) - \sqrt{2R(\mathbf{x})} \min\{|S|, |V - S|\} \leq 0$$

meaning that, for that for both the set $S$ and its complement, we have $\phi(S) \leq \sqrt{2R(\mathbf{x})}$; at least one of the sets has size at most $n/2$, and so we are done. (Basically we are using the fact that, for random variables $X, Y$ over the same sample space, although it might not be

true that $\frac{\mathbb{E}X}{\mathbb{E}Y} = \mathbb{E}\frac{X}{Y}$, we always have $\mathbb{P}[\frac{X}{Y} \leq \frac{\mathbb{E}X}{\mathbb{E}Y}] > 0$, provided that $Y > 0$ over the entire sample space.)

From now on, we will assume that

1. $x_{\lceil n/2 \rceil} = 0$, that is, the median of the entries of $\mathbf{x}$ is zero
2. $x_1^2 + x_n^2 = 1$

which can be done without loss of generality because, if $\mathbf{x} \perp \mathbf{1}$, adding a fixed constant $c$ to all entries of $\mathbf{x}$ can only reduce the Rayleigh quotient:

$$\begin{aligned} R(\mathbf{x} + (c, \ldots, c)) &= \frac{\sum_{\{u,v \in E\}} |(x_u + c) - (x_v + x)|^2}{d \sum_v (x_v + c)^2} \\ &= \frac{\sum_{\{u,v \in E\}} |x_u - x_v|^2}{d \sum_v x_v^2 - 2dc \sum_v x_v + nc^2} \\ &= \frac{\sum_{\{u,v \in E\}} |x_u - x_v|^2}{d \sum_v x_v^2 + nc^2} \\ &\leq R(\mathbf{x}) \end{aligned}$$

Multiplying all the entries by a fixed constant does not change the value of $R(\mathbf{x})$, nor does it change the property that $x_1 \leq \cdots \leq x_n$. The reason for these choices is that they allow us to define a distribution $D$ over sets such that

$$\mathop{\mathbb{E}}_{S \sim D} \min\{|S|, |V - S|\} = \sum_i x_i^2 \tag{2.6}$$

We define the distribution $D$ over sets of the form $\{1, \ldots, i\}$, $i \leq n - 1$, as the outcome of the following probabilistic process:

- We pick a real value $t$ in the range $[x_1, x_n]$ with probabily density function $f(t) = 2|t|$. That is, for $x_1 \leq a \leq b \leq x_n$, $\mathbb{P}[a \leq t \leq b] = \int_a^b 2|t|dt$.

  Doing the calculation, this means that $\mathbb{P}[a \leq t \leq b] = |a^2 - b^2|$ if $a, b$ have the same sign, and $\mathbb{P}[a \leq t \leq b] = a^2 + b^2$ if they have different signs.
- We let $S := \{i : x_i \leq t\}$

According to this definition, the probability that an element $i \leq n/2$ belongs to the smallest of the sets $S, V - S$ is the same as the probability that it belongs to $S$, which is the probability that the threshold $t$ is in the range $[x_i, 0]$, and that probability is $x_i^2$. Similarly, the probability that an element $i > n/2$ belongs to the smallest of $S, V - S$ is the same as the probability that it belongs to $V - S$, which is the probability that $t$ is in the range $[0, x_i]$, which is again $x_i^2$. So we have established (2.6).

We will now estimate the expected number of edges between $S$ and $V - S$.

$$\mathbb{E}\,\frac{1}{d}E(S, V-S) = \frac{1}{2}\sum_{i,j} M_{i,j}\,\mathbb{P}[(i,j) \text{ is cut by } (S, V-S)]$$

The event that the edge $(i,j)$ is cut by the partition $(S, V-S)$ happens when the value $t$ falls in the range between $x_i$ and $x_j$. This means that

- If $x_i, x_j$ have the same sign,

$$\mathbb{P}[(i,j) \text{ is cut by } (S, V-S)] = |x_i^2 - x_j^2|$$

- If $x_i, x_j$ have different sign,

$$\mathbb{P}[(i,j) \text{ is cut by } (S, V-S)] = x_i^2 + x_j^2$$

Some attempts, show that a good expression to upper bound both cases is

$$\mathbb{P}[(i,j) \text{ is cut by } (S, V-S)] \leq |x_i - x_j| \cdot (|x_i| + |x_j|)$$

Plugging into our expression for the expected number of cut edges, and applying Cauchy-Schwarz

$$\begin{aligned}\mathbb{E}\,E(S, V-S) &\leq \sum_{\{i,j\}\in E} |x_i - x_j| \cdot (|x_i| + |x_j|) \\ &\leq \sqrt{\sum_{\{i,j\}\in E} (x_i - x_j)^2} \cdot \sqrt{\sum_{\{i,j\}\in E} (|x_i| + |x_j|)^2}\end{aligned}$$

Finally, it remains to study the expression $\sum_{\{i,j\}\in E}(|x_i|+|x_j|)^2$. By applying the inequality $(a+b)^2 \leq 2a^2 + 2b^2$ (which follows by noting that $2a^2 + 2b^2 - (a+b)^2 = (a-b)^2 \geq 0$), we derive

$$\sum_{\{i,j\}\in E} (|x_i| + |x_j|)^2 \leq \sum_{\{i,j\}\in E} (2x_i^2 + 2x_j^2) = 2d\sum_i x_i^2$$

Putting all the pieces together we have

$$\frac{\mathbb{E}\,E(S, V-S)}{d\,\mathbb{E}\min\{|S|, |V-S|\}} \leq \frac{\sqrt{\sum_{\{i,j\}\in E} |x_i - x_j|^2} \cdot \sqrt{2d\sum_i x_i^2}}{d\sum_i x_i^2} = \sqrt{2R(\mathbf{x})} \tag{2.7}$$

which, together with (2.6) gives (2.5), which, as we already discussed, implies the Main Lemma 2.3.

Chapter 3

# More on Eigenvalues and Cut Problems

---

## 3.1 Spectral graph theory in irregular graphs

Let $G = (V, E)$ be an undirected graph in which every vertex has positive degree and $A$ be the adjacency matrix of $G$. We want to define a Laplacian matrix $L$ and a Rayleigh quotient such that the $k$-th eigenvalue of $L$ is the minimum over all $k$-dimensional spaces of the maximum Rayleigh quotient in the space, and we want the conductance of a set to be the same as the Rayleigh quotient of the indicator vector of the set. All the facts that we have proved in the regular case essentially reduce to these two properties of the Laplacian and the Rayleigh quotient.

Let $d_v$ be the degree of vertex $v$ in $G$. We define the Rayleigh quotient of a vector $\mathbf{x} \in \mathbb{R}^V$ as

$$R_G(\mathbf{x}) := \frac{\sum_{\{u,v\} \in E} |x_u - x_v|^2}{\sum_v d_v x_v^2}$$

Let $D$ be the diagonal matrix of degrees such that $D_{u,v} = 0$ if $u \neq v$ and $D_{v,v} = d_v$. Then define the Laplacian of $G$ as

$$L_G := I - D^{-1/2} A D^{-1/2}$$

Note that in a $d$-regular graph we have $D = dI$ and $L_G = I - \frac{1}{d}A$, which is the standard definition.

Since $L = L_G$ is a real symmetric matrix, the $k$-th smallest eigenvalue of $L$ is

$$\lambda_k = \min_{k-\text{dimensional } S} \max_{\mathbf{x} \in S} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

Now let us do the change of variable $\mathbf{y} \leftarrow D^{-1/2}\mathbf{x}$. We have

$$\lambda_k = \min_{k-\text{dimensional } S'} \max_{\mathbf{y} \in S'} \frac{\mathbf{y}^T D^{1/2} L D^{1/2} \mathbf{y}}{\mathbf{y}^T D \mathbf{y}}$$

In the numerator, $\mathbf{y}^T D \mathbf{y} = \sum_v d_v y_v^2$, and in the denominator a simple calculation shows

$$\mathbf{y}^T D^{1/2} L D^{1/2} \mathbf{y} = \mathbf{y}^T (D - A) \mathbf{y} = \sum_{\{u,v\}} |y_v - y_u|^2$$

so indeed

$$\lambda_k = \min_{k-\text{dimensional } S} \max_{\mathbf{y} \in S} R_G(\mathbf{y})$$

For two vectors $\mathbf{y}, \mathbf{z}$, define the inner product

$$\langle \mathbf{y}, \mathbf{z} \rangle_G := \sum_v d_v \overline{y_v} z_v$$

Then we can prove that

$$\lambda_2 = \min_{\mathbf{y}: \langle \mathbf{y}, (1,\ldots,1) \rangle_G = 0} R_G(\mathbf{y})$$

With these definitions and observations in place, it is now possible to repeat the proof of Cheeger's inequality step by step (replacing the condition $\sum_v x_v = 0$ with $\sum_i d_v x_v = 0$, adjusting the definition of Rayleigh quotient, etc.) and prove that if $\lambda_2$ is the second smallest eigenvalue of the Laplacian of an irregular graph $G$, and $\phi(G)$ is the conductance of $G$, then

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

## 3.2 Higher-order Cheeger inequality

The Cheeger inequality gives a "robust" version of the fact that $\lambda_2 = 0$ if and only if $G$ is disconnected. It is possible to also give a robust version of the fact that $\lambda_k = 0$ if and only if $G$ has at least $k$ connected components. We will restrict the discussion to regular graphs.

For a size parameter $s \leq |V|/2$, denote the size-$s$ *small-set expansion* of a graph

$$SSE_s(G) := \min_{S \subseteq V:\ |S| \leq s} \phi(S)$$

So that $SSE_{\frac{n}{2}}(G) = \phi(G)$. This is an interesting optimization problem, because in many settings in which non-expanding sets correspond to clusters, it is more interesting to find small non-expanding sets (and, possibly, remove them and iterate to find more) than to find large ones. It has been studied very intensely in the past five years because of its connection

with the *Unique Games Conjecture*, which is in turn one of the key open problems in complexity theory.

If $\lambda_k = 0$, then we know that are at least $k$ connected components, and, in particular, there is a set $S \subseteq V$ such that $\phi(S) = 0$ and $|S| \leq \frac{n}{k}$, meaning that $SSE_{\frac{n}{k}} = 0$. By analogy with the Cheeger inequality, we may look for a robust version of this fact, of the form $SSE_{\frac{n}{k}} \leq O(\sqrt{\lambda_k})$. Unfortunately there are counterexamples, but Arora, Barak and Steurer have proved that, for every $\delta$,

$$SSE_{\frac{n^{1+\delta}}{k}} \leq O\left(\sqrt{\frac{\lambda_k}{\delta}}\right)$$

To formulate a "higher-order" version of the Cheeger inequality, we need to define a quantity that generalize expansion in a different way. For an integer parameter $k \geq 2$, define "order $k$ expansion" as

$$\phi_k(G) = \min_{S_1,\ldots S_k \subseteq V \text{ disjoint}} \max_{i=1,\ldots,k} \phi(S_i)$$

Note that $\phi_2(G) = \phi(G)$. Then Lee, Oveis-Gharan and Trevisan prove that

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^2) \cdot \sqrt{\lambda_k}$$

and

$$\phi_{.9 \cdot k}(G) \leq O(\sqrt{\lambda_k \cdot \log k})$$

(which was also proved by Louis, Raghavendra, Tetali and Vempala). The upper bounds are algorithmic, and given $k$ orthogonal vectors all of Rayleigh quotient at most $\lambda$, there are efficient algorithms that find at least $k$ disjoint sets each of expansion at most $O(k^2\sqrt{\lambda})$ and at least $.9 \cdot k$ disjoint sets each of expansion at most $O(\sqrt{\lambda \log k})$.

## 3.3 A Cheeger-type inequality for $\lambda_n$

We proved that $\lambda_n = 2$ if and only if $G$ has a bipartite connected component. What happens when $\lambda_n$ is, say, 1.999?

We can define a "bipartite" version of expansion as follows:

$$\beta(G) := \min_{\mathbf{x} \in \{-1,0,1\}^V} \frac{\sum_{\{u,v\} \in E} |x_u + x_v|}{\sum_v d_v |x_v|}$$

The above quantity has the following combinatorial interpretation: take a set $S$ of vertices, and a partition of $S$ into two disjoint sets $A, B$. Then define

$$\beta(S, A, B) := \frac{2E(A) + 2E(B) + 2E(S, V - S)}{vol(S)}$$

where $E(A)$ is the number of edges entirely contained in $A$, and $E(S, V - S)$ is the number of edges with one endpoint in $S$ and one endpoint in $V - S$. We can think of $\beta(S, A, B)$ as measuring what fraction of the edges incident on $S$ we need to delete in order to make $S$ disconnected from the rest of the graph and $A, B$ be a bipartition of the subgraph induced by $S$. In other words, it measure how close $S$ is to being a bipartite connected component. Then we see that

$$\beta(G) = \min_{S \subseteq V,\ A,B \text{ partition of } S} \beta(S, A, B)$$

Trevisan proves that

$$\frac{1}{2} \cdot (2 - \lambda_n) \leq \beta(G) \leq \sqrt{2 \cdot (2 - \lambda_n)}$$

Chapter 4

# The Power Method

---

In the previous section, we showed that, if $G = (V, E)$ is a $d$-regular graph, and $L$ is its normalized Laplacian matrix with eigenvalues $0 = \lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$, given an eigenvector of $\lambda_2$, the algorithm SpectralPartition finds, in nearly-linear time $O(|E| + |V| \log |V|)$, a cut $(S, V - S)$ such that $\phi(S) \leq 2\sqrt{\phi(G)}$.

More generally, if, instead of being given an eigenvector $\mathbf{x}$ such that $L\mathbf{x} = \lambda_2\mathbf{x}$, we are given a vector $\mathbf{x} \perp \mathbf{1}$ such that $\mathbf{x}^T L\mathbf{x} \leq (\lambda_2 + \epsilon)\mathbf{x}^T\mathbf{x}$, then the algorithm finds a cut such that $\phi(S) \leq \sqrt{4\phi(G) + 2\epsilon}$. In this lecture we describe and analyze an algorithm that computes such a vector using $O((|V| + |E|) \cdot \frac{1}{\epsilon} \cdot \log \frac{|V|}{\epsilon})$ arithmetic operations.

A symmetric matrix is *positive semi-definite* (abbreviated PSD) if all its eigenvalues are nonnegative. We begin by describing an algorithm that approximates the largest eigenvalue of a given symmetric PSD matrix. This might not seem to help very much because because we want to compute the second smallest, not the largest, eigenvalue. We will see, however, that the algorithm is easily modified to accomplish what we want.

## 4.1 The Power Method to Approximate the Largest Eigenvalue

The algorithm works as follows

Algorithm *Power*

Input: PSD matrix $M$, parameter $k$

- Pick uniformly at random $\mathbf{x}_0 \sim \{-1, 1\}^n$
- for $i := 1$ to $k$
  $\mathbf{x}_i := M \cdot \mathbf{x}_{i-1}$
- return $\mathbf{x}_k$

That is, the algorithm simply picks uniformly at random a vector $\mathbf{x}$ with $\pm 1$ coordinates, and outputs $M^k\mathbf{x}$.

Note that the algorithm performs $O(k \cdot (n+m))$ arithmetic operations, where $m$ is the number of non-zero entries of the matrix $M$.

**Theorem 4.1** *For every PSD matrix $M$, positive integer $k$ and parameter $\epsilon > 0$, with probability $\geq 3/16$ over the choice of $\mathbf{x}_0$, the algorithm Power outputs a vector $\mathbf{x}_k$ such that*

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_1 \cdot (1-\epsilon) \cdot \frac{1}{1+4n(1-\epsilon)^{2k}}$$

*where $\lambda_1$ is the largest eigenvalue of $M$.*

Note that, in particular, we can have $k = O(\log n/\epsilon)$ and $\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq (1 - O(\epsilon)) \cdot \lambda_1$.

Let $\lambda_1 \geq \cdots \lambda_n$ be the eigenvalues of $M$, with multiplicities, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a system of orthonormal eigenvectors such that $M\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Theorem 4.1 is implied by the following two lemmas

**Lemma 4.2** *Let $\mathbf{v} \in \mathbb{R}^n$ be a vector such that $||\mathbf{v}|| = 1$. Sample uniformly $\mathbf{x} \sim \{-1,1\}^n$. Then*

$$\mathbb{P}\left[|\langle \mathbf{x}, \mathbf{v}\rangle| \geq \frac{1}{2}\right] \geq \frac{3}{16}$$

**Lemma 4.3** *Let $\mathbf{x} \in \mathbb{R}^n$ be a vector such that $|\langle \mathbf{x}, \mathbf{v}_1\rangle| \geq \frac{1}{2}$. Then, for every positive integer $t$ and positive $\epsilon > 0$, if we define $\mathbf{y} := M^k\mathbf{x}$, we have*

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1-\epsilon) \cdot \frac{1}{1+4||\mathbf{x}||^2(1-\epsilon)^{2k}}$$

It remains to prove the two lemmas.

Proof: (Of Lemma 4.2) Let $\mathbf{v} = (v_1, \ldots, v_n)$. The inner product $\langle \mathbf{x}, \mathbf{v}\rangle$ is the random variable

$$S := \sum_i x_i v_i$$

Let us compute the first, second, and fourth moment of $S$.

$$\mathbb{E}\, S = 0$$

$$\mathbb{E}\, S^2 = \sum_i v_i^2 = 1$$

$$\mathbb{E}\, S^4 = 3\left(\sum_i v_i^2\right) - 2\sum_i v_i^4 \leq 3$$

Recall that the Paley-Zygmund inequality states that if $Z$ is a non-negative random variable with finite variance, then, for every $0 \leq \delta \leq 1$, we have

$$\mathbb{P}[Z \geq \delta\, \mathbb{E}\, Z] \geq (1-\delta)^2 \cdot \frac{(\mathbb{E}\, Z)^2}{\mathbb{E}\, Z^2} \tag{4.1}$$

which follows by noting that

$$\mathbb{E}\, Z = \mathbb{E}[Z \cdot 1_{Z<\delta\, \mathbb{E}\, Z}] + \mathbb{E}[Z \cdot 1_{Z\geq\delta\, \mathbb{E}\, Z}]\ ,$$

that

$$\mathbb{E}[Z \cdot 1_{Z<\delta\, \mathbb{E}\, Z}] \leq \delta\, \mathbb{E}\, Z\ ,$$

and that

$$\begin{aligned}\mathbb{E}[Z \cdot 1_{Z\geq\delta\, \mathbb{E}\, Z}] &\leq \sqrt{\mathbb{E}\, Z^2} \cdot \sqrt{\mathbb{E}\, 1_{Z\geq\delta\, \mathbb{E}\, Z}} \\ &= \sqrt{\mathbb{E}\, Z^2}\sqrt{\mathbb{P}[Z \geq \delta\, \mathbb{E}\, Z]}\end{aligned}$$

We apply the Paley-Zygmund inequality to the case $Z = S^2$ and $\delta = 1/4$, and we derive

$$\mathbb{P}\left[S^2 \geq \frac{1}{4}\right] \geq \left(\frac{3}{4}\right)^2 \cdot \frac{1}{3} = \frac{3}{16}$$

□

**Remark 4.4** *The proof of Lemma 4.2 works even if* $\mathbf{x} \sim \{-1,1\}^n$ *is selected according to a 4-wise independent distribution. This means that the algorithm can be derandomized in polynomial time.*

Proof: (Of Lemma 4.3) Let us write $\mathbf{x}$ as a linear combination of the eigenvectors

$$\mathbf{x} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$$

where the coefficients can be computed as $a_i = \langle \mathbf{x}, \mathbf{v}_i \rangle$. Note that, by assumption, $|a_1| \geq .5$, and that, by orthonormality of the eigenvectors, $||\mathbf{x}||^2 = \sum_i a_i^2$.

We have

$$\mathbf{y} = a_1\lambda_1^k\mathbf{v}_1 + \cdots + a_n\lambda_n^k\mathbf{v}_n$$

and so

$$\mathbf{y}^T M \mathbf{y} = \sum_i a_i^2 \lambda_i^{2k+1}$$

and

$$\mathbf{y}^T \mathbf{y} = \sum_i a_i^2 \lambda_i^{2k}$$

We need to prove a lower bound to the ratio of the above two quantities. We will compute a lower bound to the numerator and an upper bound to the denominator in terms of the same parameter.

Let $\ell$ be the number of eigenvalues larger than $\lambda_1 \cdot (1-\epsilon)$. Then, recalling that the eigenvalues are sorted in non-increasing order, we have

$$\mathbf{y}^T M \mathbf{y} \geq \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k+1} \geq \lambda_1 (1-\epsilon) \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k}$$

We also see that

$$\begin{aligned}
&\sum_{i=\ell+1}^{n} a_i^2 \lambda_i^{2k} \\
&\leq \lambda_1^{2k} \cdot (1-\epsilon)^{2k} \sum_{i=\ell+1}^{n} a_i^2 \\
&\leq \lambda_1^{2k} \cdot (1-\epsilon)^{2k} \cdot ||\mathbf{x}||^2 \\
&\leq 4 a_1^2 \lambda_1^{2k} (1-\epsilon)^{2t} ||\mathbf{x}||^2 \\
&\leq 4 ||\mathbf{x}||^2 (1-\epsilon)^{2k} \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k}
\end{aligned}$$

So we have

$$\mathbf{y}^T \mathbf{y} \leq (1 + 4||\mathbf{x}||^2 (1-\epsilon)^{2k}) \cdot \sum_{i=1}^{\ell} a_i^2$$

giving

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1-\epsilon) \cdot \frac{1}{1 + 4||\mathbf{x}||^2 (1-\epsilon)^{2k}}$$

□

**Remark 4.5** *Where did we use the assumption that $M$ is positive semidefinite? What happens if we apply this algorithm to the adjacency matrix of a bipartite graph?*

## 4.2 Approximating the Second Largest Eigenvalue

Suppose now that we are interested in finding the second largest eigenvalue of a given PSD matrix $M$. If $M$ has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$, and we know the eigenvector $\mathbf{v}_1$ of $\lambda_2$, then $M$ is a PSD linear map from the orthogonal space to $\mathbf{v}_1$ to itself, and $\lambda_2$ is the largest eigenvalue of this linear map. We can then run the previous algorithm on this linear map.

Algorithm *Power2*

Input: PSD matrix $M$, vector $\mathbf{v}_1$ parameter $k$

- Pick uniformly at random $\mathbf{x} \sim \{-1, 1\}^n$
- $\mathbf{x}_0 := \mathbf{x} - \mathbf{v}_1 \cdot \langle \mathbf{x}, \mathbf{v}_1 \rangle$
- for $i := 1$ to $k$
  $\mathbf{x}_i := M \cdot \mathbf{x}_{i-1}$
- return $\mathbf{x}_k$

If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an orthonormal basis of eigenvectors for the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ of $M$, then, at the beginning, we pick a random vector

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$$

that, with probability at least $3/16$, satisfies $|a_2| \geq 1/2$. (Cf. Lemma 4.2.) Then we compute $\mathbf{x}_0$, which is the projection of $\mathbf{x}$ on the subspace orthogonal to $\mathbf{v}_1$, that is

$$\mathbf{x}_0 = a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$$

Note that $||\mathbf{x}||^2 = n$ and that $||\mathbf{x}_0||^2 \leq n$.

The output is the vector $\mathbf{x}_k$

$$\mathbf{x}_k = a_2\lambda_2^k\mathbf{v}_2 + \cdots a_n\lambda_n^k\mathbf{v}_n$$

If we apply Lemma 4.3 to subspace orthogonal to $\mathbf{v}_1$, we see that when $|a_2| \geq 1/2$ we have that, for every $0 < \epsilon < 1$,

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_2 \cdot (1-\epsilon) \cdot \frac{1}{4n(1-\epsilon)^{2k}}$$

We have thus established the following analysis.

**Theorem 4.6** *For every PSD matrix $M$, positive integer $k$ and parameter $\epsilon > 0$, if $\mathbf{v}_1$ is a length-1 eigenvector of the largest eigenvalue of $M$, then with probability $\geq 3/16$ over the choice of $\mathbf{x}_0$, the algorithm Power2 outputs a vector $\mathbf{x}_k \perp \mathbf{v}_1$ such that*

$$\frac{\mathbf{x}_k^T M \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}_k} \geq \lambda_2 \cdot (1-\epsilon) \cdot \frac{1}{1 + 4n(1-\epsilon)^{2k}}$$

*where $\lambda_2$ is the second largest eigenvalue of $M$, counting multiplicities.*

## 4.3 The Second Smallest Eigenvalue of the Laplacian

Finally, we come to the case in which we want to compute the second smallest eigenvalue of the normalized Laplacian matrix $L = I - \frac{1}{d}A$ of a $d$-regular graph $G = (V, E)$, where $A$ is the adjacency matrix of $G$.

Consider the matrix $M := 2I - L = I + \frac{1}{d}A$. Then if $0 = \lambda_1 \leq \ldots \leq \lambda_n \leq 2$ are the eigenvalues of $L$, we have that

$$2 = 2 - \lambda_1 \geq 2 - \lambda_2 \geq \cdots \geq 2 - \lambda_n \geq 0$$

are the eigenvalues of $M$, and that $M$ is PSD. $M$ and $L$ have the same eigenvectors, and so $\mathbf{v}_1 = \frac{1}{\sqrt{n}}(1, \ldots, 1)$ is a length-1 eigenvector of the largest eigenvalue of $M$.

By running algorithm Power2, we can find a vector $\mathbf{x}$ such that

$$\mathbf{x}^T M \mathbf{x}^T \geq (1 - \epsilon) \cdot (2 - \lambda_2) \cdot \mathbf{x}^T \mathbf{x}$$

and

$$\mathbf{x}^T M \mathbf{x}^T = 2\mathbf{x}^T \mathbf{x} - \mathbf{x}^T L \mathbf{x}$$

so, rearranging, we have

$$\frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_2 + 2\epsilon$$

If we want to compute a vector whose Rayleigh quotient is, say, at most $2\lambda_2$, then the running time will be $\tilde{O}((|V| + |E|)/\lambda_2)$, because we need to set $\epsilon = \lambda_2/2$, which is not nearly linear in the size of the graph if $\lambda_2$ is, say $O(1/|V|)$.

For a running time that is nearly linear in $n$ for all values of $\lambda_2$, one can, instead, apply the power method to the *pseudoinverse* $L^+$ of $L$. (Assuming that the graph is connected, $L^+\mathbf{x}$ is the unique vector $\mathbf{y}$ such that $L\mathbf{y} = \mathbf{x}$, if $\mathbf{x} \perp (1, \ldots, 1)$, and $L^+\mathbf{x} = 0$ if $\mathbf{x}$ is parallel to $(1, \ldots, 1)$.) This is because $L^+$ has eigenvalues $0, 1/\lambda_2, \ldots, 1/\lambda_n$, and so $L^+$ is PSD and $1/\lambda_2$ is its largest eigenvalue.

Although computing $L^+$ is not known to be doable in nearly linear time, there are nearly linear time algorithms that, given $\mathbf{x}$, solve in $\mathbf{y}$ the linear system $L\mathbf{y} = \mathbf{x}$, and this is the same as computing the product $L^+\mathbf{x}$, which is enough to implement algorithm *Power* applied to $L^+$.

In time $O((V + |E|) \cdot (\log |V|/\epsilon)^{O(1)})$, we can find a vector $\mathbf{y}$ such that $\mathbf{y} = (L^+)^k \mathbf{x}$, where $\mathbf{x}$ is a random vector in $\{-1, 1\}^n$, shifted to be orthogonal to $(1, \ldots, 1)$ and $k = O(\log |V|/\epsilon)$. What is the Rayleigh quotient of such a vector with respect to $L$?

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of orthonormal eigenvectors for $L$ and $L^+$. If $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of $L$, then we have

$$L\mathbf{v}_1 = L^+\mathbf{v}_1 = \mathbf{0}$$

and, for $i = 1, \ldots, n$, we have

$$L\mathbf{v}_i = \lambda_i \quad L^+\mathbf{v}_i = \frac{1}{\lambda_i}$$

Write $\mathbf{x} = a_2\mathbf{v}_2 + \cdots a_n\mathbf{v}_n$, where $\sum_i a_i^2 \leq n$, and ssume that, as happens with probability at least 3/16, we have $a_2^2 \geq \frac{1}{4}$. Then

$$\mathbf{y} = \sum_{i=2}^{n} a_i \frac{1}{\lambda_i^k}$$

and the Rayleigh quotient of $\mathbf{y}$ with respect to $L$ is

$$\frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T\mathbf{y}} = \frac{\sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}}}{\sum_i a_i^2 \frac{1}{\lambda_i^{2k}}}$$

and the analysis proceeds similarly to the analysis of the previous section. If we let $\ell$ be the index such that $\lambda_\ell \leq (1+\epsilon)\cdot\lambda_2 \leq \lambda_{\ell+1}$ then we can upper bound the numerator as

$$\begin{aligned}
\sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}} &\leq \sum_{i\leq\ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1}\lambda_2^{2k-1}} \sum_{i>\ell} a_i^2 \\
&\leq \sum_{i\leq\ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1}\lambda_2^{2k-1}} \cdot n \\
&\leq \sum_{i\leq\ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1}\lambda_2^{2k-1}} \cdot 4na_2^2 \\
&\leq \left(1 + \frac{4n}{(1+\epsilon)^{2k-1}}\right) \cdot \sum_{i\leq\ell} a_i^2 \frac{1}{\lambda_i^{2k-1}}
\end{aligned}$$

and we can lower bound the denominator as

$$\begin{aligned}
\sum_i a_i^2 \frac{1}{\lambda_i^{2k}} &\geq \sum_{i\leq\ell} a_i^2 \frac{1}{\lambda_i^{2k}} \\
&\geq \frac{1}{(1+\epsilon)\lambda_2} \cdot \sum_{i\leq\ell} a_i^2 \frac{1}{\lambda_i^{2k-1}}
\end{aligned}$$

and the Rayleigh quotient is

$$\frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T\mathbf{y}} \leq \lambda_2 \cdot (1+\epsilon) \cdot \left(1 + \frac{4n}{(1+\epsilon)^{2k-1}}\right) \leq (1+2\epsilon)\cdot\lambda_2$$

when $k = O\left(\frac{1}{\epsilon}\log\frac{n}{\epsilon}\right)$.

An $O((|V|+|E|)\cdot(\log|V|)^{O(1)})$ algorithm to solve in $\mathbf{y}$ the linear system $L\mathbf{y} = \mathbf{x}$ was first developed by Spielman and Teng. Faster algorithms (with a lower exponent in the

$(\log |V|)^{O(1)}$ part of the running time, and smaller constants) were later developed by Koutis, Miller and Peng, and, very recently, by Kelner, Orecchia, Sidford, and Zhu.

Chapter 5

# Eigenvalues of Cayley Graphs

---

So far we have proved that

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2 \cdot \lambda_2}$$

and that the SpectralPartitioning algorithm, when given an eigenvector of $\lambda_2$, finds a cut $(S, V - S)$ such that $\phi(S) \leq 2\sqrt{\phi(G)}$. In this section we will show that all such results are tight, up to constants, by proving that

- The dimension-$d$ hypercube $H_d$ has $\lambda_2 = \frac{2}{d}$ and $\phi(H_d) = \frac{1}{d}$, giving an infinite family of graphs for which $\frac{\lambda_2}{2} = \phi(G)$, showing that the first Cheeger inequality is exactly tight.
- The $n$-cycle $C_n$ has $\lambda_2 = O(n^{-2})$, and $\phi(C_n) = \frac{2}{n}$, giving an infinite family of graphs for which $\phi(G) = \Omega(\sqrt{\lambda_2})$, showing that the second Cheeger inequality is tight up to a constant.
- There is an eigenvector of the 2nd eigenvalue of the hypercube $H_d$, such that the SpectralPartitioning algorithm, given such a vector, outputs a cut $(S, V - S)$ of expansion $\phi(S) = \Omega(1/\sqrt{d})$, showing that the analysis of the SpectralPartitioning algorithm is tight up to a constant.

We will develop some theoretical machinery to find the eigenvalues and eigenvectors of *Cayley graphs of finite Abelian groups*, a class of graphs that includes the cycle and the hypercube, among several other interesting examples. This theory will also be useful later, as a starting point to talk about algebraic constructions of expanders.

For readers familiar with the Fourier analysis of Boolean functions, or the discrete Fourier analysis of functions $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, or the standard Fourier analysis of periodic real functions, this theory will give a more general, and hopefully interesting, way to look at what they already know.

## 5.1 Characters

We will use additive notation for groups, so, if $\Gamma$ is a group, its unit will be denoted by 0, its group operation by $+$, and the inverse of element $a$ by $-a$. Unless, noted otherwise, however, the definitions and results apply to non-abelian groups as well.

**Definition 5.1 (Character)** *Let $\Gamma$ be a group (we will also use $\Gamma$ to refer to the set of group elements). A function $f : \Gamma \to \mathbb{C}$ is a character of $\Gamma$ if*

- *$f$ is a group homomorphism of $\Gamma$ into the multiplicative group $\mathbb{C} - \{0\}$.*
- *for every $x \in \Gamma$, $|f(x)| = 1$*

Though this definition might seem to not bear the slightest connection to our goals, the reader should hang on because we will see next time that finding the eigenvectors and eigenvalues of the cycle $C_n$ is immediate once we know the characters of the group $\mathbb{Z}/n\mathbb{Z}$, and finding the eigenvectors and eigenvalues of the hypercube $H_d$ is immediate once we know the characters of the group $(\mathbb{Z}/2\mathbb{Z})^d$.

**Remark 5.2 (About the Boundedness Condition)** *If $\Gamma$ is a finite group, and $a$ is any element, then*

$$\underbrace{a + \cdots + a}_{|\Gamma| \text{ times}} = 0$$

*and so if $f : \Gamma \to \mathbb{C}$ is a group homomorphism then*

$$1 = f(0) = f(\underbrace{a + \cdots + a}_{|\Gamma| \text{ times}}) = f(a)^{|\Gamma|}$$

*and so $f(a)$ is a root of unity and, in particular, $|f(a)| = 1$. This means that, for finite groups, the second condition in the definition of character is redundant.*

*In certain infinite groups, however, the second condition does not follow from the first, for example $f : \mathbb{Z} \to \mathbb{C}$ defined as $f(n) = e^n$ is a group homomorphism of $(\mathbb{Z}, +)$ into $(\mathbb{C} - \{0\}, \cdot)$ but it is not a character.*

Just by looking at the definition, it might look like a finite group might have an infinite number of characters; the above remark, however, shows that a character of a finite group $\Gamma$ must map into $|\Gamma|$-th roots of unity, of which there are only $|\Gamma|$, showing a finite $|\Gamma|^{|\Gamma|}$ upper bound to the number of characters. Indeed, a much stronger upper bound holds, as we will prove next, after some preliminaries.

**Lemma 5.3** *If $\Gamma$ is finite and $\chi$ is a character that is not identically equal to 1, then $\sum_{a \in \Gamma} \chi(a) = 0$*

PROOF: Let $b$ be such that $\chi(b) \neq 1$. Note that

$$\chi(b) \cdot \sum_{a \in \Gamma} \chi(a) = \sum_{a \in \Gamma} \chi(b+a) = \sum_{a \in \Gamma} \chi(a)$$

where we used the fact that the mapping $a \to b + a$ is a permutation. (We emphasize that even though we are using additive notation, the argument applies to non-abelian groups.) So we have

$$(\chi(b) - 1) \cdot \sum_{a \in \Gamma} \chi(a) = 0$$

and since we assumed $\chi(b) \neq 1$, it must be $\sum_{a \in \Gamma} \chi(a) = 0$. □

If $\Gamma$ is finite, given two functions $f, g : \Gamma \to \mathbb{C}$, define the inner product

$$\langle f, g \rangle := \sum_{a \in \Gamma} f(a) g^*(a)$$

**Lemma 5.4** *If $\chi_1, \chi_2 : \Gamma \to \mathbb{C}$ are two different characters of a finite group $\Gamma$, then*

$$\langle \chi_1, \chi_2 \rangle = 0$$

We will prove Lemma 5.4 shortly, but before doing so we note that, for a finite group $\Gamma$, the set of functions $f : \Gamma \to \mathbb{C}$ is a $|\Gamma|$-dimensional vector space, and that Lemma 5.4 implies that characters are orthogonal with respect to an inner product, and so they are linearly independent. In particular, we have established the following fact:

**Corollary 5.5** *If $\Gamma$ is a finite group, then it has at most $|\Gamma|$ characters.*

It remains to prove Lemma 5.4, which follows from the next two statements, whose proof is immediate from the definitions.

**Fact 5.6** *If $\chi_1, \chi_2$ are characters of a group $\Gamma$, then the mapping $x \to \chi_1(x) \cdot \chi_2(x)$ is also a character.*

**Fact 5.7** *If $\chi$ is a character of a group $\Gamma$, then the mapping $x \to \chi^*(x)$ is also a character, and, for every $x$, we have $\chi(x) \cdot \chi^*(x) = 1$.*

To complete the proof of Lemma 5.4, observe that:

- the function $\chi(x) := \chi_1(x) \cdot \chi_2^*(x)$ is a character;
- the assumption of the lemma is that there is an $a$ such that $\chi_1(a) \neq \chi_2(a)$, and so, for the same element $a$, $\chi(a) = \chi_1(a) \cdot \chi_2^*(a) \neq \chi_2(a) \cdot \chi_2^*(a) = 1$

- thus $\chi$ is a character that is not identically equal to 1, and so

$$0 = \sum_a \chi(a) = \langle \chi_1, \chi_2 \rangle$$

Notice that, along the way, we have also proved the following fact:

**Fact 5.8** *If $\Gamma$ is a group, then the set of characters of $\Gamma$ is also a group, with respect to the group operation of pointwise multiplication. The unit of the group is the character mapping every element to 1, and the inverse of a character is the pointwise conjugate of the character. The group of characters is called the Pontryagin dual of $\Gamma$, and it is denoted by $\hat{\Gamma}$.*

We now come to the punchline of this discussion.

**Theorem 5.9** *If $\Gamma$ is a finite abelian group, then it has exactly $|\Gamma|$ characters.*

PROOF: We give a constructive proof. We know that every finite abelian group is isomorphic to a product of cyclic groups

$$(\mathbb{Z}/n_1\mathbb{Z}) \times (\mathbb{Z}/n_2\mathbb{Z}) \times \cdots \times (\mathbb{Z}/n_k\mathbb{Z})$$

so it will be enough to prove that

1. the cyclic group $\mathbb{Z}/n\mathbb{Z}$ has $n$ characters;
2. if $\Gamma_1$ and $\Gamma_2$ are finite abelian groups with $|\Gamma_1|$ and $|\Gamma_2|$ characters, respectively, then their product has $|\Gamma_1| \cdot |\Gamma_2|$ characters.

For the first claim, consider, for every $r \in \{0, \ldots, n-1\}$, the function

$$\chi_r(x) := e^{2\pi i r x/n}$$

Each such function is clearly a character (0 maps to 1, $\chi_r(-x)$ is the multiplicative inverse of $\chi_r(x)$, and, recalling that $e^{2\pi i k} = 1$ for every integer $k$, we also have $\chi_r(a + b \bmod n) = e^{2\pi i r a/n} \cdot e^{2\pi i r b/n}$), and the values of $\chi_r(1)$ are different for different values of $r$, so we get $n$ distinct characters. This shows that $\mathbb{Z}/n\mathbb{Z}$ has at least $n$ characters, and we already established that it can have at most $n$ characters.

For the second claim, note that if $\chi_1$ is a character of $\Gamma_1$ and $\chi_2$ is a character of $\Gamma_2$, then it is easy to verify that the mapping $(x, y) \to \chi_1(x) \cdot \chi_2(y)$ is a character of $\Gamma_1 \times \Gamma_2$. Furthermore, if $(\chi_1, \chi_2)$ and $(\chi_1', \chi_2')$ are two distinct pairs of characters, then the mappings $\chi(x, y) := \chi_1(x) \cdot \chi_2(y)$ and $\chi'(x, y) := \chi_1'(x) \cdot \chi_2'(y)$ are two distinct characters of $\Gamma_1 \times \Gamma_2$, because we either have an $a$ such that $\chi_1(a) \neq \chi_1'(a)$, in which case $\chi(a, 0) \neq \chi'(a, 0)$, or we have a $b$ such that $\chi_2(b) \neq \chi_2'(b)$, in which case $\chi(0, b) \neq \chi'(0, b)$. This shows that $\Gamma_1 \times \Gamma_2$

has at least $|\Gamma_1| \cdot |\Gamma_2|$ characters, and we have already established that it can have at most that many $\square$

This means that the characters of a finite abelian group $\Gamma$ form an orthogonal basis for the set of all functions $f : \Gamma \to \mathbb{C}$, so that any such function can be written as a linear combination

$$f(x) = \sum_{\chi} \hat{f}(\chi) \cdot \chi(x)$$

For every character $\chi$, $\langle \chi, \chi \rangle = |\Gamma|$, and so the characters are actually a scaled-up orthonormal basis, and the coefficients can be computed as

$$\hat{f}(\chi) = \frac{1}{|\Gamma|} \sum_{x} f(x)\chi^*(x)$$

**Example 5.10 (The Boolean Cube)** *Consider the case $\Gamma = (\mathbb{Z}/2\mathbb{Z})^n$, that is the group elements are $\{0,1\}^n$, and the operation is bitwise xor. Then there is a character for every bit-vector $(r_1, \ldots, r_n)$, which is the function*

$$\chi_{r_1,\ldots,r_n}(x_1, \ldots, x_n) := (-1)^{r_1 x_1 + \cdots r_n x_n}$$

*Every boolean function $f : \{0,1\}^n \to \mathbb{C}$ can thus be written as*

$$f(x) = \sum_{r \in \{0,1\}^n} \hat{f}(r) \cdot (-1)^{\sum_i r_i x_i}$$

*where*

$$\hat{f}(r) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) \cdot (-1)^{\sum_i r_i x_i}$$

*which is the boolean Fourier transform.*

**Example 5.11 (The Cyclic Group)** *To work out another example, consider the case $\Gamma = \mathbb{Z}/N\mathbb{Z}$. Then every function $f : \{0, \ldots, N-1\} \to \mathbb{C}$ can be written as*

$$f(x) = \sum_{r \in \{0,\ldots,N-1\}} \hat{f}(r) e^{2\pi i r x/n}$$

*where*

$$\hat{f}(x) = \frac{1}{N} \sum_{x} f(x) e^{-2\pi i r x/n}$$

*which is the discrete Fourier transform.*

## 5.2 A Look Beyond

Why is the term "Fourier transform" used in this context? We will sketch an answer to this question, although what we say from this point on is not needed for our goal of finding the eigenvalues and eigenvectors of the cycle and the hypercube.

The point is that it is possible to set up a definitional framework that unifies both what we did in the previous section with finite Abelian groups, and the Fourier series and Fourier transforms of real and complex functions.

In the discussion of the previous section, we started to restrict ourselves to finite groups $\Gamma$ when we defined an inner product among functions $f : \Gamma \to \mathbb{C}$.

If $\Gamma$ is an infinite abelian group, we can still define an inner product among functions $f : \Gamma \to \mathbb{C}$, but we will need to define a measure over $\Gamma$ and restrict ourselves in the choice of functions. A measure $\mu$ over (a sigma-algebra of subsets of) $\Gamma$ is a Haar measure if, for every measurable subset $A$ and element $a$ we have $\mu(a+A) = \mu(A)$, where $a+A = \{a+b : b \in A\}$. For example, if $\Gamma$ is finite, $\mu(A) = |A|$ is a Haar measure. If $\Gamma = (\mathbb{Z}, +)$, then $\mu(A) = |A|$ is also a Haar measure (it is ok for a measure to be infinite for some sets), and if $\Gamma = (\mathbb{R}, +)$ then the Lebesgue measure is a Haar measure. When a Haar measure exists, it is more or less unique up to multiplicative scaling. All *locally compact topological* abelian groups have a Haar measure, a very large class of abelian groups, that include all finite ones, $(\mathbb{Z}, +)$, $(\mathbb{R}, +)$, and so on.

Once we have a Haar measure $\mu$ over $\Gamma$, and we have defined an integral for functions $f : \Gamma \to \mathbb{C}$, we say that a function is an element of $L^2(\Gamma)$ if

$$\int_\Gamma |f(x)|^2 d\mu(x) < \infty$$

For example, if $\Gamma$ is finite, then all functions $f : \Gamma \to \mathbb{C}$ are in $L^2(\Gamma)$, and a function $f : \mathbb{Z} \to \mathbb{C}$ is in $L^2(\mathbb{Z})$ if the series $\sum_{n\in\mathbb{Z}} |f(n)|^2$ converges.

If $f, g \in L^2(\Gamma)$, we can define their inner product

$$\langle f, g\rangle := \int_\Gamma f(x)g^*(x)d\mu(x)$$

and use Cauchy-Schwarz to see that $|\langle f, g\rangle| < \infty$.

Now we can repeat the proof of Lemma 5.4 that $\langle \chi_1, \chi_2\rangle = 0$ for two different characters, and the only step of the proof that we need to verify for infinite groups is an analog of Lemma 5.3, that is we need to prove that if $\chi$ is a character that is not always equal to 1, then

$$\int_\Gamma \chi(x)d\mu(x) = 0$$

and the same proof as in Lemma 5.3 works, with the key step being that, for every group element $a$,

$$\int_\Gamma \chi(x+a)d\mu(x) = \int_\Gamma \chi(x)d\mu(x)$$

because of the property of $\mu$ being a Haar measure.

We don't have an analogous result to Theorem 5.9 showing that $\Gamma$ and $\hat{\Gamma}$ are isomorphic, however it is possible to show that $\hat{\Gamma}$ itself has a Haar measure $\hat{\mu}$, that the dual of $\hat{\Gamma}$ is isomorphic to $\Gamma$, and that if $f : \Gamma \to \mathbb{C}$ is continuous, then it can be written as the "linear combination"

$$f(x) = \int_{\hat{\Gamma}} \hat{f}(\chi)\chi(x)d\hat{\mu}(x)$$

where

$$\hat{f}(\chi) = \int_\Gamma f(x)\chi^*(x)d\mu(x)$$

In the finite case, the examples that we developed before correspond to setting $\mu(A) := |A|/|\Gamma|$ and $\hat{\mu}(A) = |A|$.

**Example 5.12 (Fourier Series)** *The set of characters of the group $[0,1)$ with the operation of addition modulo 1 is isomorphic to $\mathbb{Z}$, because for every integer $n$ we can define the function $\chi_n : [0,1) \to \mathbb{C}$*

$$\chi_n(x) := e^{2\pi i x n}$$

*and it can be shown that there are no other characters. We thus have the Fourier series for continuous functions $f : [0,1) \to \mathbb{C}$,*

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}(n)e^{2\pi i x n}$$

*where*

$$\hat{f}(n) = \int_0^1 f(x)e^{-2\pi i x n}dx$$

## 5.3 Cayley Graphs and Their Spectrum

Let $\Gamma$ be a finite group. We will use additive notation, although the following definition applies to non-commutative groups as well. A subset $S \subseteq \Gamma$ is *symmetric* if $a \in S \Leftrightarrow -a \in S$.

**Definition 5.13** *For a group $\Gamma$ and a symmetric subset $S \subseteq \Gamma$, the Cayley graph $Cay(\Gamma, S)$ is the graph whose vertex set is $\Gamma$, and such that $(a,b)$ is an edge if and only if $b - a \in S$. Note that the graph is undirected and $|S|$-regular.*

We can also define Cayley *weighted* graphs: if $w : \Gamma \to \mathbb{R}$ is a function such that $w(a) = w(-a)$ for every $a \in \Gamma$, then we can define the weighted graph $Cay(G, w)$ in which the edge $(a, b)$ has weight $w(b - a)$. We will usually work with unweighted graphs.

**Example 5.14 (Cycle)** *The $n$-vertex cycle can be constructed as the Cayley graph $Cay(\mathbb{Z}/n\mathbb{Z}, \{-1, 1\})$.*

**Example 5.15 (Hypercube)** *The $d$-dimensional hypercube can be constructed as the Cayley graph*

$$Cay((\mathbb{Z}/2\mathbb{Z})^d, \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\})$$

*where the group is the set $\{0, 1\}^d$ with the operation of bit-wise xor, and the set $S$ is the set of bit-vectors with exactly one 1.*

If we construct a Cayley graph from a finite abelian group, then the eigenvectors are the characters of the groups, and the eigenvalues have a very simple description.

**Lemma 5.16** *Let $\Gamma$ be a finite abelian group, $\chi : \Gamma \to \mathbb{C}$ be a character of $\Gamma$, $S \subseteq \Gamma$ be a symmetric set. Let $A$ be the adjacency matrix of the Cayley graph $G = Cay(\Gamma, S)$. Consider the vector $\mathbf{x} \in \mathbb{C}^\Gamma$ such that $x_a = \chi(a)$.*

*Then $\mathbf{x}$ is an eigenvector of $A$, with eigenvalue*

$$\sum_{s \in S} \chi(s)$$

Proof: Consider the $a$-th entry of $M\mathbf{x}$:

$$\begin{aligned}(A\mathbf{x})_a &= \sum_b A_{a,b} x_b \\ &= \sum_{b : b-a \in S} \chi(b) \\ &= \sum_{s \in S} \chi(a + s) \\ &= x_a \cdot \cdot \sum_{s \in S} \chi(s)\end{aligned}$$

And so

$$A\mathbf{x} = \left( \sum_{s \in S} \chi(s) \right) \cdot \mathbf{x}$$

□

The eigenvalues of the form $\sum_{s \in S} \chi(s)$, where $\chi$ is a character, enumerate all the eigenvalues of the graph, as can be deduced from the following observations:

1. Every character is an eigenvector;
2. The characters are linearly independent (as functions $\chi : \Gamma \to \mathbb{C}$ and, equivalently, as vectors in $\mathbb{C}^\Gamma$);
3. There are as many characters as group elements, and so as many characters as nodes in the corresponding Cayley graphs.

It is remarkable that, for a Cayley graph, a system of eigenvectors can be determined based solely on the underlying group, independently of the set $S$.

Once we have the eigenvalues of the adjacency matrix, we can derive the eigenvalues of the Laplacian, because to every eigenvalue $\lambda$ of the adjacency there corresponds an eigenvalue $1 - \frac{1}{|S|}\lambda$ of the Laplacian, with the same eigenvector.

### 5.3.1 The Cycle

The $n$-cycle is the Cayley graph $Cay(\mathbb{Z}/n\mathbb{Z}, \{-1, +1\})$. Recall that, for every $n \in \{0, \ldots, n-1\}$, the group $\mathbb{Z}/n\mathbb{Z}$ has a character $\chi_r(x) = e^{2\pi i r x/n}$.

This means that for every $r \in \{0, \ldots, n-1\}$ we have the eigenvalue of the adjacency matrix

$$\lambda_r = e^{2\pi i r/n} + e^{-2\pi i r/n} = 2\cos(2\pi r/n)$$

where we used the facts that $e^{ix} = \cos(x) + i\sin(x)$, that $\cos(x) = \cos(-x)$, and $\sin(x) = -\sin(-x)$.

For $r = 0$ we have the eigenvalue 2. For $r = 1$ we have the second largest eigenvalue $2\cos(2\pi/n) = 2 - \Theta(1/n^2)$, meaning the second smallest eigenvalue of the Laplacian is $\Theta(1/n^2)$.

The expansion of the cycle is $\phi(C_n) \geq 2/n$, and so the cycle is an example in which the second Cheeger inequality is tight.

### 5.3.2 The Hypercube

The group $\{0,1\}^d$ with bitwise xor has $2^d$ characters; for every $r \in \{0,1\}^d$ there is a character $\chi_r : \{0,1\}^d \to \{-1, 1\}$ defined as

$$\chi_r(x) = (-1)^{\sum_i r_i x_i}$$

Let us denote the set $S$ by $\{e^1, \ldots, e^d\}$, where we let $e^j \in \{0,1\}^d$ denote the bit-vector that has a 1 in the $j$-th position, and zeroes everywhere else. This means that, for every bit-vector $r \in \{0,1\}^d$, the hypercube has the eigenvalue

$$\sum_j \chi_r(e^j) = \sum_j (-1)^{r_j} = (-|r| + d - |r|) = d - 2\frac{|r|}{d}$$

where we denote by $|r|$ the *weight* of $r$, that is, the number of ones in $r$.

Corresponding to $r = (0, \ldots, 0)$, we have the eigenvalue $d$.

For each of the $d$ vectors $r$ with exactly one 1, we have the second largest eigenvalue$d - 2$, corresponding to an eigenvalue $2/d$ of the Laplacian.

Let us compute the expansion of the hypercube. Consider "dimension cuts" of the form $S_i := \{x \in \{0,1\}^n : x_i = 0\}$. The set $S_i$ contains half of the vertices, and the number of edges that cross the cut $(S_i, V - S_i)$ is also equal to half the number of vertices (because the edges form a perfect matching), so we have $\phi(S_i) = \frac{1}{d}$.

These calculations show that the first Cheeger inequality $\lambda_2/2 \leq \phi(G)$ is tight for the hypercube.

Finally, we consider the tightness of the approximation analysis of the spectral partitioning algorithm.

We have seen that, in the $d$-dimensional hypercube, the second eigenvalue has multiplicity $d$, and that its eigenvectors are vectors $\mathbf{x}^j \in \mathbb{R}^{2^d}$ such that $x^j_a = (-1)^{a_j}$. Consider now the vector $\mathbf{x} := \sum_j \mathbf{x}^j$; this is still clearly an eigenvector of the second eigenvalue. The entries of the vector $\mathbf{x}$ are

$$x_a = \sum_j (-1)^{a_j} = d - 2|a|$$

Suppose now that we apply the spectral partitioning algorithm using $\mathbf{x}$ as our vector. This is equivalent to considering all the cuts $(S_t, V - S_t)$ in the hypercube in which we pick a threshold $t$ and define $S_t := \{a \in \{0,1\}^n : |a| \geq t\}$.

Some calculations with binomial coefficients show that the best such "threshold cut" is the "majority cut" in which we pick $t = n/2$, and that the expansion of $S_{n/2}$ is

$$\phi(S_{n/2}) = \Omega\left(\frac{1}{\sqrt{d}}\right)$$

This gives an example of a graph and of a choice of eigenvector for the second eigenvalue that, given as input to the spectral partitioning algorithm, result in the output of a cut $(S, V - S)$ such that $\phi(S) \geq \Omega(\sqrt{\phi(G)})$. Recall that we proved $\phi(S) \leq 2\sqrt{\phi(G)}$, which is thus tight.

## 5.4 Expanders of Logarithmic Degree

Let $p$ be a prime and $t < p$. We'll construct a $p^2$-regular multigraph $LD_{p,t}$ with $p^{t+1}$ vertices. The vertex set of the graph will be the $(t+1)$-dimensional vextor space $\mathbb{F}_p^{t+1}$ over $\mathbb{F}_p$.

For each vertex $x \in \mathbb{F}_p^{t+1}$, and every two scalars $a, b \in \mathbb{F}$, we have the edges $(x, x + (b, ab, a^2b, \ldots, a^tb)$.

In other words, the graph $LD_{p,t}$ is a Cayley graph of the additive group of $\mathbb{F}_p^{t+1}$, constructed using the generating multiset

$$S := \{(b, ab, \ldots, a^t b) : a, b \in \mathbb{F}_p^{t+1}\}$$

Note that the generating set is symmetric, that is, if $s \in S$ then $-s \in S$ (with the same multiplicity), and so the resulting multigraph is undirected.

Let $A_{p,t}$ be the adjacency matrix of $LD_{p,t}$ and $M_{p,t} := p^{-2}A_{p,t}$ be the normalized adjacency matrix. We will prove the following bound on the eigenvalues of $M_{p,t}$.

**Theorem 5.17** *For every prime $p$ and every $t < p$, if we let $1 = \lambda_1 \geq \lambda_2 \geq \cdots \lambda_n$ be the eigenvalues of $M$ with multiplicities, then, for every $i \in \{2, \ldots, n\}$*

$$0 \leq \lambda_i \leq \frac{t}{p}$$

For example, setting $t = \lfloor p/2 \rfloor$ gives us a family of graphs such that $\lambda_2 \leq 1/2$ for each graph in the family, and hence $h \geq 1/8$, and the number of vertices is $p^{p/2}$, while the degree is $p^2$, meaning the degree is $O((\log n / \log\log n)^2)$.

PROOF: We will compute the eigenvalues of the adjacency matrix of $A_{p,t}$, and prove that, except the largest one which is $p^2$, all the others are non-negative and at most $pt$.

Recall our characterization of the eigenvalues of the adjacency matrix of a Cayley multigraph $Cay(\Gamma, S)$ of an abelian group $\Gamma$ with generating multiset $S$: we have one eigenvector for each character $\chi$ of the group, and the corresponding eigenvalue is $\sum_{s \in S} \chi(s)$.

What are the characters of the additive group of $\mathbb{F}_p^{t+1}$? It is the product of $t+1$ copies of the additive group of $\mathbb{F}_p$, or, equivalently, the product of $t+1$ copies of the cyclic group $\mathbb{Z}/p\mathbb{Z}$. Following our rules for constructing the character of the cyclic group and of products of groups, we see that the additive group of $\mathbb{F}_p^{t+1}$ has one character for each $(c_0, \ldots, c_t) \in \mathbb{F}_p^{t+1}$, and the corresponding character is

$$\chi_{c_0,\ldots,c_t}(x_0, \ldots, x_t) := \omega^{\sum_{i=0}^t c_i x_i}$$

where

$$\omega := e^{\frac{2\pi i}{p}}$$

Thus, for each $(c_0, \ldots, c_t) \in \mathbb{F}_t^p$, we have an eigenvalue

$$\lambda_{c_0,\ldots,c_t} := \sum_{a,b \in \mathbb{F}_p} \omega^{\sum_{i=0}^t c_i b a^i}$$

When $(c_0, \ldots, c_t) = (0, \ldots, 0)$ then the corresponding character is always equal to one, and the corresponding eigenvalue is $p^2$.

Now consider any $(c_0, \ldots, c_t) \neq (0, \ldots, 0)$, and define the polynomial $q(x) = \sum_{i=0}^t c_i x^i \in \mathbb{F}_p[x]$. Note that it is a non-zero polynomial of degree at most $t$, and so it has at most $t$ roots. The eigenvalue corresponding to $(c_0, \ldots, c_t)$ is

$$\lambda_{c_0,\ldots,c_t} = \sum_{a,b\in\mathbb{F}_p} \omega^{\sum_{i=0}^{t} b\cdot q(a)}$$

$$= \sum_{a:q(a)=0} \sum_{b} \omega^0 + \sum_{a:q(a)\neq 0} \sum_{b} \omega^{b\cdot q(a)}$$

$$= p\cdot|\{a\in\mathbb{F}_p : q(a)=0\}|$$

where we use the fact that, for every $q\neq 0$, the sum $\sum_b \omega^{b\cdot q}$ equals zero, since it is the sum of the values of the non-trivial character $x\to\omega^{x\cdot q}$, and we proved that, for every non-trivial character, the sum is zero.

In conclusion, we have

$$0\leq \lambda_{c_0,\ldots,c_t}\leq pt$$

□

## Exercises

1. Find the eigenvalues of a clique with $n$ vertices.
2. Consider the bipartite complete graph $K_{n,n}$ with $2n$ vertices. Express it as a Cayley graph and find its eigenvalues.
3. Show that

$$\phi(G) = \min_{\mathbf{x}\in\mathbb{R}^n,\mathrm{med}(\mathbf{x})=0} \frac{\sum_{\{u,v\}\in E}|x_u - x_v|}{d\sum_v |x_v|}$$

   Hint: use an algorithm similar to the proof of the difficult direction of the Cheeger inequality, but pick $t$ uniformly at random.

# Chapter 6

# Expander Constructions

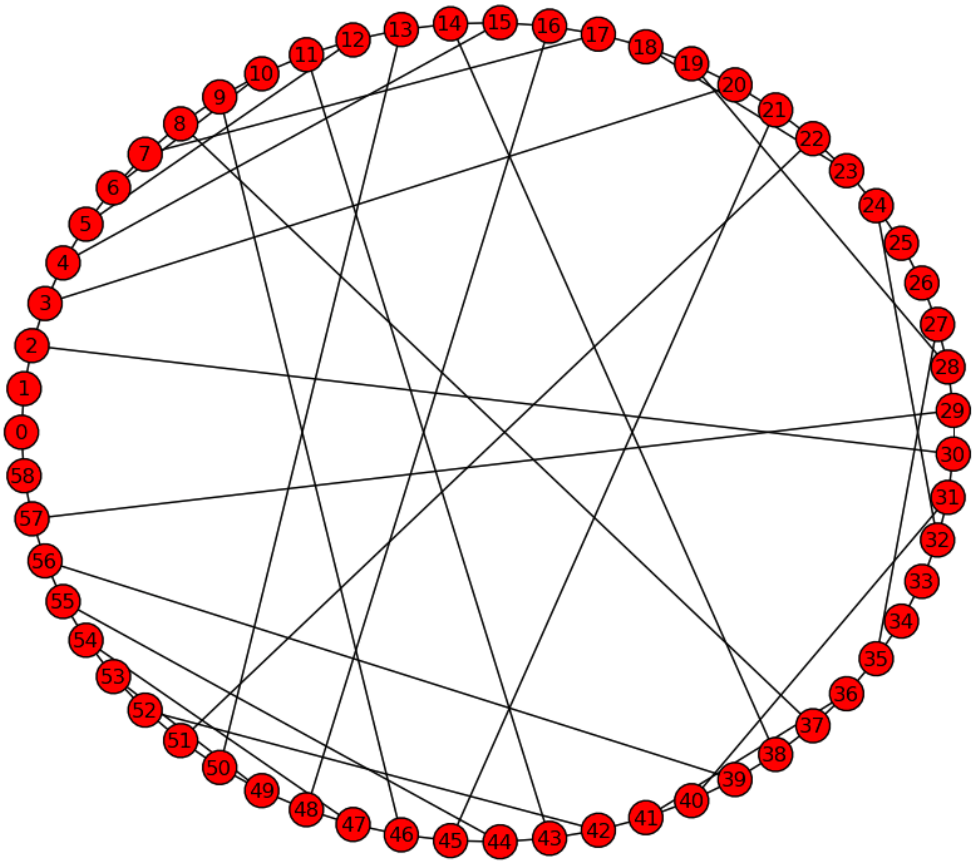
---

A *family of expanders* is a family of graphs $G_n = (V_n, E_n)$, $|V_n| = n$, such that each graph is $d_n$-regular, and the edge-expansion of each graph is at least $h$, for an absolute constant $h$ independent of $n$. Ideally, we would like to have such a construction for each $n$, although it is usually enough for most applications that, for some constant $c$ and every $k$, there is an $n$ for which the construction applies in the interval $\{k, k+1, \ldots, ck\}$, or even the interval $\{k, \ldots, ck^c\}$. We would also like the degree $d_n$ to be slowly growing in $n$ and, ideally, to be bounded above by an explicit constant. Today we will see a simple construction in which $d_n = O(\log^2 n)$ and a more complicated one in which $d_n = O(1)$.

An *explicit construction* of a family of expanders is a construction in which $G_n$ is "efficiently computable" given $n$. The weakest sense in which a construction is said to be explicit is when, given $n$, the (adjacency matrix of the) graph $G_n$ can be constructed in time polynomial in $n$. A stronger requirement, which is necessary for several applications, is that given $n$ and $i \in \{1, \ldots, n\}$, the list of neighbors of the $i$-th vertex of $G_n$ can be computed in time polynomial in $\log n$.

In many explicit constructions of constant-degree expanders, the construction is extremely simple, and besides satisfying the stricter definition of "explicit" above, it is also such that the adjacency list of a vertex is given by a "closed-form formula." The analysis of such constructions, however, usually requires very sophisticated mathematical tools.

**Example 6.1** *Let $p$ be a prime, and define the graph $G_p = (V_p, E_p)$ in which $V_p = \{0, \ldots, p-1\}$, and, for $a \in V_p - \{0\}$, the vertex $a$ is connected to $a+1 \bmod p$, to $a-1 \bmod p$ and to its multiplicative inverse $a^{-1} \bmod p$. The vertex $0$ is connected to $1$, to $p-1$, and has a self-loop. Counting self-loops, the graph is 3-regular: it is the union of a cycle over $V_p$ and of a matching over the $p-3$ vertices $V_p - \{0, 1, p-1\}$; the vertices $0$, $1$, $p-1$ have a self-loop each. There is a constant $h > 0$ such that, for each $p$, the graph $G_p$ has edge expansion at least $h$. Unfortunately, no elementary proof of this fact is known. The graph $G_{59}$ is shown in the picture below.*




Constructions based on the *zig-zag graph product*, which we shall see next, are more complicated to describe, but much simpler to analyze.

## 6.1 The Zig-Zag Graph Product

Given a regular graph $G$ with normalized adjacency matrix $M$, if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of $M$ with multiplicities we define

$$\lambda(G) := \max_{i=2,\ldots,n} \{|\lambda_i|\}$$

In particular, $\lambda(G) \geq \lambda_2$, and if we are able to construct a family of graphs such that $\lambda(G)$ is at most a fixed constant bounded away from one, then we have a family of expanders. (Our construction will be inductive and, as often happens with inductive proofs, it will be

easier to maintain this stronger property than the property that $\lambda_2$ is bounded away from one.)

Given graphs $G$ and $H$ of compatible sizes, with small degree and large edge expansion, the zig zag product $G ⓩ H$ is a method of constructing a larger graph also with small degree and large edge expansion.

If:

- $G$ a $D$-regular graph on $n$ vertices, with $\lambda(G) \leq \alpha$
- $H$ a $d$-regular graph on $D$ vertices, with $\lambda(H) \leq \beta$

Then:

- $G ⓩ H$ a $d^2$-regular graph on $nD$ vertices, with $\lambda(G ⓩ H) \leq \alpha + \beta + \beta^2$.

We will see the construction and analysis of the zig zag product in the next lecture.

For the remainder of today, we'll see how to use the zig zag product to construct arbitrarily large graphs of fixed degree with large edge expansion.

Fix a large enough constant $d$. ($1369 = 37^2$ will do.) Construct a $d$-regular graph $H$ on $d^4$ vertices with $\lambda_2(H) \leq d/5$. (For example $LD_{37,7}$ is a degree $37^2$ graph on $37^{(7+1)} = (37^2)^4$ vertices with $\lambda_2 \leq 37 \times 7 < 37^2/5$.)

For any graph $G$, let $G^2$ represent the graph on the same vertex set whose edges are the paths of length two in $G$. Thus $G^2$ is the graph whose adjacency matrix is the square of the adjacency matrix of $G$. Note that if $G$ is $r$-regular then $G^2$ is $r^2$-regular

Using the $H$ from above we'll construct inductively, a family of progressively larger graphs, all of which are $d^2$-regular and have $\lambda \leq d^2/2$.

Let $G_1 = H^2$. For $k \geq 1$ let $G_{k+1} = (G_k^2) ⓩ H$.

**Theorem 6.2** *For each $k \geq 1$, $G_k$ has degree $d^2$ and $\lambda(G_k) \leq 1/2$.*

Proof: We'll prove this by induction.
Base case: $G_1 = H^2$ is $d^2$-regular. Also, $\lambda(H^2) = (\lambda(H))^2 \leq d^2/25$.

Inductive step: Assume the statement for $k$, that is, $G_k$ has degree $d^2$ and $\lambda(G_k) \leq d^2/2$. Then $G_k^2$ has degree $d^4 = |V(H)|$, so that the product $(G_k^2) ⓩ H$ is defined. Moreover, $\lambda(G_k^2) \leq d^4/4$. Applying the construction, we get that $G_{k+1}$ has degree $d^2$ and $\lambda(G_{k+1}) \leq (\frac{1}{4} + \frac{1}{5} + \frac{1}{25})d^2 = \frac{46}{100}d^2$ This completes the proof. □

Finally note that $G_k$ has $d^{4k}$ vertices.

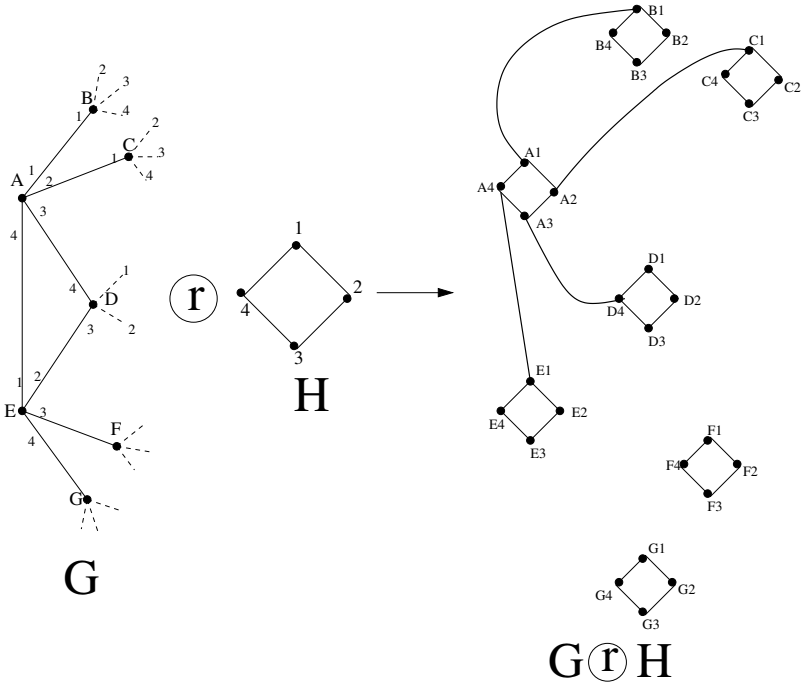
### 6.1.1 Replacement product of two graphs

We first describe a simpler product for a "small" $d$-regular graph on $D$ vertices (denoted by $H$) and a "large" $D$-regular graph on $N$ vertices (denoted by $G$). Assume that for each

vertex of $G$, there is some ordering on its $D$ neighbors. Then we construct the replacement product (see figure) $G \textcircled{r} H$ as follows:

- Replace each vertex of $G$ with a copy of $H$ (henceforth called a *cloud*). For $v \in V(G), i \in V(H)$, let $(v, i)$ denote the $i^{th}$ vertex in the $v^{th}$ cloud.
- Let $(u, v) \in E(G)$ be such that $v$ is the $i$-th neighbor of $u$ and $u$ is the $j$-th neighbor of $v$. Then $((u, i), (v, j)) \in E(G \textcircled{r} H)$. Also, if $(i, j) \in E(H)$, then $\forall u \in V(G)$ $((u, i), (u, j)) \in E(G \textcircled{r} H)$.

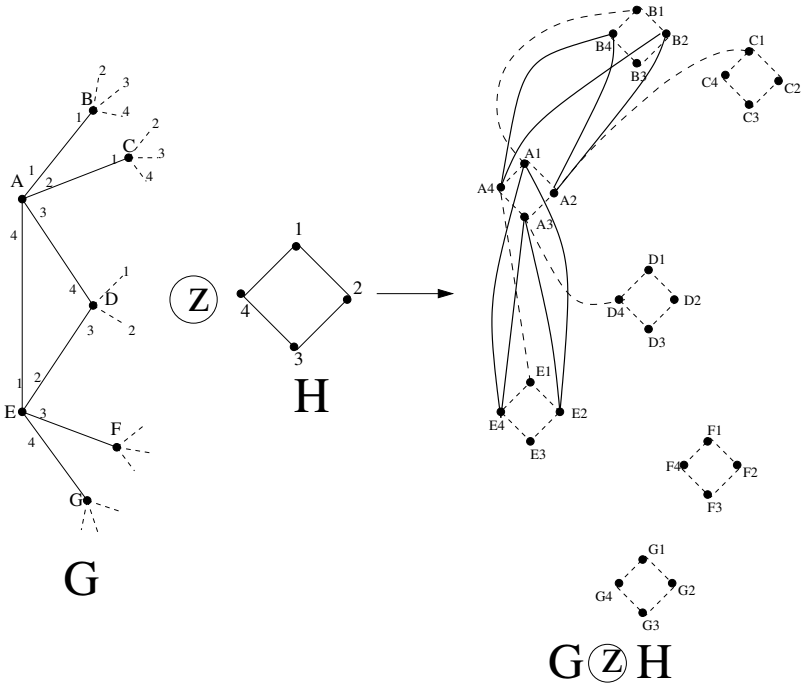
Note that the replacement product constructed as above has $ND$ vertices and is $(d + 1)$-regular.




### 6.1.2 Zig-zag product of two graphs

Given two graphs $G$ and $H$ as above, the zig-zag product $G \textcircled{z} H$ is constructed as follows (see figure):

- The vertex set $V(G \textcircled{z} H)$ is the same as in the case of the replacement product.
- $((u, i), (v, j)) \in E(G \textcircled{z} H)$ if there exist $\ell$ and $k$ such that $((u, i)(u, \ell)$, $((u, \ell), (v, k))$ and $((v, k), (v, j))$ are in $E(G \textcircled{r} H)$ i.e. $(v, j)$ can be reached from $(u, i)$ by taking a step in the first cloud, then a step between the clouds and then a step in the second cloud (hence the name!).




It is easy to see that the zig-zag product is a $d^2$-regular graph on $ND$ vertices.

Let $M \in \mathbb{R}^{([N]\times[D])\times([N]\times[D])}$ be the normalized adjacency matrix of $G \textcircled{z} H$. Using the fact that each edge in $G \textcircled{r} H$ is made up of three steps in $G \textcircled{r} H$, we can write $M$ as $BAB$, where

$$B[(u,i),(v,j)] = \begin{cases} 0 & if\ u \neq v \\ M_H[i,j] & if\ u = v \end{cases}$$

And $A[(u,i),(v,j)] = 1$ if $u$ is the $j$-th neighbor of $v$ and $v$ is the $i$-th neighbor of $u$, and $A[(u,i),(v,j)] = 0$ otherwise.

Note that $A$ is the adjacency matrix for a matching and is hence a permutation matrix.

### 6.1.3 Preliminaries on Matrix Norms

Recall that, instead of bounding $\lambda_2$, we will bound the following parameter (thus proving a stronger result).

**Definition 6.3** *Let $M$ be the normalized adjacency matrix of a graph $G = (V, E)$, and $\lambda_1 \geq \ldots \geq \lambda_n$ be its eigenvalues with multiplicities. Then we use the notation*

$$\lambda(M) := \max_{i=2,\ldots,n} \{|\lambda_i|\} = \max\{\lambda_2, -\lambda_n\}$$

The parameter $\lambda$ has the following equivalent characterizations.

**Fact 6.4**

$$\lambda(M) = \max_{\mathbf{x}\in\mathbb{R}^V-\{\mathbf{0}\},\mathbf{x}\perp\mathbf{1}} \frac{||Mx||}{||x||} = \max_{\mathbf{x}\in\mathbb{R}^v,\mathbf{x}\perp\mathbf{1},||\mathbf{x}||=1} ||Mx||$$

Another equivalent characterization, which will be useful in several contexts, can be given using the following matrix norm.

**Definition 6.5 (Spectral Norm)** *The spectral norm of a matrix $M \in \mathbb{R}^{n\times n}$ is defined as*

$$||M|| = \max_{\mathbf{x}\in\mathbb{R}^V,||\mathbf{x}||=1} ||Mx||$$

If $M$ is symmetric with eigenvalues $\lambda_1, \ldots, \lambda_n$, then the spectral norm is $\max_i |\lambda_i|$. Note that $M$ is indeed a norm, that is, for every two square real matrices $A, B$ we have $||A+B|| \leq ||A|| + ||B||$ and for every matrix $A$ and scalar $\alpha$ we have $||\alpha A|| = \alpha||A||$. In addition, it has the following useful property:

**Fact 6.6** *For every two matrices $A, B \in \mathbb{R}^{n\times n}$ we have*

$$||AB|| \leq ||A|| \cdot ||B||$$

PROOF: For every vector $x$ we have

$$||AB\mathbf{x}|| \leq ||A|| \cdot ||B\mathbf{x}|| \leq ||A|| \cdot ||B|| \cdot ||\mathbf{x}||$$

where the first inequality is due to the fact that $||Az|| \leq ||A|| \cdot ||\mathbf{z}||$ for every vector $\mathbf{z}$, and the second inequality is due to the fact that $||B\mathbf{x}|| \leq ||B|| \cdot ||\mathbf{x}||$. So we have

$$\min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}}$$

□

We can use the spectral norm to provide another characterization of the parameter $\lambda(M)$ of the normalized adjacency matrix of a graph.

**Lemma 6.7** *Let $G$ be a regular graph and $M \in \mathbb{R}^{n\times n}$ be its normalized adjacency matrix. Then*

$$\lambda(M) = ||M - \frac{1}{n}J||$$

*where $J$ is the matrix with a 1 in each entry.*

PROOF: Let $\lambda_1 = 1 \geq \lambda_2 \geq \cdots \lambda_n$ be the eigenvalues of $M$ and $\mathbf{v}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$, $\mathbf{v}_2$, …, $\mathbf{v}_n$ a corresponding system of orthonormal eigenvector. Then we can write

$$M = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T$$

Noting that $\mathbf{v}_1 \mathbf{v}_1^T = \frac{1}{n} J$, we have

$$M - \frac{1}{n} J = 0 \cdot \mathbf{v}_1 \mathbf{v}_2^T + \sum_{i=2}^{n} \lambda_i \mathbf{v}_i \mathbf{v}_i^T$$

and so $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is also a system of eigenvectors for $M - \frac{1}{n} J$, with corresponding eigenvalues $0, \lambda_2, \ldots, \lambda_n$, meaning that

$$||M - \frac{1}{n} J|| = \max\{0, \lambda_2, \ldots, \lambda_n\} = \lambda(M)$$

□

The above lemma has several applications. It states that, according to a certain definition of distance, when a graph is a good expander then it is close to a clique. (The matrix $\frac{1}{n} J$ is the normalized adjacency matrix of a clique with self-loops.) The proof of several results about expanders is based on noticing that the result is trivial for cliques, and then on "approximating" the given expander by a clique using the above lemma.

We need one more definition before we can continue with the analysis of the zig-zag graph product.

**Definition 6.8 (Tensor Product)** *Let $A \in \mathbb{R}^{N \times N}$ and $B \in \mathbb{R}^{D \times D}$ be two matrices. Then $A \otimes B \in \mathbb{R}^{ND \times ND}$ is a matrix whose rows and columns are indexed by pairs $(u, i) \in [N] \times [D]$ such that*

$$(A \otimes B)_{(u,i),(v,j)} = A_{u,v} \cdot B_{i,j}$$

For example $I \otimes M$ is a block-diagonal matrix in which every block is a copy of $M$.

### 6.1.4 Analysis of the Zig-Zag Product

Suppose that $G$ and $H$ are identical cliques with self-loops, that is, are both $n$-regular graphs with self-loops. Then the zig-zag product of $G$ and $H$ is well-defined, because the degree of $G$ is equal to the number of vertices of $H$. The resulting graph $G ⓩ H$ is a $n^2$-regular graph with $n^2$ vertices, and an inspection of the definitions reveals that $G ⓩ H$ is indeed a clique (with self-loops) with $n^2$ vertices.

The intuition for our analysis is that we want to show that the zig-zag graph product "preserves" distances measured in the matrix norm, and so if $G$ is close (in matrix norm) to a clique and $H$ is close to a clique, then $G ⓩ H$ is close to the zig-zag product of two cliques, that is, to a clique. (Strictly speaking, what we just said does not make sense, because we cannot take the zig-zag product of the clique that $G$ is close to and of the clique that $H$ is close to, because they do not have the right degree and number of vertices. The proof, however, follows quite closely this intuition.)

**Theorem 6.9** *If $\lambda(M_G) = a$ and $\lambda(M_H) = b$, then*

$$\lambda(G ⓩ H) \leq a + 2b + b^2$$

PROOF: Let $M$ be the normalized adjacency matrix of $G ⓩ H$, and let $\mathbf{x}$ be a unit vector such that $\mathbf{x} \perp \mathbf{1}$ and

$$\lambda(M) = ||M\mathbf{x}||$$

Recall that we defined a decomposition

$$M = BAB$$

where $A$ is a permutation matrix, and $B = I \otimes M_H$. Let us write $E := M_H - \frac{1}{D}J$, then $B = I \otimes \frac{1}{D}J + I \otimes E$. Let us call $\bar{J} := I \otimes \frac{1}{D}J$ and $\bar{E} := I \otimes E$.

First, we argue that the matrix norm of $\bar{E}$ is small. Take any vector $\mathbf{z} \in \mathbb{R}^{ND}$ and write is as $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_N)$, where, for each $u \in [N]$, $\mathbf{z}_u$ is the $D$-dimensional restriction of $\mathbf{z}$ to the coordinates in the cloud of $u$. Then

$$||(I \otimes E)\mathbf{z}||^2 = \sum_u ||E\mathbf{z}_u||^2 \leq \sum_u ||E||^2 \cdot ||\mathbf{z}_u||^2 = ||E||^2 \cdot ||\mathbf{z}||^2$$

and so we have

$$||I \otimes E|| \leq ||E|| \leq b$$

Then we have

$$\begin{aligned} BAB &= (\bar{J} + \bar{E})A(\bar{J} + \bar{E}) \\ &= \bar{J}A\bar{J} + \bar{J}A\bar{E} + \bar{E}A\bar{J} + \bar{E}A\bar{A} \end{aligned}$$

and so, using the triangle inequality and the property of the matrix norm, we have

$$||BAB\mathbf{x}|| \leq ||\bar{J}A\bar{J}\mathbf{x}|| + ||\bar{E}A\bar{J}|| + ||\bar{J}A\bar{E}|| + ||\bar{E}A\bar{E}||$$

where

$$\begin{aligned} ||\bar{E}A\bar{J}|| &\leq ||\bar{E}|| \cdot ||A|| \cdot ||\bar{J}|| \leq ||\bar{E}|| \leq b \\ ||\bar{J}A\bar{E}|| &\leq ||\bar{J}|| \cdot ||A|| \cdot ||\bar{E}|| \leq ||\bar{E}|| \leq b \\ ||\bar{E}A\bar{E}|| &\leq ||\bar{E}|| \cdot ||A|| \cdot ||\bar{E}|| \leq ||\bar{E}||^2 \leq b^2 \end{aligned}$$

It remains to prove that $||\bar{J}A\bar{J}\mathbf{x}|| \leq a$. If we let $A_G = DM_G$ be the adjacency matrix of $G$, then we can see that

$$(\bar{J}A\bar{J})_{(u,i),(v,j)} = \frac{1}{D^2}(A_G)_{u,v} = \frac{1}{D}(M_G)_{u,v} = (M_G \otimes \frac{1}{D}J)_{(u,i),(v,j)}$$

That is,

$$\bar{J}A\bar{J} = M_G \otimes \frac{1}{D}J$$

Finally, we write $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, where $\mathbf{x}_u$ is the $D$-dimensional vector of entries corresponding to the cloud of $u$, we call $y_u := \sum_i \mathbf{x}_u(i)/D$, and we note that, by Cauchy-Schwarz:

$$||\mathbf{y}||^2 = \sum_u \left( \sum_i \frac{1}{D} \mathbf{x}_{u,i} \right)^2 \leq \sum_u \left( \sum_i \frac{1}{D}^2 \right) \cdot \left( \sum_i \mathbf{x}_{u,i}^2 \right) = \frac{1}{D} ||\mathbf{x}||^2$$

The final calculation is:

$$\begin{aligned}
||\bar{J}A\bar{J}\mathbf{x}||^2 &= \left| | \left( M_G \otimes \frac{1}{D}J \right) \mathbf{x} \right| |^2 \\
&= \sum_{u,i} \left( \sum_{v,j} \frac{1}{D} (M_G)_{u,v} \mathbf{x}_{u,i} \right)^2 \\
&= \sum_{u,i} \left( \sum_v (M_G)_{u,v} y_u \right)^2 \\
&= D \cdot \sum_u \left( \sum_v (M_G)_{u,v} y_u \right)^2 \\
&= D \cdot ||M_G \mathbf{y}||^2 \\
&\leq D \cdot a^2 \cdot ||\mathbf{y}||^2 \\
&\leq a^2 \cdot ||\mathbf{x}^2||^2
\end{aligned}$$

□

## 6.2 The Margulis-Gabber-Galil Expanders

We present a construction of expander graphs due to Margulis, which was the first explicit construction of expanders, and its analysis due to Gabber and Galil. The analysis presented here includes later simplifications, and it follows an exposition of James Lee.

For every $n$, we construct graphs with $n^2$ vertices, and we think of the vertex set as $\mathbb{Z}_n \times \mathbb{Z}_n$, the group of pairs from $\{0, \ldots, n-1\} \times \{0, \ldots, n-1\}$ where the group operation is coordinate-wise addition modulo $n$.

Define the functions $S(a, b) := (a, a + b)$ and $T(a, b) := (a + b, b)$, where all operations are modulo $n$. Then the graph $G_n(V_n, E_n)$ has vertex set $V_n := \mathbb{Z}_n \times \mathbb{Z}_n$ and the vertex $(a, b)$ is connected to the vertices

$$(a+1,b),(a-1,b),(a,b+1),(a,b-1),S(a,b),S^{-1}(a,b),T(a,b),T^{-1}(a,b)$$

so that $G_n$ is an 8-regular graph. (The graph has parallel edges and self-loops.)

We will prove that there is a constant $c > 0$ such that $\lambda_2(G_n) \geq c$ for every $n$.

The analysis will be in three steps.

First, we show that $\lambda_2(G_n)$ is bounded from below, up to a constant, by the "spectral gap" of an infinite graph $R_n$, whose vertex set is $[0,n]^2$. We write "spectral gap" in quote because we will not define a Laplacian and argue about the existence of a spectrum, but just study the infimum of an expression that looks like a Rayleigh quotient, and prove that $\lambda_2(G_n)$ is at least a constant times this infimum. This is proved by showing that for every test function $f$ that sums to zero defined over $\mathbb{Z}_n^2$ we can define a function $g$ over $[0,n]$ whose integral is zero and whose Rayleigh quotient for $R_n$ is the same, up to a constant factor, as the Rayleigh quotient of $f$ for $G_n$

Then we consider another infinite graph $G_\infty$, with vertex set $\mathbb{Z} \times \mathbb{Z}$, and again define a formal "spectral gap" by considering the infimum of a Rayleigh quotient and we prove that, for every $n$, the spectral gap of $R_n$ is bounded from below, up to a constant, by the spectral gap of $G_\infty$. This is proved by showing that if $f$ is a test function whose integral is zero and whose Rayleigh quotient for $R_n$ is small, the Fourier transform of $f$ is a test function of small Rayleigh quotient for $G_\infty$.

Finally, we define a notion of expansion for graphs with a countable number of vertices, such as $G_\infty$. We prove that for infinite graphs with a countable set of vertices there is a Cheeger inequality relating expansion and spectral gap, we prove that $G_\infty$ has constant expansion, and we use the Cheeger inequality to conclude that $G_\infty$ has constant spectral gap. (From the previous steps, it follows that $R_n$, and hence $G_n$ also have spectral gap bounded from below by an absolute constant.)

### 6.2.1 First Step: The Continuous Graph

For every $n$, we consider a graph $R_n$ with vertex set $[0,n)^2$ and such that every vertex $(x,y)$ is connected to

$$S(x,y), S^{-1}(x,y), T(x,y), T^{-1}(x,y)$$

where, as before, $S(x,y) = (x, x+y)$ and $T(x,y) = T(x+y,y)$ and the operations are done modulo $n$.

(We are thinking of $[0,n)$ as a group with the operation of addition modulo $n$, that is, the group quotient $\mathbb{R}/n\mathbb{Z}$, where $n\mathbb{Z}$ is the group of multiples of $n$ with the operation of addition, just like $\mathbb{Z}_n$ is the quotient $\mathbb{Z}/n\mathbb{Z}$.)

Let $\ell_2([0,n)^2)$ be set of functions $f : [0,n)^2 \to \mathbb{R}$ such that $\int_{[0,n)^2} (f(x,y))^2 \mathrm{d}x\mathrm{d}y$ is well defined and finite. Then we define the following quantity, that we think of as the spectral gap of $R_n$:

$$\lambda_2(R_n) := \inf_{f \in \ell_2([0,n)^2) \,:\, \int_{[0,n)^2} f = 0} \frac{\int_{[0,n)^2} |f(x,y) - f(S(x,y))|^2 + |f(x,y) - f(T(x,y))|^2 \mathrm{d}x\mathrm{d}y}{\int_{[0,n)^2} (f(x,y))^2 \mathrm{d}x\mathrm{d}y}$$

We could define a Laplacian operator and show that the above quantity is indeed the second smallest eigenvalue, but it will not be necessary for our proof.

We have the following bound.

**Theorem 6.10** $\lambda_2(G_n) \geq \frac{1}{12} \cdot \lambda_2(R_n)$.

PROOF: Let $f$ be the function such that

$$\lambda_2(G) = \frac{\sum_{c \in \mathbb{Z}_n^2} |f(c) - f(S(c))|^2 + |f(c) - f(T(c))|^2 + |f(c) - f(c+(0,1))|^2 + |f(c) - f(c+(1,0))|^2}{8 \sum_{c \in \mathbb{Z}_n^2} f^2(c)}$$

For a point $(x,y) \in [0,n)^2$, define $floor(x,y) := (\lfloor x \rfloor, \lfloor y \rfloor)$. We extend $f$ to a function $\tilde{f} : [0,n)^2 \to \mathbb{R}$ by defining

$$\tilde{f}(z) := f(floor(z))$$

This means that we tile the square $[0,n)^2$ into unit squares whose corners are integer-coordinate, and that $\tilde{f}$ is constant on each unit square, and it equals the value of $f$ at the left-bottom corner of the square.

It is immediate to see that

$$\int_{[0,n)^2} \tilde{f}^2(z) \mathrm{d}z = \sum_{c \in \mathbb{Z}_n^2} f^2(c)$$

and so, up to a factor of 8, the denominator of the Rayleigh quotient of $f$ is the same as the denominator of the Rayleigh quotient of $\tilde{f}$.

It remains to bound the numerators.

Observe that for every $z \in [0,1)^2$, we have that $floor(S(z))$ equals either $S(floor(z))$ or $S(floor(z)) + (0,1)$, and that $floor(T(z))$ equals either $T(floor(z)$ or $T(floor(z)$. Also, $floor(z + (0,1)) = floor(z) + (0,1)$, and the same is true for $(1,0)$. The numerator of the Rayleigh quotient of $\tilde{f}$ is

$$\sum_{c=(a,b) \in \mathbb{Z}_n^2} \int_{[a,a+1) \times [b,b+1)} |\tilde{f}(z) - \tilde{f}(S(z))|^2 + |\tilde{f}(z) - \tilde{f}(T(z))|^2 \mathrm{d}z$$

$$= \frac{1}{2} \sum_{c \in \mathbb{Z}_n^2} |f(c) - f(S(c))|^2 + |f(c) - f(S(c)+(0,1))|^2 + |f(c) - f(T(c))|^2 + |f(c) - f(T(c)+(1,0))|^2$$

because for a $(x, y)$ randomly chosen in the square $[a, a+1) \times [b, b+1)$, there is probability $1/2$ that $\lfloor x+y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$ and probability $1/2$ that $\lfloor x+y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor + 1$.

Now we can use the "triangle inequality"

$$|\alpha - \beta|^2 \leq 2|\alpha - \gamma|^2 + 2|\gamma - \beta|^2$$

to bound the above quantity

$$\begin{aligned}
&\leq \frac{1}{2} \sum_{c \in \mathbb{Z}_n^2} |f(c) - f(S(c))|^2 + \\
&2|f(c) - f(c+(0,1))|^2 + 2|f(c+(0,1)) - f(S(c)+(0,1))|^2 + \\
&|f(c) - f(T(c))|^2 + \\
&2|f(c) - f(c+(1,0))|^2 + 2|f(c+(1,0)) - f(T(c)+(1,0))|^2
\end{aligned}$$

which simplifies to

$$= \frac{1}{2} \sum_{c \in \mathbb{Z}_n^2} 3|f(c)-f(S(c))|^2 + 3|f(c)-f(T(c))|^2 + 2|f(c)-f(c+(0,1))|^2 + 2|f(c)-f(c+(1,0))|^2$$

which is at most $3/2$ times the numerator of the Rayleigh quotient of $f$. $\square$

### 6.2.2 Second Step: The Countable Graph

We now define the graph $Z$ of vertex set $\mathbb{Z} \times \mathbb{Z} - \{(0,0)\}$, where each vertex $(a, b)$ is connected to

$$(a, a+b), (a, a-b), (a+b, a), (a-b, a)$$

Note

For a graph $G = (V, E)$ with an countably infinite set of vectors, define $\ell_2(V)$ to be the set of functions $f : V \to \mathbb{R}$ such that $\sum_{v \in V} f^2(v)$ is finite, and define the spectral gap of $G$ as

$$\lambda_2(G) := \inf_{f \in \ell_2(V)} \frac{\sum_{(u,v) \in V} |f(u) - f(v)|^2}{\sum_v f^2(v)}$$

So that

$$\lambda_2(Z) := \inf_{f \in \ell_2(\mathbb{Z} \times \mathbb{Z} - \{(0,0)\})} \frac{\sum_{a,b} |f(a,b) - f(a,a+b)|^2 + |f(a,b) - f(a+b,a)|^2}{\sum_{a,b} f^2(a,b)}$$

We want to show the following result.

**Theorem 6.11** *For every $n$, $\lambda_2(R_n) \geq \lambda_2(Z)$.*

PROOF: This will be the most interesting part of the argument. Let $f \in \ell_2([0,n)^2)$ be any function such that $\int f = 0$, we will show that the Fourier transform $\hat{f}$ of $f$ has a Rayleigh quotient for $Z$ that is at most the Rayleigh quotient of $f$ for $R_n$.

First, we briefly recall the definitions of Fourier transforms. If $f : [0,n)^2 \to \mathbb{R}$ is such that

$$\int_{z\in[0,n)^2} f^2(z)\mathrm{d}z < \infty$$

then we can write the linear combination

$$f(z) = \sum_{c\in\mathbb{Z}\times\mathbb{Z}} \hat{f}(c)\cdot\chi_c(z)$$

where the basis functions are

$$\chi_{a,b}(x,y) = \frac{1}{n}e^{2\pi i\cdot(ax+by)}$$

and the coefficients are

$$\hat{f}(c) = \langle f, \chi_{a,b}\rangle := \int_{[0,n)^2} f(z)\chi_c(z)\mathrm{d}z$$

The condition $\int f = 0$ gives

$$\hat{f}(0,0) = 0$$

and the Parseval identity gives

$$\sum_{c\neq(0,0)} \hat{f}^2(c) = \sum_c \hat{f}^2(c) = \int f^2(z)\mathrm{d}z$$

and so we have that the denominator of the Rayleigh quotient of $f$ for $R_n$ and of $\hat{f}$ for $Z$ As usual, the numerator is more complicated.

We can break up the numerator of the Rayleigh quotient of $f$ as

$$\int s^2(z)\mathrm{d}z + \int t^2(z)\mathrm{d}z$$

where $s(z) := f(z) - f(S(z))$ and $t(z) := f(z) - f(T(z))$, and we can use Parseval's identity to rewrite it as

$$\sum_c \hat{s}^2(c) + \hat{t}^2(c)$$

$$= \sum_c |\hat{f}(c) - \widehat{(f\circ S)}(c)|^2 + |\hat{f}(c) - \widehat{(f\circ T)}(c)|^2$$

The Fourier coefficients of the function $(f\circ S)(z) = f(S(z))$ can be computed as

$$\widehat{(f\circ S)}(a,b) = \frac{1}{n}\int f(S(x,y))e^{2\pi i(ax+by)}$$

$$= \frac{1}{n} \int f(x, x+y) e^{2\pi i(ax+by)}$$
$$= \frac{1}{n} \int f(x, y') e^{2\pi i(ax+by'-bx)}$$
$$= \hat{f}(a-b, b)$$

where we used the change of variable $y' \leftarrow x + y$.

Similarly, $\widehat{(f \circ T)}(a, b) = \hat{f}(a, b-a)$. This means that the numerator of the Rayleigh quotient of $f$ for $R_n$ is equal to the numerator of the Rayleigh quotient of $\hat{f}$ for $Z$. □

### 6.2.3 Third Step: Proving a Spectral Gap for $\mathbb{Z}$

Now we need to prove that $\lambda_2(Z) \geq \Omega(1)$. We will prove that $Z$ has constant edge expansion, and then we will use a Cheeger inequality for countable graphs to deduce a spectral gap.

Define the edge expansion of a graph $G = (V, E)$ with a countably infinite set of vertices as

$$\phi(G) = \inf_{A \subseteq V, A \text{ finite}} \frac{E(A, \bar{A})}{|A|}$$

Note that the edge expansion can be zero even if the graph is connected.

We will prove the following theorems

**Theorem 6.12 (Cheeger inequality for countable graphs)** *For every $d$-regular graph $G = (V, E)$ with a countably infinite set of vertices we have*

$$\phi(G) \leq \sqrt{2 \cdot d \cdot \lambda_2(G)}$$

**Theorem 6.13 (Expansion of $Z$)** $\phi(Z) \geq 1.25$.

Putting it all together we have that $\lambda_2(Z) \geq \frac{\phi(Z)^2}{2d} > .195$, $\lambda_2(R_n) > .195$, and $\lambda_2(G_n) > .0162$.

#### Cheeger inequality for countable graphs

Proof:[Of Theorem 6.12] This is similar to the proof for finite graphs, with the simplification that we do not need to worry about constructing a set containing at most half of the vertices.

Let $f \in \ell_2(\mathbb{Z}^2)$ be any function. We will show that $\phi$ is at most $\sqrt{2r}$ where

$$r := \frac{\sum_{(u,v) \in E} |f(u) - f(v)|^2}{\sum_{v \in V} f^2(v)}$$

is the Rayleigh quotient of $f$.

For every threshold $t \geq t_{\min} := \inf_{v \in V} f^2(v)$, define the set $S_t \subseteq V$ as

$$S_t := \{v : f^2(v) > t\}$$

and note that each set is finite because $\sum_v f^2(v)$ is finite. We have, for $t > t_{\min}$,

$$\phi(G) \leq \frac{E(S_t, \bar{S}_t)}{|S_t|}$$

and, for all $t \geq 0$

$$|S_t| \cdot \phi(G) \leq E(S_t, \bar{S}_t)$$

Now we compute the integral of the numerator and denominator of the above expression, and we will find the numerator and denominator of the Rayleigh quotient $r$.

$$\int_0^\infty |S_t| \mathrm{d}t = \sum_{v \in V} \int_0^\infty I_{f^2(v) > t} \mathrm{d}t = \sum_{v \in V} f^2(v)$$

and

$$\int_0^\infty |E(S_t, \bar{S}_t) \mathrm{d}t = \sum_{(u,v) \in E} \int_0^\infty I_{t \text{ between } f^2(u), f^2(v)} \mathrm{d}t = \sum_{(u,v)} |f^2(u) - f^2(v)|$$

Which means

$$\phi \leq \frac{\sum_{u,v} |f(u) - f(v)|^2}{\sum_v f^2(v)}$$

Now we proceed with Cauchy Schwarz:

$$\begin{aligned}
&\sum_{(u,v) \in E} |f^2(u) - f^2(v)| \\
&= \sum_{(u,v) \in E} |f(u) - f(v)| \cdot |f(u) + f(v)| \\
&\leq \sqrt{\sum_{(u,v) \in E} |f(u) - f(v)|^2} \cdot \sqrt{\sum_{(u,v) \in E} |f(u) + f(v)|^2} \\
&\leq \sqrt{\sum_{(u,v) \in E} |f(u) - f(v)|^2} \cdot \sqrt{\sum_{(u,v) \in E} 2f^2(u) + 2f^2(v)} \\
&= \sqrt{\sum_{(u,v) \in E} |f(u) - f(v)|^2} \cdot \sqrt{\sum_{v \in V} 2df(v)^2}
\end{aligned}$$

And we have

$$\phi \leq \frac{\sqrt{\sum_{(u,v)\in E} |f(u) - f(v)|^2} \cdot \sqrt{2d}}{\sqrt{\sum_{v\in V} f(v)^2}} = \sqrt{2d \cdot r}$$

□

**Expansion of $\mathbb{Z}$**

After all these reductions, we finally come to the point where we need to prove that something is an expander.

PROOF:[Of Theorem 6.13] Let $A$ be a finite subset of $\mathbb{Z} \times \mathbb{Z} - \{(0,0\}$.

Let $A_0$ be the set of elements of $A$ that have one 0 coordinate. Let $A_1, A_2, A_3, A_4$ be the set of elements of $A$ with nonzero coordinate that belong to the 1st, 2nd, 3rd and 4th quadrant. (Starting from the quadrant of points having both coordinates positive, and numbering the remaining ones clockwise.)

**Claim 6.14** $E(A - A_0, \bar{A}) \geq |A - A_0| = |A| - |A_0|$.

PROOF: Consider the sets $S(A_1)$ and $T(A_1)$; both $S()$ and $T()$ are permutations, and so $|S(A_1)| = |T(A_1)| = |A_1|$. Also, $S(A_1)$ and $T(A_1)$ are disjoint, because if we had $(a, a+b) = (a'+b', b')$ then we would have $b = -a'$ while all the coordinates are strictly positive. Finally, $S(A_1)$ and $T(A_1)$ are also contained in the first quadrant, and so at least $|A_1|$ of the edges leaving $A_1$ lands outside $A$. We can make a similar argument in each quadrant, considering the sets $S^{-1}(A_2)$ and $T^{-1}(A_2)$ in the second quadrant, the sets $S(A_3)$ and $T(A_3)$ in the third, and $S^{-1}(A_4)$ and $T^{-1}(A_4)$ in the fourth. □

**Claim 6.15** $E(A_0, \bar{A}) \geq 4|A_0| - 3|A - A_0| = 7|A_0| - 3|A|$

PROOF: All the edges that have one endpoint in $A_0$ have the other endpoint outside of $A_0$. Some of those edges, however, may land in $A - A_0$. Overall, $A - A_0$ can account for at most $4|A - A_0|$ edges, and we have already computed that at least $|A - A_0|$ of them land into $\bar{A}$, so $A - A_0$ can absorb at most $3|A - A_0|$ of the outgoing edges of $A_0$. □

Balancing the two equalities (adding 7/8 times the first plus 1/8 times the second) gives us the theorem. □

Chapter 7

# Properties of Expanders

## 7.1 Quasirandomness of Expander Graphs

Recall that if $G$ is a $d$-regular graph, $A$ is its adjacency matrix, and $M = \frac{1}{d}A$ is its normalized adjacency matrix, then, if we call $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ the eigenvalues of $M$ with repetitions, we call $\lambda(G) := \max_{i=2,\ldots,n}\{|\lambda_i|\}$, and we have

$$\lambda(G) = ||M - \frac{1}{n}J||$$

where $J$ is the matrix with a one in each entry, and $||\cdot||$ is the matrix norm $||M|| := \max_{x,||x||=1} ||Mx||$.

Our fist result today is to show that, when $\lambda(G)$ is small, the graph $G$ has the following *quasirandomness* property: for every two disjoint sets $S, T$, the number of edges between $S$ and $T$ is close to what we would expect in a random graph of average degree $d$, that is, approximately $\frac{d}{|V|}|S||T|$.

**Lemma 7.1 (Expander Mixing Lemma)** *Let $G = (V, E)$ be a $d$-regular graph, and let $S$ and $T$ be two disjoint subsets of vertices. Then*

$$\left| edges_G(S,T) - \frac{d}{|V|} \cdot |S| \cdot |T| \right| \leq \lambda(G) \cdot d \cdot \sqrt{|S| \cdot |T|}$$

Proof: We have

$$edges_G(S,T) = d\mathbf{1}_S^\top M \mathbf{1}_T$$

and

$$|S||T| = \mathbf{1}_S^\top J \mathbf{1}_T$$

so

$$\begin{aligned}
&\left|edges_G(S,T) - \frac{d}{|V|}\cdot|S|\cdot|T|\right| \\
&= d\cdot\left|\mathbf{1}_S^\top M\mathbf{1}_T - \frac{1}{|V|}\mathbf{1}_S^\top J\mathbf{1}_T\right| \\
&= d\cdot\left|\mathbf{1}_S^\top\left(M - \frac{1}{|V|}J\right)\mathbf{1}_T\right| \\
&\leq d\cdot||\mathbf{1}_S||\cdot\left\|M - \frac{1}{|V|}J\right\|\cdot\|\mathbf{1}_T\| \\
&= d\cdot\sqrt{|S|}\cdot\lambda(G)\cdot\sqrt{|T|}
\end{aligned}$$

□

Note that, for every disjoint $S, T$, we have $\sqrt{|A|\cdot|B|} \leq |V|/2$, and so the right-hand side in the expander mixing lemma is at most $\lambda(G)\cdot|E|$, which is a small fraction of the total number of edges if $\lambda$ is small.

## 7.2 Random Walks in Expanders

A *$t$-step random walk* is the probabilistic process in which we start at a vertex, then we pick uniformly at random one of the edges incident on the vertices and we move to the other endpoint of the edge, and then repeat this process $t$ times.

If $M$ is the normalized adjacency matrix of an undirected regular graph $G$, then $M(u,v)$ is the probability that, in one step, a random walk started at $u$ reaches $v$. This is why the normalized adjacency matrix of a regular graph is also called its *transition matrix*.

Suppose that we start a random walk at a vertex chosen according to a probability distribution $\mathbf{p}$, which we think of as a vector $\mathbf{p} \in \mathbb{R}^V$ such that $\mathbf{p}(u) \geq 0$ for every $u$ and $\sum_u \mathbf{p}(u) = 1$. After taking one step, the probability of being at vertex $v$ is $\sum_u \mathbf{p}(u)M(u,v)$, which means that the probability distribution after one step is described by the vector $\mathbf{p}^\top \cdot M$, and because of the symmetric of $M$, this is the same as $M\mathbf{p}$.

Iterating the above reasoning, we see that, after a $t$-step random walk whose initial vertex is chosen according to distribution $\mathbf{p}$, the last vertex reached by the walk is distributed according to $M^t\mathbf{p}$.

The parameter $\lambda$ of $M^t$ is equal to $(\lambda(G))^t$, and so if $G$ has a parameter $\lambda$ bounded away from one, and if $t$ is large enough, we have that the parameter $\lambda$ of $M^t$ is very small, and so $M^t$ is close to $\frac{1}{n}J$ in matrix norm. If $M^t$ was actually equal to $\frac{1}{n}J$, then $M^t\cdot\mathbf{p}$ would be equal to the uniform distribution, for every distribution $\mathbf{p}$. We would thus expect $M^t\cdot\mathbf{p}$ to be close to the uniform distribution for large enough $t$.

Before formalizing the above intuition, we need to fix a good measure of distance for distributions. If we think of distributions as vectors, then a possible notion of distance between

two distributions is the Euclidean distance between the corresponding vectors. This definition, however, has various shortcoming and, in particular, can assign small distance to distributions that are intuitively very different. For example, suppose that $\mathbf{p}$ and $\mathbf{q}$ are distributions that are uniform over a set $S$, and over the complement of $S$, respectively, where $S$ is a set of size $|V|/2$. Then all the entries of $\mathbf{p}-\mathbf{q}$ are $\pm 2/n$ and so $||\mathbf{p}-\mathbf{q}|| = 2/\sqrt{n}$, which is vanishingly small even though distributions over disjoint supports should be considered as maximally different distributions.

A very good measure is the *total variation distance*, defined as

$$\max_{S \subseteq V} \left| \sum_{v \in S} \mathbf{p}(v) - \sum_{v \in S} \mathbf{q}(v) \right|$$

that is, as the maximum over all events of the difference between the probability of the event happening with respect to one distribution and the probability of it happening with respect to the other distribution. This measure is usually called *statistical distance* in computer science. It is easy to check that the total variation distance between $\mathbf{p}$ and $\mathbf{q}$ is precisely $\frac{1}{2} \cdot ||\mathbf{p} - \mathbf{q}||_1$. Distributions with disjoint support have total variation distance 1, which is largest possible.

**Lemma 7.2 (Mixing Time of Random Walks in Expanders)** *Let $G$ be a regular graph, and $M$ be its normalized adjacency matrix. Then for every distribution $\mathbf{p}$ over the vertices and every $t$, we have*

$$||\mathbf{u} - M^t \mathbf{p}||_1 \leq \sqrt{|V|} \cdot (\lambda(G))^t$$

*where $\mathbf{u}$ is the uniform distribution.*

*In particular, if $t > \frac{c}{1-\lambda(G)} \cdot \ln \frac{|V|}{\epsilon}$, then $||\mathbf{u} - M^t \mathbf{p}||_1 \leq \epsilon$, where $c$ is an absolute constant.*

PROOF: Let $\bar{J} = J/|V|$ be the normalized adjacency matrix of a clique with self-loops. Then, for every distribution $\mathbf{p}$, we have $\bar{J}\mathbf{p} = \mathbf{u}$. Recall also that $\lambda(G) = ||M - \bar{J}||$.

We have

$$\begin{aligned} & ||\mathbf{u} - M^t \mathbf{p}||_1 \\ \leq & \sqrt{|V|} \cdot ||\mathbf{u} - M^t \mathbf{p}|| \\ \leq & \sqrt{|V|} \cdot ||\bar{J}\mathbf{p} - M^t \mathbf{p}|| \\ \leq & \sqrt{|V|} \cdot ||\bar{J} - M^t|| \cdot ||\mathbf{p}|| \\ \leq & \sqrt{|V|} \cdot (\lambda(G))^t \end{aligned}$$

□

The last result that we discussed today is one more instantiation of the general phenomenon that "if $\lambda(G)$ is small then a result that is true for the clique is true, within some approximation, for $G$."

Suppose that we take a $(t-1)$-step random walk in a regular graph $G$ starting from a uniformly distributed initial vertex. If $G$ is a clique with self-loops, then the sequence of $t$ vertices encountered in the random walk is a sequence of $t$ independent, uniformly distributed, vertices. In particular, if $f : V \to [0.1]$ is a bounded function, the Chernoff-Hoeffding bounds tell us that the empirical average of $f()$ over the $t$ points of the random walk is very close to the true average of $f()$, except with very small probability, that is, if we denote by $v_1, \ldots, v_t$ the set of vertices encountered in the random walk, we have

$$\mathbb{P}\left[\frac{1}{t}\sum_i f(v_i) \geq \mathbb{E}\, f + \epsilon\right] \leq e^{-2\epsilon^2 t}$$

where $n := |V|$. A corresponding Chernoff-Hoeffding bound can be proved for the case in which the random walk is taken over a regular graph such that $\lambda(G)$ is small.

**Lemma 7.3 (Chernoff-Hoeffding Bound for Random Walks in Expanders)** *Let $G = (V, E)$ be a regular graph, and $(v_1, \ldots, v_t)$ the distribution of $t$-tuples constructed by sampling $v_1$ independently, and then performing a $(t-1)$-step random walk starting at $v_1$. Let $f : V \to [0, 1]$ be any bounded function. Then*

$$\mathbb{P}\left[\frac{1}{t}\sum_i f(v_i) \geq \mathbb{E}\, f + \epsilon + \lambda(G)\right] \leq e^{-\Omega(\epsilon^2 t)}$$

We will not prove the above result, but we briefly discuss one of its many applications.

Suppose that we have a polynomial-time probabilistic algorithm $A$ that, on inputs of length $n$, uses $r(n)$ random bits and then outputs the correct answer with probability, say, at least $2/3$. One standard way to reduce the error probability is to run the algorithm $t$ times, using independent randomness each time, and then take the answer that comes out a majority of the times. (This is for problems in which we want to compute a function exactly; in combinatorial optimization we would run the algorithm $t$ times and take the best solutions, and in an application in which the algorithm performs an approximate function evaluation we would run the algorithm $t$ times and take the median. The reasoning that follows for the case of exact function computation can be applied to the other settings as well.)

On average, the number of iterations of the algorithms that give a correct answer is $\geq 2t/3$, and the cases in which the majority is erroneous correspond to cases in which the number of iterations giving a correct answer is $\leq t/2$. This means that the case in which the modified algorithm makes a mistake correspond to the case in which the empirical average of $t$ independent 0/1 random variables deviates from its expectation by more than $2/3 - 1/2 = 1/6$, which can happen with probability at most $e^{-t/18}$, which becomes vanishingly small for large $t$.

This approach uses $t \cdot r(n)$ random bits. Suppose, instead, that we consider the following algorithm: pick $t$ random strings for the algorithm by performing a $t$-step random walk in an expander graph of degree $O(1)$ with $2^{r(n)}$ vertices and such that $\lambda(G) \leq 1/12$, and then take the majority answer. A calculation using the Chernoff bound for expander graphs

show that the error probability is $e^{-\Omega(t)}$, and it is achieved using only $r(n) + O(t)$ random bits instead of $t \cdot r(n)$.

Chapter 8

# The Non-Uniform Sparsest Cut Problem

---

Let $G = (V, E)$ be an undirected graph. Unlike past lectures, we will not need to assume that $G$ is regular. Recall that, for a subset $S \subseteq V$, we defined the *sparsity* of the partition $(S, V - S)$ as

$$\sigma_G(S) := \frac{\mathop{\mathbb{E}}_{(u,v)\sim E} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\mathop{\mathbb{E}}_{(u,v)\sim V^2} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}$$

which is the ratio between the fraction of edges that are cut by $(S, V - S)$ and the fraction of pairs of vertices that are disconnected by the removal of those edges.

More generally, give two (possibly weighted) undirected graphs $G = (V, E_G)$ and $H = (V, E_H)$ over the same set of vertices, we define the non-uniform sparsity of $S \subseteq V$ as

$$\sigma_{G,H}(S) := \frac{\mathop{\mathbb{E}}_{(u,v)\sim E_G} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}{\mathop{\mathbb{E}}_{(u,v)\sim E_H} |\mathbf{1}_S(u) - \mathbf{1}_S(v)|}$$

For graphs $G, H$ with the same vertex set $V$, the non-uniform sparsest cut problem is to find

$$\sigma(G, H) := \min_{S \subseteq V} \quad \sigma_{G,H}(S)$$

Notice that $\sigma_G(S)$ is the same as $\sigma_{G,H}$ where $H$ is a clique in which every vertex has a self-loop of weight $1/2$.

If $d_v$ is the degree of $v$ in $G$, then define $H$ to be the graph in which the edge $\{u, v\}$ has weight $d_u \cdot d_v$. Then, $|E_G| = vol_G(V)/2$, $|E_H| = \frac{1}{2}\sum_u \cdot (\sum_v d_u d_v) = \frac{1}{2}(\sum_v d_v)^2$, $E_H(S, V - S) = vol_G(S) \cdot vol_G(V - S)$, and the sparsity of $S$ is

$$\sigma_{G,H}(S) = \left(\sum_v d_v\right) \cdot \frac{E(S,\bar{S})}{vol_G(S) \cdot vol_G(V-S)}$$

and if $S$ is such that $vol(S) \leq vol(V-S)$, then $\phi(G) \leq \sigma_{G,H}(S) \leq 2\phi(G)$.

Thus the non-uniform sparsest cut problem generalizes the (uniform) sparsest cut problem that we described before, and, for a proper choice of $H$, is a 2-approximation of the conductance of $G$.

Notice also that if $H$ is a graph that has only the one edge $\{s,t\}$, then $\sigma(G,H)$ is the $(s,t)$-min-cut problem for the graph $G$.

## 8.1 A Linear Programming relaxation

Another way to formulate the sparsest cut problem is

$$\sigma(G,H) := \frac{|E_H|}{|E_G|} \cdot \min_{\mathbf{x}\in\{0,1\}^n} \frac{\sum_{\{u,v\}} G_{u,v}|x_u - x_v|}{\sum_{\{u,v\}} H_{u,v}|x_u - x_v|}$$

where $G_{u,v}$ is the weight of the edge $\{u,v\}$ in $G$ and $H_{u,v}$ is the weight of the edge $\{u,v\}$ in $H$.

The observation that led us to see $\lambda_2$ as the optimum of a continuous relaxation of $\sigma(G)$ was to observe that, for a boolean vector $\mathbf{x}$, $|x_u - x_v| = |x_u - x_v|^2$, and then relax the problem by allowing arbitrary vectors $\mathbf{x}$ instead of just boolean vectors.

The Leighton-Rao relaxation of sparsest cut is obtained using, instead, the following observation: if, for a set $S$, $\mathbf{x}$ is the boolean indicator vector of $S$ and we define $d_S(u,v) := |x_u - x_v|$, then $d_S(\cdot,\cdot)$ defines a semi-metric over the set $V$, because $d_S$ is symmetric, $d_S(v,v) = 0$, and the triangle inequality holds. So we could think about allowing *arbitrary semi-metrics* in the expression for $\sigma$, and define

$$LR(G,H) := \min_{\substack{d: V\times V \to \mathbb{R} \\ d \text{ semi-metric}}} \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u,v)}{\sum_{\{u,v\}} H_{u,v} d(u,v)} \tag{8.1}$$

This might seem like such a broad relaxation that there could be graphs on which $LR(G,H)$ bears no connection to $\sigma(G,H)$. Instead, we will prove the fairly good estimate

$$LR(G,H) \leq \phi(G,H) \leq O(\log|V|) \cdot LR(G,H) \tag{8.2}$$

Furthermore, we will show that $LR(G,H)$, and an optimal solution $d(\cdot,\cdot)$ can be computed in polynomial time, and the second inequality above has a constructive proof, from which we derive a polynomial time $O(\log|V|)$-approximate algorithm for sparsest cut.

The value $LR(G,H)$ and an optimal $d(\cdot,\cdot)$ can be computed in polynomial time by solving the following linear program

$$\begin{array}{ll} \text{minimize} & \sum_{\{u,v\}} G_{u,v} d_{u,v} \\ \text{subject to} & \\ & \sum_{\{u,v\}} H_{u,v} d_{u,v} = \frac{|E_H|}{|E_G|} \\ & d_{u,w} \leq d_{u,w} + d_{w,v} \qquad \forall u,v,w \in V \\ & d_{u,v} \geq 0 \qquad \forall u,v \in V \end{array} \tag{8.3}$$

that has a variable $d_{u,v}$ for every unordered pair of distinct vertices $\{u,v\}$. Clearly, every solution to the linear program (8.3) is also a solution to the right-hand side of the definition (8.1) of the Leighton-Rao parameter, with the same cost. Also every semi-metric can be normalized so that $\sum_{\{u,v\}} H_{u,v} d(u,v) = \frac{|E_H|}{|E_G|}$ by multiplying every distance by a fixed constant, and the normalization does not change the value of the right-hand side of (8.1); after the normalization, the semimetric is a feasible solution to the linear program (8.3), with the same cost.

In the rest of this lecture, we will show how to round a solution to (8.3) into a cut, achieving the logarithmic approximation promised in (8.2).

## 8.2 An L1 Relaxation of Sparsest Cut

In the Leighton-Rao relaxation, we relax distance functions of the form $d(u,v) = |x_u - x_v|$, where $\mathbf{x}$ is a boolean vector, to completely arbitrary distance functions. Let us consider an intermediate relaxation, in which we allow distance functions that can be realized by *an embedding of the vertices in an $\ell_1$ space.*

Recall that, for a vector $\mathbf{x} \in \mathbb{R}^n$, its $\ell_1$ norm is defined as $||\mathbf{x}||_1 := \sum_i |x_i|$, and that this norm makes $\mathbb{R}^n$ into a metric space with the $\ell_1$ distance function

$$||\mathbf{x} - \mathbf{y}||_1 = \sum_i |x_i - y_i|$$

The distance function $d(u,v) = |x_u - x_v|$ is an example of a distance function that can be realized by mapping each vertex to a real vector, and then defining the distance between two vertices as the $\ell_1$ norm of the respective vectors. Of course it is an extremely restrictive special case, in which the dimension of the vectors is one, and in which every vertex is actually mapping to either zero or one. Let us consider the relaxation of sparsest cut to arbitrary $\ell_1$ mappings, and define

$$\sigma'(G,H) := \inf_{m, f:V\to\mathbb{R}^m} \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} \cdot ||f(u) - f(v)||_1}{\sum_{\{u,v\}} H_{u,v} \cdot ||f(u) - f(v)||_1}$$

This may seem like another very broad relaxation of sparsest cut, whose optimum might bear no correlation with the sparsest cut optimum. The following theorem shows that this

is not the case.

**Theorem 8.1** *For every graphs $G, H$, $\sigma(G,H) = \sigma'(G,H)$.*

*Furthermore, there is a polynomial time algorithm that, given a mapping $f : V \to \mathbb{R}^m$, finds a cut $S$ such that*

$$\sigma_{G,H}(S) \leq \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} ||f(u) - f(v)||_1}{\sum_{u,v} H_{u,v} ||f(u) - f(v)||_1} \tag{8.4}$$

PROOF: We use ideas that have already come up in the proof the difficult direction of Cheeger's inequality. First, we note that for every nonnegative reals $a_1, \ldots, a_m$ and positive reals $b_1, \ldots, b_m$ we have

$$\frac{a_1 + \cdots a_m}{b_1 + \cdots b_m} \geq \min_i \frac{a_i}{b_i} \tag{8.5}$$

as can be seen by noting that

$$\sum_j a_j = \sum_j b_j \cdot \frac{a_j}{b_j} \geq \left( \min_i \frac{a_i}{b_i} \right) \cdot \sum_j b_j$$

Let $f_i(v)$ be the $i$-th coordinate of the vector $f(v)$, thus $f(v) = (f_1(v), \ldots, f_m(v))$. Then we can decompose the right-hand side of (8.4) by coordinates, and write

$$\begin{aligned}
&\frac{\sum_{\{u,v\}} G_{u,v} ||f(u) - f(v)||_1}{\sum_{\{u,v\}} H_{u,v} ||f(u) - f(v)||_1} \\
&= \frac{\sum_i \sum_{\{u,v\}} G_{u,v} |f_i(u) - f_i(v)|}{\sum_i \sum_{\{u,v\}} H_{u,v} |f_i(u) - f_i(v)|} \\
&\geq \min_i \frac{\sum_{\{u,v\}} G_{u,v} |f_i(u) - f_i(v)|}{\sum_{\{u,v\}} H_{u,v} |f_i(u) - f_i(v)|}
\end{aligned}$$

This already shows that, in the definition of $\phi'$, we can map, with no loss of generality, to 1-dimensional $\ell_1$ spaces.

Let $i^*$ be the coordinate that achieves the minimum above. Because the cost function is invariant under the shifts and scalings (that is, the cost of a function $x \to f(x)$ is the same as the cost of $x \to af(x) + b$ for every two constants $a \neq 0$ and $b$) there is a function $g : V \to \mathbb{R}$ such that $g$ has the same cost function as $f_{i*}$ and its range is such that

$$\max_v g(v) - \min_v g(v) = 1$$

Let us now pick a threshold $t$ uniformly at random from the interval $[\min_v g(v), \max_v g(v)]$, and define the random variables

$$S_t := \{v : g(v) \leq t\}$$

We observe that for every pairs of vertices $u, v$ we have

$$\mathbb{E}\,|1_{S_t}(u) - 1_{S_t}(v)| = |g(u) - g(v)|$$

and so we get

$$\begin{aligned}
&\frac{\sum_{\{u,v\}} G_{u,v}||f(u) - f(v)||_1}{\sum_{\{u,v\}} H_{u,v}||f(u) - f(v)||_1} \\
&\geq \frac{\sum_{\{u,v\}} G_{u,v}|g(u) - g(v)|}{\sum_{\{u,v\}} H_{u,v}|g(u) - g(v)|} \\
&= \frac{\mathbb{E}\sum_{\{u,v\}} G_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|}{\mathbb{E}\sum_{\{u,v\}} H_{u,v}|1_{S_t}(u) - 1_{S_t}(v)|}
\end{aligned}$$

Finally, by an application of (8.5), we see that there must be a set $S$ among the possible values of $S_t$ such that (8.4) holds.

Notice that the proof was completely constructive: we simply took the coordinate $f_{i^*}$ of $f$ with the lowest cost function, and then the "threshold cut" given by $f_{i^*}$ with the smallest sparsity. □

## 8.3 A Theorem of Bourgain

We will derive our main result (8.2) from the L1 "rounding" process of the previous section, and from the following theorem of Bourgain (the efficiency considerations are due to Linial, London and Rabinovich).

**Theorem 8.2 (Bourgain)** *Let $d : V \times V \to \mathbb{R}$ be a semimetric defined over a finite set $V$. Then there exists a mapping $f : V \to \mathbb{R}^m$ such that, for every two elements $u, v \in R$,*

$$||f(u) - f(v)||_1 \leq d(u, v) \leq ||f(u) - f(v)||_1 \cdot c \cdot \log |V|$$

*where $c$ is an absolute constant. Given $d$, the mapping $f$ can be found with high probability in randomized polynomial time in $|V|$.*

To see that the above theorem of Bourgain implies (8.2), consider graphs $G, H$, and let $d$ be the optimal solution of the Leighton-Rao relaxation of the sparsest cut problem on $G, H$, and let $f : V \to \mathbb{R}$ be a mapping as in Bourgain's theorem applied to $d$. Then

$$LR(G, H) = \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u, v)}{\sum_{\{u,v\}} H_{u,v} d(u, v)}$$

$$\geq \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v}||f(u) - f(v)||_1}{c \cdot \log|V| \cdot \sum_{\{u,v\}} H_{u,v}||f(u) - f(v)||_1}$$

$$\geq \frac{1}{c \cdot \log|V|} \cdot \sigma(G, H)$$

The theorem has a rather short proof, but there is an element of "magic" to it. We will discuss several examples and we will see what approaches are suggested by the examples. At the end of the discussion, we will see the final proof as, hopefully, the "natural" outcome of the study of such examples and failed attempts.

### 8.3.1 Preliminaries and Motivating Examples

A first observation is that embeddings of finite sets of points into L1 can be equivalently characterized as probabilistic embeddings into the real line.

**Fact 8.3** *For every finite set $V$, dimension $m$, and mapping $F : V \to \mathbb{R}^m$, there is a finitely-supported probability distribution $D$ over functions $f : V \to \mathbb{R}$ such that for every two points $u, v \in V$:*

$$\mathop{\mathbb{E}}_{f \sim D} |f(u) - f(v)| = ||F(u) - F(v)||_1$$

*Conversely, for every finite set $V$ and finitely supported distribution $D$ over functions $f : V \to \mathbb{R}$, there is a dimension $m$ and a mapping $F : V \to \mathbb{R}^m$ such that*

$$\mathop{\mathbb{E}}_{f \sim D} |f(u) - f(v)| = ||F(u) - F(v)||_1$$

Proof: For the first claim, we write $F_i(v)$ for the $i$-th coordinate of $F(v)$, that is $F(v) = (F_1(v), \ldots, F_m(v))$, and we define $D$ to be the uniform distribution over the $m$ functions of the form $x \to m \cdot F_i(x)$.

For the second claim, if the support of $D$ is the set of functions $\{f_1, \ldots, f_m\}$, where function $f_i$ has probability $p_i$, then we define $F(v) := (p_1 f_1(v), \ldots, p_m f_m(v))$. □

It will be easier to reason about probabilistic mappings into the line, so we will switch to the latter setting from now on.

Our task is to associate a number to every point $v$, and the information that we have about $v$ is the list of distances $\{d(u, v)\}$. Probably the first idea that comes to mind is to pick a random reference vertex $r \in V$, and work with the mapping $v \to d(r, v)$, possibly scaled by a multiplicative constant. (Equivalently, we can think about the deterministic mapping $V \to \mathbb{R}^{|V|}$, in which the vertex $v$ is mapped to the sequence $(d(u_1, v), \ldots, d(u_n, v)$, for some enumeration $u_1, \ldots, u_n$ of the elements of $V$.)

This works in certain simple cases.

**Example 8.4 (Cycle)** *Suppose that $d(\cdot,\cdot)$ is the shortest-path metric on a cycle, we can see that, for every two points on the cycle, $\mathbb{E}_{r\in V}|d(r,u)-d(r,v)|$ is within a constant factor of their distance $d(u,v)$. (Try proving it rigorously!)*

**Example 8.5 (Simplex)** *Suppose that $d(u,v)=1$ for every $u\neq v$, and $d(u,u)=0$. Then, for every $u\neq v$, we have $\mathbb{E}_{r\in V}|d(r,u)-d(r,v)|=\mathbb{P}[r=u\vee r=v]=2/n$, so, up to scaling, the mapping incurs no error at all.*

But there are also simple examples in which this works very badly.

**Example 8.6 (1-2 Metric)** *Suppose that for every $u\neq v$ we have $d(u,v)\in\{1,2\}$ (any distance function satisfying this property is always a metric) and that, in particular, there is a special vertex $z$ at distance 2 from all other vertices, while all other vertices are at distance 1 from each other. Then, for vertices $u,v$ both different from $z$ we have, as before*

$$\mathbb{E}[|d(r,u)-d(r,v)|]=\frac{2}{n}$$

*but for every $v$ different from $z$ we have*

$$\mathbb{E}[|d(r,z)-d(r,v)|]=\frac{n-2}{n}\cdot|2-1|+\frac{1}{n}\cdot|2-0|+\frac{1}{n}\cdot|0-2|=1+\frac{2}{n}$$

*and so our error is going to be $\Omega(n)$ instead of the $O(\log n)$ that we are trying to establish.*

Maybe the next simplest idea is that we should pick at random several reference points $r_1,\ldots,r_k$. But how do we combine the information $d(r_1,u),\ldots,d(r_k,u)$ into a single number to associate to $u$? If we just take the sum of the distances, we are back to the case of sampling a single reference point. (We are just scaling up the expectation by a factor of $k$.)

The next simplest way to combine the information is to take either the maximum or the minimum. If we take the minimum, we see that we have the very nice property that we immediately guarantee that our distances in the L1 embedding are no bigger than the original distances, so that it "only" remains to prove that the distances don't get compressed too much.

**Fact 8.7** *Let $d:V\times V\to\mathbb{R}$ be a semimetric and $A\subseteq V$ be a non-empty subset of points. Define $f_A:V\to\mathbb{R}$ as*

$$f_A(v):=\min_{r\in A}d(r,v)$$

*Then, for every two points $u,v$ we have*

$$|f_A(u)-f_A(v)|\leq d(u,v)$$

PROOF: Let $a$ be the point such that $d(a,u) = f_A(u)$ and $b$ be the point such that $d(b,v) = f_A(v)$. (It's possible that $a = b$.) Then

$$f_A(u) = d(a,u) \geq d(v,a) - d(u,v) \geq d(v,b) - d(u,v) = f_A(v) - d(u,v)$$

and, similarly,

$$f_A(v) = d(b,v) \geq d(u,b) - d(u,v) \geq d(u,a) - d(u,v) = f_A(u) - d(u,v)$$

□

Is there a way to sample a set $A = \{r_1, \ldots, r_k\}$ such that, for every two points $u, v$, the expectation $\mathbb{E}\,|f_A(u) - f_A(v)|$ is not too much smaller than $d(u,v)$? How large should the set $A$ be?

**Example 8.8 (1-2 Metric Again)** *Suppose that for every $u \neq v$ we have $d(u,v) \in \{1,2\}$, and that we pick a subset $A \subseteq V$ uniformly at random, that is, each event $r \in A$ has probability $1/2$ and the events are mutually independent.*

*Then for every $u \neq v$:*

$$\frac{1}{4} \cdot d(u,v) \leq |\,\mathbb{E}\,|f_A(u) - f_A(v)| \leq d(u,v)$$

*because with probability $1/2$ the set $A$ contains exactly one of the elements $u, v$, and conditioned on that event we have $|f_A(u) - f_A(v)| \geq 1$ (because one of $f_A(u), f_A(v)$ is zero and the other is at least one), which is at least $d(u,v)/2$.*

If we pick $A$ uniformly at random, however, we incur an $\Omega(n)$ distortion in the case of the shortest path metric on the cycle. In all the examples seen so far, we can achieve constant distortion if we "mix" the distribution in which $A$ is a random set of size 1 and the one in which $A$ is a chosen uniformly at random among all sets, say by sampling from the former probability with probability $1/2$ and from the latter with probability $1/2$.

**Example 8.9 (Far-Away Clusters)** *Suppose now that $d(\cdot,\cdot)$ has the following structure: $V$ is partitioned into clusters $B_1, \ldots, B_k$, where $|B_i| = i$ (so $k \approx \sqrt{2n}$), and we have $d(u,v) = 1$ for vertices in the same cluster, and $d(u,v) = n$ for vertices in different clusters.*

*If $u, v$ are in the same cluster, then $d(u,v) = 1$ and*

$$\mathbb{E}\,|f_A(u) - f_A(v)| = \mathbb{P}[A \text{ contains exactly one of } u, v]$$

*If $u, v$ are in different clusters $B_i, B_j$, then $d(u,v) = n$ and*

$$\mathbb{E}\,|f_A(u) - f_A(v)| \approx n\,\mathbb{P}[A \text{ intersects exactly one of } B_i, B_j]$$

If we want to stick to this approach of picking a set $A$ of reference elements according to a certain distribution, and then defining the map $f_A(v) := \min_{r \in A} d(r, v)$, then the set $A$ must have the property that for every two sets $S, T$, there is at least a probability $p$ that $A$ intersects exactly one of $S, T$, and we would like $p$ to be as large as possible, because the distortion caused by the mapping will be at least $1/p$.

This suggest the following distribution $D$:

1. Sample a power of two $t$ uniformly at random in $\{1, 2, 4, \ldots, 2^{\lfloor \log_2 n \rfloor}\}$
2. Sample $A \subseteq V$ by selecting each $v \in V$, independently, to be in $A$ with probability $1/t$ and to be in $V - A$ with probability $1 - 1/t$.

This distribution guarantees the above property with $p = 1/O(\log n)$.

Indeed, the above distribution guarantees a distortion at most $O(\log n)$ in all the examples encountered so far, including the tricky example of the clusters of different size. In each example, in fact, we can prove the following claim: for every two vertices $u, v$, there is a scale $t$, such that conditioned on that scale being chosen, the expectation of $|f_A(u), f_A(v)|$ is at least a constant times $d(u, v)$. We could try to prove Bourgain's theorem by showing that this is true in every semimetric.

Let us call $D_t$ the conditional distribution of $D$ conditioned on the choice of a scale $t$. We would like to prove that for every semimetric $d(\cdot, \cdot)$ and every two points $u, v$ there is a scale $t$ such that

$$\mathop{\mathbb{E}}_{A \sim D_t} |f_A(u) - f_A(v)| \geq \Omega(d(u, v))$$

which, recalling that $|f_A(u) - f_A(v)| \leq d(u, v)$ for every set $A$, is equivalent to arguing that

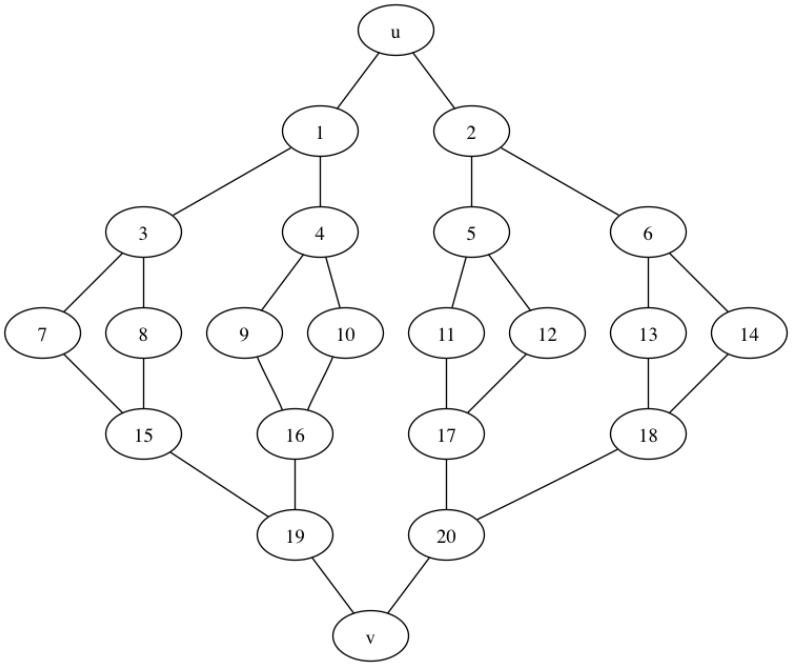
$$\mathop{\mathbb{P}}_{A \sim D_t} [|f_A(u) - f_A(v)| \geq \Omega(d(u, v))] \geq \Omega(1)$$

For this to be true, there must be distances $d_1, d_2$ such that $d_1 - d_2 \geq \Omega(d(u, v))$ and such that, with constant probability according to $D_t$, we have $f_A(u) \geq d_1$ and $f_A(v) \leq d_2$ (or vice-versa). For this to happen, there must be a constant probability that $A$ avoids the set $\{r : d(u, r) < d_1\}$ and intersects the set $\{r : d(v, r) \leq d_2\}$. For this to happen, both sets must have size $\approx t$.

This means that if we want to make this "at least one good scale for every pair of points" argument work, we need to show that for every two vertices $u, v$ there is a "large" distance $d_1$ and a "small" distance $d_2$ (whose difference is a constant times $d(u, v)$) such that a large-radius ball around one of the vertices and a small-radius ball around the other vertex contain roughly the same number of elements of $V$.

Consider, however, the following example.

**Example 8.10 (Joined Trees)** *Consider the graph obtained by taking two complete binary trees of the same size and identifying their leaves, as in the picture below.*




*Consider the shortest-path metric $d(\cdot,\cdot)$ in the above graph. Consider the "root" vertices $u$ and $v$. Their distance $d(u,v)$ is $\approx \log n$, but, at every scale $t$, both $f_A(u)$ and $f_A(v)$ are highly concentrated around $t$ and, it can be calculated that, at every scale $t$, we have*

$$\mathop{\mathbb{E}}_{A\sim D_t}[|f_A(u) - f_A(v)|] = \Theta(1)$$

*This is still good, because averaging over all scales we still get*

$$\mathop{\mathbb{E}}_{A\sim D}[|f_A(u) - f_A(v)|] \geq \Omega(1) = \frac{1}{O(\log n)} \cdot d(u,v)$$

*but this example shows that the analysis cannot be restricted to one good scale but has, in some cases, to take into account the contribution to the expectation coming from all the scales.*

In the above example, the only way to get a ball around $u$ and a ball around $v$ with approximately the same number of points is to get balls of roughly the same radius. No scale could then give a large contribution to the expectation $\mathbb{E}_{A\sim D}[|f_A(u) - f_A(v)|]$; every scale, however, gave a noticeable contribution, and adding them up we had a bounded distortion. The above example will be the template for the full proof, which will do an "amortized analysis" of the contribution to the expectation coming from each scale $t$, by looking at the radii that define a ball around $u$ and a ball around $v$ with approximately $t$ elements.

### 8.3.2 The Proof of Bourgain's Theorem

Given Fact 8.3 and Fact 8.7, proving Bourgain's theorem reduces to proving the following theorem.

**Theorem 8.11** *For a finite set of points $V$, consider the distribution $D$ over subsets of $V$ sampled by uniformly picking a scale $t \in \{1, \ldots, 2^{\lfloor \log_2 |V| \rfloor}\}$ and then picking independently each $v \in V$ to be in $A$ with probability $1/t$. Let $d : V \times V \to \mathbb{R}$ be a semimetric. Then for every $u, v \in V$,*

$$\mathop{\mathbb{E}}_{A \sim D}[|f_A(u) - f_A(v)|] \geq \frac{1}{c \log_2 |V|} \cdot d(u, v)$$

*where $c$ is an absolute constant.*

Proof: Fix two vertices $u$ and $v$

For each $t$, let $ru_t$ be the distance from $u$ to the $t$-th closest point to $u$ (counting $u$), or $d(u, v)/3$, whichever is smaller, and define $rv_t$ similarly. By definition, we have

$$|\{w : d(u, w) < ru_t\}| < t$$

Call $t^*$ the minimum of $|\{w : d(u, w) < d(u, v)/3\}|$ and $|\{w : d(v, w) < d(u, v)/3\}|$. Then, for $t \leq t^*$ we have that both $ru_t$ and $rv_t$ are $< d(u, v)/3$, but for $t \geq t^* + 1$ we have that at least one of $ru_t$ or $rv_t$ (possibly, both) equals $d(u, v)/3$. Note also that for $t \leq t^*$ we have

$$|\{w : d(u, w) \leq ru_t\}| \geq t$$

and similarly for $v$.

We claim that there is an absolute constant $c$ such that for every scale $t \leq t^*$, we have

$$\mathop{\mathbb{E}}_{A \sim D_t} |f_A(u) - f_A(v)| \geq c \cdot (ru_{2t} + rv_{2t} - ru_t - rv_t) \tag{8.6}$$

We prove the claim by showing that there are two disjoint events, each happening with probability $\geq c$, such that in one event $|f_A(u) - f_A(v)| \geq ru_{2t} - rv_t$, and in the other event $|f_A(u) - f_A(v)| \geq rv_{t2t} - ru_t$.

1. The first event is that $A$ avoids the set $\{z : d(u, z) < ru_{2t}\}$ and intersects the set $\{z : d(v, z) \leq rv_t\}$. The former set has size $< 2t$, and the latter set has size $\leq t$; the sets are disjoint because we are looking at balls or radius $\leq d(u, v)/3$ around $u$ and $v$; so the event happens with a probability that is at least an absolute constant. When the event happens,

$$|f_A(u) - f_A(v)| \geq f_A(u) - f_A(v) \geq ru_{2t} - rv_t$$

2. The second event is that $A$ avoids the set $\{z : d(v,z) < rv_{2t}\}$ and intersects the set $\{z : d(u,z) \leq ru_t\}$. The former set has size $< 2t$, and the latter set has size $\leq t$; the sets are disjoint because we are looking at balls or radius $\leq d(u,v)/3$ around $u$ and $v$; so the event happens with a probability that is at least an absolute constant. When the event happens,

$$|f_A(u) - f_A(v)| \geq f_A(v) - f_A(u) \geq rv_{2t} - ru_t$$

So we have established (8.6). Summing over all scales up to the largest power of two $t' \leq t^*$, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{A \sim D} |f_A(u) - f_A(v)| \\ &\geq \frac{c}{1 + \log_2 n} \cdot (ru_{2t'} + rv_{2t'} - ru_1 - rv_1) \\ &\geq \frac{c}{1 + \log_2 n} \cdot \frac{d(u,v)}{3} \end{aligned}$$

□

There is one remaining point to address. In Fact 8.3, we proved that a distribution over embeddings on the line can be turned into an L1 embeddings, in which the number of dimensions is equal to the size of the support of the distribution. In our proof, we have used a distribution that ranges over $2^{|V|}$ possible functions, so this would give rise to an embedding that uses a superpolynomial number of dimensions.

To fix this remaining problem, we sample $m = O(\log^3 |V|)$ sets $A_1, \ldots, A_m$ and we define the embedding $f(u) := (m^{-1} \cdot f_{A_1}(u), \ldots, m^{-1} \cdot f_{A_m}(u))$. It remains to prove that this randomized mapping has low distortion with high probability, which is an immediate consequence of the Chernoff bounds. Specifically, we use the following form of the Chernoff bound:

**Lemma 8.12** *Let $Z_1, \ldots, Z_m$ be independent nonnegative random variables such that, with probability 1, $0 \leq Z_i \leq M$. Let $Z := \frac{1}{m}(Z_1 + \cdots + Z_m)$. Then*

$$\mathbb{P}[\mathbb{E}\, Z - Z \geq t] \leq e^{-2mt^2/M^2}$$

Let us look at any two vertices $u, v$. Clearly, for every choice of $A_1, \ldots, A_m$, we have $||f(u) - f(v)||_1 \leq d(u,v)$ so it remains to prove a lower bound to their L1 distance. Let us call $Z$ the random variable denoting their L1 distance, that is

$$Z := ||f(u) - f(v)|| = \sum_{i=1}^{m} \frac{1}{m} |f_{A_i}(u) - f_{A_i}(v)|$$

We can write $Z = \frac{1}{m} \cdot (Z_1 + \cdots + Z_m)$ where $Z_i := |f_{A_i}(u) - f_{A_i}(v)|$, so that $Z$ is the sum of identically distributed nonnegative random variables, such that

$$Z_i \leq d(u,v)$$
$$\mathbb{E}\, Z_i \geq \frac{c}{\log |V|} d(u,v)$$

Applying the Chernoff bound with $M = d(u,v)$ and $t = \frac{c}{2\log|V|} d(u,v)$, we have

$$\begin{aligned}
&\mathbb{P}\left[Z \leq \frac{c}{2\log|V|} d(u,v)\right] \\
&\leq \mathbb{P}\left[Z \leq \mathbb{E}\, Z - \frac{c}{2\log|V|} d(u,v)\right] \\
&\leq 2^{-2mc^2/(2\log|V|)^2}
\end{aligned}$$

which is, say, $\leq 1/|V|^3$ if we choose $m = c' \log^3 |V|$ for an absolute constant $c'$.

By taking a union bound over all pairs of vertices,

$$\mathbb{P}\left[\forall u, v.\ ||f(u) - f(v)||_1 \geq \frac{c}{2\log|V|} \cdot d(u,v)\right] \geq 1 - \frac{1}{|V|}$$

## 8.4 Tightness of the Analysis of the Leighton-Rao Relaxation

If $(X,d)$ and $(X',d')$ are metric spaces, we say that a mapping $f : X \to X'$ is an *embedding of* $(X,d)$ *into* $(X',d)$ *with distortion at most* $c$ if there are parameters $c_1, c_2$, with $c = c_1 c_2$ such that, for every $u, v \in X$, we have

$$\frac{1}{c_1} \cdot d'(u,v) \leq d(u,v) \leq c_2 \cdot d'(u,v)$$

The metric space $\mathbb{R}^m$ with distance $||u - v|| = \sqrt{\sum_i (u_i - v_i)^2}$ is denoted by $\ell_m^2$, and the metric space $\mathbb{R}^m$ with distance $||u - v||_1 = \sum_i |u_i - v_i|$ is denoted by $\ell_m^1$. We just proved the following result.

**Theorem 8.13 (Bourgain)** *There is an absolute constant $c$ such that every finite metric space $(V,d)$ embeds into $\ell_m^1$ with distortion at most $c \log |V|$, where $m = O(\log^3 |V|)$.*

If we solve the Leighton-Rao linear programming relaxation to approximate the sparsest cut of a graph $G = (V,E)$, and we let $d(\cdot,\cdot)$ be an optimal solution, we note that, if we weigh each edge $(u,v) \in E$ by $d(u,v)$, and then compute shortest paths in this weighted graph, then, for every two vertices $x, y$, the distance $d(x,y)$ is precisely the length of the shortest path from $x$ to $y$. In particular, if we are using the Leighton-Rao relaxation in order to

approximate the sparsest cut in a given *planar* graph, for example, then the solution $d(\cdot,\cdot)$ that we need to round is not an arbitrary metric space, but it is the shortest path metric of a weighted planar graph. It is conjectured that, in this case, the Leighton-Rao relaxation could deliver a constant-factor approximation.

**Question 8.1** *Is there an absolute constant $c$ such that every metric space $(X,d)$ constructed as the shortest-path metric over the vertices of a planar graph can be embedded into $\ell_m^1$ with distortion at most $c$, where $m = |V|^{O(1)}$?*

So far, it is known that *$k$-outerplanar* graphs, for constant $k$, embed in $\ell^1$ with constant distortion.

This is just an example of a large family of questions that can be asked about the embeddability of various types of metric spaces into each other.

For general finite metric spaces, the logarithmic distortion of Bougain's theorem is best possible.

In order to prove the optimality of Bourgain's theorem, we recall a theorem that we proved earlier.

**Theorem 8.14 (Existence of Expanders)** *There are absolute constants $d$ and $c$ such that, for infinitely many $n$, there is an $n$-vertex $d$-regular graph $G_n$ such that $\phi(G_n) \geq c$.*

On such graphs, the Leighton-Rao relaxation is $LR(G_n) \leq O(1/\log n)$, showing that our proof that $LR(G) \geq \phi(G)O(\log n)$ is tight.

For every two vertices $u, v$, define $d(u,v)$ as the length of (that is, the number of edges in) a shortest path from $u$ to $v$ in $G_n$.

Then

$$\sum_{u,v} A_{u,v} d(u,v) = 2|E|$$

Because each graph $G_n$ is $d$-regular, it follows that for every vertex $v$ there are $\leq 1 + d + \cdots + d^k < d^{k+1}$ vertices at distance $\leq k$ from $v$. In particular, at least half of the vertices have distance $\geq t$ from $v$, where $t = \lfloor \log_d n/2 \rfloor - 1$, which implies that

$$\sum_{u,v} d(u,v) \geq n \cdot \frac{n}{2} \cdot t = \Omega(n^2 \log n)$$

Recall that

$$LR(G) = \min_{d \text{ semimetric}} \frac{|V|^2}{2|E|} \frac{\sum_{u,v} A_{u,v} d(u,v)}{\sum_{u,v} d(u,v)}$$

and so

$$LR(G_n) \leq O\left(\frac{1}{\log n}\right)$$

even though

$$\phi(G_n) \geq \Omega(1)$$

Note that we have also shown that every embedding of the shortest-path metric $d(\cdot,\cdot)$ on $G_n$ into $\ell^1$ requires distortion $\Omega(\log n)$, and so we have proved the tightness of Bourgain's theorem.

## Exercises

1. Let $G = (V, E_G), H = (V, E_H)$ be an instance of the non-uniform sparsest cut problem. Let $d(u, v)$ a feasible solution to the Leighton-Rao relaxation

$$LR(G, H) := \min_{\substack{d : V \times V \to \mathbb{R} \\ d \text{ semi-metric}}} \frac{|E_H|}{|E_G|} \cdot \frac{\sum_{\{u,v\}} G_{u,v} d(u, v)}{\sum_{\{u,v\}} H_{u,v} d(u, v)}$$

   Let $d'(u, v)$ be the length of the shortest path from $u$ to $v$ in the graph that has the edges of $G$, and each edge $(u, v) \in E_G$ is weighted by $d(u, v)$. Show that $d'(u, v)$ is a feasible solution whose cost is smaller than or equal to the cost $d(u, v)$.

2. Using the above fact, show that if $G$ is a cycle and $H$ is a clique, then the solution in which $d(u, v)$ is the length of the shortest path from $u$ to $v$ in $G$ is an optimal solution.

   [Hint: start from an optimal solution, derive from it another solution of the same cost in which $d(u, v)$ is the same for every $u, v$ that are adjacent in $G$, then apply the fact proved in the previous exercise.]