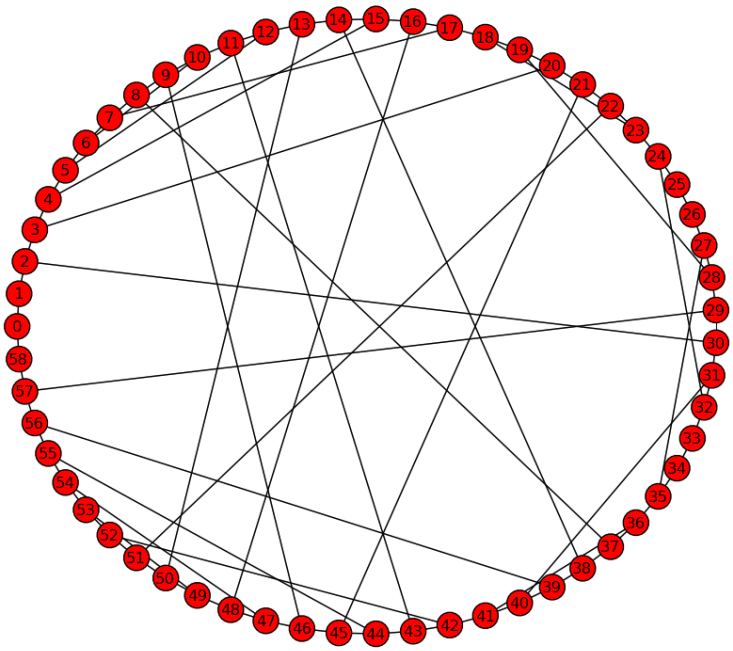
Lecture Notes on Graph Partitioning, Expanders and Spectral Methods



Luca Trevisan

University of California, Berkeley

Foreword

These notes are a lightly edited revision of notes written for the course "Graph Partitioning, Expanders and Spectral Methods" offered at offered at U.C. Berkeley in Spring 2016.

This material is based upon work supported by the National Science Foundation under Grants No. 1216642, 1540685 and 1655215, 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, June 2017.

Luca Trevisan

Contents

Foreword i

1 Linear Algebra Background **1**

1.1 Basic Definitions 1

1.2 The Spectral Theorem 2

1.3 Variational Characterization of Eigenvalues 5

2 Introduction **7**

2.1 Overview 7

2.2 Spectral Graph Algorithms 8

2.3 Constructions and Applications of Expander Graphs 11

2.4 Mixing time of random walks 12

2.5 Linear Systems, Electrical Flows, and Applications 13

3 The Basics of Spectral Graph Theory **15**

3.1 The Laplacian Matrix 15

3.2 Some Facts About Laplacian Eigenvalues 16

4 Cheeger's Inequalities **19**

4.1 Irregular Graphs 19

4.2 Edge Expansion, Fiedler's Algorithm and Cheeger's Inequalities 21

4.3 Proof that $\frac{\lambda_2}{2} \leq \phi(G)$ 22

4.4 First Part of the Analysis of Fiedler's Algorithm 23

5 Cheeger's Inequalities cont'd **26**

6 Cheeger-type Inequalities for λ_n **28**
6.1 Cheeger-type Inequalities for λ_n . . . 28
6.1.1 Proof of Main Lemma . . . 30
6.2 Application to Max Cut . . . 31

7 Cheeger-type Inequalities for λ_k **34**
7.1 Cheeger-type Inequalities for λ_k . . . 34
7.2 The Easy Direction . . . 35
7.3 The Difficult Direction: Main Lemma . . . 36
7.4 The Spectral Embedding . . . 37
7.5 Overview of the Proof of the Main Lemma . . . 38

8 Cheeger-type Inequalities for λ_k, cont'd **40**
8.1 Review . . . 40
8.2 Some Preliminaries . . . 41
8.3 Localization . . . 42
8.3.1 Proof Ideas . . . 42
8.3.2 Proof . . . 43
8.4 Well-Separated Sets . . . 44

9 Spectral Algorithms Wrap-up **47**
9.1 Irregular Graphs . . . 47
9.2 More Cheeger-type Bounds . . . 48
9.3 The Power Method . . . 49
9.3.1 The Power Method to Approximate the Largest Eigenvalue . . . 50
9.3.2 Approximating the Second Largest Eigenvalue . . . 53
9.3.3 The Second Smallest Eigenvalue of the Laplacian . . . 54

10 The Sparsest Cut Problem **57**
10.1 The Uniform Sparsest Cut problem, Edge Expansion and λ_2 . . . 57
10.2 The Non-Uniform Sparsest Cut Problem . . . 59
10.3 The Leighton-Rao Relaxation . . . 59
10.4 An L1 Relaxation of Sparsest Cut . . . 60
10.5 A Theorem of Bourgain . . . 62

11 Proof of Bourgain's Theorem **64**
11.1 Preliminary and Motivating Examples 64
11.2 The Proof of Bourgain's Theorem 69

12 ARV **72**
12.1 The Goemans-Linial Relaxation 72
12.2 Polynomial Time Solvability 75
12.3 Rounding when H is a clique 78

13 ARV Analysis **83**
13.1 Bottlenecks 84
13.2 Gaussian Projections 86
13.3 The Algorithm to Refine L and R 87

14 ARV Analysis cont'd **89**
14.1 Concentration of Measure 89
14.2 Reworking the Assumption 91

15 ARV Analysis, part 3 **93**
15.1 An Inductive Proof that Gives a Weaker Result 93
15.2 The Tight Bound 95

16 Cayley Graphs of Abelian Groups **99**
16.1 Characters 100
16.2 A Look Beyond 104
16.3 Cayley Graphs and Their Spectrum 106
16.4 The Cycle 107
16.5 The Hypercube 108

17 Constructions of Expanders via the Zig-Zag Graph Product **110**
17.1 Expanders of Logarithmic Degree 112
17.2 The Zig-Zag Graph Product 113

18 Analysis of the zig-zag graph product **115**
18.1 Replacement Product and Zig-Zag Product 115
18.2 Zig-zag product of two graphs 116

18.3 A Technical Preliminary . . . 117
18.4 Analysis of the zig-zag Product . . . 118

19 Algebraic construction of expanders **121**
19.1 The Marguli-Gabber-Galil Expanders . . . 121
19.2 First Step: The Continuous Graph R_n . . . 122
19.3 Second Step: The Countable Graph . . . 124
19.4 Third Step: A Cheeger Inequality for countable graphs . . . 126
19.5 Expansion of Z . . . 127

20 Probabilistic construction of expanders **129**

21 Properties of Expanders **132**
21.1 Quasirandomness of Expander Graphs . . . 132
21.2 Random Walks in Expanders . . . 133

22 Approximate Counting **137**
22.1 Counting vs Sampling . . . 137
22.2 Approximate Sampling . . . 137

Chapter 1

Linear Algebra Background

In which we review linear algebra prerequisites.

The following background from linear algebra will be sufficient for the sake of this course: to know what is an eigenvalue and an eigenvector, to know that real symmetric matrices have real eigenvalues and their real eigenvectors are orthogonal, and to know the variational characterization of eigenvalues.

1.1 Basic Definitions

If $x = a + ib$ is a complex number, then we let $\bar{x} = a - ib$ denote its *conjugate*. Note that a complex number x is real if and only if $x = \bar{x}$. 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 the entries of M are real, then $M^* = M^T$, where M^T is the *transpose* of M, that is, the matrix such that $(M^T)_{i,j} = M_{j,i}$.

We say that a matrix M is *Hermitian* if $M = M^*$. In particular, real symmetric matrices are Hermitian.

If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ are two vectors, then their inner product is defined as

$$\langle \mathbf{v}, \mathbf{w} \rangle := \mathbf{v}^* \mathbf{w} = \sum_i \overline{v_i} \cdot w_i \tag{1.1}$$

Notice that, by definition, we have $\langle \mathbf{v}, \mathbf{w} \rangle = (\langle \mathbf{w}, \mathbf{v} \rangle)^*$ and $\langle \mathbf{v}, \mathbf{v} \rangle = ||\mathbf{v}||^2$. Note also that, for two matrices A, B, we have $(A \cdot B)^* = B^* \cdot A^*$, and that for every matrix M and every two vectors $\mathbf{x}$, $\mathbf{y}$, we have

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

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

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

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

1.2 The Spectral Theorem

We want to prove

Theorem 1.1 (Spectral Theorem) *Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix with real-valued entries, then there are n real numbers (not necessarily distinct) $\lambda_1, \ldots, \lambda_n$ and n orthonormal real vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^n$ such that $\mathbf{x}_i$ is an eigenvector of λ_i.*

Assuming the fundamental theorem of algebra (that every polynomial has a complex root) and basic properties of the determinant, the cleanest proof of the spectral theorem is to proceed by induction on n, and to show that M must have a real eigenvalue λ_1 with a real eigenvector $\mathbf{v}_1$, and to show that M maps vectors orthogonal to $\mathbf{v}_1$ to vectors orthogonal to $\mathbf{v}_1$. Then one applies the inductive hypothesis to M restricted to the $(n-1)$-dimensional space of vectors orthogonal to $\mathbf{v}_1$ and one recovers the remaining $(n-1)$ eigenvalues and eigenvectors.

The cleanest way to formalize the above proof is to give all definitions and results in terms of linear operators $T : V \to V$ where V is an arbitrary vector space over the reals. This way, however, we would be giving several definitions that we would never use in the future, so, instead, the inductive proof will use a somewhat inelegant change of basis to pass from M to an $(n-1) \times (n-1)$ matrix M'.

We begin by showing that a real symmetric matrix has real eigenvalues and eigenvectors.

Theorem 1.2 *If $M \in \mathbb{R}^{n \times n}$ is symmetric, then there is a real eigenvalue $\lambda \in \mathbb{R}$ and a real eigenvector $\mathbf{v} \in \mathbb{R}^n$ such that $M\mathbf{v} = \lambda\mathbf{v}$.*

We begin by noting that every matrix has a complex eigenvalue.

Lemma 1.3 *For every matrix $M \in \mathbb{C}^{n \times n}$, there is an eigenvalue $\lambda \in \mathbb{C}$ and an eigenvector $\mathbf{v} \in \mathbb{C}^n$ such that $M\mathbf{v} = \lambda\mathbf{v}$.*

PROOF: Note that λ is an eigenvalue for M if and only if

$$\exists \mathbf{x} \neq \mathbf{0}.\ (M - \lambda I)\mathbf{x} = \mathbf{0}$$

which is true if and only if the rows of $M - \lambda I$ are not linearly independent, which is true if and only if

$$\det(M - \lambda I) = 0$$

Now note that the mapping $t \to \det(M - tI)$ is a univariate polynomial of degree n in t, and so it must have a root λ by the fundamental theorem of algebra. □

Next we show that if M is real and symmetric, then its eigenvalues are real.

Lemma 1.4 *If M is Hermitian, then, for every $\mathbf{x}$ and $\mathbf{y}$,*

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

PROOF:

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

□

Lemma 1.5 *If M is Hermitian, then all the eigenvalues of M are real.*

PROOF: Let M be an Hermitian matrix and let λ 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 λ is a real number.

We note that

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

and

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

and by the fact that $\langle M\mathbf{x}, \mathbf{x}\rangle = \langle \mathbf{x}, M\mathbf{x}\rangle$, we have $\lambda = \lambda^*$. □

In order to prove Theorem 1.2, it remains to argue that, for a real eigenvalue of a real symmetric matrix, we can find a real eigenvector.

PROOF:[Of Theorem 1.2] Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, then M has a real eigenvalue λ and a (possibly complex valued) eigenvector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are real vectors. We have

$$M\mathbf{x} + iM\mathbf{y} = \lambda\mathbf{x} + i\lambda\mathbf{y}$$

from which (recalling that the entries of M and the scalar λ are real) it follows that $M\mathbf{x} = \lambda\mathbf{x}$ and that $M\mathbf{y} = \lambda\mathbf{y}$; since $\mathbf{x}$ and $\mathbf{y}$ cannot both be zero, it follows that λ has a real eigenvector. □

We are now ready to prove the spectral theorem

PROOF:[Of Spectral Theorem] We proceed by induction on n. The case $n = 1$ is trivial.

Assume that the statement is true for dimension $n-1$. Let λ_1 be a real eigenvalue of M and $\mathbf{x}_1$ be a real eigenvector λ_1.

Now we claim that for every vector $\mathbf{y}$ that is orthogonal to $\mathbf{x}_1$, then $M\mathbf{y}$ is also orthogonal to $\mathbf{x}_1$. Indeed, the inner product of $M\mathbf{y}$ and $\mathbf{x}_1$ is

$$\langle \mathbf{x}_1, M\mathbf{y} \rangle = \langle M\mathbf{x}_1, \mathbf{y} \rangle = \langle \lambda \mathbf{x}_1, \mathbf{y} \rangle = 0$$

Let V be the $n-1$-dimensional subspace of $\mathbb{R}^n$ that contains all the vectors orthogonal to $\mathbf{x}_1$. We want to apply the inductive hypothesis to M restricted to V; we cannot literally do that, because the theorem is not stated in terms of arbitrary linear operators over vector spaces, so we will need to do that by fixing an appropriate basis for V.

let $B \in \mathbb{R}^{n \times (n-1)}$ be a matrix that computes a bijective map from $\mathbb{R}^{n-1}$ to V. (If $\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ is an orthonormal 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 $\mathbf{y} \in V$, $BB'\mathbf{y} = \mathbf{y}$. (We can set $B' = B^T$ where B is as described above.) We apply the inductive hypothesis to the matrix

$$M' := B'MB \in \mathbb{R}^{(n-1) \times (n-1)}$$

and we find eigenvalues $\lambda_2, \ldots, \lambda_n$ and orthonormal eigenvectors $\mathbf{y}_2, \ldots, \mathbf{y}_n$ for M'.

For every $i = 2, \ldots, n$, we have

$$B'MB\mathbf{y}_i = \lambda_i \mathbf{y}_i$$

and so

$$BB'MB\mathbf{y}_i = \lambda_i B\mathbf{y}_i$$

Since $B\mathbf{y}_i$ is orthogonal to $\mathbf{x}_1$, it follows that $MB\mathbf{y}_i$ is also orthogonal to $\mathbf{x}_1$, and so $BB'MB\mathbf{y}_i = MB\mathbf{y}_i$, so we have

$$MB\mathbf{y}_i = \lambda_i B\mathbf{y}_i$$

and, defining $\mathbf{x}_i := B\mathbf{y}_i$, we have

$$M\mathbf{x}_i = \lambda_i \mathbf{x}_i$$

Finally, we observe that the vectors $\mathbf{x}_i$ are orthogonal. By construction, $\mathbf{x}_1$ is orthogonal to $\mathbf{x}_2, \ldots, \mathbf{x}_n$, and, for every $2 \leq i < j \leq n$, we have that

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \langle B\mathbf{y}_i, B\mathbf{y}_j \rangle = \langle \mathbf{y}_i, B^T B\mathbf{y}_j \rangle = \langle \mathbf{y}_i, \mathbf{y}_j \rangle = 0$$

We have not verified that the vectors $\mathbf{x}_i$ have norm 1 (which is true), but we can scale them to have norm 1 if not. □

1.3 Variational Characterization of Eigenvalues

We conclude these notes with the variational characterization of eigenvalues for real symmetric matrices.

Theorem 1.6 *Let $M \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ be the eigenvalues of M in non-increasing order. Then*

$$\lambda_k = \min_{k-\dim V} \quad \max_{\mathbf{x}\in V-\{\mathbf{0}\}} \quad \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

The quantity $\frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is called the *Rayleigh quotient* of $\mathbf{x}$ with respect to M, and we will denote it by $R_M(\mathbf{x})$.

PROOF: Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be orthonormal eigenvectors of the eigenvalues $\lambda_1, \ldots, \lambda_n$, as promised by the spectral theorem. Consider the k-dimensional space spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_k$. For every vector $\mathbf{x} = \sum_{i=1}^k a_i\mathbf{v}_i$ in such a space, the numerator of the Rayleigh quotient is

$$\sum_{i,j} a_i a_j \mathbf{v}_i^T M\mathbf{v}_j = \sum_{i,j} a_i a_j \lambda_j \mathbf{v}_i^T\mathbf{v}_j = \sum_{i=1}^k \lambda_i a_i^2 \leq \lambda_k \cdot \sum_{i=1}^k a_i^2$$

The denominator is clearly $\sum_{i=1}^k a_j^2$, and so $R_M(\mathbf{x}) \leq \lambda_k$. This shows that

$$\lambda_k \geq \min_{k-\dim V} \quad \max_{\mathbf{x}\in V-\{\mathbf{0}\}} \quad \frac{\mathbf{x}^T M\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

For the other direction, let V be any k-dimensional space: we will show that V must contain a vector of Rayleigh quotient $\geq \lambda_k$. Let S be the span of $\mathbf{v}_k, \ldots, \mathbf{v}_n$; since S has dimension $n-k+1$ and V has dimension k, they must have some non-zero vector in common. Let $\mathbf{x}$ be one such vector, and let us write $\mathbf{x} = \sum_{i=k}^n a_i\mathbf{v}_i$. The numerator of the Rayleigh quotient of $\mathbf{x}$ is

$$\sum_{i=k}^n \lambda_i a_i^2 \geq \lambda_k \sum_i a_i^2$$

and the denominator is $\sum_i a_i^2$, so $R_M(\mathbf{x}) \geq \lambda_k$. □

We have the following easy consequence.

Fact 1.7 *If λ_1 is the smallest eigenvalue of a real symmetric matrix M, then*

$$\lambda_1 = \min_{\mathbf{x}\neq\mathbf{0}} R_M(\mathbf{x})$$

Furthermore, every minimizer is an eigenvector of λ_1.

PROOF: The identity is the $k = 1$ case of the previous theorem. For the furthermore part, let $\lambda_1 \leq \cdots \lambda_n$ be the list of eigenvalues of M in non-decreasing order, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be corresponding eigenvectors. If $\mathbf{x} = \sum_i a_i \mathbf{v}_i$ is any vector, then

$$R_M(\mathbf{x}) = \frac{\sum_i \lambda_i a_i^2}{\sum_i a_i^2}$$

If $R_M(\mathbf{x}) = \lambda_1$, then $a_i = 0$ for every i such that $\lambda_i > \lambda_1$, that is, $\mathbf{x}$ is a linear combination of eigenvectors of λ_1, and hence it is an eigenvector of λ_1. □

Fact 1.8 *If λ_n is the largest eigenvalue of a real symmetric matrix M, then*

$$\lambda_n = \max_{\mathbf{x} \neq \mathbf{0}} R_M(\mathbf{x})$$

Furthermore, every maximizer is an eigenvector of λ_n.

PROOF: Apply Fact 1.7 to the matrix $-M$. □

Fact 1.9 *If λ_1 is the smallest eigenvalue of a real symmetric matrix M, and $\mathbf{x}_1$ is an eigenvector of λ_1, then*

$$\lambda_2 = \min_{\mathbf{x} \neq \mathbf{0},\ \mathbf{x} \perp \mathbf{x}_1} R_M(\mathbf{x})$$

Furthermore, every minimizer is an eigenvector of λ_2.

PROOF: A more conceptual proof would be to consider the restriction of M to the space orthogonal to $\mathbf{x}_1$, and then apply Fact 1.7 to such a linear operator. But, since we have not developed the theory for general linear operators, we would need to explicitly reduce to an $(n-1)$-dimensional case via a projection operator as in the proof of the spectral theorem.

Instead, we will give a more hands-on proof. Let $\lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ be the list of eigenvalues of M, with multiplicities, and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be orthonormal vectors as given by the spectral theorem. We may assume that $\mathbf{v}_1 = \mathbf{x}_1$, possibly by changing the orthonormal basis of the eigenspace of λ_1. For every vector $\mathbf{x} = \sum_{i=2}^{k} a_i \mathbf{v}_i$ orthogonal to $\mathbf{v}_1$, its Rayleigh quotient is

$$R_M(\mathbf{x}) = \frac{\sum_{i=2}^{n} \lambda_i a_i^2}{\sum_i a_i^2} \geq \lambda_2$$

and the minimum is achieved by vectors $\mathbf{x}$ such that $a_i = 0$ for every $\lambda_i > \lambda_2$, that is, for vectors $\mathbf{x}$ which are linear combinations of the eigenvectors of λ_2, and so every minimizer is an eigenvector of λ_2. □

Chapter 2
Introduction

In which we describe what this course is about.

2.1 Overview

This is class is about applications of linear algebra to graph theory and to graph algorithms. In the finite-dimensional case, linear algebra deals with vectors and matrices, and with a number of useful concepts and algorithms, such as determinants, eigenvalues, eigenvectors, and solutions to systems of linear equations.

The application to graph theory and graph algorithms comes from associating, in a natural way, a matrix to a graph $G = (V, E)$, and then interpreting the above concepts and algorithms in graph-theoretic language. The most natural representation of a graph as a matrix is via the $|V| \times |V|$ *adjacency matrix* of a graph, and certain related matrices, such as the *Laplacian* and *normalized Laplacian* matrix will be our main focus. We can think of $|V|$-dimensional Boolean vectors as a representing a partition of the vertices, that is, a *cut* in the graph, and we can think of arbitrary vectors as *fractional* cuts. From this point of view, eigenvalues are the optima of continuous relaxations of certain cut problems, the corresponding eigenvectors are optimal solutions, and connections between spectrum and cut structures are given by rounding algorithms converting fractional solutions into integral ones. Flow problems are dual to cut problems, so one would expect linear algebraic techniques to be helpful to find flows in networks: this is the case, via the theory of electrical flows, which can be found as solutions to linear systems.

The course can be roughly subdivided into three parts: in the first part of the course we will study *spectral graph algorithms*, that is, graph algorithms that make use of eigenvalues and eigenvectors of the normalized Laplacian of the given graph. In the second part of the course we will look at constructions of expander graphs, and their applications. In the third part of the course, we will look at fast algorithms for solving systems of linear equations of

the form $L\mathbf{x} = \mathbf{b}$, where L is Laplacian of a graph, their applications to finding electrical flows, and the applications of electrical flows to solving the max flow problem.

2.2 Spectral Graph Algorithms

We will study approximation algorithms for the *sparsest cut* problem, in which one wants to find a cut (a partition into two sets) of the vertex set of a given graph so that a minimal number of edges cross the cut compared to the number of pairs of vertices that are disconnected by the removal of such edges.

This problem is related to estimating the edge expansion of a graph and to find *balanced separators*, that is, ways to disconnect a constant fraction of the pairs of vertices in a graph after removing a minimal number of edges.

Finding balanced separators and sparse cuts arises in *clustering* problems, in which the presence of an edge denotes a relation of similarity, and one wants to partition vertices into few clusters so that, for the most part, vertices in the same cluster are similar and vertices in different clusters are not. For example, sparse cut approximation algorithms are used for *image segmentation*, by reducing the image segmentation problem to a graph clustering problem in which the vertices are the pixels of the image and the (weights of the) edges represent similarities between nearby pixels.

Balanced separators are also useful in the design of divide-and-conquer algorithms for graph problems, in which one finds a small set of edges that disconnects the graph, recursively solves the problem on the connected components, and then patches the partial solutions and the edges of the cut, via either exact methods (usually dynamic programming) or approximate heuristic. The sparsity of the cut determines the running time of the exact algorithms and the quality of approximation of the heuristic ones.

We will study a spectral algorithm first proposed by Fiedler in the 1970s, and to put its analysis into a broader context, we will also study the Leighton-Rao algorithm, which is based on linear programming, and the Arora-Rao-Vazirani algorithm, which is based on semidefinite programming. We will see how the three algorithms are based on conceptually similar continuous relaxations.

Before giving the definition of sparsest cut, it is helpful to consider examples of graphs that have very sparse cuts, in order to gain intuition.

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 the 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 of view.

For simplicity, here we will give definitions that apply only to the case of regular graphs.

Definition 2.1 (Edge expansion of a set) *Let $G = (V, E)$ be a d-regular graph, and $S \subseteq V$ a subset of vertices. The edge expansion of S is*

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

where $E(S, V-S)$ is the number of edges in E that have one endpoint in S and one endpoint in $V - S$.

$d|S|$ is a trivial upper bound to the number of edges that can leave S, and so $\phi(S)$ measures how much smaller the actual number of edges is than this upper bound. We can also think of $\phi(S)$ as the probability that, if we pick a random node v in S and then a random neighbor w of v, the node w happens to be outside of S.

The quantity $1 - \phi(S)$ is the average fraction of neighbors that vertices in S have within S. For example, if G represents a social network, and S is a subset of users of expansion .3, this means that, on average, the users in S have 70% of their friends within S.

If $(S, V-S)$ is a cut of the graph, and $|S| \leq |V-S|$, then $\phi(S)$ is, within a factor of two, the ratio between the fraction $E(S, V-S)/|E| = 2E(S, V-S)/dn$ of edges that we have to remove to disconnect S from $V - S$, and the fraction $|S| \cdot |V-S| / \binom{n}{2}$ of pairs of vertices

that become unreachable if we do so. We define the edge expansion of a cut as

$$\phi(S, V - S) := \max\{\phi(S), \phi(V - S)\}$$

The edge expansion of a graph is the minimum of the edge expansion of all cuts.

Definition 2.2 (Edge expansion of a graph) *Let $G = (V, E)$ be a d-regular graph, its edge expansion is*

$$\phi(G) := \min_{S:0<|S|<|V|} \phi(S, V - S) = \min_{S:0<|S|\leq\frac{|V|}{2}} \phi(S)$$

If A is the adjacency matrix of a d-regular graph $G = (V, E)$, then the *normalized Laplacian* of G is the matrix $L := I - \frac{1}{d}A$. We will prove the Cheeger inequalities: that if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of L, counted with multiplicities and sorted in nondecreasing order, then

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

The lower bound $\phi(G) \geq \frac{\lambda_2}{2}$ follows by using the variational characterization of eigenvalues to think of λ_2 as the optimum of a continuous optimization problem, and then realizing that, from this point of view, λ_2 is actually the optimum of a *relaxation* of $\phi(G)$.

The upper bound $\phi(G) \leq \sqrt{2\lambda_2}$ has a constructive proof, showing that the set S_F returned by Fiedler's algorithm has size $\leq |V|/2$ and satisfies $\phi(S_F) \leq 2\sqrt{\lambda_2}$. The two inequalities, combined, show that $\phi(S_F) \leq 2\sqrt{\phi(G)}$ and provide a (tight) worst-case analysis of the quality of the cut found by Fiedler-s algorithm, compared with the optimal cut.

To put this result in a broader context, we will see the Leighton-Rao approximation algorithm, based on linear programming, which finds a cut of expansion $\leq \phi(G) \cdot O(\log |V|)$, and the Arora-Rao-Vazirani algorithm, based on semidefinite programming, which finds a cut of expansion $\leq \phi(G) \cdot O(\sqrt{\log |V|})$. The spectral, linear programming, and semidefinite programming relaxation can all be seen as very related.

We will then consider combinatorial characterizations, and algorithms for other laplacian eigenvalues.

We will prove a "higher order" Cheeger inequality that characterizes λ_k for $k \geq 2$ similarly to how the standard Cheeger inequality characterizes λ_2, and the proof will provide a worst-case analysis of spectral partitioning algorithms similarly to how the proof of the standard Cheeger inequality provides a worst-case analysis of Fiedler's algorithm.

The outcome of these results is that small Laplacian eigenvalues characterize the presence of sparse cuts in the graph. Analogously, we will show that the value of λ_n characterizes large cuts, and the proof a Cheeger-type inequality for λ_n will lead to the worst-case analysis of a spectral algorithm for max cut.

2.3 Constructions and Applications of Expander Graphs

A family of constant-degree expanders is a collection of arbitrarily large graphs, all of degree $O(1)$ and edge expansion $\Omega(1)$. Expanders are useful in several applications, and a common theme in such applications is that even though they are sparse, they have some of the "connectivity" properties of a complete graph.

For example, if one removes a $o(1)$ fraction of edges from an expander, one is left with a connected component that contains a $1 - o(1)$ fraction of vertices.

Lemma 2.3 *Let $G = (V, E)$ be a regular graph of expansion ϕ. 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 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.

Another way in which expander graphs act similarly to a complete graph is the following. Suppose that, given a graph $G = (V, E)$, we generate a sequence $v_1, \ldots, v_k$ by choosing $v_1 \in V$ uniformly at random and then performing a $(k-1)$-step random walk. If G is a complete graph (in which every vertex has a self-loop), this process uses $k \log |V|$ random bits and generates k uniform and independent random vertices. In an expander of constant degree, the process uses only $\log |V| + O(k)$ random bits, and the resulting

sequence has several of the useful statistical properties of a sequence generated uniformly at random. Especially in the case in which k is of the order of $\log |V|$, using $O(\log |V|)$ instead of $O(\log^2 |V|)$ random bits can be a significant saving in certain application. (Note, in particular, that the sample space has polynomial size instead of quasi-polynomial size.)

Constructions of constant-degree expanders are useful in a variety of applications, from the design of data structures, to the derandomization of algorithms, from efficient cryptographic constructions to being building blocks of more complex quasirandom objects.

There are two families of approaches to the explicit (efficient) construction of bounded-degree expanders. One is via algebraic constructions, typically ones in which the expander is constructed as a Cayley graph of a finite group. Usually these constructions are easy to describe but rather difficult to analyze. The study of such expanders, and of the related group properties, has become a very active research program. There are also combinatorial constructions, which are somewhat more complicated to describe but considerably simpler to analyze.

2.4 Mixing time of random walks

If one takes a random walk in a regular graph that is connected and not bipartite, then, regardless of the starting vertex, the distribution of the t-th step of the walk is close to the uniform distribution over the vertices, provided that t is large enough. It is always sufficient for t to be quadratic in the number of vertices; in some graphs, however, the distribution is near-uniform even when t is just poly-logarithmic, and, indeed, the time is at most $O\left(\frac{1}{\lambda_2} \log |V|\right)$, and thus it is at most logarithmic in expander graphs.

Among other applications, the study of the "mixing time" (the time that it takes to reach the uniform distribution) of random walks has applications to analyzing the convergence time of certain randomized algorithms.

The design of approximation algorithms for *combinatorial counting* problems, in which one wants to count the number of solutions to a given NP-type problem, can be reduced to the design of *approximately uniform sampling* in which one wants to approximately sample from the set of such solutions. For example, the task of approximately counting the number of perfect matchings can be reduced to the task of sampling almost uniformly from the set of perfect matchings of a given graph. One can design approximate sampling algorithms by starting from an arbitrary solution and then making a series of random local changes. The behavior of the algorithm then corresponds to performing a random walk in the graph that has a vertex for every possible solution and an edge for each local change that the algorithm can choose to make. Although the graph can have an exponential number of vertices in the size of the problem that we want to solve, it is possible for the approximate sampling algorithm to run in polynomial time, provided that a random walk in the graph converges to uniform in time poly-logarithmic in its size.

The study of the mixing time of random walks in graphs is thus a main analysis tool to bound the running time of approximate sampling algorithms (and, via reductions, of approximate counting algorithms).

As a way of showing applications of results proved so far, we will show that, because of Cheeger's inequality, the mixing time is upper-bounded by $O\left(\frac{1}{\phi^2}\log|V|\right)$, and then we will use the dual of the Leighton-Rao relaxation to show that $1/\phi$ can be upper-bounded by the congestion of a certain flow problem. We will apply this theory to the analysis of an algorithm that approximates the number of perfect matchings in a given dense bipartite graph.

2.5 Linear Systems, Electrical Flows, and Applications

In the last part of the course, we will turn to connections between graph theory and a different aspect of linear algebra, namely the solution of systems of linear equations. If we have a system of linear equations of the form

$$A\mathbf{x} = \mathbf{b}$$

we can solve it (or determine that it has no solution) in polynomial time using Gaussian elimination. Sometimes, it is possible to develop faster and more numerically stable algorithms by thinking of the problem has an *optimization*, such as, for example,

$$\min_{\mathbf{x}} ||A\mathbf{x} - \mathbf{b}||$$

for an appropriate choice of norm.

If A is positive definite (that is, all the eigenvalues are strictly positive), then another way of turning a linear system into an optimization problem is to consider the problem

$$\min \ \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} \tag{2.1}$$

The problem is strictly convex, because the Hessian of the function $f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x}$, that is, the matrix of partial second derivatives of $f(\cdot)$, is, at every point, the matrix A itself, which we assumed to be positive definite.

The strongly convex optimization problem (2.1) has a unique minimum, achieved at a point $\mathbf{x}^*$. The gradient of $f(\cdot)$ at a point $\mathbf{x}$, that is, the vector of partial derivates at $\mathbf{x}$, is $\nabla f(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$. The gradient has to be equal to the $\mathbf{0}$ vector at the optimum $\mathbf{x}^*$, and so we have $A\mathbf{x}^* = \mathbf{b}$.

If we want to solve the linear system $A\mathbf{x} = \mathbf{b}$, and A is positive definite, then a possible strategy is to solve the convex optimization problem (2.1) using gradient descent, or similar local-search algorithms for convex optimization. The running time of such algorithms will be determined by the smallest eigenvalue A. In order to deal with matrix having small eigenvalues, one resorts to *preconditioning*, which is a technique that reduces the $A\mathbf{x} = \mathbf{b}$ system to a $B\mathbf{y} = \mathbf{b}'$ system in which B has a larger smallest eigenvalue. In the interesting special case in which A is the Laplacian matrix of an undirected graph, the running time is determined by the expansion of the graph, and preconditioning can be understood in graph-theoretic terms.

(Technically, the Laplacian is not positive definite. What we mean above is that we are interested in solving an equation of the form $L\mathbf{x} = \mathbf{b}$ where L is a Laplacian matrix, and $\mathbf{x}$ is further constrained to be orthogonal to the eigenspace of zero.)

Efficiently solving "Laplacian systems" of the form $L\mathbf{x} = \mathbf{b}$ is closely related to the problem of finding *sparsifiers* of graphs, and we will see nearly linear time algorithms for both problems.

One application of finding solutions to systems of the form $L\mathbf{x} = \mathbf{b}$ is to find *electrical flows* in networks. We will then see how to use fast algorithms for finding electrical flows and turn them into algorithm for the Max Flow problem.

Chapter 3

The Basics of Spectral Graph Theory

In which we introduce the Laplacian matrix and we prove our first results in spectral graph theory.

3.1 The Laplacian Matrix

Given an undirected graph $G = (V, E)$, the approach of spectral graph theory is to associate a symmetric real-valued matrix to G, and to related the eigenvalues of the matrix to combinatorial properties of G.

For the sake of this lecture, we will restrict ourselves to the case in which G is a d-regular graph, and we will then see how to extend our results to apply to irregular graphs as well.

The most natural matrix to associate to G is the adjacency matrix A such that $A_{i,j} = 1$ if $\{i, j\} \in E$ and $A_{i,j} = 0$ otherwise. In the second part of the course, in which we will study expander graphs, the adjacency matrix will indeed be the most convenient matrix to work with. For the sake of the algorithms that we will analyze in the first part of the course, however, a slight variation called the *normalized Laplacian* is more convenient.

There are a few ways to motivate the definition of the Laplacian. One way is the following: the variational characterization of the eigenvalues of real symmetric matrices tells us that we can think of the eigenvalues of M as optima of min-max optimization problems in which vectors $\mathbf{x} \in \mathbb{R}^V$ are feasible solutions and the cost function is the Rayleigh quotient

$$R_M(x) = \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

We know that every homogeneous polynomial of degree 2 can be realized as $x^T M x$ for some matrix M, and,if we want to study cuts in a graph $G = (V, E)$, it makes sense to choose a matrix M such that

$$\mathbf{x}^T M \mathbf{x} = \sum_{\{u,v\} \in E} (x_u - x_v)^2$$

because, if $\mathbf{x} \in \{0,1\}^V$ is a Boolean vector, representing a cut in the graph, then the right-hand-side expression above is counting the number of edges that cross the cut, and so optimization problems with the above cost functions will be relaxations of cut problems.

Some calculations show that the matrix having such a property is $dI - A$, which is called the Laplacian matrix of G. Indeed, we can verify that

$$\mathbf{x}^T(dI - A)\mathbf{x} = \sum_{\{u,v\}\in E} (x_u - x_v)^2$$

because both expressions are easily seen to be equal to

$$\sum_v dx_v^2 - 2\sum_{\{u,v\}\in E} x_u x_v$$

As we will see in a moment, the eigenvalues of $dI - A$ are in the range $[0, 2d]$, and it is not hard to see that their sum is dn, so it is convenient to divide the Laplacian matrix by d so that the range and the average values of the eigenvalues of the resulting matrix are independent of the degree. (This degree independence will make it possible to generalize results to the irregular case.)

We have thus reached the following definition.

Definition 3.1 (Normalized Laplacian) *The normalized Laplacian matrix of an undirected d-regular graph $G = (V, E)$ is $L := I - \frac{1}{d}A$.*

3.2 Some Facts About Laplacian Eigenvalues

We shall now prove the following relations between the eigenvalues of L and certain purely combinatorial properties of G.

Theorem 3.2 *Let G be a d-regular undirected graph, let A be the adjacency matrix of G, and $L = I - \frac{1}{d} \cdot A$ be the normalized Laplacian matrix of G. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the real eigenvalues of L with multiplicities, in nondecreasing order. 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.

PROOF: By the characterization of the Rayleigh quotient of L that we established above, and from the variational characterization of eigenvalues, we have

$$\lambda_1 = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}} \frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{d\sum_v x_v^2}$$

and so $\lambda_1 \geq 0$ because the Rayleigh quotient, being a ratio of sums of squares, is always non-negative.

If we take $\mathbf{1} = (1, \ldots, 1)$ to be the all-one vector, we see that its Rayleigh quotient is 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 λ_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 $\mathbf{x} \in S$ and every $\{u, v\} \in E$, 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 $\mathbf{x} \in S$ must be constant within each connected component of G, and so the dimension of S 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 λ_n, we first note that we have the formula

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

from the variational characterization of eigenvalues (see Handout 0).

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

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

and so

$$\lambda_n = 2 - \min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}} \frac{\sum_{\{u,v\} \in E} (x_u + x_v)^2}{d \sum_v x_v^2}$$

from which it follows that

$$\lambda_n \leq 2$$

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 us now define $A := \{v : x_v > 0\}$ and $B := \{v : x_v < 0\}$. The set $A \cup B$ is non-empty (otherwise we would have $\mathbf{x} = \mathbf{0}$) and is either the entire graph, or else it is disconnected from the rest of the graph, because otherwise an edge with an endpoint in $A \cup B$ and an endpoint in $V - (A \cup B)$ would give a positive contribution to $\sum_{\{u,v\} \in E} (x_u - x_v)^2$; 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. $\square$

Chapter 4

Cheeger's Inequalities

In which we generalize the notion of normalized Laplacian to irregular graphs, we extend the basic spectral graph theory results from last lecture to irregular graphs, and we prove the easy direction of Cheeger's inequalities.

4.1 Irregular Graphs

Let $G = (V, E)$ be an undirected graph, not necessarily regular. We will assume that every vertex has non-zero degree. We would like to define a normalized Laplacian matrix associated to G so that the properties we proved last time are true: that the multiplicity of 0 as an eigenvalue is equal to the number of connected components of G, that the largest eigenvalue is at most 2, and that it is 2 if and only if (a connected component of) the graph is bipartite.

In order to have a matrix such that zero is the smallest eigenvalue, and such that multiplicity of zero is the number of connected component, we want a matrix such that the numerator of the Rayleigh quotient is (a multiple of)

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

and the matrix M such that $\mathbf{x}^T M \mathbf{x}$ is the above expression is the matrix $M = D - A$, where D is the diagonal matrix such that $D_{v,v} = d_v$, the degree of v. The matrix $D - A$ is called the *Laplacian* matrix of G. Note that there is no fixed constant upper bound to the largest eigenvalue of $D - A$; for example, if G is a d-regular bipartite graph, the largest eigenvalue is $2d$, as proved in the last lecture.

Some calculations shows that the right analog of the normalization that we did in the regular case (in which we divided by the degree d) would be to have a matrix whose Rayleigh

quotient is

$$\frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{\sum_v d_v x_v^2} = 2 - \frac{\sum_{\{u,v\}\in E}(x_u + x_v)^2}{\sum_v d_v x_v^2} \tag{4.1}$$

and it's clear that the above expression is at most 2 for every $\mathbf{x}$, and it is possible to find an $\mathbf{x}$ for which the above expression is 2 if and only if G has a bipartite connected component.

Unfortunately, there is no matrix whose Rayleigh quotient equals (4.1), because the denominator of a Rayleigh quotient is, by definition, $\sum_v x_v^2$ regardless of the matrix.

One way to work around this problem would be to give a more general form of the variational characterization of eigenvalues, in which we have an arbitrary inner product $\langle \cdot, \cdot \rangle$, and the Rayleigh quotient is defined as $\frac{\langle \mathbf{x}, M\mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle}$.

Here we will proceed in a way that is essentially equivalent, but without introducing this additional definitional framework.

The point is that, if we look at the Rayleigh quotient of the vector $D^{1/2}\mathbf{x}$, where $D^{1/2}$ is the diagonal matrix such that $D_{v,v}^{1/2} = \sqrt{d_v}$, then the denominator will indeed be $\mathbf{x}^T D\mathbf{x} = \sum_v d_v x_v^2$, and that we can find a matrix L such that the numerator of the Rayleigh quotient of $D^{1/2}\mathbf{x}$ is $\mathbf{x}^T D^{1/2} L D^{1/2}\mathbf{x} = \sum_{\{u,v\}\in E}(x_u - x_v)^2$, so that the Rayleigh quotient $R_L(D^{1/2}\mathbf{x})$ is indeed the expression in (4.1).

This matrix L is called the *normalized Laplacian* of G and, by the above observation, it has to be $L = D^{-1/2}(D - A)D^{-1/2} = I - D^{-1/2}AD^{-1/2}$. Note that, in a d-regular graph, we get $L = I - \frac{1}{d}A$, consistent with our definition from the last lecture.

Now the point is that the mapping $\mathbf{x} \to D^{1/2}\mathbf{x}$ is linear and bijective, so it maps the set of all possible vectors to the set of all possible vectors, and it maps a k-dimensional space to a (possibly different) k-dimensional space.

If we let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $L = I - D^{-1/2}AD^{-1/2}$, counting repetitions, the variational characterization gives us

$$\lambda_1 = \min_{\mathbf{x}\neq\mathbf{0}} R_L(\mathbf{x}) = \min_{\mathbf{x}\neq\mathbf{0}} R_L(D^{1/2}\mathbf{x}) = \min_{\mathbf{x}\neq\mathbf{0}} \frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{\sum_v d_v x_v^2}$$

and

$$\begin{aligned}
\lambda_k &= \min_{X \ k-\text{dimensional}} \ \max_{\mathbf{x}\in X-\{\mathbf{0}\}} R_L(\mathbf{x}) \\
&= \min_{X \ k-\text{dimensional}} \ \max_{\mathbf{x}\in X-\{\mathbf{0}\}} R_L(D^{1/2}\mathbf{x}) \\
&= \min_{X \ k-\text{dimensional}} \ \max_{\mathbf{x}\in X-\{\mathbf{0}\}} \frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{\sum_v d_v x_v^2}
\end{aligned}$$

from which we have that $\lambda_1 = 0$ and that the multiplicity of zero is equal to the number of connected components.

We also have

$$\lambda_n = \max_{\mathbf{x}\neq\mathbf{0}} R_L(\mathbf{x}) = \max_{\mathbf{x}\neq\mathbf{0}} R_L(D^{1/2}\mathbf{x})$$
$$= 2 - \min_{\mathbf{x}\neq\mathbf{0}} \frac{\sum_{\{u,v\}\in E}(x_u + x_v)^2}{\sum_v d_v x_v^2}$$

from which we see that $\lambda_n \leq 2$ and that $\lambda_n = 2$ if and only if one of the connected components of G is bipartite.

4.2 Edge Expansion, Fiedler's Algorithm and Cheeger's Inequalities

We will now return, for simplicity, to the regular case.

If $G = (V, E)$ is an undirected d-regular graph, and $S \subseteq V$ is a set of vertices, we call the quantity

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

the *edge expansion* of S. The quantity $\phi(S)$ is the average fraction of neighbors outside of S for a random element of S, and it compares the actual number of edges crossing the cut $(S, V - S)$ with the trivial upper bound $d|S|$.

We define the expansion of a cut $(S, V - S)$ as

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

The edge expansion of the graph G is defined as

$$\phi(G) := \min_S \phi(S, V - S) = \min_{S:1\leq|S|\leq\frac{|V|}{2}} \phi(S)$$

(Note: it is common in the literature to use the notation $\phi(S)$ to refer to the quantity that we call $\phi(S, V - S)$.)

Finding cuts of small expansion is a problem with several applications. It is an open question if there is a polynomial-time approximation with a constant-factor approximation ratio; a positive answer would refute the "small-set expansion conjecture" which is closely related to the unique games conjecture.

The following algorithm was proposed by Fiedler, and it works well in practice when $\mathbf{x}$ is the eigenvector of λ_2.

- Input: graph $G = (V, E)$, vector $\mathbf{x} \in \mathbb{R}^V$
  - Sort the vertices according the values x_v, and let $v_1, \ldots, v_n$ be the sorted order
  - Find a k that minimizes $\phi(\{v_1, \ldots, v_k\}, \{v_{k+1}, \ldots, v_n\})$, and output such a cut

Note that Fiedler's algorithm can be implemented in time $O(|E| + |V| \log |V|)$, because it takes time $O(|V| \log |V|)$ to sort the vertices, and the cut of minimal expansion that respects the sorted order can be found in time $O(E)$. (To see that this is the case, consider that, in order to find such a cut, we just need to compute the numbers $e_k := E(\{v_1, \ldots, v_k\}, \{v_{k+1}, \ldots, v_n\})$ for each $k = 1, \ldots, n-1$. We see that e_1 is equal to the degree of v_1, and that, given e_{k-1}, the value of e_k can be computed by just adding to e_{k-1} the number of neighbors of v_k in $\{v_{k+1}, \ldots, v_n\}$, and subtracting the number of neighbors of v_k in $\{v_1, \ldots, v_k\}$, on operation that can be done in time $O(d_{v_k})$. Thus the total running time is of the order of $\sum_v d_v$, that is, $O(|E|)$.)

We will prove the following result

Theorem 4.1 (Cheeger's Inequalities) *Let G be an undirected regular graph and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of the normalized Laplacian, with repetitions, then*

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

Furthermore, if $(S, V-S)$ is the cut found by Fiedler's algorithm given the eigenvector of λ_2, then

$$\phi(S, V-S) \leq \sqrt{2\lambda_2}$$

Note that, from the *furthermore* part of the Theorem, it follows that, if $(S, V-S)$ is the cut found by Fiedler's algorithm given an eigenvector of λ_2, we have

$$\phi(S, V-S) \leq 2\sqrt{\phi(G)}$$

which is a worst-case guarantee of the quality of the solution found by the algorithm.

4.3 Proof that $\frac{\lambda_2}{2} \leq \phi(G)$

Let S be a set of vertices such that $\phi(S, V-S) = \phi(G)$. Recall that for every set S, we have that expansion of S is the same as the Rayleigh quotient of the indicator vector $\mathbf{1}_S$. (The indicator vector of a set S is the 0/1 vector $\mathbf{1}_S$ whose v-th coordinate is 1 if and only if $v \in S$.) So we have

$$R_L(\mathbf{1}_S) \leq \phi(G)$$
$$R_L(\mathbf{1}_{V-S}) \leq \phi(G)$$

also recall that, from the variational characterization of eigenvalues, we have

$$\lambda_2 = \min_{X \text{ 2-dimensional}} \max_{\mathbf{x} \in X - \{\mathbf{0}\}} R_L(\mathbf{x})$$

We will prove the inequality $\lambda_2 \leq 2\phi(G)$ by showing that all the vectors in the 2-dimensional space X of linear combinations of the orthogonal vectors $\mathbf{1}_S, \mathbf{1}_{V-S}$ have Rayleigh quotient at most $2\phi(G)$. This is a consequence of the following useful fact.

Lemma 4.2 *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be two orthogonal vectors, and let* M *be a positive semidefinite matrix. Then*

$$R_M(\mathbf{x}+\mathbf{y}) \leq 2 \cdot \max\{R_M(\mathbf{x}), R_M(\mathbf{y})\}$$

PROOF: Let $0 \leq \lambda_1 \leq \cdots \lambda_n$ be the eigenvalues of M and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a corresponding basis of eigenvectors. Let us write $\mathbf{x} = \sum_i a_i \mathbf{v}_i$ and $\mathbf{y} = \sum_i b_i \mathbf{v}_i$.

The Rayleigh quotient of $\mathbf{x}+\mathbf{y}$ is

$$\frac{\sum_i \lambda_i (a_i+b_i)^2}{||\mathbf{x}+\mathbf{y}||^2} \leq \frac{\sum_i \lambda_i 2(a_i^2+b_i^2)}{||\mathbf{x}||^2+||\mathbf{y}||^2}$$

$$= \frac{2R_M(\mathbf{x}) \cdot ||\mathbf{x}||^2 + 2R_M(\mathbf{y}) \cdot ||\mathbf{y}||^2}{||\mathbf{x}||^2+||\mathbf{y}||^2} \leq 2\max\{R_M(\mathbf{x}), R_M(\mathbf{y})\}$$

In the first inequality, we used orthogonality of $\mathbf{x}$ and $\mathbf{y}$ to derive $||\mathbf{x}+\mathbf{y}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2$ and we used the Cauchy-Schwarz inequality $(a+b)^2 \leq 2a^2 + 2b^2$. □

4.4 First Part of the Analysis of Fiedler's Algorithm

The vector $\mathbf{1} = (1, \ldots, 1)$ is an eigenvector for 0, which is the smallest eigenvalue of the normalized Laplacian of G, and so, from the variational characterization of eigenvalues, we have that

$$\lambda_2 = \min_{\mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{d \sum_v x_v^2}$$

and that any eigenvector $\mathbf{x}$ of λ_2 is a minimizer of the above expression. We will prove that $\phi(G) \leq \sqrt{2\lambda_2}$ and that the *Furthermore* part of Theorem 4.1 is true by showing the following stronger result:

Lemma 4.3 *Let* $\mathbf{x}$ *be a vector orthogonal to* $\mathbf{1}$ *and let* $(S, V-S)$ *be the cut found by Fiedler's algorithm given* $\mathbf{x}$*. Then*

$$\phi(S, V-S) \leq \sqrt{2R_L(\mathbf{x})}$$

This stronger form is useful, because often one runs Fiedler's algorithm on an approximate eigenvector, and Lemma 4.3 shows that one gets a guarantee on the quality of the resulting cut that does not require $\mathbf{x}$ to be an eigenvector, as long as its Rayleigh quotient is small.

We divide the proof of Lemma 4.3 into two parts: we analyze the performance of the algorithm given a vector $\mathbf{x}$ that, instead of being orthogonal to $\mathbf{1}$, has the property of having non-negative entries and at most $|V|/2$ non-zero entries, and we show that analyzing the performance of the algorithm on vectors of the former type reduces to analyzing the performance on vectors of the latter type.

Lemma 4.4 *Let $\mathbf{y} \in \mathbb{R}^V_{\geq 0}$ be a vector with non-negative entries. Then there is a $0 < t \leq \max_v\{y_v\}$ such that*

$$\phi(\{v : y_v \geq t\}) \leq \sqrt{2R_L(\mathbf{y})}$$

Lemma 4.5 *Let $\mathbf{x} \in R^V$ be orthogonal to $\mathbf{1}$. Then there is a vector $\mathbf{y} \in \mathbb{R}^V_{\geq 0}$ with at most $|V|/2$ non-zero entries such that*

$$R_L(\mathbf{y}) \leq R_L(\mathbf{x})$$

Furthermore, for every $0 < t \leq \max_v\{y_v\}$, the cut $(\{v : y_v \geq t\}, \{v : y_v < t\})$ is one of the cuts considered by Fiedler's algorithm on input $\mathbf{x}$.

Let us quickly see how to prove Lemma 4.3 given Lemma 5.1 and Lemma 4.5. Let $\mathbf{x}$ be a vector orthogonal to $\mathbf{1}$, and let $(S_F, V - S_F)$ be the cut found by Fiedler's algorithm given $\mathbf{x}$. Let $\mathbf{y}$ be the non-negative vector with at most $|V|/2$ positive entries and such that $R_L(\mathbf{y}) \leq R_L(\mathbf{x})$ as promised by Lemma 4.5. Let $0 < t \leq \max_v\{y_v\}$ be a threshold such that

$$\phi(\{v : y_v \geq t\}) \leq \sqrt{2R_L(\mathbf{y})} \leq \sqrt{2R_L(\mathbf{x})}$$

as promised by Lemma 4.5. The set $S_t := \phi(\{v : y_v \geq t\})$ contains at most $|V|/2$ vertices, and the cut $(S_t, V - S_t)$ is one of the cuts considered by Fiedler's algorithm on input $\mathbf{x}$, and so

$$\phi(S_F, V - S_F) \leq \phi(S_t, V - S_t) = \phi(S_t) \leq \sqrt{2R_L(\mathbf{x})}$$

We will prove Lemma 5.1 next time. We conclude this lecture with a proof of Lemma 4.5.

Proof: (Of Lemma 4.5) First we observe that, for every constant c,

$$R_L(\mathbf{x} + c\mathbf{1}) \leq R_L(\mathbf{x})$$

because the numerator of $R_L(\mathbf{x} + c\mathbf{1})$ and the numerator of $R_L(\mathbf{x})$ are the same, and the denominator of $R_L(\mathbf{x} + c\mathbf{1})$ is $||\mathbf{x} + c\mathbf{1}||^2 = ||\mathbf{x}||^2 + ||c\mathbf{1}||^2 \geq ||\mathbf{x}^2||$.

Let m be the median value of the entries of $\mathbf{x}$, and call $\mathbf{x}' := \mathbf{x} - m\mathbf{1}$. Then we have $R_L(\mathbf{x}') \leq R_L(\mathbf{x})$, and the median of the entries of $\mathbf{x}'$ is zero, meaning that $\mathbf{x}'$ has at most $|V|/2$ positive entries and at most $|V|/2$ negative entries. We will refer to the vertices v such that $x'_v > 0$ as the *positive* vertices, and the vertices v such that $x'_v < 0$ as the *negative* vertices.

We write

$$\mathbf{x}' = \mathbf{x}^+ - \mathbf{x}^-$$

where $x^+_v = x'_v$ if v is positive and $x^+_v = 0$ otherwise; similarly, $x^-_v = -x'_v$ if v is negative, and $x^-_v = 0$ otherwise. Note that $\mathbf{x}^+$ and $\mathbf{x}^-$ are orthogonal, non-negative, and each of them has at most $|V|/2$ nonzero entries. Note also that, for every t, the cut defined by the set $\{v : x^+_v \geq t\}$ is one of the cuts considered by Fiedler's algorithm on input $\mathbf{x}$, because it is the cut

$$(\{v : x_v < t + m\}, \{v : x_v \geq t + m\})$$

Similarly, for every t, the cut defined by the set $\{v : x_v^- \geq t\}$ is one of the cuts considered by Fiedler's algorithm on input $\mathbf{x}$, because it is the cut

$$(\{v : x_v \leq m - t\}, \{v : x_v > m - t\})$$

It remains to show that at least one of $\mathbf{x}^+$ or $\mathbf{x}^-$ has Rayleigh quotient smaller than or equal to the Rayleigh quotient of $\mathbf{x}'$ (and, hence, of $\mathbf{x}$). We claim that

$$R_L(\mathbf{x}') = \frac{\sum_{\{u,v\}}(x_u - x_v)^2}{||\mathbf{x}'||^2} = \frac{\sum_{\{u,v\}}((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2}{||\mathbf{x}^+||^2 + ||\mathbf{x}^-||^2}$$

$$\geq \frac{\sum_{\{u,v\}}(x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2}{||\mathbf{x}^+||^2 + ||\mathbf{x}^-||^2}$$

$$= \frac{R_L(\mathbf{x}^+) \cdot ||\mathbf{x}^+||^2 + R_L(\mathbf{x}^-) \cdot ||\mathbf{x}^-||^2}{||\mathbf{x}^+||^2 + ||\mathbf{x}^-||^2} \geq \min\{R_L(\mathbf{x}^+), R_L(\mathbf{x}^-)\}$$

The only step that we need to justify is that for every edge $\{u, v\}$ we have

$$((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2 \geq (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2$$

If $\{u, v\}$ is an edge between two non-positive vertices, or between two non-negative vertices, then the left-hand side and the right-hand side are clearly equal. If it is an edge between a positive vertex u and a negative vertex v, then the left-hand side is equal to $(x_u^+ + x_v^-)^2$, and the right-hand side is equal to $(x_u^+)^2 + (x_v^-)^2$. $\square$

Chapter 5

Cheeger's Inequalities cont'd

In which we finish the proof of Cheeger's inequalities.

It remains to prove the following statement.

Lemma 5.1 *Let $\mathbf{y} \in \mathbb{R}^V_{\geq 0}$ be a vector with non-negative entries. Then there is a $0 < t \leq \max_v\{y_v\}$ such that*

$$\phi(\{v : y_v \geq t\}) \leq \sqrt{2R_L(\mathbf{y})}$$

We will provide a probabilistic proof. Without loss of generality (multiplication by a scalar does not affect the Rayleigh quotient of a vector) we may assume that $\max_v y_v = 1$. We consider the probabilistic process in which we pick $t > 0$ in such a way that t^2 is uniformly distributed in $[0, 1]$ and then define the non-empty set $S_t := \{v : y_v \geq t\}$.

We claim that

$$\frac{\mathbb{E}\, E(S_t, V - S_t)}{\mathbb{E}\, d|S_t|} \leq \sqrt{2R_L(\mathbf{y})} \tag{5.1}$$

Notice that Lemma 5.1 follows from such a claim, because of the following useful fact.

Fact 5.2 *Let X and Y be random variables such that $\mathbb{P}[Y > 0] = 1$. Then*

$$\mathbb{P}\left[\frac{X}{Y} \leq \frac{\mathbb{E}\, X}{\mathbb{E}\, Y}\right] > 0$$

Proof: Call $r := \frac{\mathbb{E}\, X}{\mathbb{E}\, Y}$. Then, using linearity of expectation, we have $\mathbb{E}\, X - rY = 0$, from which it follows $\mathbb{P}[X - rY \leq 0] > 0$, but, whenever $Y > 0$, which we assumed to happen with probability 1, the conditions $X - rY \leq 0$ and $\frac{X}{Y} \leq r$ are equivalent. □

It remains to prove (5.1).

To bound the denominator, we see that

$$\mathbb{E}\, d|S_t| = d \cdot \sum_{v \in V} \mathbb{P}[v \in S_t] = d \sum_v y_v^2$$

because

$$\mathbb{P}[v \in S_t] = \mathbb{P}[y_v \geq t] = \mathbb{P}[y_v^2 \geq t^2] = y_v^2$$

To bound the numerator, we say that an edge is *cut* by S_t if one endpoint is in S_t and another is not. We have

$$\mathbb{E}\, E(S_t, V - S_t) = \sum_{\{u,v\} \in E} \mathbb{P}[\{u,v\} \text{ is cut}]$$

$$= \sum_{\{u,v\} \in E} |y_v^2 - y_u^2| = \sum_{\{u,v\} \in E} |y_v - y_u| \cdot (y_u + y_v)$$

Applying Cauchy-Schwarz, we have

$$\mathbb{E}\, E(S_t, V - S_t) \leq \sqrt{\sum_{\{u,v\} \in E} (y_v - y_u)^2} \cdot \sqrt{\sum_{\{u,v\} \in E} (y_v + y_u)^2}$$

and applying Cauchy-Schwarz again (in the form $(a+b)^2 \leq 2a^2 + 2b^2$) we get

$$\sum_{\{u,v\} \in E} (y_v + y_u)^2 \leq \sum_{\{u,v\} \in E} 2y_v + 2y_u^2 = 2d \sum_v y_v^2$$

Putting everything together gives

$$\frac{\mathbb{E}\, E(S_t, V - S_t)}{\mathbb{E}\, d|S_t|} \leq \sqrt{2 \frac{\sum_{\{u,v\} \in E} (y_v - y_u)^2}{d \sum_v y_v^2}}$$

which is (5.1).

Chapter 6

Cheeger-type Inequalities for λ_n

In which we prove an analog of Cheeger's inequalities for the largest Laplacian eigenvalue and we show how to use it to develop a spectral approximation algorithm for Max Cut.

6.1 Cheeger-type Inequalities for λ_n

Let $G = (V, E)$ be an undirected graph (not necessarily regular), D its diagonal matrix of degrees, A its adjacency matrix, $L = I - D^{-1/2}AD^{-1/2}$ its normalized Laplacian matrix, and $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$ be the eigenvalues of L, counted with multiplicities and listed in non-decreasing order.

In Handout 2, we proved that $\lambda_k = 0$ if and only if G has at least k connected component and $\lambda_n = 2$ if and only if there is a connected component of G (possibly, all of G) that is bipartite.

A special case of the former fact is that $\lambda_2 = 0$ if and only if the graph is disconnected, and the Cheeger inequalities give a "robust" version of this fact, showing that λ_2 can be small if and only if the expansion of the graph is small. In these notes we will see a robust version of the latter fact; we will identify a combinatorial parameter that is zero if and only if the graph has a bipartite connected component, and it is small if and only if the graph is "close" (in an appropriate sense) to having a bipartite connected components, and we will show that this parameter is small if and only if $2 - \lambda_n$ is small.

Recall that

$$2 - \lambda_n = \min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}} \frac{\sum_{\{u,v\} \in E} (x_u + x_v)^2}{\sum_{v \in V} d_v x_v^2}$$

We will study the following combinatorial problem, which formalizes the task of finding an "almost bipartite connected component:" we are looking for a non-empty subset of vertices

$S \subseteq V$ (we allow $S = V$) and a bipartition (A, B) of S such that there is a small number of "violating edges" compared to the number of edges incident on S, where an edge $\{u, v\}$ is *violating* if it is in the cut $(S, V - S)$, if it has both endpoints in A, or if it has both endpoints in B. (Note that if there are no violating edges, then S is a bipartite connected component of G.)

It will be convenient to package the information about A, B, S as a vector $\mathbf{y} \in \{-1, 0, 1\}^n$, where the non-zero coordinates correspond to S, and the partition of S is given by the positive versus negative coordinates. We define the "bipartiteness ratio" of $\mathbf{y}$ as

$$\beta(\mathbf{y}) := \frac{\sum_{\{u,v\}\in E} |y_u + y_v|}{\sum_{v\in V} d_v |y_v|}$$

Note that in the numerator we have the number of violating edges, with edges contained in A or in B counted with a weight of 2, and edges from S to $V - S$ counted with a weight of 1. In the denominator we have the sum of the degrees of the vertices of S (also called the *volume* of S, and written $vol(S)$) which is, up to a factor of 2, the number of edges incident on S.

(Other definitions would have been reasonable, for example in the numerator we could just count the number of violating edges without weights, or we could have the expression $\sum_{\{u,v\}\in E}(y_u + y_v)^2$. Those choices would give similar bounds to the ones we will prove, with different multiplicative constants.)

We define the bipartiteness ratio of G as

$$\beta(G) = \min_{\mathbf{y}\in\{-1,0,1\}^n - \{\mathbf{0}\}} \beta(\mathbf{y})$$

We will prove the following analog of Cheeger's inequalities:

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

The first inequality is the easy direction

$$\begin{aligned} 2 - \lambda_n &= \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}} \frac{\sum_{\{u,v\}\in E}(x_u + x_v)^2}{\sum_{v\in V} d_v x_v^2} \\ &\leq \min_{\mathbf{y}\in\{-1,0,1\}^n - \{\mathbf{0}\}} \frac{\sum_{\{u,v\}\in E} |y_u + y_v|^2}{\sum_{v\in V} d_v |y_v|^2} \\ &\leq \min_{\mathbf{y}\in\{-1,0,1\}^n - \{\mathbf{0}\}} \frac{\sum_{\{u,v\}\in E} 2 \cdot |y_u + y_v|}{\sum_{v\in V} d_v |y_v|} \end{aligned}$$

The other direction follows by applying the following lemma to the case in which $\mathbf{x}$ is the eigenvector of λ_n.

Lemma 6.1 (Main) *For every $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$ there is a threshold t, $0 < t \leq \max_v |x_v|$, such that, if we define $\mathbf{y}^{(t)} \in \{-1, 0, 1\}^n$ as*

$$y_v^{(t)} = \begin{cases} -1 & \text{if } x_v \leq -t \\ 0 & \text{if } -t < x_v < t \\ 1 & \text{if } x_v \geq t \end{cases}$$

we have

$$\beta(\mathbf{y}^{(t)}) \leq \sqrt{2 \cdot \frac{\sum_{\{u,v\} \in E} (x_u + x_v)^2}{\sum_{v \in V} d_v x_v^2}}$$

Note that the Lemma is giving the analysis of an algorithm that is the "bipartite analog" of Fiedler's algorithm. We sort vertices according to $|x_v|$, and then we consider all sets S which are suffixes of the sorted order and cut S into (A, B) according to sign. We pick the solution, among those, with smallest bipartiteness ratio. Given $\mathbf{x}$, such a solution can be found in time $O(|E| + |V| \log |V|)$ as in the case of Fiedler's algorithm.

6.1.1 Proof of Main Lemma

We will assume without loss of generality that $\max_v |x_v| = 1$. (Scaling $\mathbf{x}$ by a multiplicative constant does not change the Rayleigh quotient and does not change the set of $\mathbf{y}$ that can be obtained from $\mathbf{x}$ over the possible choices of thresholds.)

Consider the following probabilistic experiment: we pick t at random in $[0, 1]$ such that t^2 is uniformly distributed in $[0, 1]$, and we define the vector $\mathbf{y}^{(t)}$ as in the statement of the lemma. We claim that

$$\frac{\mathbb{E} \sum_{\{u,v\} \in E} |y_u^{(t)} + y_v^{(t)}|}{\mathbb{E} \sum_{v \in V} d_v |y_v^{(t)}|} \leq \sqrt{2 \cdot \frac{\sum_{\{u,v\} \in E} (x_u + x_v)^2}{\sum_{v \in V} d_v x_v^2}} \tag{6.1}$$

and we note that the Main Lemma follows from the above claim and from the fact, which we have used before, that if X and Y are random variables such that $\mathbb{P}[Y > 0] = 1$, then there is a positive probability that $\frac{X}{Y} \leq \frac{\mathbb{E}\, X}{\mathbb{E}\, Y}$.

We immediately see that

$$\mathbb{E} \sum_{v \in V} d_v |y_v^{(t)}| = \sum_v d_v \, \mathbb{P}[\ |x_v| \geq t\ ] = \sum_v d_v x_v^2$$

To analyze the numerator, we distinguish two cases

1. If x_u and x_v have the same sign, and, let's say, $x_u^2 \leq x_v^2$ then there is a probability x_u^2 that both $y_u^{(t)}$ and $y_v^{(t)}$ are non-zero (and have the same sign), meaning that $|y_u^{(t)} + y_v^{(t)}| = 2$; and there is an additional probability $x_v^2 - x_u^2$ that $y_u^{(t)} = 0$ and $y_v^{(t)} = \pm 1$, so that $|y_u^{(t)} + y_v^{(t)}| = 1$. Overall we have

$$\mathbb{E}\, |y_u^{(t)} + y_v^{(t)}| = 2x_u^2 + x_v^2 - x_u^2 = x_u^2 + x_v^2$$

since the last expression is symmetric with respect to u and v, the equation

$$\mathbb{E}\,|y_u^{(t)} + y_v^{(t)}| = x_u^2 + x_v^2$$

holds also if $x_u^2 \geq x_v^2$;

2. If x_u and x_v have opposite signs, and, let's say, $x_u^2 \leq x_v^2$, there is probability $x_v^2 - x_u^2$ that $y_u^{(t)} = 0$ and $y_v^{(t)} = \pm 1$, in which case $|y_u^{(t)} + y_v^{(t)}| = 1$, and otherwise we have $|y_u^{(t)} + y_v^{(t)}| = 0$. If $x_u^2 \geq x_v^2$, then $|y_u^{(t)} + y_v^{(t)}|$ equals 1 with probability $x_u^2 - x_v^2$ and it equals zero otherwise. In either case, we have

$$\mathbb{E}\,|y_u^{(t)} + y_v^{(t)}| = |x_u^2 - x_v^2|$$

In both cases, the inequality

$$\mathbb{E}\,|y_u^{(t)} + y_v^{(t)}| \leq |x_u + x_v| \cdot (|x_u| + |x_v|)$$

is satisfied.

Applying Cauchy-Schwarz as in the proof of Cheeger's inequalities we have

$$\begin{aligned}\mathbb{E} \sum_{\{u,v\}\in E} |y_u^{(t)} + y_v^{(t)}| &\leq \sum_{\{u,v\}\in E} |x_u + x_v| \cdot (|x_u| + |x_v|) \\ &\leq \sqrt{\sum_{\{u,v\}\in E} (x_u + x_v)^2} \cdot \sqrt{\sum_{\{u,v\}\in E} (|x_u| + |x_v|)^2}\end{aligned}$$

and

$$\sum_{\{u,v\}\in E} (|x_u| + |x_v|)^2 \leq \sum_{\{u,v\}\in E} 2x_u^2 + x_v^2 = 2\sum_v d_v x_v^2$$

and, combining all the bounds, we get (6.1).

6.2 Application to Max Cut

In the Max Cut problem, given an undirected graph $G = (V, E)$ we want to find a cut $(C, V - C)$ maximizing the number of cut edges $E(C, V - C)$. We call $maxcut(G)$ the fraction of edges of G cut by the optimal solution. We see that if $maxcut(G) = 1 - \epsilon$, then $2 - \lambda_n \leq 2\epsilon$, as seen by taking the boolean vector $\mathbf{x} \in \{-1, 1\}^n$ such that $x_v = 1$ iff $v \in C$.

This means that, in polynomial time, we can find a $\mathbf{y} \in \{-1, 0, 1\}$ such that $\beta(\mathbf{y}) \leq 2\sqrt{\epsilon}$. Unfortunately, the number of non-zero vertices in $\mathbf{y}$ could be very small, and so $\mathbf{y}$ would not help find a way to partition all vertices.

We can, however, apply the algorithm of the previous section recursively: after we find $\mathbf{y}$, we remove the non-zero vertices from G (because $\mathbf{y}$ gives us a way to partition them with few violating edges) and then we can recurse on the residual graph. If the (volume of) the residual graph is very small, then we already almost have a global cut, and if the volume

of the residual graph is large, then the optimal cut of G is still a good cut in the residual graph, and it implies that λ_n is close to 2 and that we can find a $\mathbf{y}$ of small bipartiteness ratio, and so on.

The overall algorithm is as follows:

RecursiveSpectralCut $(G = (V, E))$

- Use algorithm of previous section to find disjoint sets of vertices A, B such that

$$2E(A) + 2E(B) + E(A \cup B, V - A \cup B) \leq \sqrt{2 - 2\lambda_n} \cdot vol(A \cup B)$$

- If $V = A \cup B$, then return (A, B)
- Else
  - Let $V' := V - (A \cup B)$, and let $G' = (V', E')$ be the subgraph of G induced by V'
  - $(C, V' - C) :=$ **RecursiveSpectralCut** $(G' = (V', E'))$
  - Return best cut between $(C \cup A, (V' - C) \cup B)$ and $(C \cup B, (V' - C) \cup A)$

We prove the following bound.

Lemma 6.2 *If* $maxcut(G) = 1 - \epsilon$, *then* **RecursiveSpectralCut** G *finds a cut crossed by at least* $(1 - 4\sqrt{\epsilon})|E|$ *edges.*

PROOF: We proceed by induction on the number of recursive steps.

If there is no recursive call, then $A \cup B = V$, and so (A, B) is already a cut. The number of edges of G that *do not* cross the cut is

$$E(A) + E(B) \leq \frac{1}{2}\sqrt{2 \cdot (2 - \lambda_n)} \cdot vol(V) \leq 2\sqrt{\epsilon} \cdot |E| \leq 4\sqrt{\epsilon}|E|$$

because $2 - \lambda_n \leq 2\epsilon$ and $vol(V) = 2|E|$. This settles the base case.

For the inductive step, the number of edges not cut by the algorithm is at most

$$E(A) + E(B) + \frac{1}{2}E(A \cup B, V') + (|E'| - E'(C, V' - C))$$

because we should count all the edges with both endpoints in A and both endpoints in B, all the edges of G' not cut in the recursive step, and half of the edges from $A \cup B$ to V', because the best way of combining the cuts loses at most half of those edges.

By using the fact that $2 - \lambda_n \leq 2\epsilon$ and the inductive hypothesis we have

$$E(A) + E(B) + \frac{1}{2}E(A \cup B, V') \leq \sqrt{\epsilon} \cdot vol(A \cup B)$$

$$|E'| - E'(C, V' - C) \leq 4\sqrt{\epsilon'} \cdot |E'|$$

where ϵ' is defined so that $1 - \epsilon' = maxcut(G')$.

Let us call $\rho := \frac{|E|-|E'|}{|E|}$ the fraction of edges of G that is not in G'. Then we have

$$vol(A \cup B) \leq 2|E - E'| = 2\rho|E|$$

and

$$|E'| = |E| \cdot (1 - \rho)$$

We also have

$$\epsilon'|E'| \leq \epsilon|E|$$

because the number of edges not cut by the optimal cut of G' is at most the number of edges not cut by the optimal cut of G, given that G' is a subgraph of G. So we have

$$\epsilon' \leq \epsilon \cdot \frac{|E|}{|E'|} = \epsilon \cdot \frac{1}{1 - \rho}$$

Putting everything together, the number of uncut edges is at most

$$\begin{aligned} &\sqrt{\epsilon} \cdot vol(A \cup B) + 4\sqrt{\epsilon'} \cdot |E'| \\ &\leq \sqrt{\epsilon} \cdot 2\rho|E| + 4\sqrt{\epsilon \cdot (1 - \rho)}|E| \\ &\leq 4\sqrt{\epsilon}|E| \end{aligned}$$

where the last line is equivalent to

$$2\rho + 4\sqrt{1 - \rho} \leq 4$$

which follows from the fact that

$$\sqrt{1 - \rho} \leq \sqrt{\left(1 - \frac{\rho}{2}\right)^2} = 1 - \frac{\rho}{2}$$

□

For small ϵ, the bound of the previous lemma is close to the bound achieved by the Goemans-Williamson semidefinite programming-based algorithm, that, under the assumption that $maxcut(G) = 1 - \epsilon$, finds a cut crossed by about a $1 - \frac{2}{\pi}\sqrt{\cdot}\epsilon$ fraction of edges, which is best possible for polynomial time algorithms assuming the unique games conjecture.

The bound of the lemma is not very good for large ϵ, however: if $\epsilon > \frac{1}{64}$, the lemma guarantees a number of cut edges that is less than half the number of edges, which is worse than the performance of a simple greedy algorithm, and if $\epsilon > \frac{1}{16}$ the lemma does not guarantee than any edge at all is cut.

If one always chooses the best between the cut returned by the recursion and a greedy cut, it is possible to derive an analysis that works well even in graphs in which the value of $maxcut(G)$ is small, and show that the algorithm finds a cut crossed by at least a $.531 \cdot maxcut(G)$ fraction of edges.

Chapter 7

Cheeger-type Inequalities for λ_k

In which we state an analog of Cheeger's inequalities for the k-th smallest Laplacian eigenvalue, and we discuss the connection between this result and the analysis of spectral partitioning algorithms

7.1 Cheeger-type Inequalities for λ_k

Let $G = (V, E)$ be an undirected d-regular graph, A its adjacency matrix, $L = I - \frac{1}{d}A$ its normalized Laplacian matrix, and $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$ be the eigenvalues of L, counted with multiplicities and listed in non-decreasing order.

In Handout 2, we proved that $\lambda_k = 0$ if and only if G has at least k connected components, that is, if and only if there are k disjoint sets $S_1, \ldots, S_k$ such that $\phi(S_i) = 0$ for $i = 1, \ldots, k$. In this lecture and the next one we will prove a robust version of this fact.

First we introduce the notion of higher-order expansion. If $S_1, \ldots, S_k$ is a collection of disjoint sets, then their order-k expansion is defined as

$$\phi_k(S_1, \ldots, S_k) = \max_{i=1,\ldots,k} \quad \phi(S_i)$$

and the order-k expansion of a graph G is

$$\phi_k(G) = \min_{S_1,\ldots,S_k \text{ disjoint}} \quad \phi(S_1, \ldots, S_k)$$

If the edges of a graph represent a relation of similarity of affinity, a low-expansion collection of sets $S_1, \ldots, S_k$ represents an interesting notion of clustering, because the vertices in each set S_i are more related to each other than to the rest of the graph. (Additional properties are desirable in a good clustering, and we will discuss this later.)

We will prove the following higher-order Cheeger inequalities:

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^{3.5})\sqrt{\lambda_k}$$

Stronger upper bounds are known, but the bound above is easier to prove from scratch. It is known that $\phi_k(G) \leq O(k^2)\sqrt{\lambda_k}$ and that $\phi_k(G) \leq O_\epsilon(\sqrt{\log k}) \cdot \sqrt{\lambda_{(1+\epsilon)\cdot k}}$.

7.2 The Easy Direction

As usual, the direction $\frac{\lambda_k}{2} \leq \phi_k(G)$ is the easy one, and it comes from viewing λ_k as a sort of continuous relaxation of the problem of minimizing order-k expansion.

Recall that, in order to prove the easy direction of Cheeger's inequality for λ_2, we proved that if $\mathbf{x}$ and $\mathbf{y}$ are two orthogonal vectors, both of Rayleigh quotient at most ϵ, then the Rayleigh quotient of their sum is at most 2ϵ. A similar argument could be made to show that the Rayleigh quotient of the sum of k such vectors is at most $k\epsilon$. Such results hold for every positive semidefinite matrix.

In the special case of the Laplacian of a graph, and of vectors that are not just orthogonal but actually *disjointly supported*, then we can lose only a factor of 2 instead of a factor of k. (The *support* of a vector is the set of its non-zero coordinates; two vectors are disjointly supported if their supports are disjoint.)

Lemma 7.1 *Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ be disjointly supported vectors. Then*

$$R_L\left(\sum_i \mathbf{x}^{(i)}\right) \leq 2 \cdot \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)})$$

Proof: We just have to prove that, for every edge $\{u, v\}$,

$$\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2 \leq 2\sum_i (x_u^{(i)} - x_v^{(i)})^2$$

The support disjointness implies that there is an index j such that $x_u^{(i)} = 0$ for $i \neq j$, and an index k such that $x_v^{(i)} = 0$ for $i \neq k$. If $j = k$, then

$$\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2 = (x_u^{(j)} - x_v^{(j)})^2 = \sum_i (x_u^{(i)} - x_v^{(i)})^2$$

and, if $j \neq k$, then

$$\left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2 = (x_u^{(j)} - x_v^{(k)})^2$$

$$\leq 2(x_u^{(j)})^2 + 2(x_v^{(k)})^2 = 2\sum_i (x_u^{(i)} - x_v^{(i)})^2$$

and now, using also the fact that disjoint support implies orthogonality, we have

$$R_L\left(\sum_i \mathbf{x}^{(i)}\right) = \frac{\sum_{\{u,v\}} \left(\sum_i x_u^{(i)} - x_v^{(i)}\right)^2}{\left\|\sum_i \mathbf{x}^{(i)}\right\|^2}$$

$$\leq 2\frac{\sum_i \sum_{\{u,v\}\in E}(x_u^{(i)} - x_v^{(i)})^2}{\sum_i ||\mathbf{x}^{(i)}||^2}$$

$$\leq 2 \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)})$$

□

To finish the proof of the easy direction, let $S_1, \ldots, S_k$ be sets such that $\phi(S_i) \leq \phi(G)$ for every i. Consider the k-dimensional space X of linear combinations of the indicator vectors $\mathbf{1}_{S_i}$ of such sets. The indicator vectors have Rayleigh quotient at most $\phi(G)$ and are disjointly supported, so all their linear combinations have Rayleigh quotient at most $2\phi(G)$. We have found a k-dimensional space of vectors all of Rayleigh quotient $\leq 2\phi(G)$, which proves $\lambda_k \leq 2\phi(G)$.

7.3 The Difficult Direction: Main Lemma

We will prove the following result

Lemma 7.2 (Main) *Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ be orthonormal vectors. Then we can find disjointly supported non-negative vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ such that for every $i = 1, \ldots, k$*

$$R_L(\mathbf{y}^{(i)}) \leq O(k^7) \cdot \max_{j=1,\ldots,k} R_L(\mathbf{x}^{(j)})$$

By applying the Main Lemma to the eigenvectors of $\lambda_1, \ldots, \lambda_k$, we get disjointly supported vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ all of Rayleigh quotient at most $O(k^7) \cdot \lambda_k$. In a past lecture, we proved that for every non-negative vector $\mathbf{y}$ there is a subset S of its support such that $\phi(S) \leq \sqrt{2R_L(\mathbf{y})}$, and applying this fact to the vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ we find k disjoint sets all of expansion at most $O(k^{3.5}) \cdot \sqrt{\lambda_k}$, proving

$$\phi_k(G) \leq O(k^{3.5}) \cdot \sqrt{\lambda_k}$$

It is possible, with a more involved proof, to improve the $O(k^7)$ factor in the conclusion of the Main Lemma to $O(k^6)$, implying that $\phi_k(G) \leq O(k^3) \cdot \sqrt{\lambda_k}$. A different approach, which we will not discuss, is used to show that, given k orthonormal vectors, one can find k disjoint sets $S_1, \ldots, S_k$ such that, for all i,

$$\phi(S_i) \leq O(k^2) \cdot \sqrt{\max_{j=1,\ldots,k} R_L(\mathbf{x}^{(j)})}$$

implying $\phi_k(G) \leq O(k^2) \cdot \sqrt{\lambda_k}$, which is the best known bound.

Note that, in all the known arguments, the bounds still hold if one replaces λ_k by the (possibly smaller) quantity

$$\inf_{\mathbf{x}^{(1)},\ldots,\mathbf{x}^{(k)} \text{ orthonormal}} \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)}) \tag{7.1}$$

There are graphs, however, in which

$$\phi_k(G) \geq \Omega(\sqrt{k}) \cdot \sqrt{\inf_{\mathbf{x}^{(k)},\ldots,\mathbf{x}^{(k)} \text{ orthonormal}} \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)})}$$

so, if a bound of the form $\phi_k(G) \leq (\log k)^{O(1)} \cdot \sqrt{\lambda_k}$ is true, then, in order to prove it, we need to develop new techniques that distinguish between λ_k and the quantity (7.1).

7.4 The Spectral Embedding

Given orthonormal vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ as in the premise of the Main Lemma, we define the mapping $F : V \to \mathbb{R}^k$

$$F(v) := (x_v^{(1)}, \ldots, x_v^{(k)}) \tag{7.2}$$

If $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$ are the eigenvectors of the k smallest Laplacian eigenvalues of L, then $F(\cdot)$ is called the *spectral embedding* of G into $\mathbb{R}^k$. *Spectral clustering* algorithms compute such an embedding, and then find clusters of nodes by clustering the points $\{F(v) : v \in V\}$ using geometric clustering algorithms, such as k-means, according either to Euclidian distance, or to the normalized distance function

$$dist(u, v) := \left\| \frac{F(u)}{||F(u)||} - \frac{F(v)}{||F(v)||} \right\| \tag{7.3}$$

Our construction of disjointly supported vectors with small Rayleigh quotient will proceed similarly, by working only with the points $\{v : F(v)\}$ and forgetting the edge structure of the graph, and by making use of the above distance function.

To develop some intuition about the spectral mapping, we introduce a notion of Laplacian Rayleigh quotient for a mapping $f : V \to \mathbb{R}^k$, defined, by formally replacing absolute values with norms, as

$$R_L(f) := \frac{\sum_{\{u,v\}} ||f(u) - f(v)||^2}{d \sum_v ||f(v)||^2}$$

For a mapping $F : V \to \mathbb{R}^k$ defined in terms of k orthonormal vectors $\mathbf{x}^{(i)}$ as in (8.1), we have

$$
\begin{aligned}
R_L(F) &= \frac{\sum_{\{u,v\}} \sum_i (x_u^{(i)} - x_v^{(i)})^2}{d \sum_v \sum_i (x_v^{(i)})^2} \\
&= \frac{\sum_i \sum_{\{u,v\}} (x_u^{(i)} - x_v^{(i)})^2}{dk} \\
&= \frac{1}{k} \sum_i \frac{\sum_{\{u,v\}} (x_u^{(i)} - x_v^{(i)})^2}{d} \\
&= \frac{1}{k} \sum_i R_L(\mathbf{x}^{(i)}) \\
&\leq \max_{i=1,\ldots,k} R_L(\mathbf{x}^{(i)})
\end{aligned}
$$

In particular, if $\mathbf{x}^{(i)}$ are the eigenvectors of the k smallest Laplacian eigenvalues, then $R_L(F) \leq \lambda_k$.

Let us use this setup to prove again that if $\lambda_k = 0$ then G has at least k connected components. If $\lambda_k = 0$, and we construct $F(\cdot)$ using the eigenvectors of the smallest Laplacian eigenvalues, then $R_L(F) = 0$, which means that $F(u) = F(v)$ for every edge $\{u, v\}$, and so $F(u) = F(v)$ for every u and v which are in the same connected component. Equivalently, if $F(u) \neq F(v)$, then u and v are in different connected component. For every point in the range $\{F(v) : v \in V\}$ in the range of $F(\cdot)$, let us consider its pre-image, and let $S_1, \ldots, S_t$ be the sets constructed in this way. Clearly, every set has expansion zero.

How many sets do we have? We claim that the range of $F(\cdot)$ must contain at least k distinct points, and so $t \geq k$ and G has at least k connected component. To prove the claim, consider the matrix X whose rows are the vectors $\mathbf{x}^{(i)}$; since the rows of X are linearly independent, X has full rank k; but if the range of $F(\cdot)$ contained $\leq k-1$ distinct points, then X would have $\leq k-1$ distinct columns, and so its rank would be $\leq k-1$.

Our proof of the higher-order Cheeger inequality will be somewhat analogous to the previous argument: we will use the fact that, if the Rayleigh quotient of $F(\cdot)$ is small, then the endpoints of edges $\{u, v\}$ are typically close, in the sense that the distance defined in (8.2) between u and v will typically be small; we will also use the fact that, because the $\mathbf{x}^{(i)}$ are orthonormal, $F(\cdot)$ tends to "spread out" vertices across $\mathbb{R}^k$, so that we can find k regions of $\mathbb{R}^k$ each containing a large (in a certain weighted sense) number of vertices, and such that the regions are well-separated according to the distance (8.2), implying that there are few edges crossing from one region to the other, so that the vertices in each region are a non-expanding set. (This is an imprecise description of the argument, but it conveys the basic intuition.)

7.5 Overview of the Proof of the Main Lemma

We will break up the proof of the Main Lemma into the following two Lemmas.

Lemma 7.3 (Well-Separated Sets) *Given a function $F : V \to \mathbb{R}^k$ defined in terms of k orthonormal vectors as in (8.1), we can find k disjoint subsets of vertices $A_1, \ldots, A_k$ such that*

- *For every $i = 1, \ldots, k$, $\sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{4}$*
- *For every u and v belonging to different sets, $dist(u, v) \geq \Omega(k^{-3})$*

Lemma 7.4 (Localization) *Given a function $F : V \to \mathbb{R}^k$ defined in terms of k orthonormal vectors as in (8.1), and t sets $A_1, \ldots, A_t$ such that, for every $i = 1, \ldots, t$, $\sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{4}$ and,for every u, v in different sets $dist(u, v) \geq \delta$, we can construct t disjointly supported vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(t)}$ such that for every $i = 1, \ldots, t$, we have*

$$R_L(\mathbf{y}^{(t)}) \leq O(k \cdot \delta^{-2}) \cdot R_L(F)$$

Chapter 8

Cheeger-type Inequalities for λ_k, cont'd

In which we state an analog of Cheeger's inequalities for the k-th smallest Laplacian eigenvalue, and we discuss the connection between this result and the analysis of spectral partitioning algorithms

8.1 Review

Let $G = (V, E)$ be a d-regular undirected graph, L its normalized Laplacian, $0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2$ be the Laplacian eigenvalues, and $\phi_k(G)$ be the order-k expansion of G.

We want to prove

$$\phi_k(G) \leq O(k^{3.5}) \cdot \sqrt{\lambda_k}$$

We will prove the somewhat stronger result that, given k orthonormal vectors $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(k)}$, we can find k disjointly supported vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(k)}$ such that, for every $i = 1, \ldots, k$,

$$R_L(\mathbf{y}^{(i)}) \leq O(k^7) \cdot \max_{j=1,\ldots,k} R_L(\mathbf{x}^{(i)})$$

In order to do that, we define the mapping

$$F(v) := (x_v^{(1)}, \ldots, x_v^{(k)}) \tag{8.1}$$

of vertices to $\mathbb{R}^k$ and the normalized distance

$$dist(u, v) := \left\| \frac{F(u)}{||F(u)||} - \frac{F(v)}{||F(v)||} \right\| \tag{8.2}$$

between vertices, and we are going to prove the following two lemmas.

Lemma 8.1 (Localization) *Given t sets $A_1, \ldots, A_t$ such that, for every $i = 1, \ldots, t$, $\sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{2}$ and,for every u, v in different sets $dist(u, v) \geq \delta$, we can construct t disjointly supported vectors $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(t)}$ such that for every $i = 1, \ldots, t$, we have*

$$R_L(\mathbf{y}^{(t)}) \leq O(k \cdot \delta^{-2}) \cdot R_L(F)$$

Lemma 8.2 (Well-Separated Sets) *There are k disjoint subsets of vertices $A_1, \ldots, A_k$ such that*

- *For every $i = 1, \ldots, k$, $\sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{2}$*
- *For every u and v belonging to different sets, $dist(u, v) \geq \Omega(k^{-3})$*

Recall that, for a function $f : V \to \mathbb{R}^k$ the Rayleigh quotient of f is defined as

$$R_L(f) := \frac{\sum_{\{u,v\}} ||f(u) - f(v)||^2}{d \sum_v ||f(v)||^2}$$

and, by definition of F, we have

$$R_L(F) = \frac{1}{k} \sum_i R_L(\mathbf{x}^{(i)})$$

8.2 Some Preliminaries

We will prove some simple properties of the embedding $F(\cdot)$ and of the distance function $dist(\cdot, \cdot)$.

First, we observe that

$$\sum_{v \in V} ||F(v)||^2 = \sum_v \sum_i (x_v^{(i)})^2 = \sum_i ||\mathbf{x}^{(i)}||^2 = k$$

Next, we prove the sense in which $F(\cdot)$ "spreads out" vertices across $\mathbb{R}^k$.

Lemma 8.3 *For every unit vector $\mathbf{w} \in \mathbb{R}^k$,*

$$\sum_{v \in V} \langle F(v), \mathbf{w} \rangle^2 = 1$$

PROOF: Consider the $k \times n$ matrix X whose rows are the vectors $\mathbf{x}^{(i)}$ and whose columns are the points $F(v)$. Then we have

$$\sum_{v \in V} \langle F(v), \mathbf{w} \rangle^2 = ||X^T \mathbf{w}||^2 = \mathbf{w}^T X X^T \mathbf{w} = \mathbf{w}^T \mathbf{w} = 1$$

where we used the fact that the rows of X are orthogonal and so XX^T is the identity. $\square$

This means that, for every direction, the points $F(v)$ correlate with that direction in the same way, regardless of the direction itself.

In the proof of the localization lemma, we will make use of the following inequality: for every two vectors $\mathbf{x}, \mathbf{y}$,

$$| \, ||\mathbf{x}|| - ||\mathbf{y}|| \, | \leq ||\mathbf{x} - \mathbf{y}||$$

which is a consequence of Cauchy-Schwarz:

$$\begin{aligned}(||\mathbf{x}|| - ||\mathbf{y}||)^2 &= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2||\mathbf{x}|| \cdot ||\mathbf{y}|| \\ &\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle \\ &= ||\mathbf{x} - \mathbf{y}||^2\end{aligned}$$

8.3 Localization

In this section we prove Lemma 8.1.

8.3.1 Proof Ideas

The basic idea is that we would like to define the vectors $\mathbf{y}^{(i)}$ as

$$y_v^{(i)} := \mathbf{1}_{A_i}(v) \cdot ||F(v)||$$

The denominator of the Rayleigh quotient of such a vector is, by definition, at least $1/2$, and we might hope to upper bound the numerator of the Rayleigh quotient of $\mathbf{y}^{(i)}$ in terms of the numerator of the Rayleigh quotient of F, which is $kR_L(F)$.

Indeed, every edge $\{u, v\}$ with both endpoints outside of A_i contributes zero to the numerator of the Rayleigh quotient of $\mathbf{y}^{(i)}$, and every edge $\{u, v\}$ with both endpoints in A_i contributes

$$(\ ||F(u)|| - ||F(v)||\ )^2 \leq ||F(u) - F(v)||^2$$

to the numerator of the Rayleigh quotient of $\mathbf{y}^{(i)}$, and the right-hand-side above is the contribution of the edge to the numerator of the Rayleigh quotient of F.

So far, so good, but the problem comes from edges $\{u, v\}$ with one endpoint $u \in A_i$ and one endpoint $v \notin A_i$. Such an edge contributes $||F(u)||^2$ to the Rayleigh quotient of $\mathbf{y}^{(i)}$ and $||F(u) - F(v)||^2$ to the Rayleigh quotient of F, and the former quantity could be much larger than the latter. If $dist(u, v)$ is large, however, $||F(u)||^2$ cannot be much larger than $||F(u) - F(v)||^2$, because of the following fact

Lemma 8.4

$$||F(v)|| \cdot dist(u,v) \leq 2||F(u) - F(v)|| \tag{8.3}$$

PROOF:

$$\begin{aligned}
||F(v)|| \cdot dist(u,v) &= ||F(v)|| \cdot \left\| \frac{F(u)}{||F(u)||} - \frac{F(v)}{||F(v)||} \right\| \\
&= \left\| F(u) \cdot \frac{||F(v)||}{||F(u)||} - F(v) \right\| \\
&\leq \left\| F(u) \cdot \frac{||F(v)||}{||F(u)||} - F(u) \right\| + \|F(u) - F(v)\| \\
&= \left\| F(u) \cdot \left(\frac{||F(v)||}{||F(u)||} - 1 \right) \right\| + \|F(u) - F(v)\| \\
&= ||F(u)|| \cdot \left| \frac{||F(v)||}{||F(u)||} - 1 \right| + \|F(u) - F(v)\| \\
&= |\ ||F(v)|| - ||F(u)||\ | + \|F(u) - F(v)\| \\
&\leq 2\|F(u) - F(v)\|
\end{aligned}$$

□

We can conclude that the only problem comes from edges $\{u, v\}$ such that $u \in A_i$, $v \notin A_i$, and $dist(u, v)$ is small. To deal with such edges, we will modify the definition of $\mathbf{y}^{(i)}$, and use a "smoothed" version of the indicator function of A_i instead of the actual indicator.

8.3.2 Proof

If v is a vertex and A is a set of vertices, we define

$$dist(v, A) = \min_{u \in A} dist(v, u)$$

For each i, we define the following *smooth indicator function* of A_i:

$$\tau_i(v) = \begin{cases} 1 & \text{if } v \in A_i \\ 0 & \text{if } dist(v, A_i) \geq \frac{\delta}{2} \\ 1 - \frac{2}{\delta} \cdot dist(v, A_i) & \text{otherwise} \end{cases}$$

Notice that the functions $\tau_i(\cdot)$ are disjointly supported: there cannot be a vertex v such that $\tau_i(v) > 0$ and $\tau_j(v) > 0$ for $i \neq j$, otherwise we would have $dist(v, A_i) < \frac{\delta}{2}$ and $dist(v, A_j) < \frac{\delta}{2}$, contradicting the well-separated condition on the sets A_i.

We define the vectors $\mathbf{y}^{(i)}$ as

$$y_v^{(i)} = \tau_i(v) \cdot ||F(v)||$$

The vectors $\mathbf{y}^{(i)}$ are disjointly supported, and it remains to understand their Rayleigh quotient.

The denominator of the Rayleigh quotient of $\mathbf{y}^{(i)}$ is

$$\sum_{v \in V} \tau_i^2(v) \cdot ||F(v)||^2 \geq \sum_{v \in A_i} ||F(v)||^2 \geq \frac{1}{2}$$

The contribution of an edge $\{u, v\}$ to the numerator is the square of

$$\begin{aligned}
|y_v^{(i)} - y_u^{(i)}| &= |\ \tau_i(v) \cdot ||F(v)|| - \tau_i(u) \cdot ||F(u)||\ | \\
&\leq |\ \tau_i(v) \cdot ||F(v)|| - \tau_i(v) \cdot ||F(u)||\ | + |\ \tau_i(v) \cdot ||F(u)|| - \tau_i(u) \cdot ||F(u)||\ | \\
&= \tau_i(v) \cdot ||F(v) - F(u)|| + ||F(u)|| \cdot |\tau_i(v) - \tau_i(u)| \\
&\leq ||F(v) - F(u)|| + ||F(u)|| \cdot \frac{2}{\delta} \cdot dist(v, u) \\
&\leq ||F(v) - F(u)|| \cdot \left(1 + \frac{4}{\delta}\right)
\end{aligned}$$

where we used the inequality

$$|\tau_i(v) - \tau_i(u)| \leq \frac{2}{\delta} |dist(v, A_i) - dist(u, A_i)| \leq \frac{2}{\delta} dist(v, u)$$

The numerator of the Rayleigh quotient of $\mathbf{y}^{(i)}$ is thus

$$\sum_{\{u,v\} \in E} |y_v^{(i)} - y_u^{(i)}|^2 \leq O(\delta^{-2}) \sum_{\{u,v\} \in E} ||F(v) - F(u)||^2 = O(\delta^{-2}) \cdot kR_L(F)$$

and this proves Lemma 8.1.

8.4 Well-Separated Sets

In this section we prove Lemma 8.2, which follows easily from the following result.

Lemma 8.5 *We can find disjoint sets of vertices $T_1, \ldots, T_m$ such that*

- $\sum_{i=1}^{m} \sum_{v \in T_i} ||F(v)||^2 \geq k - \frac{1}{4}$
- *For every u, v in different sets, $dist(u, v) \geq \Omega(k^{-3})$*
- *For every set T_i, $\sum_{v \in T_i} ||F(v)||^2 \leq 1 + \frac{1}{4k}$*

We can derive Lemma 8.2 from Lemma 8.5 as follows. Let us call the quantity $\sum_{v\in A} ||F(v)||^2$ the *mass* of a set A. Starting from the sets $T_1, \ldots, T_m$ promised by Lemma 8.5 we run the following process: as long as there are two sets both of mass $< \frac{1}{2}$ we merge them. Call $A_1, \ldots, A_t$ the sets of mass $\geq \frac{1}{2}$ obtained at the end of this process; in addition, there may be one more set of mass $< \frac{1}{2}$. Every set has mass $\leq 1 + \frac{1}{4k}$. This means that the total mass of the sets is at most $\frac{1}{2} + t \cdot \left(1 + \frac{1}{4k}\right) \geq k - \frac{1}{4}$, which implies $t \geq k$. Thus we have found k sets of vertices, each of mass at least $1/2$, and such that any two sets have separation $\Omega(k^{-3})$.

Now we turn to the proof of Lemma 8.5. We are going to use the fact that, for every small cone in $\mathbb{R}^k$, the mass of vertices v such that $F(v)$ is in the cone is also small and, in particular, it can made at most $1 + \frac{1}{4k}$. We will prove the Lemma by covering almost all the points $F(v)$ using a collection of well-separated small cones.

We first formalize the above intuition about cones. If R (for *r*egion) is a subset of the unit sphere in $\mathbb{R}^n$, then the diameter of R is

$$diam(R) := \sup_{\mathbf{w},\mathbf{z}\in R} ||\mathbf{w} - \mathbf{z}||$$

and the *cone* generated by R is the set $\{\alpha\mathbf{w} : \alpha \in \mathbb{R}_{\geq 0}, \mathbf{w} \in R\}$ of non-negative scalar multiples of elements of R. The set of vertices covered by R, denoted $V(R)$ is the set of vertices v such that $F(v)$ is in the cone generated by R or, equivalently,

$$V(R) := \left\{ v \in V : \frac{F(v)}{||F(v)||} \in R \right\}$$

If R has small diameter, then $V(R)$ has small mass.

Lemma 8.6 *For every subset R of the unit sphere,*

$$\sum_{v\in V(R)} ||F(v)||^2 \leq \left(1 - \frac{1}{2}(diam(R))^2\right)^{-2}$$

PROOF: For every two unit vectors $\mathbf{w}$ and $\mathbf{z}$, we have

$$\langle \mathbf{z}, \mathbf{w} \rangle = 1 - \frac{1}{2}||\mathbf{w} - \mathbf{z}||^2$$

For every vertex v, call

$$\bar{F}(v) := \frac{F(v)}{||F(v)||}$$

Let $\mathbf{w}$ be a vector in R. Then we have

$$1 \geq \sum_{v\in V(R)} \langle F(v), \mathbf{w} \rangle^2$$

$$= \sum_{v \in V(R)} ||F(v)||^2 \cdot \langle \bar{F}(v), \mathbf{w} \rangle^2$$

$$= \sum_{v \in V(R)} ||F(v)||^2 \cdot \left(1 - \frac{1}{2}||\bar{F}(v) - \mathbf{w}||^2\right)^2$$

$$\geq \sum_{v \in V(R)} ||F(v)||^2 \cdot \left(1 - \frac{1}{2}(diam(R))^2\right)^2$$

□

In particular, if $diam(R) \leq \frac{1}{\sqrt{5}k}$, then the mass of $V(R)$ is at most

$$\left(1 - \frac{1}{10k}\right)^{-2} \leq \left(1 - \frac{1}{5k}\right)^{-1} = 1 + \frac{1}{5k-1} \leq 1 + \frac{1}{4k}$$

Another observation is that, for every two subsets R_1, R_2 of the unit sphere,

$$\min_{u \in V(R_1), v \in V(R_2)} dist(u, v) \geq \min_{\mathbf{w} \in R_1, \mathbf{z} \in R_2} ||\mathbf{w} - \mathbf{z}||$$

Our approach will be to find disjoint subsets $R_1, \ldots, R_m$ of the unit sphere, each of diameter at most $1/2\sqrt{k}$, such that the total mass of the sets $V(R_1), \ldots, V(R_m)$ is at least $k - \frac{1}{4}$ and such that the separation between any two R_i, R_j is at least $\Omega(k^{-3})$.

To do this, we tile $\mathbb{R}^k$ with axis-parallel cubes of side $L = \frac{1}{2k}$, which clearly have diameter at most $\frac{1}{2\sqrt{k}}$, and, for every cube C, we define its *core* $\tilde{C}$ to be a cube with the same center as C and of side $L \cdot \left(1 - \frac{1}{4k^2}\right)$. Note two points in the core of two different cubes have distance at least $\frac{1}{8k^3}$. Let now $R_1, R_2, \ldots$ be the intersections of the cube cores with the unit sphere. Since each R_i is a subset of a core of a cube, it has diameter at most $\frac{1}{2\sqrt{k}}$, and the distance between any two points in different regions is at least $\frac{1}{8k^3}$. We claim that there is a way to choose the location of the centers of the cubes so that $\sum_i \sum_{v \in V(R_i)} ||F(v)||^2 \geq k - \frac{1}{4}$.

Let us start by a fixed configuration of the cubes and then apply an axis-parallel random shift (by adding to each coordinate, a random real in the range $[0, L]$. Then, for each fixed point in $\mathbb{R}^n$ and, in particular, for each point $\bar{F}(v)$, the probability that it falls in the core of a cube after the shift is at least $1 - \frac{1}{4k}$, so the average mass of the vertices covered by the regions is at least $k - \frac{1}{4}$, and there must exist a shift that is at least as good.

Chapter 9

Spectral Algorithms Wrap-up

In which we talk about even more generalizations of Cheeger's inequalities, and we analyze the power method to find approximate eigenvectors, thus having a complete description of a polynomial-time approximation algorithm for sparsest cut

9.1 Irregular Graphs

For simplicity, we proved our results on λ_2 and λ_k for regular graphs. Those results extend, essentially with the same proofs, to the case of irregular undirected graphs. In an irregular graph $G = (V, E)$, the notion that generalizes edge expansion is called *conductance*. If d_v is the degree of vertex v, then the conductance of set S of vertices is

$$\phi(S) := \frac{E(S, V - S)}{\sum_{v \in S} d_v}$$

We will call the sum of the degrees of a set of vertices the *volume* of the set, and denote it $vol(S) := \sum_{v \in S} d_v$. The conductance of the graph G is

$$\phi(G) := \min_{S : vol(S) \leq \frac{1}{2} vol(V)} \phi(S)$$

Higher-order conductance is defined as higher-order expansion, but with conductance replacing expansion in the definition.

The Cheeger inequalities

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

still hold, with the same proof. With some abuse of notation, we will call the following quantity the Rayleigh quotient of $\mathbf{x}$

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

even if, technically, it is the Rayleigh quotient of $D^{1/2}\mathbf{x}$, where D is the diagonal matrix of degrees.

We can also adapt the proof of the higher-order Cheeger inequality to show

$$\frac{\lambda_k}{2} \leq \phi_k(G) \leq O(k^{3.5}) \cdot \sqrt{\lambda_k}$$

9.2 More Cheeger-type Bounds

We proved that if $(S_F, V - S_F)$ is the cut found by Fiedler's algorithm using the eigenvector of λ_2, then

$$\phi(S_F, V - S_F) \leq 2\sqrt{\phi(G)}$$

which is a good bound, although it usually underestimates the quality of the solutions found in practice. (There are graphs, however, for which the above inequality is tight within a constant factor.)

One case in which we can improve the analysis is when there are not too many eigenvalues close to λ_2

Theorem 9.1 *There is a constant c such that, if $(S_F, V - S_F)$ is the cut obtained by Fiedler's algorithm using an eigenvector for λ_2, then, for every $k \geq 2$,*

$$\phi(S_F, V - S_F) \leq c \cdot k \cdot \frac{\lambda_2}{\sqrt{\lambda_k}}$$

So we have

$$\phi(S_F, V - S_F) \leq 2c \cdot \phi(G) \cdot \min_{k \geq 2} \frac{k}{\sqrt{\lambda_k}}$$

which is a better bound for families of graphs in which, for some k, $\lambda_k >> k^2\lambda_2$.

We will not have time to prove Theorem 9.1, but we will state the two main pieces of its proof.

Lemma 9.2 *Let $\mathbf{x} \in \mathbb{R}^V_{\geq 0}$ be a non-negative vector. Then, for every k, there is a non-negative vector $\mathbf{y} \in \mathbb{R}^V_{\geq 0}$ whose entries take at most $2k$ distinct values and such that*

$$||\mathbf{x} - \mathbf{y}||^2 \leq \frac{R_L(\mathbf{x})}{\lambda_k} ||\mathbf{x}||^2$$

That is, if $R_L(\mathbf{x}) >> \lambda_k$, then there are $2k$ values such that most entries of $\mathbf{x}$ are close to one of those $2k$ values.

Lemma 9.3 *There is a constant c' such that, for every non-negative vectors $\mathbf{x} \in \mathbb{R}^V_{\geq 0}$ and $\mathbf{y} \in \mathbb{R}^V_{\geq 0}$, if $\mathbf{y}$ is such that its entris contain only k distinct values, then there is a threshold $t > 0$ such that*

$$\phi(\{v : x_v \geq t\}) \leq c' \cdot k \cdot \left(R_L(\mathbf{x}) + \sqrt{R_L(\mathbf{x})} \cdot \frac{||\mathbf{x} - \mathbf{y}||}{||\mathbf{x}||} \right)$$

The above lemma should be compared to the fact, which was a major piece in the proof of Cheeger's inequality, that if $\mathbf{x} \in \mathbb{R}^V_{\geq 0}$ is an arbitrary non-negative vector, then there is a threshold $t > 0$ such that

$$\phi(\{v : x_v \geq t\}) \leq \sqrt{2R_L(\mathbf{x})}$$

One obtains Theorem 9.1 in the following. Start from an eigenvector $\mathbf{x}$ of λ_2 and, using the first step in the proof of Cheeger's inequality, obtain a vector $\mathbf{x}' \in \mathbb{R}^V_{\geq 0}$ with non-negative entries such that $R_L(\mathbf{x}') \leq R_L(\mathbf{x}) = \lambda_2$ and such that the support of $\mathbf{x}$ contains at most $|V|/2$ vertices.

Use Lemma 9.2 to find a vector $\mathbf{y}$ with non-negative entries and with at most $2k$ distinct values among its entries such that $||\mathbf{x}' - \mathbf{y}||^2 \leq \frac{\lambda_2}{\lambda_k}||\mathbf{x}'||^2$. Then use Lemma 9.3 and the fact that $\lambda_k \leq 2$ to conclude that there exists at $t > 0$ such that

$$\phi(\{v : x'_v \geq t\}) \leq O(k) \cdot \frac{\lambda_2}{\sqrt{\lambda_k}}$$

The set $\{v : x'_v \geq t\}$ contains at most $|V|/2$ vertices, it is one of the cuts considered by Fiedler's algorithm on input $\mathbf{x}$.

Another property of graphs in which λ_k is large for small k is that they contain large expanders as induced subgraphs.

Theorem 9.4 *There is a constant c such that, for every graph G and every k, there exists a partition of the vertices into $\ell \leq k$ sets $(S_1 \ldots, S_\ell)$ such that, if we call G_i the subgraph induced by the vertex set S_i, we have*

$$\phi_{G_i} \geq c\frac{\lambda_k}{k^2}$$

Theorem 9.5 *If $\phi_{k+1} > (1 + \epsilon)\phi_k$, then there is a partition of the vertices into k subsets $(S_1, \ldots, S_k$ such that*

$$\forall i \in \{1, \ldots, k\} : \quad \phi_{G_i} \geq \Omega\left(\frac{\epsilon}{k}\right) \cdot \phi_{k+1}, \quad \phi(S_i) \leq k\phi_k$$

9.3 The Power Method

Earlier in this class, 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 λ_2, Fiedler's algorithm 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 $\phi(S) \leq \sqrt{4\phi(G) + 2\epsilon}$. We will now see how to compute 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.

9.3.1 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 ± 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 9.6 *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 λ_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 9.6 is implied by the following two lemmas

Lemma 9.7 *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 9.8 *For every $\mathbf{x} \in \mathbb{R}^n$, for every positive integer k and positive $\epsilon > 0$, if we define $\mathbf{y} := M^k \mathbf{x}$, we have*

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1 - \epsilon) \cdot \left(1 + \frac{||\mathbf{x}||^2}{\langle \mathbf{x}, \mathbf{v}_1 \rangle^2} (1 - \epsilon)^{2k}\right)^{-1}$$

It remains to prove the two lemmas.

Proof: (Of Lemma 9.7) 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

$$\mathbb{P}[Z \geq \delta\, \mathbb{E}\, Z] \geq (1 - \delta)^2 \cdot \frac{(\mathbb{E}\, Z)^2}{\mathbb{E}\, Z^2} \tag{9.1}$$

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 9.9 *The proof of Lemma 9.7 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 9.8) 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 $a_i = \langle \mathbf{x}, \mathbf{v}_i \rangle$. 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 ℓ 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

$$\sum_{i=\ell+1}^{n} a_i^2 \lambda_i^{2k}$$

$$\leq \lambda_1^{2k} \cdot (1-\epsilon)^{2k} \sum_{i=\ell+1}^{n} a_i^2$$

$$\leq \lambda_1^{2k} \cdot (1-\epsilon)^{2k} \cdot ||\mathbf{x}||^2$$

$$\leq a_1^2 \lambda_1^{2k} (1-\epsilon)^{2t} \cdot \frac{||\mathbf{x}||^2}{a_1^2}$$

$$\leq \frac{||\mathbf{x}||^2}{a_1^2} (1-\epsilon)^{2k} \sum_{i=1}^{\ell} a_i^2 \lambda_i^{2k}$$

So we have

$$\mathbf{y}^T \mathbf{y} \leq \left(1 + \frac{||\mathbf{x}||^2}{a_1^2}(1-\epsilon)^{2k}\right) \cdot \sum_{i=1}^{\ell} a_i^2$$

giving

$$\frac{\mathbf{y}^T M \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \geq \lambda_1 \cdot (1-\epsilon) \cdot \left(1 + \frac{||\mathbf{x}||^2}{a_1^2}(1-\epsilon)^{2k}\right)^{-1}$$

□

Remark 9.10 *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?*

9.3.2 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 λ_2, then M is a PSD linear map from the orthogonal space to $\mathbf{v}_1$ to itself, and λ_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 9.7.) 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 9.8 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 9.11 *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 λ_2 is the second largest eigenvalue of M, counting multiplicities.

9.3.3 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 λ_2 is, say $O(1/|V|)$.

For a running time that is nearly linear in n for all values of λ_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^+.

(Such algorithms will be discussed in the third part of the course. The algorithms will find an approximate solution $\mathbf{y}$ to the linear system $L\mathbf{y} = \mathbf{x}$, but this will be sufficient. In the following, we proceed as if the solution was exact.)

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 ℓ 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

$$\begin{aligned}
\sum_i a_i^2 \frac{1}{\lambda_i^{2k-1}} &\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \sum_{i > \ell} a_i^2 \\
&\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \cdot n \\
&\leq \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}} + \frac{1}{(1+\epsilon)^{2k-1} \lambda_2^{2k-1}} \cdot 4n a_2^2 \\
&\leq \left(1 + \frac{4n}{(1+\epsilon)^{2k-1}}\right) \cdot \sum_{i \leq \ell} a_i^2 \frac{1}{\lambda_i^{2k-1}}
\end{aligned}$$

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

Chapter 10

The Sparsest Cut Problem

In which we introduce the sparsest cut problem and the Leighton-Rao relaxation.

10.1 The Uniform Sparsest Cut problem, Edge Expansion and λ_2

Let $G = (V, E)$ be an undirected graph with $n := |V|$ vertices.

We define the uniform sparsity of a cut $(S, V - S)$ as

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

(we will omit the subscript when clear from the context) and the uniform sparsest cut of a graph is

$$usc(G) := \min_S usc_G(S)$$

In d-regular graphs, approximating the uniform sparsest cut is equivalent (up to a factor of 2 in the approximation) to approximating the edge expansion, because, for every cut $(S, V - S)$, we have

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

and, noting that, for every, S,

$$\frac{1}{n}|S| \cdot |V - S| \leq \min\{|S|, |V - S|\} \leq \frac{2}{n}|S| \cdot |V - S|$$

we have, for every S,

$$\phi(S, V-S) \leq \frac{n}{d} \cdot usc(S) \leq 2\phi(S, V-S)$$

and so

$$\phi(G) \leq \frac{n}{d} \cdot usc(G) \leq 2\phi(G)$$

It will be instructive to see that, in d-regular graphs, λ_2 is a relaxation of $\frac{n}{d} usc(G)$, a fact that gives an alternative proof of the easy direction $\lambda_2 \leq 2\phi(G)$ of Cheeger's inequalities.

If G is d-regular, then λ_2 satisfies

$$\begin{aligned}
\lambda_2 &= \min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{\sum_v d x_v^2} \\
&= \min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}, \mathbf{x} \perp \mathbf{1}} \frac{n}{d} \cdot \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{\sum_{\{u,v\}} (x_u - x_v)^2} \\
&= \min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}} \frac{n}{d} \cdot \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{\sum_{\{u,v\}} (x_u - x_v)^2}
\end{aligned}$$

where the first identity above comes from the fact that

$$\begin{aligned}
\sum_{\{u,v\}} (x_u - x_v)^2 = \frac{1}{2} \sum_{(u,v) \in V^2} (x_u - x_v)^2 &= n \sum_v x_v^2 - \sum_{u,v} x_u x_v \\
&= n \sum_v x_v^2 - \left(\sum_v x_v \right)^2 \\
&= n \sum_v x_v^2 - \langle \mathbf{x}, \mathbf{1} \rangle^2 \\
&= n \sum_v x_v^2
\end{aligned}$$

and the second identity follows by noticing that the cost function is invariant by addition of a multiple of $\mathbf{1}$, and so optimizing over all non-zero vectors gives the same result as optimizing over all vectors orthogonal to $\mathbf{1}$.

On the other hand, the uniform sparsest cut problem can be formulated as

$$usc(G) = \min_{\mathbf{x} \in \{0,1\}^n - \{\mathbf{0}, \mathbf{1}\}} \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{\sum_{\{u,v\}} (x_u - x_v)^2}$$

(because the square of a number in $\{-1, 0, 1\}$ is the same as its absolute value) and we see that λ_2 can be considered a continuous relaxation of $\frac{n}{d} usc(G)$.

10.2 The Non-Uniform Sparsest Cut Problem

In the non-uniform sparsest cut problem, we are given two graphs $G = (V, E_G)$ and $H = (V, E_H)$, over the same set of vertices; the non-uniform sparsity of a cut $(S, V - S)$ is defined as

$$nsc_{G,H}(S) := \frac{E_G(S, V - S)}{E_H(S, V - S)}$$

and the non-uniform sparsest cut problem is the optimization problem

$$nsc(G, H) := \min_S \quad nsc_{G,H}(S)$$

Note that the non-uniform sparsest cut problem generalizes the sparsest cut problem (consider the case in which H is a clique).

If H is the graph that contains the single edge $\{s, t\}$, then $nsc(G, H)$ is the undirected min-st-cut problem, in which we want to find the cut that separates two vertices s and t and that minimizes the number of crossing edges.

10.3 The Leighton-Rao Relaxation

We can write the non-uniform sparsity of a set as

$$nsc_{G,H}(S) = \frac{\sum_{\{u,v\} \in E_G} |1_S(u) - 1_S(v)|}{\sum_{\{u,v\} \in E_H} |1_S(u) - 1_S(v)|}$$

The observation that led us to see λ_2 as the optimum of a continuous relaxation of $\frac{n}{d} nsc_{G,K_n}$ was to observe that $|1_S(u) - 1_S(v)| = |1_S(u) - 1_S(v)|^2$, and then relax the problem by allowing arbitrary functions $x : V \to \mathbb{R}$ instead of indicator functions $1_S : V \to \{0, 1\}$.

The Leighton-Rao relaxation of sparsest cut is obtained using, instead, the following observation: if, for a set S, we define $d_S(u, v) := |1_S(i) - 1_S(j)|$, 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 nsc, and define

$$LR(G, H) := \min_{\substack{d : V \times V \to \mathbb{R} \\ d \text{ semi-metric}}} \frac{\sum_{\{u,v\} \in E_G} d(u, v)}{\sum_{\{u,v\} \in E_H} d(u, v)} \tag{10.1}$$

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

$$LR(G, H) \leq nsc(G, H) \leq O(\log |V|) \cdot LR(G, H) \tag{10.2}$$

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

$$\begin{array}{lll} \text{minimize} & \sum_{\{u,v\}\in E_G} d_{u,v} & \\ \text{subject to} & & \\ & \sum_{\{u,v\}\in E_H} d_{u,v} = 1 & \\ & d_{u,v} \le d_{u,z} + d_{z,v} & \forall u,v,z \in V \\ & d_{u,v} \ge 0 & \forall u,v \in V \end{array} \tag{10.3}$$

that has a variable $d_{u,v}$ for every unordered pair of distinct vertices $\{u,v\}$. Clearly, every solution to the linear program (10.3) is also a solution to the right-hand side of the definition (10.1) of the Leighton-Rao parameter, with the same cost. Also every semi-metric can be normalized so that $\sum_{\{u,v\}\in E_H} d(u,v) = 1$ by multiplying every distance by a fixed constant, and the normalization does not change the value of the right-hand side of (10.1); after the normalization, the semimetric is a feasible solution to the linear program (10.3), with the same cost.

10.4 An L1 Relaxation of Sparsest Cut

In the Leighton-Rao relaxation, we relax distance functions of the form $d(i,j) = |1_S(i) - 1_S(j)|$ 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 ℓ_1 space*.

Recall that, for a vector $\mathbf{x} \in \mathbb{R}^n$, its ℓ_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 ℓ_1 distance function

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

The distance function $d(u,v) = |1_S(u) - 1_S(v)|$ 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 ℓ_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 ℓ_1 mappings, and define

$$\ell_1 nsc(G,H) := \inf_{m, f:V\to\mathbb{R}^m} \frac{\sum_{\{u,v\}\in E_G} ||f(u) - f(v)||_1}{\sum_{\{u,v\}\in E_H} ||f(u) - f(v)||_1}$$

This may seem like another very broad relaxation of sparsest cut, whose optimum might be much smaller than the sparsest cut optimum. The following theorem shows that this is not the case.

Theorem 10.1 *For every graphs G, H, $nsc(G,H) = \ell_1 nsc(G,H)$.*

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

$$\frac{\sum_{\{u,v\}\in E_G} |1_S(u) - 1_S(v)|}{\sum_{\{u,v\}\in E_H} |1_S(u) - 1_S(v)|} \leq \frac{\sum_{\{u,v\}\in E_G} ||f(u) - f(v)||_1}{\sum_{\{u,v\}\in E_H} ||f(u) - f(v)||_1} \tag{10.4}$$

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

$$\frac{a_1 + \cdots a_m}{b_1 + \cdots b_m} \geq \min_i \frac{a_i}{b_i} \tag{10.5}$$

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 (10.4) by coordinates, and write

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

This already shows that, in the definition of ϕ', we can map, with no loss of generality, to 1-dimensional ℓ_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 it has a unit-length range $\max_v g(v) - \min_v g(v) = 1$.

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

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

Finally, by an application of (10.5), we see that there must be a set S among the possible values of S_t such that (10.4) 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. □

10.5 A Theorem of Bourgain

We will derive our main result (10.2) 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 10.2 (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 $u, v \in V$,*

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

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

$$\begin{aligned}
LR(G, H) &= \frac{\sum_{\{u,v\}\in E_G} d(u, v)}{\sum_{\{u,v\}\in E_H} d(u, v)} \\
&\geq \frac{\sum_{\{u,v\}\in E_G} ||f(u) - f(v)||_1}{c \cdot \log |V| \cdot \sum_{\{u,v\}\in E_H} ||f(u) - f(v)||_1}
\end{aligned}$$

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

Chapter 11

Proof of Bourgain's Theorem

In which we prove Bourgain's theorem.

Today we prove the following theorem.

Theorem 11.1 (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 $|V|$.

Together with the results that we proved in the last lecture, this implies that an optimal solution to the Leighton-Rao relaxation can be rounded to an $O(\log n)$-approximate solution to the sparsest cut problem. This was the best known approximation algorithm for sparsest cut for 15 years, until the Arora-Rao-Vazirani algorithm, which will be our next topic.

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.

11.1 Preliminary 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 11.2 *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 11.3 (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 11.4 (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 11.5 (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 11.6 *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 11.7 (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 11.8 (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 t uniformly at random in $\{0, \ldots, \log_2 n\}$
2. Sample $A \subseteq V$ by selecting each $v \in V$, independently, to be in A with probability 2^{-t} and to be in $V - A$ with probability $1 - 2^{-t}$.

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 $\approx 2^t$.

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 11.9 (Tree) *Consider a complete binary tree, and the shortest-path metric $d(\cdot,\cdot)$ in the tree. Take any two vertices u and v at distance $\frac{1}{2}\log n$. If we look at the ball of radius d_1 around u and the ball of radius $d_2 = d_1 + \epsilon \log n$ around v, we see that the former has 2^{d_1} points in it, and the latter has $2^{d_1} \cdot n^\epsilon$ points: it is clearly hopeless to have constant probability of hitting the former and of missing the latter.*

For every $t < \frac{1}{2}\log n$, however, we have

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

because there is a constant probability of hitting one of the 2^{t+1} points at distance $\leq t$ from u, so that $f_A(u) \leq t$ and also a constant probability of missing the 2^{t+2} points at distance $\geq t+1$ from v, in which case $f_A(v) \geq t+1$. 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, in some cases, we have 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 2^t elements.

11.2 The Proof of Bourgain's Theorem

Given Fact 11.2 and Fact 11.6, proving Bourgain's theorem reduces to proving the following theorem.

Theorem 11.10 *For a finite set of points V, consider the distribution D over subsets of V sampled by uniformly picking a scale $t \in \{0, \ldots, \log_2 |V|\}$ and then picking independently each $v \in V$ to be in A with probability 2^{-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: For each t, let ru_t be the distance from u to the 2^t-th closest point to u (counting u). That is,

$$|\{w : d(u, w) < ru_t\}| < 2^t$$

and

$$|\{w : d(u, w) \leq ru_t\}| \geq 2^t$$

and define rv_t similarly. Let t^* be the scale such that both ru_{t^*} and rv_{t^*} are smaller than $d(u, v)/3$, but at least one of ru_{t^*+1} or rv_{t^*+1} are $\geq d(u, v)/3$.

Define

$$ru'_t := \min\{ru_t, d(u, v)/3\}$$

and similarly

$$rv'_t := \min\{rv_t, d(u, v)/3\}$$

We claim that there is an absolute constant c such that for every scale $t \in \{0, \ldots, t^*\}$, we have

$$\mathop{\mathbb{E}}_{A\sim D_t} |f_A(u) - f_A(v)| \geq c \cdot (ru'_{t+1} + rv'_{t+1} - ru'_t - rv'_t) \tag{11.1}$$

We prove the claim by showing that there are two disjoint events, each happening with probability $\geq c$, such that in one event $|f_A(u) - f_A(v)| \geq ru'_{t+1} - rv'_t$, and in the other event $|f_A(u) - f_A(v)| \geq rv'_{t+1} - ru'_t$.

1. The first event is that A avoids the set $\{z : d(u,z) < ru'_{t+1}\}$ and intersects the set $\{z : d(v,z) \leq rv'_t\}$. The former set has size $< 2^{t+1}$, and the latter set has size $\leq 2^t$; 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,

$$|f_A(u) - f_A(v)| \geq f_A(u) - f_A(v) \geq ru'_{t+1} - rv'_t$$

2. The second event is that A avoids the set $\{z : d(v,z) < rv'_{t+1}\}$ and intersects the set $\{z : d(u,z) \leq ru'_t\}$. The former set has size $< 2^{t+1}$, and the latter set has size $\leq 2^t$; 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,

$$|f_A(u) - f_A(v)| \geq f_A(v) - f_A(u) \geq rv'_{t+1} - ru'_t$$

So we have established (11.1). Averaging over all scales, we have

$$\begin{aligned} &\mathop{\mathbb{E}}_{A \sim D} |f_A(u) - f_A(v)| \\ &\geq \frac{c}{1 + \log_2 n} \cdot (ru'_{t^*+1} + rv'_{t^*+1} - ru'_0 - rv'_0) \\ &\geq \frac{c}{1 + \log_2 n} \cdot \frac{d(u,v)}{3} \end{aligned}$$

□

There is one remaining point to address. In Fact 11.2, 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 11.11 *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

$$\mathbb{P}\left[Z \leq \frac{c}{2 \log |V|} d(u, v) \right]$$
$$\leq \mathbb{P}\left[Z \leq \mathbb{E}\, Z - \frac{c}{2 \log |V|} d(u, v) \right]$$
$$\leq 2^{-2mc^2/(2 \log |V|)^2}$$

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,

$$\mathbb{P}\left[\forall u, v.\ ||f(u) - f(v)||_1 \geq \frac{c}{2 \log |V|} \cdot d(u, v) \right] \geq 1 - \frac{1}{|V|}$$

Chapter 12

ARV

In which we introduce semi-definite programming and a semi-definite programming relaxation of sparsest cut, and we reduce its analysis to a key lemma that we will prove in the next lecture(s)

12.1 The Goemans-Linial Relaxation

Recall that, for two undirected graphs G, H, the sparsest cut problem is to optimize

$$nsc(G,H) := \min_{S \subseteq V} \frac{\sum_{\{u,v\} \in E_G} |1_S(u) - 1_S(v)|}{\sum_{\{u,v\} \in E_H} |1_S(u) - 1_S(v)|}$$

and the Leighton-Rao relaxation is obtained by noting that if we define $d(u,v) := |1_S(u) - 1_S(v)|$ then $d(\cdot,\cdot)$ is a semimetric over V, meaning that the following quantity is a relaxation of (G,H):

$$LR(G,H) = \min_{\substack{d : V \times V \to R \\ d \text{ semimetric}}} \frac{\sum_{\{u,v\} \in E_G} d(u,v)}{\sum_{\{u,v\} \in E_H} d(u,v)}$$

If G is r-regular, H is a clique, and $0 = \lambda_1 \leq \lambda_2 \leq \cdots \geq \lambda_n$ are the eigenvalues of the normalized Laplacian of G, then

$$\frac{r}{n}\lambda_2 = \min_{f : V \to \mathbb{R}} \frac{\sum_{\{u,v\} \in E_G} |f(u) - f(v)|^2}{\sum_{\{u,v\} \in E_{K_n}} |f(u) - f(v)|^2} \tag{12.1}$$

which is a relaxation of $nsc(G, K_n)$, because, for every S, every u and every v, $|1_S(u) - 1_S(v)| = |1_S(u) - 1_S(v)|^2$.

We note that if we further relax (12.1) by allowing V to be mapped into a higher dimension space $\mathbb{R}^m$ instead of $\mathbb{R}$, and we replace $|\cdot - \cdot|$ by $||\cdot - \cdot||^2$, the optimum remains the same.

Fact 12.1

$$\lambda_2 = \inf_{m, F:V\to\mathbb{R}^m} \frac{\sum_{\{u,v\}\in E_G} ||F(u) - F(v)||^2}{\sum_{\{u,v\}\in E_{K_n}} ||F(u) - F(v)||^2}$$

PROOF: For every $F : V \to \mathbb{R}^m$, if we write $F(v) = (f_1(v), \ldots, f_n(v))$, we have

$$\begin{aligned}
&\frac{\sum_{\{u,v\}\in E_G} ||F(u) - F(v)||^2}{\sum_{\{u,v\}\in E_{K_n}} ||F(u) - F(v)||^2} \\
&= \frac{\sum_i \sum_{\{u,v\}\in E_G} (f_i(u) - f_i(v))^2}{\sum_i \sum_{\{u,v\}\in E_{K_n}} (f_i(u) - f_i(v))^2} \\
&\geq \min_{i=1,\ldots,m} \frac{\sum_{\{u,v\}\in E_G} (f_i(u) - f_i(v))^2}{\sum_{\{u,v\}\in E_{K_n}} (f_i(u) - f_i(v))^2} \\
&\geq \lambda_2
\end{aligned}$$

□

The above observations give the following comparison between the Leighton-Rao relaxation and the spectral relaxation: both are obtained by replacing $|1_S(u) - 1_S(v)|$ with a "distance function" $d(u, v)$; in the Leighton-Rao relaxation, $d(u, v)$ is constrained to satisfy the triangle inequality; in the spectral relaxation, $d(u, v)$ is constrained to be the square of the Euclidean distance between $F(u)$ and $F(v)$ for some mapping $F : V \to \mathbb{R}^m$.

The Arora-Rao-Vazirani relaxation is obtained by *enforcing both conditions*, that is, by considering distance functions $d(u, v)$ that satisfy the triangle inequality *and* can be realized of $||F(u) - F(v)||^2$ for some mapping $F : V \to \mathbb{R}^m$.

Definition 12.2 *A semimetric $d : V \to V \to \mathbb{R}$ is called of negative type if there is a dimension m and a mapping $F : V \to \mathbb{R}^m$ such that $d(u, v) = ||F(u) - F(v)||^2$ for every $u, v \in V$.*

With the above definition, we can formulate the Goemans-Linial relaxation as

$$ARV(G, H) := \min_{\substack{d : V\times V\to R \\ d \text{ semimetric of negative type}}} \frac{\sum_{\{u,v\}\in E_G} d(u, v)}{\sum_{\{u,v\}\in E_H} d(u, v)} \tag{12.2}$$

Remark 12.3 *The relaxation (12.2) was first proposed by Goemans and Linial. Arora, Rao and Vazirani were the first to prove that it achieves an approximation guarantee which is better than the approximation guarantee of the Leighton-Rao relaxation.*

We have, by definition,

$$nsc(G,H) \leq ARV(G,H) \leq LR(G,H)$$

and, when H is a clique and G is r-regular,

$$nsc(G,K_n) \leq ARV(G,K_n) \leq \frac{r}{n}\lambda_2(G)$$

and so the approximation results that we have proved for λ_2 and LR apply to ARV. For every graphs G and H:

$$ARV(G,H) \leq O(\log|V|) \cdot nsc(G,H)$$

and for every r-regular graph G

$$\frac{n}{r}ARV(G,K_n) \leq \sqrt{8\frac{n}{r} \cdot nsc(G,K_n)}$$

Interestingly, known examples of graphs for which LR and λ_2 give poor approximation are complementary. When H is a clique, if G is a cycle, then $\frac{r}{n}\lambda_2$ is a poor approximation of $nsc(G,K_n)$, but $LR(G,K_n)$ is a good approximation of $nsc(G,K_n)$; if G is a constant-degree expander then $LR(G,K_n)$ is a poor approximation of $nsc(G,K_n)$, but $\frac{r}{n}\lambda_2$ is a good approximation.

When Goemans and Linial (separately) proposed to study the relaxation (12.2), they conjectured that it would always provide a constant-factor approximation of $nsc(G,H)$. Unfortunately, the conjecture turned out to be false, but Arora, Rao and Vazirani were able to prove that (12.2) does provide a strictly better approximation than the Leighton-Rao relaxation. In the next lectures, we will present parts of the proof of the following results.

Theorem 12.4 *There is a constant c such that, for every graph $G=(V,E)$,*

$$nsc(G,K_n) \leq c \cdot \sqrt{\log|V|} \cdot ARV(G,K_n)$$

Theorem 12.5 *There is a constant c such that, for every graphs $G=(V,E_G)$, $H=(V,E_H)$,*

$$nsc(G,H) \leq c \cdot \sqrt{\log|V|} \cdot \log\log|V| \cdot ARV(G,H)$$

Theorem 12.6 *There is a constant c and an infinite family of graphs $G_n=(V_n,E_n)$ such that*

$$nsc(G_n,K_n) \geq c \cdot \log\log|V_n| \cdot ARV(G_n,K_n)$$

Theorem 12.7 *There are families of graphs $G_n = (V_n, EG_n)$ and $H_n(V_n, EH_n)$ such that, for every $\epsilon > 0$ and every sufficiently larger n,*

$$ARV(G_n, K_n) \geq (\log |V|)^{\frac{1}{2}-\epsilon} \cdot nsc(G_n, K_n)$$

12.2 Polynomial Time Solvability

In this section we show that the Ellipsoid algorithm can compute $ARV(G, H)$ in polynomial time.

Definition 12.8 *If $C \subseteq \mathbb{R}^m$ is a set, then a separation oracle for C is a procedure that, on input $\mathbf{x} \in R^m$,*

- *If $\mathbf{x} \in C$, outputs "yes"*
- *If $\mathbf{x} \notin C$, outputs coefficients $a_1, \ldots, a_m, b$ such that*

$$\sum_i x_i a_i < b$$

but, for every $\mathbf{z} \in C$,

$$\sum_i z_i a_i \geq b$$

Note that a set can have a separation oracle only if it is convex. Under certain additional mild conditions, if C has a polynomial time computable separation oracle, then the optimization problem

$$\begin{array}{ll} \text{minimize} & \sum_i \mathbf{c}^T \mathbf{x} \\ \text{subject to} & \\ & A\mathbf{x} \geq b \\ & \mathbf{x} \in C \end{array}$$

is solvable in polynomial time using the Ellipsoid Algorithm.

It remains to see how to put the Arora-Rao-Vazirani relaxation into the above form.

Recall that a matrix $X \in \mathbb{R}^{n \times n}$ is *positive semidefinite* if all its eigenvalues are nonnegative. We will use the set of all $n \times n$ positive semidefinite matrices as our set C (thinking of them as n^2-dimensional vectors). If we think of two matrices $M, M' \in \mathbb{R}^{n \times n}$ as n^2-dimensional vectors, then their "inner product" is

$$M \bullet M' := \sum_{i,j} M_{i,j} \cdot M'_{i,j}$$

Lemma 12.9 *The set of $n \times n$ positive semidefinite matrices has a separation oracle computable in time polynomial in n.*

PROOF: Given a symmetric matrix X, its smallest eigenvalue is

$$\min_{\mathbf{z}\in\mathbb{R}^n,\ ||\mathbf{z}||=1} \mathbf{z}^T X \mathbf{z}$$

the vector achieving the minimum is a corresponding eigenvector, and both the smallest eigenvalue and the corresponding eigenvector can be computed in polynomial time.

If we find that the smallest eigenvalue of X is non-negative, then we answer "yes." Otherwise, if $\mathbf{z}$ is an eigenvector of the smallest eigenvalue we output the matrix $A = \mathbf{z}^T\mathbf{z}$. We see that we have

$$A \bullet X = \mathbf{z}^T X \mathbf{z} < 0$$

but that, for every positive semidefinite matrix M, we have

$$A \bullet M = \mathbf{z}^T M \mathbf{z} \geq 0$$

□

This implies that any optimization problem of the following form can be solved in polynomial time

$$\begin{array}{ll} \text{minimize} & C \bullet X \\ \text{subject to} & \\ & A^1 \bullet X \geq b_1 \\ & \cdots \\ & A^m \bullet X \geq b_m \\ & X \succeq 0 \end{array} \tag{12.3}$$

where $C, A^1, \ldots, A^m$ are square matrices of coefficients, $b_1, \ldots, b_m$ are scalars, and X is a square matrix of variables. An optimization problem like the one above is called a *semidefinite program.*

It remains to see how to cast the Arora-Rao-Vazirani relaxation as a semidefinite program.

Lemma 12.10 *For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, the following properties are equivalent:*

1. *M is positive semidefinite;*
2. *there are vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ such that, for all i, j, $M_{i,j} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$;*
3. *for every vector $\mathbf{z} \in \mathbb{R}^n$, $\mathbf{z}^T M \mathbf{z} \geq 0$*

PROOF: That (1) and (3) are equivalent follows from the characterization of the smallest eigenvalue of M as the minimum of $\mathbf{z}^T M \mathbf{z}$ over all unit vectors $\mathbf{z}$.

To see that (2) $\Rightarrow$ (3), suppose that vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ exist as asserted in (2), and let X be the matrix whose columns are the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, so that $X^T \cdot X = M$. Take any vector $\mathbf{z}$, and see that

$$\mathbf{z}^T M \mathbf{z} = \mathbf{z}^T X^T X \mathbf{z} = ||X\mathbf{z}||^2 \geq 0$$

Finally, to see that (1) $\Rightarrow$ (2), let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of M with multiplicities, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a corresponding orthonormal set of eigenvectors. Then

$$M = \sum_i \lambda_k \mathbf{v}_k \mathbf{v}_k^T$$

that is,

$$M_{i,j} = \sum_k \lambda_k v_k(i) v_k(j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

if we define $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as the vectors such that $x_i(k) := \sqrt{\lambda_k} v_k(i)$. $\square$

This means that the generic semidefinite program (12.4) can be rewritten as an optimization problem in which the variables are the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as in part (2) of the above lemma.

$$\begin{array}{ll} \text{minimize} & \sum_{i,j} C_{i,j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ \text{subject to} & \\ & \sum_{i,j} A^1_{i,j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \geq b_1 \\ & \cdots \\ & \sum_{i,j} A^m_{i,j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle \geq b_m \\ & \mathbf{x}_i \in \mathbb{R}^m \qquad\qquad \forall i \in \{1, \ldots, n\} \end{array} \tag{12.4}$$

where the dimension m is itself a variable (although one could fix it, without loss of generality, to be equal to n). In this view, a semidefinite program is an optimization problem in which we wish to select n vectors such that their pairwise inner products satisfy certain linear inequalities, while optimizing a cost function that is linear in their pairwise inner product.

The square of the Euclidean distance between two vectors is a linear function of inner products

$$||\mathbf{x} - \mathbf{y}||^2 = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$$

and so, in a semidefinite program, we can include expressions that are linear in the pairwise squared distances (or squared norms) of the vectors. The ARV relaxation can be written as follows

$$
\begin{array}{ll}
\text{minimize} & \sum_{\{u,v\}\in E_G} ||\mathbf{x}_u - \mathbf{x}_v||^2 \\
\text{subject to} & \\
& \sum_{\{u,v\}\in E_H} ||\mathbf{x}_u - \mathbf{x}_v||^2 = 1 \\
& ||\mathbf{x}_u - \mathbf{x}_v||^2 \le ||\mathbf{x}_u - \mathbf{x}_w||^2 + ||\mathbf{x}_w - \mathbf{x}_v||^2 \quad \forall u,v,w \in V \\
& \mathbf{x}_u \in \mathbb{R}^m \qquad\qquad\qquad\qquad\qquad\qquad\quad \forall u \in V
\end{array}
$$

and so it is a semidefinite program, and it can be solved in polynomial time.

Remark 12.11 *Our discussion of polynomial time solvability glossed over important issues about numerical precision. To run the Ellipsoid Algorithm one needs, besides the separation oracle, to be given a ball that is entirely contained in the set of feasible solutions and a ball that entirely contains the set of feasible solutions, and the running time of the algorithm is polynomial in the size of the input, polylogarithmic in the ratio of the volumes of the two balls, and polylogarithmic in the desired amount of precision. At the end, one doesn't get an optimal solution, which might not have a finite-precision exact representation, but an approximation within the desired precision. The algorithm is able to tolerate a bounded amount of imprecision in the separation oracle, which is an important feature because we do not have exact algorithms to compute eigenvalues and eigenvectors (the entries in the eigenvector might not have a finite-precision representation).*

The Ellipsoid algorithm is typically not a practical algorithm. Algorithms based on the *interior point* method have been adapted to semidefinite programming, and run both in worst-case polynomial time and in reasonable time in practice.

Arora and Kale have developed an $\tilde{O}((|V|+|E|)^2/\epsilon^{O(1)})$ time algorithm to solve the ARV relaxation within a multiplicative error $(1+\epsilon)$. The dependence on the error is worse than that of generic algorithms, which achieve polylogarithmic dependency, but this is not a problem in this application, because we are going to lose an $O(\sqrt{\log |V|})$ factor in the rounding, so an extra constant factor coming from an approximate solution of the relaxation is a low-order consideration.

12.3 Rounding when H is a clique

Given the equivalence between the sparsest cut problem and the "$L1$ relaxation" of sparsest cut, it will be enough to prove the following result.

Theorem 12.12 (Rounding of ARV) *Let $G = (V,E)$ be a graph, and $\{\mathbf{x}_v\}_{v\in V}$ be a feasible solution of the relaxation $ARV(G, K_n)$.*

Then there is a mapping $f : V \to \mathbb{R}$ such that

$$
\frac{\sum_{\{u,v\}\in E} |f(u) - f(v)|}{\sum_{\{u,v\}} |f(u) - f(v)|} \le O(\sqrt{\log |V|}) \cdot \frac{\sum_{\{u,v\}\in E} ||\mathbf{x}_u - \mathbf{x}_v||^2}{\sum_{\{u,v\}} ||\mathbf{x}_u - \mathbf{x}_v||^2}
$$

As in the rounding of the Leighton-Rao relaxation via Bourgain's theorem, we will identify a set $S \subseteq V$, and define

$$f_S(v) := \min_{s \in S} ||\mathbf{x}_s - \mathbf{x}_v||^2 \tag{12.5}$$

Recall that, as we saw in the proof of Bourgain's embedding theorem, no matter how we choose the set S we have

$$|f_S(u) - f_S(v)| \leq ||\mathbf{x}_u - \mathbf{x}_v||^2 \tag{12.6}$$

where we are not using any facts about $||\cdot - \cdot||^2$ other than the fact that, for solutions of the ARV relaxation, it is a distance function that obeys the triangle inequality.

This means that, in order to prove the theorem, we just have to find a set $S \subseteq V$ such that

$$\sum_{u,v} |f_S(u) - f_S(v)| \geq \frac{1}{O(\sqrt{\log |V|})} \cdot \sum_{u,v} ||\mathbf{x}_u - \mathbf{x}_v||^2 \tag{12.7}$$

and this is a considerable simplification because the above expression is completely independent of the graph! The remaining problem is purely one about geometry.

Recall that if we have a set of vectors $\{\mathbf{x}_v\}_{v \in V}$ such that the distance function $d(u,v) := ||\mathbf{x}_u - \mathbf{x}_v||^2$ satisfies the triangle inequality, then we say that $d(\cdot,\cdot)$ is a (semi-)metric of *negative type*.

After these preliminaries observations, our goal is to prove the following theorem.

Theorem 12.13 (Rounding of ARV – Revisited) *If $d(\cdot,\cdot)$ is a semimetric of negative type over a set V, then there is a set S such that if we define*

$$f_S(v) := \min_{s \in S}\{d(s,v)\}$$

we have

$$\sum_{u,v} |f_S(u) - f_S(v)| \geq \frac{1}{O(\sqrt{\log |V|})} \cdot \sum_{u,v} d(u,v)$$

Furthermore, the set S can be found in randomized polynomial time with high probability given a set of vector $\{\mathbf{x}_v\}_{v \in V}$ such that $d(u,v) = ||\mathbf{x}_u - \mathbf{x}_v||^2$.

Since the statement is scale-invariant, we can restrict ourselves, with no loss of generality, to the case $\sum_{u,v} d(u,v) = |V|^2$.

Remark 12.14 Let us discuss some intuition before continuing with the proof.

As our experience proving Bourgain's embedding theorem shows us, it is rather difficult to pick sets such that $|f_S(u) - f_S(v)|$ is not much smaller than $d(u,v)$. Here we have a somewhat simpler case to solve because we are not trying to preserve all distances, but only the average

pairwise distance. A simple observation is that if we find a set S which contains $\Omega(|V|)$ elements and such that $\Omega(|V|)$ elements of V are at distance $\Omega(\delta)$ from S, then we immediately get $\sum_{u,v} |f_S(u) - f_S(v)| \geq \Omega(\delta |V|^2)$, because there will be $\Omega(|V|^2)$ pairs u, v such that $f_S(u) = 0$ and $f_S(v) \geq \delta$. In particular, if we could find such a set with $\delta = 1/O(\sqrt{\log |V|})$ then we would be done. Unfortunately this is too much to ask for in general, because we always have $|f_S(u) - f_S(v)| \leq d(u, v)$, which means that if we want $\sum_{u,v} |f_S(u) - f_S(v)|$ to have $\Omega(V^2)$ noticeably large terms we must also have that $d(u, v)$ is noticeably large for $\Omega(|V|^2)$ pairs of points, which is not always true.

There is, however, the following argument, which goes back to Leighton and Rao: either there are $\Omega(|V|)$ points concentrated in a ball whose radius is a quarter (say) of the average pairwise distance, and then we can use that ball to get an ℓ_1 mapping with only *constant* error; or there are $\Omega(|V|)$ points in a ball of radius twice the average pairwise distance, such that the pairwise distances of the points in the ball account for a constant fraction of all pairwise distances. In particular, the sum of pairwise distances includes $\Omega(|V|^2)$ terms which are $\Omega(1)$.

After we do this reduction and some scaling, we are left with the task of proving the following theorem: suppose we are given an n-point negative type metric in which the points are contained in a ball of radius 1 and are such that the sum of pairwise distances is $\Omega(n^2)$; then there is a subset S of size $\Omega(n)$ such that there are $\Omega(n)$ points whose distance from the set is $1/O(\sqrt{\log n})$. This theorem is the main result of the Arora-Rao-Vazirani paper. (Strictly speaking, this form of the theorem was proved later by Lee – Arora, Rao and Vazirani had a slightly weaker formulation.)

We begin by considering the case in which a constant fraction of the points are concentrated in a small ball.

Definition 12.15 (Ball) *For a point $z \in V$ and a radius $r > 0$, the ball of radius r and center z is the set*

$$B(z, r) := \{v : d(z, v) \leq r\}$$

Lemma 12.16 *For every vertex z, if we define $S := B(z, 1/4)$, then*

$$\sum_{u,v} |f_S(u) - f_S(v)| \geq \frac{|S|}{2|V|} \sum_{u,v} d(u, v)$$

PROOF: Our first calculation is to show that the typical value of $f_S(u)$ is rather large. We note that for every two vertices u and v, if we call a a closest vertex in S to u, and b a closest vertex to v in S, we have

$$\begin{aligned} d(u, v) &\leq d(u, a) + d(a, z) + d(z, b) + d(b, v) \\ &\leq f_S(u) + f_S(v) + \frac{1}{2} \end{aligned}$$

and so

$$|V|^2 = \sum_{u,v} d(u,v) \leq 2|V| \cdot \sum_v f_S(v) + \frac{|V|^2}{2}$$

that is,

$$\sum_v f_S(v) \geq \frac{|V|}{2}$$

Now we can get a lower bound to the sum of ℓ_1 distances given by the embedding $f_S(\cdot)$.

$$\begin{aligned}
&\sum_{u,v} |f_S(u) - f_S(v)| \\
&\geq \sum_{u \in S, v \in V} |f_S(v)| \\
&= |S| \sum_v f_S(v) \\
&\geq \frac{1}{2}|S| \cdot |V|
\end{aligned}$$

□

This means that if there is a vertex z such that $|B(z,1/4)| = \Omega(|V|)$, or even $|B(z,1/4)| = \Omega(|V|/\sqrt{\log |V|})$, then we are done.

Otherwise, we will find a set of $\Omega(|V|)$ vertices such that their average pairwise distances are within a constant factor of their maximum pairwise distances, and then we will work on finding an embedding for such a set of points. (The condition that the average distance is a constant fraction of the maximal distance will be very helpful in subsequent calculations.)

Lemma 12.17 *Suppose that for every vertex z we have $|B(z,1/4)| \leq |V|/4$. Then there is a vertex w such that, if we set $S = B(w,2)$, we have*

- $|S| \geq \frac{1}{2} \cdot |V|$
- $\sum_{u,v \in S} d(u,v) \geq \frac{1}{8}|S|^2$

PROOF: Let w be a vertex that maximizes $|B(w,2)|$; then $|B(w,2)| \geq |V|/2$, because if we had $|B(u,2)| < |V|/2$ for every vertex u, then we would have

$$\sum_{u,v} d(u,v) > \sum_u 2 \cdot (|V - B(u,2)|) > |V|^2$$

Regarding the sum of pairwise distances of elements of S, we have

$$\sum_{u,v \in S} d(u,v) > \sum_{u \in S} \frac{1}{4}(|S - B(u,1/4)|) \geq |S| \cdot \frac{1}{4} \cdot \frac{|S|}{2}$$

□

The proof of the main theorem now reduces to proving the following geometric fact.

Lemma 12.18 (ARV Main Lemma) *Let d be a negative-type metric over a set V such that the points are contained in a unit ball and have constant average distance, that is,*

- *there is a vertex z such that $d(v,z) \le 1$ for every $v \in V$*
- *$\sum_{u,v \in V} d(u,v) \ge c \cdot |V|^2$*

Then there are sets $S, T \subseteq V$ such that

- *$|S|, |T| \ge \Omega(|V|)$;*
- *for every $u \in S$ and every $v \in S$, $d(u,v) \ge 1/O(\sqrt{\log |V|})$*

where the multiplicative factors hidden in the $O(\cdot)$ and $\Omega(\cdot)$ notations depend only on c.

Indeed, applying the ARV Main Lemma to $\frac{1}{2}d(u,v)$ tells us that there are subsets S, T of $B(z,2)$, both of size $\Omega(|B(z,2)|) = \Omega(n)$ such that $d(u,v) \ge 1/O(\sqrt{\log n})$ for every $u \in S$ and $v \in T$. If we consider the Frechet embedding f_S, we have

$$\begin{aligned}
\sum_{\{u,v\}} |f_S(u) - f_S(v)| &\ge \sum_{u \in S, v \in T} |f_S(u) - f_S(v)| \\
&\ge |S| \cdot |T| \cdot \frac{1}{O(\sqrt{\log n})} \\
&\ge n^2 \cdot \frac{1}{O(\sqrt{\log n})} \\
&= \frac{1}{O(\sqrt{\log n})} \cdot \sum_{\{u,v\}} d(u,v)
\end{aligned}$$

It remains the prove the ARV Main Lemma.

Chapter 13

ARV Analysis

In which we begin the analysis of the ARV rounding algorithm

We want to prove

Lemma 13.1 (ARV Main Lemma) *Let d be a negative-type metric over a set V such that the points are contained in a unit ball and have constant average distance, that is,*

- *there is a vertex z such that $d(v, z) \leq 1$ for every $v \in V$*
- *$\sum_{u,v \in V} d(u, v) \geq c \cdot |V|^2$*

Then there are sets $S, T \subseteq V$ such that

- *$|S|, |T| \geq \Omega(|V|)$;*
- *for every $u \in S$ and every $v \in T$, $d(u, v) \geq 1/O(\sqrt{\log |V|})$*

where the multiplicative factors hidden in the $O(\cdot)$ and $\Omega(\cdot)$ notations depend only on c.

In this lecture, we will show how to reduce the ARV Main Lemma to a statement of the following form: if $\{\mathbf{x}_v\}_{v \in V}$ is a set of vectors such that the metric $d(\cdot, \cdot)$ in the ARV Main Lemma can be written as $d(u, v) = ||\mathbf{x}_u - \mathbf{x}_v||^2$, and $\mathbf{g}$ is a random Gaussian vectors, and if ℓ is such that with $\Omega(1)$ probability, there are $\Omega(n)$ disjoint pairs u, v such that $d(u, v) < \ell$ and $|\langle g, \mathbf{x}_u \rangle - \langle g, \mathbf{x}_v \rangle| \geq \Omega(1)$, then $\ell \geq \Omega(1/\sqrt{\log n})$. We will then prove such a statement in the next lecture.

13.1 Bottlenecks

Before beginning with the proof, it will be useful to see that certain variations of the ARV Main Lemma are false, and that we must use the assumptions of the lemma in a certain way in order to be able to prove it.

For example, consider the variation of the lemma in which $d(\cdot,\cdot)$ is an arbitrary semi-metric, rather than being of negative type. We have the following counterexample.

Fact 13.2 *For every n, there is a metric $d(\cdot,\cdot)$ over $V = \{1,\ldots,n\}$ such that*

- $d(i,j) \leq 1$ *for all i,j*
- $\sum_{i,j} d(i,j) \geq \Omega(n^2)$
- *For every subsets S,T of size $\Omega(n)$ we have*

$$\min_{i\in S, j\in T} d(i,j) \leq O\left(\frac{1}{\log n}\right)$$

We will not provide a full proof but here is a sketch: consider a family $G_n = ([n], E_n)$ of constant-degree graphs of constant edge expansion. (We will see later in the course that such a family exists.) Consider the shortest-path distance $d_{SP}(\cdot,\cdot)$ on $[n]$. We have:

- For every pair i,j, $d_{SP}(i,j) \leq O(\log n)$, because graphs of constant expansion have logarithmic diameter (another fact that we will prove later in the course)
- $\sum_{i,j} d_{SP}(i,j) \geq \Omega(n^2 \log n)$, because, if r is the degree of the graph, then every vertex has at most r^{t+1} other vertices at distance at most t from it, and so every vertex has at least $n/2$ other vertices at distance $\Omega(\log n)$ from itself.
- For every subsets S,T of size $\Omega(n)$ we have

$$\min_{i\in S, j\in T} d_{SP}(i,j) \leq O(1)$$

  Because, if the edge expansion is $\Omega(1)$ and the degree is $O(1)$, then for every set A of $\leq n/2$, there are $\Omega(|A|)$ vertices outside A with neighbors in A, and so the number of vertices at distance at most t from S is at least $\min\{n/2, |S|\cdot 2^{\Omega(t)}\}$. If $|S| \geq \Omega(n)$, then there is a $t = O(1)$ such that more than $n/2$ vertices are at distance $\leq t$ from S, and the same is true for T, meaning that S and T are at distance at most $2t = O(1)$ from each other.

If divide $d_{SP}(\cdot,\cdot)$ by the diameter of G, which is $O(\log n)$, we obtain a metric that satisfies the conditions of the Fact above.

This means that we cannot only use the property of $d(\cdot,\cdot)$ being a semi-metric, but we have to use the fact that it is of negative type, and we need to use in the proof the vectors $\mathbf{x}_v$ such that $d(u,v) = ||\mathbf{x}_v - \mathbf{x}_u||^2$.

Fact 13.2 is tight: using Bourgain's theorem, or an earlier technique of Leighton and Rao, if $d(\cdot,\cdot)$ is a semi-metric over $[n]$ such that $\max_{i,j} d(i,j) \leq 1$ and $\sum_{i,j} d(i,j) \geq \Omega(1)$, then we can find sets S, T of size $\Omega(n)$ such that $\min_{i\in S, j\in T} d(i,j) \geq \Omega(1/\log n)$.

Fact 13.3 *For every n, there are vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ such that*

- $||\mathbf{x}_i - \mathbf{x}_j||^2 \leq 1$ *for all* i, j
- $\sum_{i,j} ||\mathbf{x}_i - \mathbf{x}_j||^2 \geq \Omega(n^2)$
- *For every subsets S, T of size $\Omega(n)$ we have*

$$\min_{i\in S, j\in T} ||\mathbf{x}_i - \mathbf{x}_j||^2 \leq O\left(\frac{\log\log n}{\log n}\right)$$

Here we will not even provide a sketch, but the idea is to use an ϵ-net of the sphere of radius $1/2$ in dimension $O(\log n/\log\log n)$, with $\epsilon = o(1)$, and the isoperimeteric inequality for the sphere.

This means that we need to use the fact that our vectors satisfy the triangle inequalities $||\mathbf{x}_i - \mathbf{x}_j||^2 \leq ||\mathbf{x}_i - \mathbf{x}_k||^2 + ||\mathbf{x}_k - \mathbf{x}_j||^2$. It is also worth noting that for all vectors, including those of Fact 13.3, we have

$$||\mathbf{x}_i - \mathbf{x}_j||^2 \leq 2||\mathbf{x}_i - \mathbf{x}_k||^2 + 2||\mathbf{x}_k - \mathbf{x}_j||^2$$

so any argument that proves the ARV Main Lemma will need to use the triangle inequalities in a way that breaks down if we substitute them with the above "factor-of-2-triangle-inequalities".

Fact 13.3 is also tight, up to the factor of $\log\log n$, as we will see later in this lecture.

Finally, we note that the ARV Main Lemma is tight, which means that every step of its proof will have to involve statements that are tight up to constant factors.

Fact 13.4 *For every n that is a power of two, there is a negative-type metric $d(\cdot,\cdot)$ over a set V of size n such that*

- $d(i,j) \leq 1$ *for all* i, j
- $\sum_{i,j} d(i,j) \geq \Omega(n^2)$
- *For every subsets S, T of size $\Omega(n)$ we have*

$$\min_{i\in S, j\in T} d(i,j) \leq O\left(\frac{1}{\sqrt{\log n}}\right)$$

Let $n = 2^t$ and $V = \{0,1\}^t$. The Hamming distance $d_H(\cdot,\cdot)$ is a negative-type metric over $\{0,1\}^t$ (let $\mathbf{x}_v$ be v itself, and notice that $d_H(u,v) = ||u - v||^2$) , and it satisfies

- $d(i,j) \leq t$ for all i, j
- $\sum_{i,j} d(i,j) \geq \Omega(t \cdot n^2)$
- For every subsets S, T of size $\Omega(n)$ we have

$$\min_{i \in S, j \in T} d(i,j) \leq O(\sqrt{t})$$

  which follows from isoperimetric results on the hypercube that we will not prove

Fact 13.4 follows by dividing the above metric by t.

13.2 Gaussian Projections

The tool of *Gaussian projections* is widely used to analyze semidefinite programs. Given vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ which are solutions to a semidefinite program of interest, we pick a random Gaussian vector $\mathbf{g} \sim \mathbb{R}^m$, and we consider the projections $Y_1, \ldots, Y_m$, where $Y_i := \langle \mathbf{x}_i, \mathbf{g} \rangle$. The vector $\mathbf{g} = (g_1, \ldots, g_m)$ is sampled so that the coordinates g_i are independent standard normal distributions.

We see that each Y_i has a Gaussian distribution with expectation 0 and variance $||\mathbf{x}_i||^2$, and each difference $Y_i - Y_j$ has a gaussian distribution with expectation 0 and variance $||\mathbf{x}_i - \mathbf{x}_j||^2 = d(i,j)$.

From standard bounds on Gaussian random variables,

$$\mathbb{P}[|Y_i - Y_j| \leq \delta \sqrt{d(i,j)}] \leq \frac{2}{\sqrt{2\pi}} \delta < \delta \tag{13.1}$$

$$\mathbb{P}[|Y_i - Y_j| \geq t \sqrt{d(i,j)}] \leq \frac{2}{\sqrt{2\pi}} e^{-t^2/2} < e^{-t^2/2} \tag{13.2}$$

And, setting $t = \sqrt{5 \log n}$ in (13.2), we get

$$\mathbb{P}[\ \forall i, j.\ |Y_i - Y_j|^2 \leq 5 \log n \cdot d(i,j)] \geq 1 - o(1) \tag{13.3}$$

Our first result is that, with $\Omega(1)$ probability, there are $\Omega(n^2)$ pairs i, j such that $|Y_i - Y_j| \geq \Omega(1)$.

Lemma 13.5 *There are constants c_1, σ that depend only on c such that with probability at least 90%, if we let L be the $c_1 n$ indices i with smallest Y_i, and R be the $c_1 n$ indices i with largest Y_i, we have*

$$\forall i \in L. \forall j \in R \quad |Y_i - Y_j| \geq \sigma$$

PROOF: A standard Markov argument shows that if $d(i,j) \leq 1$ for all pairs i,j, and $\sum_{i,j} d(i,j) \geq cn^2$, then there are at least $cn^2/2$ pairs at distance at least $c/2$. We argue that, with probability at least 90%, $\Omega(n^2)$ of those pairs are such that $|Y_i - Y_j| \geq \Omega(1)$, which implies the conclusion.

Let F be the set of "far" pairs i,j such that $d(i,j) \geq c/2$.

By setting $\delta = \frac{1}{20}$ in (13.1), we have for each $(i,j) \in F$

$$\mathbb{P}[|Y_i - Y_j| \leq \sqrt{c}/20\sqrt{2}] < \frac{1}{20}$$

so, by linearity of expectation,

$$\mathbb{E}[\ |\{(i,j) \in F.\ |Y_i - Y_j| \leq \sqrt{c}/20\sqrt{2}] < \frac{|F|}{20}$$

and by Markov inequality

$$\mathbb{P}\left[\ \left|\left\{(i,j) \in F.\ |Y_i - Y_j| \leq \frac{\sqrt{c}}{20\sqrt{2}}\right\}\right| > \frac{|F|}{2}\right] < .1$$

so, with probability $\geq 90\%$, there are at least $|F|/2 \geq cn^2/4$ pairs (i,j) such that $|Y_i - Y_j| \geq \frac{\sqrt{c}}{20\sqrt{2}}$.

If L and R are defined as above, and $\sigma = \min_{i \in L,\ j \in R} Y_j - Y_i$, then the number of pairs i,j at distance $> \sigma$ is at most

$$(1 - (1 - 2c_1)^2) \cdot n^2 \leq 4c_1 n^2$$

and the lemma follows if we set $c_1 = c/16$ and $\sigma = \sqrt{c}/20\sqrt{2}$. $\square$

Note that, with $90\% - o(1)$ probability, we have sets L, R, both of size $\geq c_1 n$, such that

$$\forall i,j \in V.\ \ |Y_i - Y_j|^2 \leq 5 \log n \cdot d(i,j)$$

$$\forall i \in L,\ j \in R,\ Y_j - Y_i \geq \sigma$$

so that

$$\forall i \in L,\ j \in R,\ d(i,j) \geq \frac{\sigma^2}{5 \log n} \geq \frac{1}{O(\log n)}$$

Since we have not used the triangle inequality, the above bound is almost best possible, given Fact 13.3 .

13.3 The Algorithm to Refine L and R

Consider the following algorithm, given $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ satisfying the assumptions of the Main Lemma, and a parameter ℓ,

- Pick a random gaussian vector $\mathbf{g} \sim \mathbb{R}^m$

- Define $Y_i := \langle \mathbf{x}_i, \mathbf{g} \rangle$ for $i = 1, \ldots, n$
- Let L be the $c_1 n$ indices i for which Y_i is smallest
- Let R be the $c_1 n$ indices i for which Y_i is largest
- while there is an $i \in L$ and $j \in L$ such that $d(i,j) < \ell$
  - remove i from L and j from R
- return L, R

Where c_1, σ are the constants (that depend only on c) of Lemma 13.5. We will prove

Lemma 13.6 *There is a constant c_2 (dependent only on c) such that, if we set $\ell \leq \frac{c_2}{\sqrt{\log n}}$, there is at least a 80% probability that the algorithm removes at most $\frac{c_1 n}{2}$ pairs (i,j) in the 'while' loop.*

Once we establish the above lemma, we have completed our proof of the ARV Main Lemma, because, with $70\% - o(1)$ probability, the output of the algorithm is a pair of sets L, R of size $\geq \frac{c_1 n}{2}$ such that for each $i \in L$ and $j \in R$ we have $d(i,j) \geq \frac{c_2}{\sqrt{\log n}}$.

We will prove the contrapositive, that is, if the algorithm has probability at least 20% of removing at least $\frac{c_1 n}{2}$ pairs (i,j) in the 'while' loop, then $\ell \geq c_2/\sqrt{\log n}$.

Call M the set of pairs (i,j) removed by the algorithm (like $Y_1, \ldots, Y_n$, L and R, M is a random variable determined by $\mathbf{g}$). If the algorithm has probability at least 20% of removing at least $\frac{c_1 n}{2}$ pairs (i,j) in the 'while' loop, then there is a probability at least 10% that the above happens, and that $\min_{i \in L, j \in R} |Y_i - Y_j| \geq \sigma$. This means that with probability at least 10% there are $\frac{c_i n}{2}$ disjoint pairs (i,j) such that $|Y_i - Y_j| \geq \sigma$ and $d(i,j) \leq \ell$.

By the above observation, the following lemma implies Lemma 13.6 and hence the ARV Main Lemma.

Lemma 13.7 *Let $d(\cdot,\cdot)$ be a negative-type metric over a set $V = \{1, \ldots, n\}$, let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ be vectors such that $d(i,j) = ||\mathbf{x}_i - \mathbf{x}_j||^2$, let $\mathbf{g} \sim \mathbb{R}^m$ be a random vector with a Gaussian distribution, and let $Y_i := \langle \mathbf{g}, \mathbf{x}_i \rangle$.*

Suppose that, for constants c', σ and a parameter ℓ, we have that there is a $\geq 10\%$ probability that there are at least $c'n$ pairs (i,j) such that $d(i,j) \leq \ell$ and $|Y_i - Y_j| \geq \sigma$.

Then there is a constant c_2, that depends only on c' and σ, such that

$$\ell \geq \frac{c_2}{\sqrt{\log n}}$$

We will prove Lemma 14.1 in the next lecture.

Chapter 14

ARV Analysis cont'd

In which we continue the analysis of the ARV rounding algorithm

We are continuing the analysis of the Arora-Rao-Vazirani rounding algorithm, which rounds a Semidefinite Programming solution of a relaxation of sparsest cut into an actual cut, with an approximation ratio $O(\sqrt{\log |V|})$.

In previous lectures, we reduced the analysis of the algorithm to the following claim.

Lemma 14.1 *Let $d(\cdot,\cdot)$ be a negative-type semimetric over a set $V = \{1,\ldots,n\}$, let $\mathbf{x}_1,\ldots,\mathbf{x}_n \in \mathbb{R}^m$ be vectors such that $d(i,j) = ||\mathbf{x}_i - \mathbf{x}_j||^2$, let $\mathbf{g} \sim \mathbb{R}^m$ be a random vector with a Gaussian distribution, and let $Y_i := \langle \mathbf{g}, \mathbf{x}_i \rangle$.*

Suppose that, for constants c', σ and a parameter ℓ, we have that there is a $\geq 10\%$ probability that there are at least $c'n$ pairs (i,j) such that $d(i,j) \leq \ell$ and $|Y_i - Y_j| \geq \sigma$.

Then there is a constant c_2, that depends only on c' and σ, such that

$$\ell \geq \frac{c_2}{\sqrt{\log n}}$$

14.1 Concentration of Measure

In the last lecture, we are have already introduced two useful properties of Gaussian distributions: that there is a small probability of being much smaller than the standard deviation in absolute value, and a very small probability of being much larger than the standard deviation in absolute value. Here we introduce a third property of a somewhat different flavor.

For a set $A \subseteq \mathbb{R}^n$ and a distance parameter D, define

$$A_D := \{\mathbf{x} \in \mathbb{R}^m : \exists \mathbf{a} \in A.\ ||\mathbf{x} - \mathbf{a}|| \leq D\}$$

the set of points at distance at most D from A. Then we have:

Theorem 14.2 (Gaussian concentration of measure) *There is a constant c_3 such that, for every $\epsilon, \delta > 0$ and for every set $A \subseteq \mathbb{R}^n$, if*

$$\mathbb{P}[A] \geq \epsilon$$

then

$$\mathbb{P}[A_D] \geq 1 - \delta$$

for every $D \geq c_3 \cdot \sqrt{\log \frac{1}{\epsilon\delta}}$, where the probabilities are taken according to the Gaussian measure in $\mathbb{R}^m$, that is $\mathbb{P}[A] = \mathbb{P}[\mathbf{g} \in A]$, where $\mathbf{g} = (g_1, \ldots, g_m)$ and the g_i are independent Gaussians of mean 0 and variance 1.

The above theorem says that if we have some property that is true with $\geq 1\%$ probability for a random Gaussian vector $\mathbf{g}$, then there is a $\geq 99\%$ probability that $\mathbf{g}$ is within distance $O(1)$ of a vector $\mathbf{g}'$ that satisfies the required property. In high dimension m, this is a non-trivial statement because, with very high probability $||\mathbf{g}||$ is about $\sqrt{m}$, and so the distance between $\mathbf{g}$ and $\mathbf{g}'$ is small relative to the length of the vector.

We will use the following corollary.

Corollary 14.3 *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be vectors in $\mathbb{R}^m$ and let $d_{\max} = \max_{j=2,\ldots,n} ||\mathbf{x}_j - \mathbf{x}_1||^2$. Let $\mathbf{g}$ be a random Gaussian vector in $\mathbb{R}^m$, and let $Y_i = \langle \mathbf{x}_i, \mathbf{g} \rangle$. If, for some k and ϵ, we have*

$$\mathbb{P}[\exists j.\ Y_j - Y_1 \geq k] \geq \epsilon$$

then

$$\mathbb{P}[\exists j.\ Y_j - Y_1 \geq k - c_3\sqrt{\log 1/(\epsilon\gamma)} \cdot \sqrt{d_{\max}}] \geq 1 - \gamma$$

Proof: Let

$$A := \{\mathbf{g} : \exists j.\ Y_j - Y_1 \geq k] \geq \epsilon$$

By assumption, we have $\mathbb{P}[A] \geq \epsilon$, and so, by concentration of measure:

$$\mathbb{P}[\exists \mathbf{g}'.\ ||\mathbf{g} - \mathbf{g}'|| \leq c_3\sqrt{\log 1/(\epsilon\gamma)}\ \wedge\ \mathbf{g}' \in A] \geq 1 - \gamma$$

The even in the above probability can be rewritten as

$$\exists \mathbf{g}' \in \mathbb{R}^m\ \exists j \in \{2, \ldots, n\}.\ ||\mathbf{g} - \mathbf{g}'|| \leq c_3\sqrt{\log \frac{1}{\epsilon\gamma}}\ \wedge \langle \mathbf{x}_j - \mathbf{x}_1, \mathbf{g}' \rangle \geq k$$

and the above condition gives us

$$\begin{aligned} k &\leq \langle \mathbf{x}_j - \mathbf{x}_1, \mathbf{g}' \rangle \\ &= \langle \mathbf{x}_j - \mathbf{x}_1, \mathbf{g} \rangle + \langle \mathbf{x}_j - \mathbf{x}_1, \mathbf{g}' - \mathbf{g} \rangle \\ &\leq \langle \mathbf{x}_j - \mathbf{x}_1, \mathbf{g} \rangle + ||\mathbf{x}_j - \mathbf{x}_1|| \cdot ||\mathbf{g}' - \mathbf{g}|| \\ &\leq Y_j - Y_1 + \sqrt{d_{\max}} \cdot c_3\sqrt{\log \frac{1}{\epsilon\gamma}} \end{aligned}$$

□

The (use of the) above statement is by far the most innovative part of the analysis of Arora, Rao and Vazirani, so it is worth developing an intuitive feeling for its meaning.

Let's say that we are interested in the distribution of $p_{\max} := \max_{j=2,\ldots,n} Y_j - Y_1$. We know that the random variables $Y_j - Y_1$ are Gaussians of mean 0 and standard deviation at most $\sqrt{d_{\max}}$, but it is impossible to say anything about, say, the average value or the median value of $p_{\max}$ without knowing something about the correlation of the random variables $Y_j - Y_1$.

Interestingly, the above Corollary says something about the *concentration* of $p_{\max}$ without any additional information. The corollary says that, for example, the first percentile of $p_{\max}$ and the 99-th percentile of $p_{\max}$ differ by at most $O(\sqrt{d_{\max}})$, and that we have a concentration result of the form

$$\mathbb{P}[|p_{\max} - \text{median}(p_{\max})| > t \cdot \sqrt{d_{\max}}] \leq e^{-\Omega(t^2)}$$

which is a highly non-trivial statement for any configuration of $\mathbf{x}_i$ for which $p_{\max} >> \sqrt{d_{\max}}$.

14.2 Reworking the Assumption

Lemma 14.4 *Under the assumptions of Lemma 14.1, there is a fixed set $C \subseteq [n]$, $|C| \geq \frac{c'}{10}n$, and a set $M_{\mathbf{g}}$ of disjoint pairs $\{i,j\}$, dependent on $\mathbf{g}$, such that, for every $\mathbf{g}$ and for every pair $\{i,j\} \in M_{\mathbf{g}}$ we have*

$$d(i,j) \leq \ell$$

and

$$|Y_i - Y_j| \geq \sigma$$

and for all $i \in C$ we have

$$\mathbb{P}[\exists j \in C.\ \{i,j\} \in M_{\mathbf{g}}] \geq \frac{c'}{20}$$

Proof: Let $M_{\mathbf{g}}$ be the set of disjoint pairs promised by the assumptions of Lemma 14.1. Construct a weighted graph $G = ([n], W)$, where the weight of the edge $\{i,j\}$ is $\mathbb{P}[\{i,j\} \in M_{\mathbf{g}}]$. The degree of every vertex is at most 1, and the sum of the degrees is twice the expectation of $|M|$, and so, by the assumptions of Lemma 14.1, it is at least $c'n/5$.

Now, repeatedly delete from G all vertices of degree at most $c'n/20$, and all the edges incident on them, until no such vertex remains. At the end we are left with a (possibly empty!) graph in which all remaining vertices have degree at most $c'n/20$; each deletion reduces the sum of the degree by at most $c'/10$, and so the residual graph has total degree at least $c'n/10$, and hence at least $c'n/10$ vertices □

By the above Lemma, the following result implies Lemma 14.1 and hence the ARV Main Lemma.

Lemma 14.5 *Let $d(\cdot,\cdot)$ be a semi-metric over a set C such that $d(u,v) \leq 1$ for all $u, v \in C$, let $\{\mathbf{x}_v\}_{v \in C}$ be a collection of vectors in $\mathbb{R}^m$, such that $d(i,j) := ||\mathbf{x}_u - \mathbf{x}_v||^2$ is a semimetric, let $\mathbf{g}$ be a random Gaussian vector in $\mathbb{R}^m$, define $Y_v := \langle \mathbf{g}, \mathbf{x}_v \rangle$, and suppose that, for every $\mathbf{g}$, we can define a set of disjoint pairs $M_{\mathbf{g}}$ such that, with probability 1 over $\mathbf{g}$,*

$$\forall \{u,v\} \in M_{\mathbf{g}}.\ |Y_u - Y_v| \geq \sigma \ \wedge \ d(u,v) \leq \ell$$

and

$$\forall u \in C.\ \mathbb{P}[\exists v. \{u,v\} \in M_{\mathbf{g}}] \geq \epsilon$$

Then

$$\ell \geq \Omega_{\epsilon,\sigma}\left(\frac{1}{\sqrt{\log |C|}}\right)$$

Chapter 15

ARV Analysis, part 3

In which we complete the analysis of the ARV rounding algorithm

We are finally going to complete the analysis of the Arora-Rao-Vazirani rounding algorithm, which rounds a Semidefinite Programming solution of a relaxation of sparsest cut into an actual cut, with an approximation ratio $O(\sqrt{\log |V|})$.

In previous lectures, we reduced the analysis of the algorithm to the following claim.

Lemma 15.1 *Let $d(\cdot,\cdot)$ be a semi-metric over a set C such that $d(u,v) \leq 1$ for all $u, v \in C$, let $\{\mathbf{x}_v\}_{v \in C}$ be a collection of vectors in $\mathbb{R}^m$, such that $d(i,j) := ||\mathbf{x}_u - \mathbf{x}_v||^2$ is a semimetric, let $\mathbf{g}$ be a random Gaussian vector in $\mathbb{R}^m$, define $Y_v := \langle \mathbf{g}, \mathbf{x}_v \rangle$, and suppose that, for every $\mathbf{g}$, we can define a set of disjoint pairs $M_{\mathbf{g}}$ such that, with probability 1 over $\mathbf{g}$,*

$$\forall \{u,v\} \in M_{\mathbf{g}}.\ |Y_u - Y_v| \geq \sigma \ \wedge \ d(u,v) \leq \ell$$

and

$$\forall u \in C.\ \mathbb{P}[\exists v. \{u,v\} \in M_{\mathbf{g}}] \geq \epsilon$$

Then

$$\ell \geq \Omega_{\epsilon,\sigma}\left(\frac{1}{\sqrt{\log |C|}}\right)$$

15.1 An Inductive Proof that Gives a Weaker Result

In this section we will prove a weaker lower bound on ℓ, of the order of $\frac{1}{(\log |C|)^{\frac{2}{3}}}$. We will then show how to modify the proof to obtain the tight result.

We begin will the following definitions. We define the ball or radius r centered at u as

$$B(u,r) := \{v \in C.\ d(u,v) \leq r\}$$

We say that a point $u \in C$ has the (p, r, δ)-Large-Projection-Property, or that it is (p, r, δ)-LPP if

$$\mathbb{P}\left[\max_{v \in B(u,r)} Y_v - Y_u \geq p\right] \geq \delta$$

Lemma 15.2 *Under the assumptions of Lemma 15.1, there is a constant $c_4 > 0$ (that depends only on ϵ and σ) such that for all $t \leq c_4 \cdot \frac{1}{\sqrt{\ell}}$, at least $\left(\frac{\epsilon}{8}\right)^t \cdot |C|$ elements of C have the $\left(t\frac{\sigma}{2}, t\ell, 1 - \frac{\epsilon}{4}\right)$ Large Projection Property.*

Proof: We will prove the Lemma by induction on t. We call C_t the set of elements of C that are $\left(t\frac{\sigma}{2}, t\ell, 1 - \frac{\epsilon}{4}\right)$-LPP

Let $M'_{\mathbf{g}}$ be the set of *ordered* pairs (u, v) such that $\{u, v\} \in M_{\mathbf{g}}$ and $Y_v > Y_u$, and hence $Y_v - Y_u \geq \sigma$. Because $\mathbf{g}$ and $-\mathbf{g}$ have the same distribution, we have that, for every $u \in C$, there is probability $\geq \epsilon/2$ that there is a $v \in C$ such that $(v, u) \in M'_{\mathbf{g}}$ (a fact that we will use in the inductive step).

For the base case $t = 0$ there is nothing to prove.

For the inductive case, define the function $F : C_t \to C$ (which will be a random variable dependent on $\mathbf{g}$) such that $F(v)$ is the lexicographically smallest $w \in B(v, t\ell)$ such that $Y_w - Y_v \geq \sigma$ if such a w exists, and $F(v) = \perp$ otherwise. The definition of C_t is that $\mathbb{P}[F(v) \neq \perp] \geq 1 - \epsilon/4$ for every $v \in C_t$, and the inductive assumption is that $|C_t| \geq |C| \cdot (\epsilon/8)^t$.

By a union bound, for every $v \in C_t$, there is probability at least $\epsilon/4$ that there is an $u \in C$ such that $(u, v) \in M'_{\mathbf{g}}$ and $F(v) = w \neq \perp$. In this case, we will define $F'(u) = w$, otherwise $F'(u) = \perp$.

Note that the above definition is consistent, because $M'_{\mathbf{g}}$ is a set of disjoint pairs, so for every u there is at most one v that could be used to define $F'(u)$. We also note that, if $F'(u) = w \neq \perp$, then

$$Y_w - Y_u \geq t \cdot \frac{\sigma}{2} + \sigma \ ,$$
$$d(u, w) \leq (t + 1) \cdot \ell$$

and

$$\sum_{u \in C} \mathbb{P}[F'(u) \neq \perp] = \sum_{v \in C_t} [F(v) \neq \perp \wedge \exists u.(u, v) \in M'_{\mathbf{g}}] \geq |C_t| \cdot \frac{\epsilon}{4}$$

Now we can use another averaging argument to say that there have to be at least $|C_t| \cdot \frac{\epsilon}{8}$ elements u of C such that

$$\mathbb{P}[F'(u) \neq \perp] \geq \frac{\epsilon}{8} \cdot \frac{|C_t|}{|C|} \geq \left(\frac{\epsilon}{8}\right)^{t+1}$$

Let us call C_{t+1} the set of such element. As required, $|C_{t+1}| \geq |C| \cdot (\epsilon/8)^{t+1}$.

By applying concentration of measure, the fact that, for every $u \in C_{t+1}$ we have

$$\mathbb{P}\left[\max_{w \in B(u,(t+1)\cdot\ell)} Y_w - Y_u \geq (t + 1)\frac{\sigma}{2} + \frac{\sigma}{2}\right] \geq \left(\frac{\epsilon}{8}\right)^{t+1}$$

implies that, for every $u \in C_{t+1}$

$$\mathbb{P}\left[\max_{w \in B(u,(t+1)\cdot\ell)} Y_w - Y_u \geq (t+1)\frac{\sigma}{2} + \frac{\sigma}{2} - c_3\sqrt{\log\frac{4\cdot 8^{t+1}}{\epsilon^{t+2}}}\sqrt{(t+1)\cdot\ell}\right] \geq 1 - \frac{\epsilon}{4}$$

and the inductive step is proved, provided

$$\frac{\sigma}{2} \geq c_3\sqrt{(t+2)\cdot\log\frac{8}{\epsilon}}\sqrt{(t+1)\cdot\ell}$$

which is true when

$$t+2 \leq \frac{\sigma}{2c_3\sqrt{\log 8/\epsilon}}\cdot\frac{1}{\sqrt{\ell}}$$

which proves the lemma if we choose c_4 appropriately. □

Applying the previous lemma with $t = c_4/\sqrt{\ell}$, we have that, with probability $\Omega(1)$, there is a pair u, v in C such that

$$Y_v - Y_u \geq \Omega(1/\sqrt{\ell})$$

and

$$d(u,v) \leq O(\sqrt{\ell})$$

but we also know that, with $1 - o(1)$ probability, for all pairs u, v in C,

$$|Y_v - Y_u|^2 \leq O(\log|C|)\cdot d(i,j)$$

and so

$$\frac{1}{\ell} \leq O(\log|C|)\sqrt{\ell}$$

implying

$$\ell \geq \Omega\left(\frac{1}{(\log|C|)^{2/3}}\right)$$

15.2 The Tight Bound

In the result proved in the previous section, we need $\frac{\sigma}{2}$, which is a constant, to be bigger than the loss incurred in the application of concentration of measure, which is of the order of $t\sqrt{\ell}$. A factor of $\sqrt{t\ell}$ simply comes from the distances between the points that we are considering; an additional factor of $\sqrt{t}$ comes from the fact that we need to push up the probability from a bound that is exponentially small in t.

The reason for such a poor probability bound is the averaging argument: each element of C_t has probability $\Omega(1)$ of being the "middle point" of the construction, so that the sum over the elements u of C of the probability that u has $F'(u) \neq \perp$ adds up to $\Omega(|C_t|)$; such overall probability, however, could be spread out over all of C, with each element of C getting a very low probability of the order of $|C_t|/|C|$, which is exponentially small in t.

Not all elements of C, however, can be a u for which $F'(u) \neq \perp$; this is only possible for elements u that are within distance ℓ from C_t. If the set $\Gamma_\ell(C_t) := \{u : \exists v \in C_t : d(u,v) \leq \ell\}$

has cardinality of the same order of C_t, then we only lose a constant factor in the probability, and we do not pay the extra $\sqrt{t}$ term in the application of concentration of measure. But what do we do if $\Gamma_\ell(C_t)$ is much bigger than C_t? In that case we may replace C_t and $\Gamma_\ell(C_t)$ and have similar properties.

Lemma 15.3 *Under the assumptions of Lemma 15.1, if $S \subseteq C$ is a set of points such that for every $v \in S$*

$$\mathbb{P}\left[\max_{w \in B(v,d)} Y_w - Y_v \geq p\right] \geq \epsilon$$

then, for every distance D, every $k > 0$, and every $u \in \Gamma_D(S)$

$$\mathbb{P}\left[\max_{w \in B(u,d+D)} Y_w - Y_u \geq p - \sqrt{D} \cdot k\right] \geq \epsilon - e^{-k^2/2}$$

That is, if all the elements of S are (p, d, ϵ)-LPP, then all the elements of $\Gamma_D(S)$ are $(p - k\sqrt{D}, d + D, \epsilon - e^{-k^2/2})$-LPP.

Proof: If $u \in \Gamma_D(S)$, then there is $v \in S$ such that $d(u, v) \leq D$, and, with probability $1 - e^{-k^2/2}$ we have $Y_u - Y_v \leq \sqrt{D} \cdot k$. The claim follows from a union bound. □

Lemma 15.4 *Under the assumptions of Lemma 15.1, there is a constant $c_5 > 0$ (that depends only on ϵ and σ) such that for all $t \leq c_5 \cdot \frac{1}{\ell}$, there is a set $C_t \subseteq C$ such that $|C_t| \geq |C| \cdot (\epsilon/8)^t$, every element of C_t is $\left(t \cdot \frac{\sigma}{4}, \left(2t + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}\right) \cdot \ell, 1 - \frac{\epsilon}{4}\right)$-LPP, and*

$$|\Gamma_\ell(C_t)| \leq \frac{8}{\epsilon}|C_t|$$

Proof: The base case $t = 0$ is proved by setting $C_0 = C$.

For the inductive step, we define $F(\cdot)$ and $F'(\cdot)$ as in the proof of Lemma 15.2. We have that if $F'(u) = w \neq \perp$, then

$$Y_w - Y_u \geq t \cdot \frac{\sigma}{4} + \sigma \ ,$$

$$d(u, w) \leq \left(2t + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}\right) \cdot \ell + \ell \ ,$$

and

$$\sum_{u \in C} \mathbb{P}[F'(u) \neq \perp] = \sum_{v \in C_t} [F(v) \neq \perp \wedge \exists u.(u, v) \in M'_{\mathbf{g}}] \geq |C_t| \cdot \frac{\epsilon}{4}$$

Now we can use another averaging argument to say that there have to be at least $|C_t| \cdot \frac{\epsilon}{8}$ elements u of C such that

$$\mathbb{P}[F'(u) \neq \perp] \geq \frac{\epsilon}{8} \cdot \frac{|C_t|}{|\Gamma_\ell(C_t)|} \geq \left(\frac{\epsilon^2}{64}\right)$$

Let us call $C^{(0)}_{t+1}$ the set of such elements.

Define $C_{t+1}^{(1)} := \Gamma_\ell(C_{t+1}^{(0)})$, $C_{t+1}^{(2)} := \Gamma_\ell(C_{t+1}^{(1)})$, and so on, and let k be the first time such that $|C_{t+1}^{(k+1)}| \leq \frac{8}{\epsilon}|C_{t+1}^{(k)}|$. We will define $C_{t+1} := C_{t+1}^{(k)}$. Note that

$$|C_{t+1}| \geq \left(\frac{8}{\epsilon}\right)^k \cdot |C_{t+1}^{(0)}| \geq \left(\frac{8}{\epsilon}\right)^{k-1} \cdot |C_t| \geq \left(\frac{8}{\epsilon}\right)^{k-1-t} |C|$$

which implies that $k \leq t+1$.

We have $|C_{t+1}| \geq |C_{t+1}^{(0)}| \geq \frac{\epsilon}{8}|C_t|$ so we satisfy the inductive claim about the size of C_t. Regarding the other properties, we note that $C_{t+1} \subseteq \Gamma_{k\ell}(C_{t+1}^{(0)})$, and that every element of $C_{t+1}^{(0)}$ is

$$\left(t\frac{\sigma}{4} + \sigma, \left(2t + 1 + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}\right) \cdot \ell, \frac{\epsilon^2}{64}\right) - \text{LPP}$$

so we also have that every element of C_{t+1} is

$$\left(t\frac{\sigma}{4} + \frac{\sigma}{2}, \left(2t + 1 + k + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}\right) \cdot \ell, \frac{\epsilon^2}{128}\right) - \text{LPP}$$

provided

$$\frac{\sigma}{2} \geq \sqrt{2\log\frac{128}{\epsilon^2}} \cdot k\ell$$

which we can satisfy with an appropriate choice of c_4, recalling that $k \leq t+1$.

Then we apply concentration of measure to deduce that every element of C_{t+1} is

$$\left(t\frac{\sigma}{4} + \frac{\sigma}{4}, \left(2t + 1 + k + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}\right) \cdot \ell, 1 - \frac{\epsilon}{4}\right) - \text{LPP}$$

provided that

$$\frac{\sigma}{4} \geq c_3\sqrt{\log\frac{512}{\epsilon^3}} \cdot \left(2t + 1 + k + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}\right) \cdot \ell$$

which we can again satisfy with an appropriate choice of c_4, because $k \leq t+1$ and $\log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|}$ is smaller than or equal to zero.

Finally,

$$2t + 1 + k + \log_{\frac{8}{\epsilon}} \frac{|C_t|}{|C|} \leq 2t + 2 + \log_{\frac{8}{\epsilon}} \frac{|C_{t+1}|}{|C|}$$

because, as we established above,

$$|C_{t+1}| \geq \left(\frac{8}{\epsilon}\right)^{k-1} |C_t|$$

□

By applying Lemma 15.4 with $t = \Omega(1/\ell)$, we find that there is $\Omega(1)$ probability that there are u, v in C such that

$$Y_j - Y_i \geq \Omega(1/\ell)$$

$$d(i,j) \leq 1$$

$$|Y_i - Y_j|^2 \leq O(\log n) \cdot d(i,j)$$

which, together, imply

$$\ell \geq \Omega\left(\frac{1}{\sqrt{\log n}}\right)$$

Chapter 16

Cayley Graphs of Abelian Groups

In which we show how to find the eigenvalues and eigenvectors of Cayley graphs of Abelian groups, we find tight examples for various results that we proved in earlier lectures, and, along the way, we develop the general theory of harmonic analysis which includes the Fourier transform of periodic functions of a real variable, the discrete Fourier transform of periodic functions of an integer variable, and the Walsh transform of Boolean functions.

Earlier, we prove the Cheeger inequalities

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

and the fact that Fiedler's algorithm, when given an eigenvector of λ_2, finds a cut $(S, V-S)$ such that $\phi(S, V-S) \leq 2\sqrt{\phi(G)}$. We will show that all such results are tight, up to constants, by proving that

- The dimension-d hypercube H_d has $\lambda_2 = 1 - \frac{2}{d}$ and $h(H_d) = \frac{1}{d}$, 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 Fiedler's algorithm, given such a vector, outputs a cut $(S, V-S)$ of expansion $\phi(S, V-S) = \Omega(1/\sqrt{d})$, showing that the analysis of the Fiedler's algorithm is tight up to a constant.

In this lecture 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 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.

16.1 Characters

We will use additive notation for groups, so, if Γ 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 16.1 (Character) *Let Γ be a group (we will also use Γ to refer to the set of group elements). A function $f : \Gamma \to \mathbb{C}$ is a character of Γ if*

- *f is a group homomorphism of Γ 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 16.2 (About the Boundedness Condition) *If Γ 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

Γ 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 16.3 *If Γ is finite and χ 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 Γ 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 16.4 *If $\chi_1, \chi_2 : \Gamma \to \mathbb{C}$ are two different characters of a finite group Γ, then*

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

We will prove Lemma 16.4 shortly, but before doing so we note that, for a finite group Γ, the set of functions $f : \Gamma \to \mathbb{C}$ is a $|\Gamma|$-dimensional vector space, and that Lemma 16.4 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 16.5 *If Γ is a finite group, then it has at most $|\Gamma|$ characters.*

It remains to prove Lemma 16.4, which follows from the next two statements, whose proof is immediate from the definitions.

Fact 16.6 *If χ_1, χ_2 are characters of a group Γ, then the mapping $x \to \chi_1(x)\cdot\chi_2(x)$ is also a character.*

Fact 16.7 *If χ is a character of a group Γ, 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 16.4, 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 χ 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 16.8 *If Γ is a group, then the set of characters of Γ 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 Γ, and it is denoted by $\hat{\Gamma}$.

We now come to the punchline of this discussion.

Theorem 16.9 *If Γ 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 Γ_1 and Γ_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 χ_1 is a character of Γ_1 and χ_2 is a character of Γ_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 (χ_1, χ_2) and (χ_1', χ_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 Γ 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 χ, $\langle \chi, \chi \rangle = |\Gamma|$, 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 16.10 (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 16.11 (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.

16.2 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 Γ when we defined an inner product among functions $f : \Gamma \to \mathbb{C}$.

If Γ 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 Γ and restrict ourselves in the choice of functions. A measure μ over (a sigma-algebra of subsets of) Γ 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 Γ 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 μ over Γ, 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 Γ 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 16.4 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 16.3, that is we need to prove that if χ 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 16.3 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 μ being a Haar measure.

We don't have an analogous result to Theorem 16.9 showing that Γ 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 Γ, 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 16.12 (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

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

16.3 Cayley Graphs and Their Spectrum

Let Γ 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 16.13 *For a group Γ and a symmetric subset $S \subseteq \Gamma$, the Cayley graph $Cay(\Gamma, S)$ is the graph whose vertex set is Γ, 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, although we will sometimes allow parallel edges (corresponding to positive integer weights).

Example 16.14 (Cycle) *The n-vertex cycle can be constructed as the Cayley graph $Cay(\mathbb{Z}/n\mathbb{Z}, \{-1, 1\})$.*

Example 16.15 (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 16.16 *Let Γ be a finite abelian group, $\chi : \Gamma \to \mathbb{C}$ be a character of Γ, $S \subseteq \Gamma$ 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 G, with eigenvalue

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

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

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

$$= x_a \cdot \sum_{s \in S} \chi(s)$$

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 χ 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.

16.4 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

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

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 $2\cos(2\pi/n) = 2 - \Theta(1/n^2)$. If λ is an eigenvalue of the adjacency matrix, then $1 - \lambda/2$ is an eigenvalue of the normalized Laplacian. From the above calculations, we have that the second smallest Laplacian eigenvalue is $\Theta(n^{-2})$.

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.

16.5 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

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

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$d - 2$. The second smallest Laplacian eigenvalue is $1 - (d-2)/d = 2/d$.

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 $\phi(S_i, V - S_i) = \frac{1}{d}$ and so $\phi(G) \leq \frac{1}{d}$.

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 Fiedler's 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 Fiedler's 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

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

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

Chapter 17

Constructions of Expanders via the Zig-Zag Graph Product

In which we give an explicit construction of expander graphs of polylogarithmic degree, state the properties of the zig-zag product of graphs, and provide an explicit construction of a family of constant-degree expanders using the zig-zag product and the polylogarithmic-degree construction.

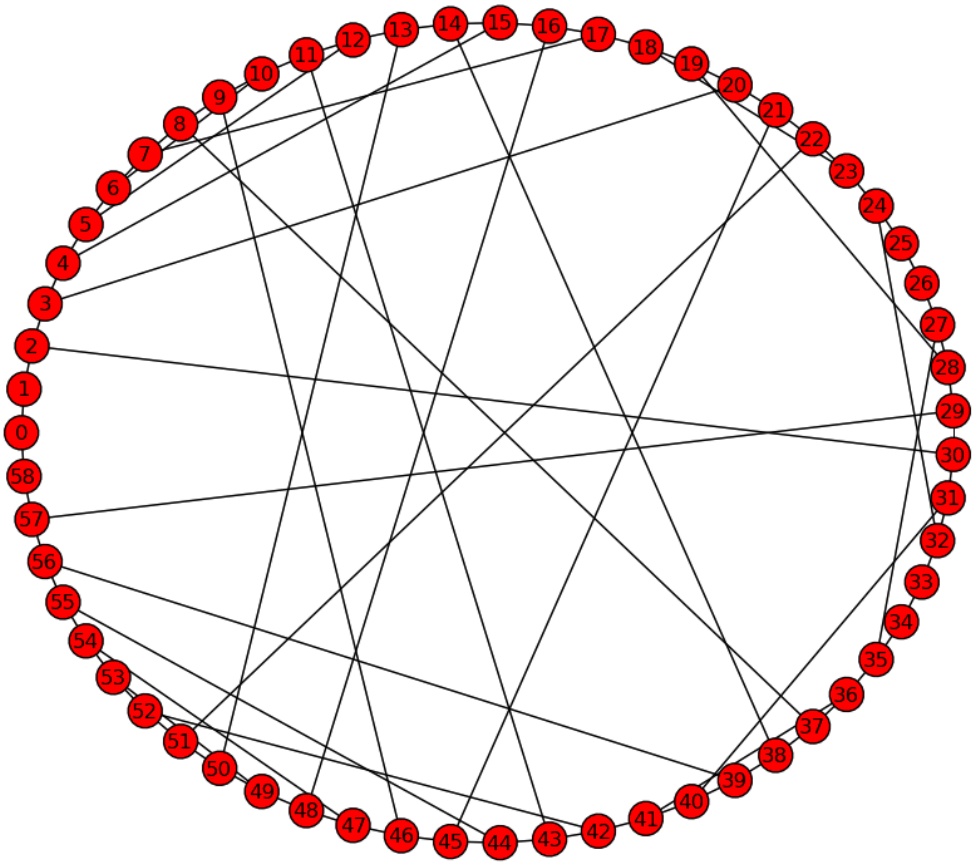
A *family of expanders* is a family of graphs $G_n = (V_n, E_n)$, $|V_n| = n$, such that each graph is d_n-regular, and the edge-expansion of each graph is at least h, for an absolute constant h independent of n. Ideally, we would like to have such a construction for each n, although it is usually enough 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 17.1 *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.

We begin by describing a building block in the construction, which is also an independently interesting construction: a family of expanders with polylogarithmic degree, which have both a very simple description and a very simple analysis.

17.1 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 vector 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

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

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 $L_{p,t} := I - p^{-2}A_{p,t}$ be the normalized Laplacian matrix. We will prove the following bound on the eigenvalues of $L_{p,t}$.

Theorem 17.2 *For every prime p and every $t < p$, if we let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \lambda_n$ be the eigenvalues of M with multiplicities, then, for every $i \in \{2, \ldots, n\}$*

$$1 - \frac{t}{p} \leq \lambda_i \leq 1$$

For example, setting $t = \lfloor p/2 \rfloor$ gives us a family of graphs such that $\lambda_2 \geq 1/2$ for each graph in the family, and hence $\phi(G) \geq 1/4$, 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 Γ with generating multiset S: we have one eigenvector for each character χ 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

$$\begin{aligned}\lambda_{c_0,\ldots,c_t} &= \sum_{a,b\in\mathbb{F}_p} \omega^{\sum_{i=0}^t b\cdot q(a)} \\ &= \sum_{a:q(a)=0} \sum_b \omega^0 + \sum_{a:q(a)\neq 0} \sum_b \omega^{b\cdot q(a)} \\ &= p\cdot|\{a\in\mathbb{F}_p : q(a)=0\}|\end{aligned}$$

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

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

□

17.2 The Zig-Zag Graph Product

Given a d-regular graph G with adjacency matrix A, if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ are the eigenvalues of A 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 times d, 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 λ_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 $\lambda(G) \leq \alpha D$
- H a d-regular graph on D vertices, with $\lambda(H) \leq \beta d$

Then:

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

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

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.

Chapter 18

Analysis of the zig-zag graph product

In which we analyze the zig-zag graph product.

In the previous lecture, we claimed it is possible to "combine" a d-regular graph on D vertices and a D-regular graph on N vertices to obtain a d^2-regular graph on ND vertices which is a good expander if the two starting graphs are. Let the two starting graphs be denoted by H and G respectively. Then, the resulting graph, called the *zig-zag product* of the two graphs is denoted by $G \textcircled{z} H$.

We will use $\lambda(G)$ to denote the eigenvalue with the second-largest absolute value of the normalized adjacency matrix $\frac{1}{d}A_G$ of a d-regular graph G. If $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$ are the eigenvalues of the normalized Laplacian of G, then $\lambda(G) = \max\{1 - \lambda_2, \lambda_n - 1\}$.

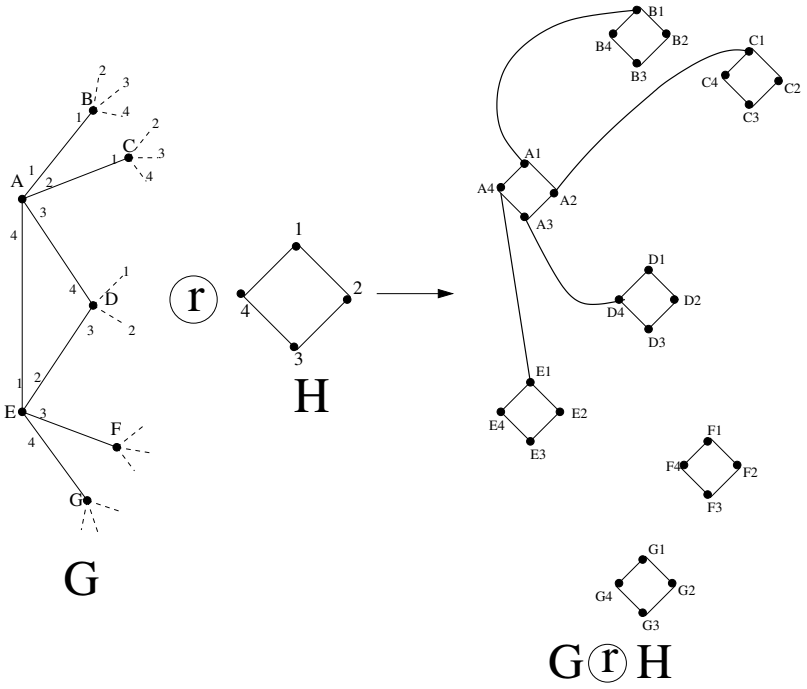
We claimed that if $\lambda(H) \leq b$ and $\lambda(G) \leq a$, then $\lambda(G \textcircled{z} H) \leq a + 2b + b^2$. In this lecture we shall recall the construction for the zig-zag product and prove this claim.

18.1 Replacement Product and Zig-Zag Product

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 $\forall u \in V(G)$ $((u, i), (u, j)) \in E(G \textcircled{r} H)$.

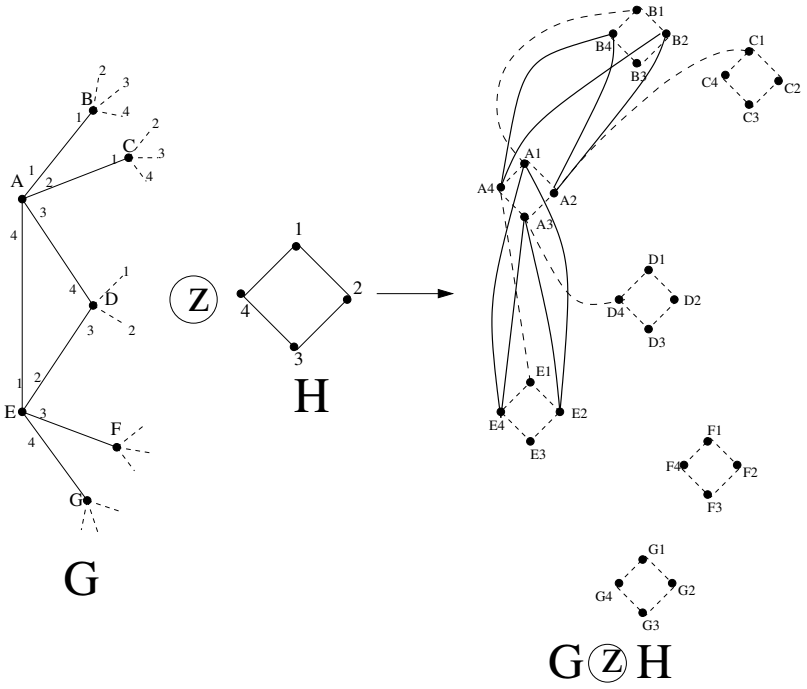
Note that the replacement product constructed as above has ND vertices and is $(d + 1)$-regular.



18.2 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 ℓ 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 $G Ⓩ H$. Using the fact that each edge in $G Ⓡ H$ is made up of three steps in $G Ⓡ H$, we can write M as BAB, where

$$B[(u,i),(v,j)] = \begin{cases} 0 & \text{if } u \neq v \\ \frac{1}{d} & \text{if } u = v \text{ and } \{i,j\} \in H \end{cases}$$

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.

18.3 A Technical Preliminary

We will use the following fact. Suppose that $M = \frac{1}{d}A_G$ is the normalized adjacency matrix of a graph G. Thus the largest eigenvalue of M is 1, with eigenvector $\mathbf{1}$; we have

$$\lambda(G) = \max_{\mathbf{x}\perp\mathbf{1}} \frac{|\mathbf{x}^T M\mathbf{x}|}{||\mathbf{x}||^2} = \max_{\mathbf{x}\perp\mathbf{1}} \frac{||M\mathbf{x}||}{||\mathbf{x}||} \tag{18.1}$$

which is a corollary of the following more general result. Recall that a vector space $S \subseteq \mathbb{R}^n$ is an invariant subspace for a matrix $M \in \mathbb{R}^{n\times n}$ if $M\mathbf{x} \in S$ for every $\mathbf{x} \in S$.

Lemma 18.1 *Let M be a symmetric matrix, and S be a k-dimensional invariant subspace for M. Thus, (from the proof of the spectral theorem) we have that S has an orthonormal*

basis of eigenvectors; let $\lambda_1 \leq \cdots \leq \lambda_k$ be the corresponding eigenvalues with multiplicities; we have

$$\max_{i=1,\ldots k} |\lambda_i| = \max_{\mathbf{x}\in S} \frac{|\mathbf{x}^T M \mathbf{x}|}{||\mathbf{x}||^2} = \max_{\mathbf{x}\in S} \frac{||M\mathbf{x}||}{||\mathbf{x}||}$$

PROOF: If the largest eigenvalue in absolute value is λ_k, then

$$\max_{i=1,\ldots k} |\lambda_i| = \lambda_k = \max_{\mathbf{x}\in S} \frac{\mathbf{x}^T M \mathbf{x}}{||\mathbf{x}||^2}$$

and if it is $-\lambda_1$ (because λ_1 is negative, and $-\lambda_1 > \lambda_n$)

$$\max_{i=1,\ldots k} |\lambda_i| = -\lambda_1 = -\min_{\mathbf{x}\in S} \frac{\mathbf{x}^T M \mathbf{x}}{||\mathbf{x}||^2} = \max_{\mathbf{x}\in S} -\frac{\mathbf{x}^T M \mathbf{x}}{||\mathbf{x}||^2}$$

so we have

$$\max_{i=1,\ldots k} |\lambda_i| \leq \max_{\mathbf{x}\in S} \frac{|\mathbf{x}^T M \mathbf{x}|}{||\mathbf{x}||^2} \tag{18.2}$$

From Cauchy-Schwarz, we have

$$|\mathbf{x}^T M \mathbf{x}| \leq ||\mathbf{x}|| \cdot ||M\mathbf{x}||$$

and so

$$\max_{\mathbf{x}\in S} \frac{|\mathbf{x}^T M \mathbf{x}|}{||\mathbf{x}||^2} \leq \max_{\mathbf{x}\in S} \frac{||M\mathbf{x}||}{||\mathbf{x}||} \tag{18.3}$$

Finally, if $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is the basis of orthonormal eigenvectors in S such that $M\mathbf{x}_i = \lambda_i$, then, for every $\mathbf{x} \in S$, we can write $\mathbf{x} = \sum_i a_i \mathbf{x}_i$ and

$$||M\mathbf{x}|| = ||\sum_i \lambda_i a_i \mathbf{x}_i|| = \sqrt{\sum_i \lambda_i^2 a_i^2} \leq \max_{i=1,\ldots,k} |\lambda_i| \cdot \sqrt{\sum_i a_i^2} = \max_{i=1,\ldots,k} |\lambda_i| \cdot ||\mathbf{x}||$$

so

$$\max_{\mathbf{x}\in S} \frac{||M\mathbf{x}|}{||\mathbf{x}||} \leq \max_{i=1,\ldots,k} |\lambda_i| \tag{18.4}$$

and the Lemma follows by combining (18.2), (18.3) and (18.4). □

18.4 Analysis of the zig-zag Product

Theorem 18.2 *Let G be a D-regular graph with n nodes, H be a d-regular graph with D nodes, and let $a := \lambda(G)$, $b := \lambda(H)$, and let the normalized adjacency matrix of $G ⓩ H$ be $M = BAB$ where A and B are as defined in Section 18.1.*

Then $\lambda(G ⓩ H) \leq a + b + b^2$

PROOF: Let $\mathbf{x}\mathbb{R}^{n\times D}$ be such that $\mathbf{x} \perp \mathbf{1}$. We refer to a set of coordinates of $\mathbf{x}$ corresponding to a copy of H as a "block" of coordinate.

We write $\mathbf{x} = \mathbf{x}_{||} + \mathbf{x}_{\perp}$, where $\mathbf{x}_{||}$ is constant within each block, and $\mathbf{x}_{\perp}$ sums to zero within each block. Note both $\mathbf{x}_{||}$ and $\mathbf{x}_{\perp}$ are orthogonal to $\mathbf{1}$, and that they are orthogonal to each other.

We want to prove

$$\frac{|\mathbf{x}^T M \mathbf{x}|}{||\mathbf{x}||^2} \leq a + b + b^2 \tag{18.5}$$

We have (using the fact that M is symmetric)

$$|\mathbf{x}^T M \mathbf{x}| =\leq |\mathbf{x}_{||}^T M \mathbf{x}_{||}| + 2|\mathbf{x}_{||}^T M \mathbf{x}_{\perp}| + |\mathbf{x}_{\perp}^T M \mathbf{x}_{\perp}|$$

And it remains to bound the three terms.

1. $|\mathbf{x}_{||}^T M \mathbf{x}_{||}| \leq a||\mathbf{x}_{||}||^2$

   Because, after writing $M = BAB$, we see that $B\mathbf{x}_{||} = \mathbf{x}_{||}$, because B is the same as $I_n \otimes \left(\frac{1}{d} A_H\right)$, the tensor product of the identity and of the normalized adjacency matrix of H. The normalized adjacency matrix of H leaves a vector parallel to all-ones unchanged, and so B leaves every vector that is constant in each block unchanged.

   Thus

   $$|\mathbf{x}_{||}^T M \mathbf{x}_{||}| = |\mathbf{x}_{||}^T A \mathbf{x}_{||}$$

   Let $\mathbf{y}$ be the vector such that y_v is equal to the value that $\mathbf{x}_{||}$ has in the block of v. Then

   $$|\mathbf{x}_{||}^T A \mathbf{x}_{||}| = 2 \sum_{\{(v,i),(w,j)\} \in E_{G(\mathbb{Z})H}} y_v y_w = \mathbf{y}^T A_G \mathbf{y} = aD||\mathbf{y}||^2 \leq a||\mathbf{x}_{||}||^2$$

   because $\mathbf{y} \perp \mathbf{1}$ and $||\mathbf{y}||^2 = \frac{1}{D}||\mathbf{x}_{||}||^2$

2. $|\mathbf{x}_{\perp}^T M \mathbf{x}_{\perp}| \leq b^2||\mathbf{x}_{\perp}||^2$ Because, from Cauchy-Schwarz and the fact that permutation matrices preserve length, we have

   $$|\mathbf{x}_{\perp}^T BAB \mathbf{x}_{\perp}| \leq ||B\mathbf{x}_{\perp}|| \cdot ||AB\mathbf{x}_{\perp}|| = ||B\mathbf{x}_{\perp}||^2$$

   Now let us call $\mathbf{x}_{\perp}^v$ the restriction of $\mathbf{x}_{\perp}$ to coordinates of the form (v, i) for $i = 1, \ldots, D$. Then each $\mathbf{x}_{\perp}^v$ is orthogonal to the all-one vector and $A_H \mathbf{x}_{\perp}^v \leq db||\mathbf{x}_{\perp}^v||$, so

   $$||B\mathbf{x}_{\perp}||^2 = \sum_v ||d^{-1} A_H \mathbf{x}_{\perp}^v||^2 \leq \sum_v b^2 ||\mathbf{x}_{\perp}^v||^2 = b^2 ||\mathbf{x}_{\perp}||^2$$

3. $2|\mathbf{x}_{||}^T M \mathbf{x}_{\perp}| \leq b||\mathbf{x}||^2$

Because, from Cauchy-Schwarz, the fact that $B\mathbf{x}_{||} = \mathbf{x}_{||}$ and the fact that permutation matrices preserve length, we have

$$|\mathbf{x}_{||}^T BAB\mathbf{x}_\perp| \leq ||B\mathbf{x}_{||}|| \cdot ||AB\mathbf{x}_\perp|| = ||\mathbf{x}_{||}|| \cdot ||B\mathbf{x}_\perp||$$

and we proved above that

$$||B\mathbf{x}_\perp|| \leq b||\mathbf{x}_\perp||$$

so

$$|\mathbf{x}_{||}^T BAB\mathbf{x}_\perp| \leq b \cdot ||\mathbf{x}_{||}|| \cdot ||\mathbf{x}_\perp|| \leq \frac{b}{2}(||\mathbf{x}_{||}||^2 + ||\mathbf{x}_\perp||^2) = \frac{b}{2}||\mathbf{x}||^2$$

□

Chapter 19

Algebraic construction of expanders

In which we present an algebraic construction of expanders.

19.1 The Marguli-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 $V_n := (\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z})$ 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 four steps, and it will refer to certain infinite "graphs."

We define an infinite family of graphs R_n, such that the vertex set of R_n is $(\mathbb{R}/n\mathbb{Z} \times \mathbb{R}/n\mathbb{Z})$, that is, every vertex of R_n is a pair (x, y), where $0 \leq x < n$ and $0 \leq y < n$, and we think of x and y as elements of the group $\mathbb{R}/n\mathbb{Z}$ in which we do addition modulo n. Every vertex of R_n is connected to the vertices

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

and is 4-regular. For each of these graphs, we will define a "spectral gap" $\lambda_2(R_n)$; we put "spectral gap" in quotes because, although it is actually the second smallest eigenvalue of a Laplacian operator, we will define it purely formally as the minimum of a certain optimization problem.

We will also define the graph Z, whose vertex set is $\mathbb{Z} \times \mathbb{Z} - \{(0,0\}$, and such that each vertex (a,b) is connected to

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

so that Z is also 4-regular. We will define a "spectral gap" $\lambda_1(Z)$, again purely formally as the infimum of a certain expression, although it is the infimum of the spectrum of a certain Laplacian operator. We will also define an "edge expansion" $\phi(Z)$ of Z.

The proof of the expansion of G_n will proceed by establishing the following four facts:

1. $\lambda_2(G_n) \geq \frac{1}{3}\lambda_2(R_n)$
2. $\lambda_2(R_n) \geq \lambda_1(Z)$
3. $\phi(Z) \leq \sqrt{2\lambda_1(Z)}$
4. $\phi(Z) \geq \frac{1}{7}$

The first step will be a discretization argument, showing that a test vector of small Rayleigh quotient for G_n can be turned into a test function of small Rayleigh quotient for R_n. The second step is the most interesting and unexpected part of the proof; we will not spoil the surprise of how it works. The third step is proved the same way as Cheeger's inequality. The fourth step is just a careful case analysis.

19.2 First Step: The Continuous Graph R_n

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:

$$\lambda_2(R_n) := \inf_{f \in \ell_2([0,n)^2) \,:\, \int_{[0,n)^2} f = 0} \frac{\int_{[0,n)^2} |f(x,y) - f(S(x,y))|^2 + |f(x,y) - f(T(x,y))|^2 \mathrm{d}x\mathrm{d}y}{4\int_{[0,n)^2}(f(x,y))^2 \mathrm{d}x\mathrm{d}y}$$

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 19.1 $\lambda_2(G_n) \geq \frac{1}{3} \cdot \lambda_2(R_n)$.

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

$$\tilde{f}(z) := f(\lfloor z \rfloor)$$

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 2, 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 $\lfloor S(z) \rfloor$ equals either $S(\lfloor z \rfloor)$ or $S(\lfloor z \rfloor)+(0,1)$, and that $floor(T(z))$ equals either $T(\lfloor z \rfloor)$ or $T(\lfloor z \rfloor)+(1,0)$. 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"

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

to bound the above quantity

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

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. $\square$

19.3 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 d-regular 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 smallest eigenvalue of G as

$$\lambda_1(G) := \inf_{f \in \ell_2(V)} \frac{\sum_{(u,v)\in V} |f(u) - f(v)|^2}{d \sum_v f^2(v)}$$

So that

$$\lambda_1(Z) := \inf_{f \in \ell_2(\mathbb{Z}\times\mathbb{Z}-\{(0,0)\})} \frac{\sum_{a,b} |f(a,b) - f(a,a+b)|^2 + |f(a,b) - f(a+b,a)|^2}{4 \sum_{a,b} f^2(a,b)}$$

We want to show the following result.

Theorem 19.2 *For every n, $\lambda_2(R_n) \geq \lambda_1(Z)$.*

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. $\square$

19.4 Third Step: A Cheeger Inequality for countable graphs

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

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

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

Theorem 19.3 (Cheeger inequality for countable graphs) *For every graph $G = (V, E)$ with a countably infinite set of vertices we have*

$$\phi(G) \leq \sqrt{2 \cdot \lambda_1(G)}$$

PROOF: 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 ϕ is at most $\sqrt{2r}$ where

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

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 $t > t_{\min}$,

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

and, for all $t \geq 0$

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

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

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

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

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

□

19.5 Expansion of Z

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

Theorem 19.4 $\phi(Z) \geq \frac{1}{7}$

PROOF: 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 19.5 $E(A - A_0, \bar{A}) \geq |A - A_0| = |A| - |A_0|$.

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 19.6 $E(A_0, \bar{A}) \geq 4|A_0| - 3|A - A_0| = 7|A_0| - 3|A|$

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 the first plus 1/7 times the second) gives us the theorem. □

Chapter 20

Probabilistic construction of expanders

In which we present a probabilistic construction of expanders.

We have seen a combinatorial construction of expanders, based on the zig-zag graph product, and an algebraic one. Today we see how to use the probabilistic method to show that random graphs, selected from an appropriate probability distribution, are expanders with high probability.

A non-trivial part of today's lecture is the choice of distribution. For us, a family of expanders is a family of regular graphs of fixed degree, but if we pick a graph at random according to the Erdös-Renyi distribution, selecting each pair $\{u, v\}$ to be an edge independently with probability p, then we do not get a regular graph and, indeed, not even a bounded-degree graph. (Even when $p = O(1/n)$, ensuring constant average degree, the maximum degree is of the order of $\log n / \log \log n$.)

This means that we have to study distributions of graphs in which there are correlations between edges, which are often difficult to reason about.

We could study the expansion of random d-regular graphs, but that is a particularly challenging distribution of graphs to analyze. Instead, the following distributions over d-regular graphs are usually considered over a vertex set V:

- Pick at random d perfect matching over V, and let E be their union
- Pick at random $d/2$ permutations $f_1, \ldots, f_{\frac{d}{2}} : V \to V$, and have an edge for each pair $\{v, f(v)\}$, for $i = 1, \ldots, d/2$.

The first method is applicable when n is even, and the second method is applicable when d is even. (When n and d are both odd, it is not possible to have an n-vertex d-regular graph, because the number of edges in such a graph is $dn/2$.)

We will study the expansion of graphs generated according to the first distribution, and show that there exists an integer d and a $c > 0$, such that a random d regular graph on n vertices has probability at least $1 - 1/n^{\Omega(1)}$ of having edge expansion at least c.

In particular, we will show that for $d = 18$, the probability that a random 18-regular graph has expansion $\geq \frac{1}{108}$ is at least $1 - O(1/n^2)$. Our bounds will be very loose, and much tighter analyses are possible.

We have to show that, with high probability over the choice of the graph, every set of vertices $S \subseteq V$ with $|S| \leq |V|/2$ has at least $c \cdot d \cdot |S|$ edges leaving it.

The common approach to prove that a random object satisfies a certain collection of properties is to prove that each property holds with high probability, and then to use a union bound to show that all properties are simultaneously true with high probability. For a d-regular graph to have expansion c, we want every set S of size $\leq n/2$ to have at least $cd|S|$ outgoing edges; a naive approach would be to show that such a property holds for every fixed set except with probability at most $<< \frac{1}{2^n}$, and then take a union bound over all 2^{n-1} sets of size $\leq n/2$.

Unfortunately the naive approach does not work, because the probability that small sets fail to expand is much higher than 2^{n-1}. For example, the probability that a fixed set of $d+1$ nodes form a clique is at least of the order of $1/n^{d^2}$. Fortunately, the number of small sets is small, and if the probability of a fixed set of size k being non-expanding is, say, at most $1/\binom{n}{k}^2$, then, by taking a union bound over all sets of size k, the probability that there is a non-expanding set of size k is at most $1/\binom{n}{k}$, and then by taking a union bound over all sizes k we get that the probability that there is a non-expanding set is at most inverse polynomial in n.

Let $\Gamma(S)$ denote the set of nodes that have at least a neighbor in S. If $|\Gamma(S) - S| \leq t$, then there are most t edges leaving from S. In order to upper bound the probability that there are $\leq \frac{1}{2}|S|$ edges leaving S, we will upper bound the probability that $|\Gamma(S) - S| \leq \frac{1}{2}|S|$.

It will be convenient to have the following model in mind for how a random perfect matchings is chosen. Let $v_1, \ldots, v_n$ be an arbirtary ordering of the vertices such that $S = \{v_1, \ldots, v_k\}$, then the following algorithm samples a random perfect matching over V:

- $M := \emptyset$, $C := \emptyset$
- while $C \neq V$
  - let v be the smallest-index unmatched vertex in $V - C$
  - let w be a randomly selected unmatched vertex in $V - (C \cup \{v\})$
  - $M := M \cup \{\{v, w\}\}$; $C := C \cup \{v, w\}$
- return M

It is easy to see that the above algorithm has $(n-1) \cdot (n-3) \cdots 3 \cdot 1$ possible outputs, each equally likely, each distinct, and that $(n-1) \cdot (n-3) \cdots 3$ is also the number of perfect matchings over a set of n vertices, so that the algorithm indeed samples a uniformly distributed perfect matching.

Now, fix a set S of size $k \leq n/2$ and a set $T \subseteq V - S$ of size $k/6$. The probability that, in a random matching, the vertices of S are all matched to vertices in $S \cup T$ is at most the

probability that, during the first $k/2$ executions of the "while" loop, the randomly selected vertex v_j is in $S \cup T$.

For $i = 1, \ldots, k/2$, conditioned on the first $i-1$ iterations picking a vertex $w \in S \cup T$, the probability that this happens on the i-th iteration is the number of unmatched vertices in $(S \cup T) - (C \cup \{v\})$ which is $\frac{7}{6}k - 2i + 1$, divided by the total number of unmatched vertices in $V - (C \cup \{v\})$, which is $n - 2i + 1$.

Thus, the probability that, in a random matching, all vertices of S are matched to vertices in $S \cup T$ is at most

$$\prod_{i=1}^{k/2} \frac{\frac{7}{6}k - 2i + 1}{n - 2i + 1} < \prod_{i=k/3+1}^{k/2} \frac{.5k}{.5n} = \left(\frac{k}{n}\right)^{k/6}$$

when we pick G as the union of d random matchings, the probability that all the neighbors of S are in $S \cup T$ is at most the above bound raised to the power of d:

$$\Pr[\Gamma(S) \subseteq S \cup T] \leq \left(\frac{k}{n}\right)^{dk/6}$$

and taking a union bound over all choices of S of size k and all choices of T of size $k/6$, we have

$$\begin{aligned}
&\Pr[\exists S \text{ such that } |S| = k \text{ and } \phi(S) \leq 1/6d] \leq \\
&\Pr[\exists S \text{ such that } |S| = k \text{ and } |\Gamma(S) - S| \leq k/6] \\
&\leq \left(\frac{k}{n}\right)^{dk/6} \cdot \binom{n}{k/6} \cdot \binom{n}{k} \\
&\leq \left(\frac{k}{n}\right)^{dk/6} \cdot \binom{n}{k}^2 \\
&\leq \binom{n}{k}^{-d/6} \cdot \binom{n}{k}^2
\end{aligned}$$

Now, for $d = 18$, taking a union bound over all $k \geq 2$ (in a graph without self-loops, every singleton set is expanding), we have

$$\Pr\left[\phi(G) \leq \frac{1}{18 \cdot 6}\right] \leq \sum_{k=2}^{n/2} \frac{1}{\binom{n}{k}} \leq O\left(\frac{1}{n^2}\right)$$

Chapter 21

Properties of Expanders

In which we prove properties of expander graphs.

21.1 Quasirandomness of Expander Graphs

Recall that if G is a d-regular graph, and A is its adjacency matrix, then, if we call $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ the eigenvalues of A with repetitions, we are interested in the parameter $\sigma_2(G) := \max_{i=2,\ldots,n}\{|\lambda_i|\}$, and we have

$$\sigma_2(G) = \left\| A - \frac{d}{n} J \right\|$$

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 $\sigma_2(G)$ 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 $\frac{d}{|V|}|S||T|$.

For two (possibly overlapping) sets of vertices S, T, we define $edges_G(S,T)$ to be the number of edges with one endpoint in S and one endpoint in T, with edges having both endpoints in $S \cap T$, if any, counted twice.

Lemma 21.1 (Expander Mixing Lemma) *Let $G = (V, E)$ be a d-regular graph, and let S and T be two disjoint subsets of vertices. Then*

$$\left| edges_G(S,T) - \frac{d}{|V|} \cdot |S| \cdot |T| \right| \leq \sigma_2(G) \cdot \sqrt{|S| \cdot |T|}$$

PROOF: We have

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

and

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

so

$$\begin{aligned}
&\left| edges_G(S,T) - \frac{d}{|V|} \cdot |S| \cdot |T| \right| \\
&= \cdot \left| \mathbf{1}_S^\top A \mathbf{1}_T - \frac{d}{|V|} \mathbf{1}_S^\top J \mathbf{1}_T \right| \\
&= \cdot \left| \mathbf{1}_S^\top \left(A - \frac{d}{|V|} J \right) \mathbf{1}_T \right| \\
&\leq ||\mathbf{1}_S|| \cdot \left\| A - \frac{d}{|V|} J \right\| \cdot \|\mathbf{1}_T\| \\
&= \sqrt{|S|} \cdot \sigma_2(G) \cdot \sqrt{|T|}
\end{aligned}$$

□

Note that, for every disjoint S, T, we have $\sqrt{|S| \cdot |T|} \leq |V|/2$, and so the right-hand side in the expander mixing lemma is at most $\frac{\sigma_2(G)}{d} \cdot |E|$, which is a small fraction of the total number of edges if σ_2 is small compared to d.

21.2 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 $P := \frac{1}{d}A$ is the normalized adjacency matrix of an undirected regular graph G, then $P(u,v)$ 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 $\sum_u \mathbf{p}(u)P(u,v)$, which means that the probability distribution after one step is described by the vector $\mathbf{p}^\top \cdot P$, and because of the symmetric of P, this is the same as $P\mathbf{p}$.

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 $P^t\mathbf{p}$.

The parameter σ_2 of P^t is equal to $(\sigma_2(G)/d)^t$, and so if G has a parameter σ_2 bounded away from d, and if t is large enough, we have that the parameter σ_2 of P^t is very small,

and so P^t is close to $\frac{1}{n}J$ in matrix norm. If P^t was actually equal to $\frac{1}{n}J$, then $P^t \cdot \mathbf{p}$ would be equal to the uniform distribution, for every distribution $\mathbf{p}$. We would thus expect $P^t \cdot \mathbf{p}$ 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 21.2 (Mixing Time of Random Walks in Expanders) *Let G be a regular graph, and P be its normalized adjacency matrix. Then for every distribution $\mathbf{p}$ over the vertices and every t, we have*

$$||\mathbf{u} - P^t\mathbf{p}||_1 \leq \sqrt{|V|} \cdot (\sigma_2(G)/d)^t$$

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

In particular, if $t > \frac{dc}{d-\sigma_2(G)} \cdot \ln \frac{|V|}{\epsilon}$, then $||\mathbf{u} - P^t\mathbf{p}||_1 \leq \epsilon$, 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 $\sigma_2(G) = ||P - \bar{J}||$.

We have

$$\begin{aligned} &||\mathbf{u} - P^t\mathbf{p}||_1 \\ &\leq \sqrt{|V|} \cdot ||\mathbf{u} - P^t\mathbf{p}|| \\ &\leq \sqrt{|V|} \cdot ||\bar{J}\mathbf{p} - P^t\mathbf{p}|| \\ &\leq \sqrt{|V|} \cdot ||\bar{J} - P^t|| \cdot ||\mathbf{p}|| \end{aligned}$$

$$\leq \sqrt{|V|} \cdot (\sigma_2(G)/d)^t$$

□

The last result that we discussed today is one more instantiation of the general phenomenon that "if $\sigma_2(G)$ 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 $\sigma_2(G)$ is small.

Lemma 21.3 (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*

$$\mathbb{P}\left[\frac{1}{t}\sum_i f(v_i) \geq \mathbb{E}\, f + \epsilon + \frac{\sigma_2(G)}{d}\right] \leq e^{-\Omega(\epsilon^2 t)}$$

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 $\sigma_2(G) \leq d/12$, 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 $t \cdot r(n)$.