List of Lectures

Chapter 1

Overview

This course begins with a review of classical results in spectral graph theory, followed by an exploration of several recent major developments, with a focus on algorithmic results.

Classical Results

After we introduce basic concepts and results from linear algebra, we study Cheeger's inequality, a foundational result in spectral graph theory. This theorem states the spectral gap[1] is large if and only if the graph expansion[2] is large. This connection between an algebraic quantity and a combinatorial property has three major applications.

1. Random Walks on Graphs

Analyzing the mixing time of random walks[3] is an important topic with numerous applications in random sampling and approximate counting [LPW06]. A basic result in spectral graph theory is that the mixing time of random walks is roughly equal to the inverse of the spectral gap. Cheeger's inequality thus implies that a graph has small mixing time if and only if it has large expansion, providing a combinatorial characterization useful for analyzing mixing time.

2. Expander Graphs

Expander graphs, typically defined as sparse graphs with large expansion, have surprisingly many applications in theoretical computer science and mathematics [HLW06]. Cheeger's inequality provides an efficient method to certify that a graph has large expansion, which is crucial in the construction of expander graphs. We will study the basic properties of expander graphs and Ramanujan graphs[4], a combinatorial construction called the zig-zag product, and several interesting applications of expander graphs.

[1] Defined as the difference between the first and second eigenvalues.

[2] Which quantifies how well a graph is connected by comparing the number of edges leaving a subset of vertices to the size of the subset.

[3] Defined as the number of steps required for the probability distribution on vertices to converge to the limiting distribution.

[4] Expander graphs with the maximum possible spectral gap, roughly speaking.

3. Graph Partitioning

The proof of Cheeger's inequality provides a fast algorithm to output a subset of vertices of approximately minimal expansion. This is known as the spectral partitioning algorithm, a widely-used heuristic in practical graph partitioning applications with good performance, e.g. [SM00].

Recent Developments

The recent developments still center around these three primary themes, but introduce significantly new ideas and techniques, extending the reach of spectral graph theory.

1. Generalizations of Cheeger's Inequality and Graph Partitioning

Spectral graph theory has a long history, but only in the last decade have researchers begun to explore graph partitioning using higher eigenvalues, inspired by the influential work on small-set expansion [ABS10]. We study the so-called higher-order Cheeger inequality [LOT14, LRTV12] and improved Cheeger inequality [KLL$^+$13], which use higher eigenvalues to design multi-way graph partitioning algorithm and to analyze the classical spectral partitioning algorithm. Until recently, all Cheeger-type inequalities apply only to edge expansion of undirected graphs. We discuss how to use reweighted eigenvalues to extend these inequalities to vertex expansion [KLT22], as well as to directed graphs and hypergraphs [LTW23].

2. Random Walks and Graph Decompositions

Random walks can be used to design local graph partitioning algorithms with a running time depending only on the output size, providing a valuable algorithmic tool for processing massive graphs [ST13]. We present a unified spectral analysis for this result and the small-set expansion result in [ABS10]. We also study the combinatorial approach used in [ST13], based on the Lovász-Simonovits curve, which can be applied to analyze other diffusion processes and to obtain improved analyses. Finally, we discuss how these results can be used to develop graph decomposition algorithms, which are crucial tools for designing both fast and approximation algorithms.

3. Spectral Sparsification and Applications

Expander graphs can be seen as sparse approximations of complete graphs. Spectral sparsification involves constructing a sparse graph that approximates the spectral properties of a dense graph [ST11]. The study of this problem has been highly productive, leading to significant results and techniques. The analysis of a natural random sampling algorithm [SS11] brought the tools of matrix concentration inequalities to the field, with important applications in designing fast algorithms for solving Laplacian linear equations []. The design of optimal spectral sparsification algorithms led to the potential function developed in [BSS14, ALO15], with applications in designing approximation algorithms and far-reaching consequences in mathematics [MSS14]. We will study these results as well as recent developments in spectral sparsification for directed graphs and a new discrepancy-theoretic approach for constructing spectral sparsifiers.

4. Semidefinite Programming and Approximation Algorithms

Expander graphs satisfy a local-to-global property that is useful for designing approximation algorithms [AKK+08]. We study the correlation rounding method for semidefinite programming developed in [BRS11, GS11] which extends this approach to low threshold rank graphs[5]. We also discuss the subspace enumeration method [Kol10] and the subexponential time approximation algorithm for Unique Games [ABS10]. It remains an open question whether these methods can be extended to obtain a subexponential time approximation algorithm for the maximum cut problem beyond the Goemans-Williamson approximation ratio.

5. Cut-Matching Game and Matrix Multiplicative Update

The cut-matching game is an iterative framework for constructing expanders through flows and cuts, originally developed to design fast algorithms for approximating edge expansion [KRV09]. It has since found unexpected applications in designing both fast and approximation algorithms for other graph problems. We study the original proof in [KRV09] and a more systematic proof using the matrix multiplicative method [AK16]. We further study an almost linear-time $O(\sqrt{\log n})$-approximation algorithm for edge expansion [She09], building on the expander flow framework in the seminal work [ARV09].

6. High-Dimensional Expanders and Mixing Time

High-dimensional expanders generalize expander graphs to higher dimensions [Lub18], with recent breakthrough applications in error correcting codes and analysis of random walks. We study how this new concept provides a local-to-global way to bound the spectral gap of the random walk matrix [Opp18, KO20], leading to an elegant solution to the matroid expansion conjecture [ALOV19]. We will also see how this approach is sharpened to develop the spectral independence framework [AL20, ALO20], a powerful method for analyzing the mixing time of random walks.

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 563–572. IEEE Computer Society, 2010. 4, 5, 79, 87

[AK16] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. *J. ACM*, 63(2):12:1–12:35, 2016. 5

[AKK+08] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi. Unique games on expanding constraint graphs are easy. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 21–28. ACM, 2008. 5

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020*, pages 1198–1211. ACM, 2020. 5

[ALO15] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. Spectral sparsification and regret minimization beyond matrix multiplicative updates. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015*, pages 237–245. ACM, 2015. 4

[5] Graphs with few eigenvalues close to the first eigenvalue.

[ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 1319–1330. IEEE, 2020. 5

[ALOV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an FPRAS for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019*, pages 1–12. ACM, 2019. 5

[ARV09] Sanjeev Arora, Satish Rao, and Umesh V. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):5:1–5:37, 2009. 5

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 472–481. IEEE Computer Society, 2011. 5

[BSS14] Joshua D. Batson, Daniel A. Spielman, and Nikhil Srivastava. Twice-ramanujan sparsifiers. *SIAM Rev.*, 56(2):315–334, 2014. 4

[GS11] Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with PSD objectives. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011*, pages 482–491. IEEE Computer Society, 2011. 5

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 3, 19, 20, 54, 57, 61, 62, 66, 69, 70, 72

[KLL+13] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee, Shayan Oveis Gharan, and Luca Trevisan. Improved cheeger's inequality: analysis of spectral partitioning algorithms through higher order spectral gap. In *Symposium on Theory of Computing Conference, STOC'13*, pages 11–20. ACM, 2013. 4

[KLT22] Tsz Chiu Kwok, Lap Chi Lau, and Kam Chuen Tung. Cheeger inequalities for vertex expansion and reweighted eigenvalues. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 366–377. IEEE, 2022. 4

[KO20] Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Comb.*, 40(2):245–281, 2020. 5

[Kol10] Alexandra Kolla. Spectral algorithms for unique games. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010*, pages 122–130. IEEE Computer Society, 2010. 5

[KRV09] Rohit Khandekar, Satish Rao, and Umesh V. Vazirani. Graph partitioning using single commodity flows. *J. ACM*, 56(4):19:1–19:15, 2009. 5

[LOT14] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37:1–37:30, 2014. 4, 79, 80

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006. 3, 33, 34, 35, 42, 43, 44

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh S. Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC*, pages 1131–1140. ACM, 2012. 4, 79, 80, 87, 88

[LTW23] Lap Chi Lau, Kam Chuen Tung, and Robert Wang. Cheeger inequalities for directed graphs and hypergraphs using reweighted eigenvalues. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC*, pages 1834–1847. ACM, 2023. 4

[Lub18] Alexander Lubotzky. High dimensional expanders. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 705–730, 2018. 5

[MSS14] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Ramanujan graphs and the solution of the Kadison Singer problem. In *Proceedings of the international Congress of Mathematicians*, volume 3, pages 363–386. 2014. 4

[Opp18] Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part I: descent of spectral gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018. 5

[She09] Jonah Sherman. Breaking the multicommodity flow barrier for o(vlog n)-approximations to sparsest cut. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 363–372. IEEE Computer Society, 2009. 5

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000. 4, 27

[SS11] Daniel A. Spielman and Nikhil Srivastava. Graph sparsification by effective resistances. *SIAM J. Comput.*, 40(6):1913–1926, 2011. 4

[ST11] Daniel A. Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM J. Comput.*, 40(4):981–1025, 2011. 4

[ST13] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.*, 42(1):1–26, 2013. 4, 42

Chapter 2

Graph Spectrum

The linear algebraic approach to algorithmic graph theory views graphs as matrices and use concepts and tools in linear algebra to design and analyze algorithms for graph problems.

Spectral graph theory focuses on using eigenvalues and eigenvectors of matrices associated with the graph to study its combinatorial properties. While it may not be clear why eigenvalues provide useful information about the combinatorial properties of graphs, they do, and a surprising amount of information can be obtained from them.

In this chapter, we consider the adjacency matrix and the Laplacian matrix of an undirected graph, and study some basic results in spectral graph theory such as characterizations of bipartiteness and connectedness. General references for this chapter include [Spi19, Tre17].

2.1 Adjacency Matrix

We start with simple graphs for simplicity. The generalization to weighted graphs is straightforward.

Definition 2.1 (Adjacency Matrix of Simple Graph)**.** *Given a simple graph $G = (V = [n], E)$, the adjacency matrix $A(G)$ is an $n \times n$ matrix where $A_{ij} = A_{ji} = 1$ if $ij \in E(G)$ and $A_{ij} = A_{ji} = 0$ otherwise.*

The adjacency matrix of an undirected graph is symmetric. Therefore, by the spectral theorem for real symmetric matrices in Theorem A.5, the adjacency matrix has an orthonormal basis of eigenvectors with real eigenvalues. We denote the eigenvalues of the adjacency matrix by

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n.$$

Let us begin with some examples and compute their spectra.

Example 2.2 (Complete Graphs)**.** *If G is a complete graph, then $A(G) = J - I$, where J denotes the all-one matrix. Any vector is an eigenvector of I with eigenvalue 1. Hence, the eigenvalues of A are one less than those of J. Since J is rank 1, there are $n - 1$ eigenvalues of 0. The all-ones vector is an eigenvector of J with eigenvalue n. Thus, $n - 1$ is an eigenvalue of A with multiplicity 1, and -1 is an eigenvalue of A with multiplicity $n - 1$.*

This example exhibits the largest gap between the largest eigenvalue and the second largest eigenvalue.

Example 2.3 (Complete Bipartite Graphs). *Let $K_{p,q}$ be the complete bipartite graph with p vertices on one side and q vertices on the other side. Its adjacency matrix $A(K_{p,q})$ is rank 2, so 0 is an eigenvalue with multiplicity $p+q-2$, and there are two non-zero eigenvalues α and β. By Fact A.35, the sum of the eigenvalues is equal to the trace of A, which is 0 since there are no self-loops. Thus, $\alpha = -\beta$. To determine α, consider the characteristic polynomial $\det(xI - A) = (x-\alpha)(x+\alpha)x^{p+q-2} = x^{p+q} - \alpha^2 x^{p+q-2}$. Using the Leibniz formula for determinants in Fact A.26, any term contributing to x^{p+q-2} must have $p+q-2$ diagonal entries, with the remaining two entries must be $-A_{ij}$ and $-A_{ji}$ for some i, j. There are totally pq such terms (one for each edge), and the sign of the corresponding permutation is -1 because it has only one inversion pair. Therefore, $\alpha^2 = pq$, and thus $|\alpha| = \sqrt{pq}$. To conclude, the spectrum is $(\sqrt{pq}, 0, \ldots, 0, -\sqrt{pq})$, where 0 is an eigenvalue with multiplicity $p+q-2$.*

In Section 2.5, you are asked to compute the spectrum of the cycles and the hypercubes.

Bipartiteness

It turns out that bipartite graphs can be characterized by the spectrum of their adjacency matrix. The following lemma says that the spectrum of a bipartite graph is symmetric around the origin on the real line.

Lemma 2.4 (Spectrum of Bipartite Graph is Symmetric). *If G is a bipartite graph and α is an eigenvalue of $A(G)$ with multiplicity k, then $-\alpha$ is also an eigenvalue of $A(G)$ with multiplicity k.*

Proof. If G is a bipartite graph, we can permute its rows and columns of G to obtain the form

$$A(G) = \begin{pmatrix} 0 & B \\ B^\top & 0 \end{pmatrix}.$$

Suppose $u = \begin{pmatrix} x \\ y \end{pmatrix}$ is an eigenvector of $A(G)$ with eigenvalue α. Then

$$\begin{pmatrix} 0 & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \alpha \begin{pmatrix} x \\ y \end{pmatrix} \qquad \iff \qquad B^\top x = \alpha y \ \text{ and } \ By = \alpha x.$$

Now consider $\begin{pmatrix} x \\ -y \end{pmatrix}$. It satisfies:

$$\begin{pmatrix} 0 & B \\ B^\top & 0 \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -By \\ B^\top x \end{pmatrix} = \begin{pmatrix} -\alpha x \\ \alpha y \end{pmatrix} = -\alpha \begin{pmatrix} x \\ -y \end{pmatrix}.$$

Thus, $\begin{pmatrix} x \\ -y \end{pmatrix}$ is an eigenvector of $A(G)$ with eigenvalue $-\alpha$. By construction, k linearly independent eigenvectors with eigenvalue α correspond to k linearly independent eigenvectors with eigenvalue $-\alpha$, so their multiplicity is the same. □

The next lemma shows that the converse is also true. The proof is by a trace argument that is commonly applied in bounding eigenvalues of random graphs.

Lemma 2.5 (Symmetric Spectrum Implies Bipartiteness). *Let G be an undirected graph and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. If $\alpha_i = -\alpha_{n-i+1}$ for each $1 \leq i \leq n$, then G is a bipartite graph.*

Proof. Let k be any positive odd number. Then $\sum_{i=1}^{n} \alpha_i^k = 0$, by the symmetry of the spectrum. Note that $\alpha_1^k \geq \alpha_2^k \geq \ldots \geq \alpha_n^k$ are the eigenvalues of A^k, because if $Av = \alpha v$ then $A^k v = \alpha^k v$. By Fact A.35, it follows that $\mathrm{Tr}(A^k) = \sum_{i=1}^{n} \alpha_i^k = 0$. Observe that $(A^k)_{i,j}$ is the number of length-k walks from i to j in G, which can be proved by a simple induction. Now suppose G has an odd cycle of length k. Then $(A^k)_{i,i} > 0$ for each vertex i in the odd cycle, and this would imply that $\mathrm{Tr}(A^k) = \sum_{i=1}^{n} (A^k)_{i,i} > 0$, since each diagonal entry of A^k is non-negative. Therefore, since $\mathrm{Tr}(A^k) = 0$, G must have no odd cycles and is thus a bipartite graph. □

Combining Lemma 2.4 and Lemma 2.5, a graph is bipartite if and only if the spectrum of its adjacency matrix is symmetric around the origin.

Proposition 2.6 (Spectral Characterization of Bipartite Graphs)**.** *Let G be an undirected graph and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. Then G is a bipartite graph if and only if $\alpha_i = -\alpha_{n-i+1}$ for each $1 \leq i \leq n$.*

When the graph is connected, the characterization is even simpler. In Section 2.5, you are asked to prove that a connected graph is bipartite if and only if $\alpha_1 = -\alpha_n$.

Largest Eigenvalue

Here we see some upper and lower bounds on the largest eigenvalue of the adjacency matrix.

Lemma 2.7 (Max Degree Upper Bound)**.** *Let $G = (V, E)$ be an undirected graph with maximum degree d, and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. Then $\alpha_1 \leq d$.*

Proof. Let v be an eigenvector with eigenvalue α_1. Let j be a vertex with $v(j) \geq v(i)$ for all $i \in V(G)$. Then

$$\alpha_1 \cdot v(j) = (Av)(j) = \sum_{i:ij \in E(G)} v(i) \leq \sum_{i:ij \in E(G)} v(j) = \deg(j) \cdot v(j) \leq d \cdot v(j),$$

which implies that $\alpha_1 \leq d$. □

Following the proof more closely, we can characterize the connected graphs for which α_1 equals the maximum degree.

Exercise 2.8 (Tight Max Degree Upper Bound)**.** *Let G be a connected undirected graph with maximum degree d and the largest eigenvalue $\alpha_1 = d$. Then G is a d-regular graph.*

The maximum degree upper bound can be far from tight. In Section 2.5, you are asked to prove that the maximum eigenvalue of a tree of maximum degree d is at most $2\sqrt{d-1}$.

On the other hand, the average degree provides a lower bound on the largest eigenvalue. More generally, the largest eigenvalue is at least the average degree of the densest induced subgraph. One corollary of this is that the largest eigenvalue is at least the size of a maximum clique minus one.

Exercise 2.9 (Average Degree Lower Bound)**.** *Let $G = (V, E)$ be an undirected graph with largest eigenvalue α_1. For a subset $S \subseteq V$ and a vertex $v \in S$, let $\deg_S(v) := \big|\{u \mid uv \in E \text{ and } u \in S\}\big|$ be the degree of v induced in S. Then*

$$\alpha_1 \geq \max_{S:S \subseteq V} \frac{1}{|S|} \sum_{v \in S} \deg_S(v).$$

The largest eigenvalue of the adjacency matrix of a connected graph is always of multiplicity one, as guaranteed by the Perron-Frobenius Theorem A.18. The spectrum of the adjacency matrix satisfies

$$d \geq \alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n \geq -d.$$

We will see a combinatorial characterization of the spectral radius of a matrix when we study the Expander Mixing Lemma later. However, there does not seem to be a simple graph-theoretic characterization of the largest eigenvalue of the adjacency matrix. See Section 2.5 for a question in this direction.

2.2 Laplacian Matrix

The Laplacian matrix plays a more important role in spectral graph theory than the adjacency matrix, as we will see some reasons shortly.

Definition 2.10 (Diagonal Degree Matrix)**.** *Let $G = (V, E)$ be an undirected graph with $V(G) = [n]$. The diagonal degree matrix $D(G)$ of G is the $n \times n$ diagonal matrix with $D_{i,i} = \deg(i)$ for each $1 \leq i \leq n$.*

Definition 2.11 (Laplacian Matrix)**.** *Let G be an undirected graph. The Laplacian matrix $L(G)$ of G is defined as $L(G) := D(G) - A(G)$, where $D(G)$ is the diagonal degree matrix in Definition 2.10 and $A(G)$ is the adjacency matrix in Definition 2.1.*

For d-regular graphs, the diagonal degree matrix $D(G)$ is simply $d \cdot I_n$, and so the spectrums of the adjacency matrix and the Laplacian matrix are essentially the same. That is, let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of the adjacency matrix, and $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be the eigenvalues of the Laplacian matrix. For d-regular graphs, it holds that $\lambda_i = d - \alpha_i$ for $1 \leq i \leq n$, and thus the i-th largest eigenvalue of A corresponds to the i-th smallest eigenvalue of L.

Throughout this course, we use the convention that the eigenvalues of A are denoted by $\{\alpha_i\}_{i=1}^n$ and those of L are denoted by $\{\lambda_i\}_{i=1}^n$. The eigenvalues of A are ordered in non-increasing order while those of L are ordered in non-decreasing order. When we refer to the k-th eigenvalue of a graph, we mean either the k-th largest eigenvalue of the adjacency matrix or the k-th smallest eigenvalue of the Laplacian matrix.

For non-regular graphs, relating the eigenvalues of the adjacency matrix and the Laplacian matrix is more challenging. On the one hand, as discussed earlier, it is not clear what is a good characterization of α_1 for non-regular graphs. On the other hand, the smallest eigenvalue λ_1 of the Laplacian matrix is always equal to zero, as we will soon demonstrate.

We define a matrix for the proof, which will also be useful later.

Definition 2.12 (Edge Incidence Matrix)**.** *Let $G = (V, E)$ be an undirected graph with $V(G) = [n]$ and $m = |E|$. For each edge $e = ij \in E$, let b_e be the n-dimensional vector with the i-th position equal to $+1$, the j-th position equal to -1, and all other positions equal to 0. Let $B(G)$ be the $n \times m$ edge incidence matrix whose columns are $\{b_e \mid e \in E\}$.*

For an edge $e \in E$, let L_e be its Laplacian matrix, where $(L_e)_{i,i} = (L_e)_{j,j} = 1$ and $(L_e)_{i,j} = (L_e)_{j,i} = -1$. Note that the Laplacian L_e of an edge e can be written as $b_e b_e^\top$, and the Laplacian of the graph G can be written as

$$L(G) = \sum_{e \in E} L_e = \sum_{e \in E} b_e b_e^\top = B(G) \cdot B(G)^\top.$$

Using this definition, we see that zero is always the smallest eigenvalue of the Laplacian matrix.

Lemma 2.13 (Smallest Eigenvalue of Laplacian Matrix)**.** *The Laplacian matrix $L(G)$ of an undirected graph G is positive semidefinite, and its smallest eigenvalue is zero with the all-ones vector being a corresponding eigenvector.*

Proof. As L can be written as $BB^\top$, as shown in Definition 2.12, it follows that L is a positive semidefinite matrix by Fact A.9. Thus, all eigenvalues of L are non-negative. It is straightforward to verify that $L\vec{1} = 0$, so 0 is the smallest eigenvalue, and $\vec{1}$ is the corresponding eigenvector. □

Having a trivial smallest eigenvalue and a simple corresponding eigenvector is one reason that Laplacian matrix is easier to work with. Another reason is that the Laplacian matrix has a quadratic form which has a nice combinatorial interpretation.

Lemma 2.14 (Quadratic Form for Laplacian Matrix)**.** *Let L be the Laplacian matrix of an undirected graph $G = (V, E)$ with $V(G) = [n]$. For any vector $x \in \mathbb{R}^n$,*

$$x^\top Lx = \sum_{ij \in E} \big(x(i) - x(j)\big)^2.$$

Proof. Using the decomposition of L in Definition 2.12,

$$x^\top Lx = x^\top \Big(\sum_{ij\in E} L_{ij}\Big)x = x^\top \Big(\sum_{ij\in E} b_{ij}b_{ij}^\top\Big)x = \sum_{ij\in E} x^\top b_{ij}b_{ij}^\top x = \sum_{ij\in E} \big(x(i) - x(j)\big)^2.$$

□

Lemma 2.13 and Lemma 2.14 will be used to derive a useful formulation for the second smallest eigenvalue of the Laplacian matrix when we study Cheeger's inequality.

Connectedness

It turns out that the second smallest eigenvalue of the Laplacian matrix can be used to determine whether the graph is connected or not.

Proposition 2.15 (Spectral Characterization of Connected Graphs)**.** *Let G be an undirected graph and let $\lambda_1 \le \ldots \le \lambda_n$ be the eigenvalues of its Laplacian matrix L. Then G is a connected graph if and only if $\lambda_2 > 0$.*

Proof. Suppose G is disconnected. Then the vertex set can be partitioned into two sets S_1 and S_2 such that there are no edges between them. For a subset $S \subseteq V$, let $\chi_S \in \mathbb{R}^n$ be the characteristic vector of S. It is easy to verify that both χ_{S_1} and χ_{S_2} are eigenvectors of L with eigenvalue 0. Since χ_{S_1} and χ_{S_2} are linearly independent, it follows that 0 is an eigenvalue with multiplicity at least 2, and thus $\lambda_2 = 0$.

Conversely, suppose G is connected. Let x be an eigenvector with eigenvalue 0. Then its quadratic form $x^T Lx = 0$, and so $\sum_{ij\in E}(x(i) - x(j))^2 = 0$ by Lemma 2.14, which implies that $x(i) = x(j)$ for every edge $ij \in E$. Since G is connected, it follows that $x = c \cdot \vec{1}$ for some c, and thus the eigenspace corresponding to eigenvalue 0 is one-dimensional. Therefore, the eigenvalue 0 has multiplicity 1, which implies $\lambda_2 > 0$. □

The proof of Proposition 2.15 can be extended to the following generalization.

Exercise 2.16 (Spectral Characterization of Number of Components)**.** *Prove that the Laplacian matrix $L(G)$ of an undirected graph G has 0 as its eigenvalue with multiplicity k if and only if G has k connected components.*

2.3 Normalized Adjacency and Laplacian Matrix

Recall that the spectrum of the adjacency matrix satisfies

$$d \geq \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -d,$$

where the upper and lower bounds depend on the maximum degree d of the graph. This often introduces a dependency on d when relating these eigenvalues to combinatorial parameters.

To remove this dependency and state the Cheeger's inequality more cleanly, we consider the normalized version of the adjacency matrix and the Laplacian matrix. These normalized matrices were popularized and systematically studied in Chung's book [Chu97] to generalize results for regular graphs to all graphs.

Definition 2.17 (Normalized Adjacency and Laplacian Matrix)**.** *Let G be an undirected graph with no isolated vertices. The normalized adjacency matrix $\mathcal{A}(G)$ of G is defined as*

$$\mathcal{A}(G) := D^{-\frac{1}{2}} A D^{-\frac{1}{2}},$$

where D is the diagonal degree matrix in Definition 2.10 and A is the adjacency matrix in Definition 2.1. The normalized Laplacian matrix $\mathcal{L}(G)$ of G is defined as

$$\mathcal{L}(G) := D^{-\frac{1}{2}} L D^{-\frac{1}{2}},$$

where L is the Laplacian matrix in Definition 2.11. Note that $\mathcal{L}(G) = I - \mathcal{A}(G)$.

We will use the same notation conventions as before. The eigenvalues of $\mathcal{A}(G)$ are denoted by $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$, and those of $\mathcal{L}(G)$ are denoted by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Since $\mathcal{L}(G) = I - \mathcal{A}(G)$ as stated in Definition 2.17, the spectra of $\mathcal{L}(G)$ and $\mathcal{A}$ are essentially equivalent such that $\lambda_i = 1 - \alpha_i$ for $1 \leq i \leq n$. After normalization, the eigenvalues are bounded as follows.

Lemma 2.18 (Normalized Eigenvalues)**.** *Let G be an undirected graph with no isolated vertices. Let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its normalized adjacency matrix and $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of its normalized Laplacian matrix. Then $1 = \alpha_1 \geq \alpha_n \geq -1$ and $0 = \lambda_1 \leq \lambda_n \leq 2$.*

Proof. First, we show that $\lambda_1 = 0$. Note that 0 is an eigenvalue of $\mathcal{L}$, as

$$\mathcal{L}\big(D^{\frac{1}{2}}\vec{1}\big) = \big(D^{-\frac{1}{2}} L D^{-\frac{1}{2}}\big)\big(D^{\frac{1}{2}}\vec{1}\big) = \big(D^{-\frac{1}{2}} L \vec{1}\big) = 0.$$

Furthermore, note that

$$\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = D^{-\frac{1}{2}} B B^\top D^{-\frac{1}{2}} = \big(D^{-\frac{1}{2}} B\big)\big(D^{-\frac{1}{2}} B\big)^\top$$

where B is the edge incidence matrix defined in Definition 2.12. It follows that $\mathcal{L} = I - \mathcal{A}$ is a positive semidefinite matrix by Fact A.9, and thus 0 is the smallest eigenvalue of $\mathcal{L}$. This implies that $\alpha_1 = 1$ as $\lambda_1 = 1 - \alpha_1$.

Next, we prove that $\alpha_n \geq -1$. We will show that $D+A$ is also a positive semidefinite matrix. Then the same argument as above implies that $I + \mathcal{A} = D^{-\frac{1}{2}}(D+A)D^{-\frac{1}{2}}$ is also positive semidefinite, and thus $1 + \alpha_n \geq 0$. There are at least two ways to see that $D + A$ is positive semidefinite. One way is to define $\bar{B}$ as the "unsigned" matrix of B, where $\bar{B}_{ij} = |B_{ij}|$ for all $i, j \in V$. Using a similar argument as in Definition 2.12, we can verify that $D + A = \bar{B}\bar{B}^\top$. Another way is to use a similar decomposition as in Definition 2.12 and see that the quadratic form of $D + A$ can be written as

$$x^\top (D+A)x = \sum_{ij \in E} \big(x_i + x_j\big)^2,$$

which is a sum of squares and thus non-negative. This implies that $\lambda_n \leq 2$, as $\lambda_n = 1 - \alpha_n$. □

2.4 Generalizations

We discuss two natural directions to generalize these basic results, one direction with many interesting results, while the other direction not much is known.

Quantitative Generalizations for Undirected Graphs

So far, we have used the graph spectrum to deduce simple combinatorial properties of the graph, such as bipartiteness and connectedness. These properties are easy to deduce directly by simple combinatorial methods, such as breadth first search or depth first search. One might wonder why these spectral characterizations are useful.

The key feature of the spectral characterizations is that they can be generalized quantitatively to prove robust generalizations of the basic results. For example:

- λ_2 is close to zero if and only if the graph is close to being disconnected. This is the content of Cheeger's inequality.
- λ_n is close to 2 if and only if the graph has a structure close to a bipartite component. This is an analog of Cheeger's inequality for λ_n.
- λ_k is close to zero if and only if the graph is close to having k connected components. This is a generalization called the higher-order Cheeger's inequality.

These results form the basis of many spectral graph algorithms. We will see precise statements and proofs of these results in later chapters.

Directed Graphs and Hypergraphs

For directed graphs, we can consider its adjacency matrix A and Laplacian matrix $L = D - A$, where D is the diagonal out-degree matrix. We may ask whether the spectrum of these matrices can be related to the combinatorial properties of the directed graph. However, since these matrices are no longer symmetric, the eigenvalues can be complex numbers, and very little is known about the relationship between the spectrum and combinatorial properties of directed graphs.

For hypergraphs, it is not even clear what the natural associated matrices should be. It has been an open direction to develop a spectral theory for directed graphs and hypergraphs.

In this course, we will keep these directions in mind and mention reasonable questions and known results whenever possible. For directed graphs, we will discuss a recent generalization of Cheeger's inequality using reweighted eigenvalues, as well as how to define directed spectral sparsifiers and solve directed Laplacian equations. For hypergraphs, we will introduce the active research area of high-dimensional expanders, which provides a promising framework for developing an interesting spectral theory for hypergraphs.

2.5 Problems

Problem 2.19 (Cycles). *Compute the Laplacian spectrum of C_n, the cycle with n vertices.*

Hint: The eigenvectors of the Laplacian matrix of C_n involve the n-th roots of unity.

Problem 2.20 (Hypercubes). *A hypercube of n-dimension is an undirected graph with 2^n vertices. Each vertex corresponds to a string of n bits. Two vertices have an edge if and only if their corresponding strings differ by exactly one bit.*

1. *Given two undirected graphs $G = (V, E)$ and $H = (U, F)$, we define $G \times H$ as the undirected graph with vertex set $V \times U$, where two vertices (v_1, u_1) and (v_2, u_2) have an edge if and only if either (1) $v_1 = v_2$ and $u_1u_2 \in F$, or (2) $u_1 = u_2$ and $v_1v_2 \in E$. Let x be an eigenvector of the Laplacian of G with eigenvalue α, and let y be an eigenvector of the Laplacian of H with eigenvalue β. Show that we can use x and y to construct an eigenvector of the Laplacian of $G \times H$ with eigenvalue $\alpha + \beta$.*

2. *Use (1), or otherwise, to compute the Laplacian spectrum of the hypercube of n dimension.*

Problem 2.21 (Spectral Characterization of Connected Bipartite Graphs). *Let G be a connected undirected graph, and let $\alpha_1 \geq \ldots \geq \alpha_n$ be the eigenvalues of its adjacency matrix. Prove that G is bipartite if and only if $\alpha_1 = -\alpha_n$.*

You may need to use the Perron-Frobenius result in Theorem A.18 and also the optimization formulation of eigenvalues in Definition A.11 to solve this problem.

Problem 2.22 (Largest Eigenvalue of a Tree). *Prove that the maximum eigenvalue of the adjacency matrix of a tree with maximum degree d is at most $2\sqrt{d-1}$.*

(This bound is important in the study of Ramanujan graphs.)

Question 2.23 (Largest Eigenvalue of Graphs of Bounded Arborcity). *A graph $G = (V, E)$ is of arboricity k if k is the minimum number of edge-disjoint forests required to cover all the edges of the graph. A classic result in combinatorial optimization by Nash-Williams states that*

$$k = \max_{S \subseteq V} \left\lceil \frac{|E(S,S)|}{|S| - 1} \right\rceil,$$

which is closely related to the density of the densest subgraph.

What is the best upper bound on the largest eigenvalue of the adjacency matrix of a graph of arboricity k, expressed in terms of k and the maximum degree d?

Problem 2.24 (Number of Spanning Trees). *Let $G = (V, E)$ be an undirected graph with $V = [n]$.*

1. *Let B be the edge incidence matrix of G as defined in Definition 2.12. Prove that the determinant of any $(n-1)\times(n-1)$ submatrix of B is ± 1 if and only if the $n-1$ edges corresponding to the columns form a spanning tree of G.*

2. *Let L be the Laplacian matrix of G and let L' be the matrix obtained from L by deleting the last row and last column. Use (1), or otherwise, to prove that $\det(L')$ is equal to the number of spanning trees in G.*

   You may use the Cauchy-Binet formula in Fact A.30 to solve this problem.

Problem 2.25 (Wilf's Theorem)**.** *Let G be an undirected graph, and let α_1 be the largest eigenvalue of its adjacency matrix. Prove that $\chi(G) \leq \lfloor \alpha_1 \rfloor + 1$, where $\chi(G)$ is the chromatic number of G.*

You may find the Cauchy interlacing Theorem A.15 useful.

References

[Chu97] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997. 14, 20

[Spi19] Daniel A. Spielman. *Spectral and Algebraic Graph Theory*. 2019. 9, 89

[Tre17] Luca Trevisan. *Lecture Notes on Graph Partitioning, Expanders and Spectral Methods*. 2017. 9, 62, 90

Chapter 3

Cheeger's Inequality

Recall that a graph G is connected if and only if $\lambda_2 > 0$, where λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix. Informally, Cheeger's inequality is a robust generalization that a graph is well-connected if and only if λ_2 is large. This connection between a combinatorial property and an algebraic quantity is important in the theory of random walks and the study of expander graphs, which we will explore in next chapters. Moreover, the proof of Cheeger's inequality provides an efficient algorithm for graph partitioning, which is useful in both theory and practice.

Cheeger's original inequality was proved in the setting of Riemannian manifolds [Che70]. The inequality in the graph setting was established in several works in the 1980s [Dod84, AM85, Alo86, SJ89]. In this chapter, we begin by motivating the formulation of Cheeger's inequality in the graph setting, using edge conductance as a measure of well-connectedness. We then interpret λ_2 as a continuous relaxation of edge conductance to establish the easy direction of Cheeger's inequality. Next, we follow the exposition of Trevisan [Tre08], which explains the hard direction through a rounding algorithm to relate the continuous relaxation to the discrete property. Finally, we present the spectral partitioning algorithm and discuss its strengths and limitations.

3.1 Cheeger's Inequality for Graphs

Cheeger's original inequality relates the isoperimetric constant of a manifold to the second smallest eigenvalue of its Laplace operator.

The discrete analog of the Laplace operator is the Laplacian matrix, as defined in Definition 2.11. For a more detailed explanation of why the Laplacian matrix is the discrete analog of the Laplace operator, please see [HLW06, page 472]. To provide a quick intuition: one application of the Laplace operator is in defining the heat equation $\partial u/\partial t = \Delta u$, where $u(x,t)$ represents the temperature at point x at time t and Δ is the Laplacian operator. The discrete analog of the heat equation is $du/dt = -Lu$, which implies that

$$\frac{du(i)}{dt} = -Lu(i) = -\sum_{j:ji\in E} (u(i) - u(j)).$$

This equation states that the rate of change of $u(i)$ is proportional to the net flow of heat from vertex i to its neighboring vertices.

The isoperimetric constant of a Riemannian manifold quantifies how well-connected the manifold is by measuring the ratio of the volume of the boundary of a subset to the volume of the subset itself.

More formally, the Cheeger constant is defined as

$$h(M) := \inf_S \frac{\mu_{n-1}(\partial S)}{\min\{\mu_n(S), \mu_n(M \setminus S)\}} = \inf_{S:\mu_n(S)\leq \frac{1}{2}\mu_n(M)} \frac{\mu_{n-1}(\partial S)}{\mu_n(S)},$$

where ∂A denotes the boundary of an open subset A, and μ_n and μ_{n-1} denotes the n-dimensional and $(n-1)$-dimensional measures respectively.

A natural way to define an isoperimetric constant of a graph is to measure the volume using the edges of the graph. (Other natural definitions will be discussed in Section 3.5.)

Definition 3.1 (Edge Conductance)**.** *Let $G = (V, E)$ be an undirected graph. The conductance of a subset $S \subseteq V$ and the conductance of the graph G are defined as*

$$\phi(S) := \frac{|\delta(S)|}{\mathrm{vol}(S)} \qquad \text{and} \qquad \phi(G) := \min_{S\subseteq V:\mathrm{vol}(S)\leq |E|} \phi(S),$$

where $\delta(S)$ denotes the set of edges with exactly one endpoint in S, and $\mathrm{vol}(S) := \sum_{v\in S} \deg(v)$ is the volume of the subset S. Note that the constraint $\mathrm{vol}(S) \leq |E|$ is equivalent to $\mathrm{vol}(S) \leq \frac{1}{2}\mathrm{vol}(V)$ as $\mathrm{vol}(V) = 2|E|$. Note also that for all $S \subseteq V$, it holds that $0 \leq \phi(S) \leq 1$, as $\phi(S)$ is the ratio of the number of edges cut by S to the total degree in S.

Cheeger's original inequality [Che70] states that for a compact Riemannian manifold M,

$$\lambda \geq \frac{h(M)^2}{4},$$

where λ is the smallest positive eigenvalue of the Laplace operator.

The corresponding inequality in the graph setting was established in several works during the 1980s [Dod84, AM85, Alo86, SJ89], with motivating applications in constructing expander graphs and analyzing random walks. We present the following version, which uses the second smallest eigenvalue of the normalized Laplacian matrix, as formulated by Chung [Chu97]. This provides the cleanest bound for non-regular graphs, without any dependency on the maximum degree.

Theorem 3.2 (Cheeger's Inequality for Graphs)**.** *Let $G = (V, E)$ be an undirected graph, and let λ_2 be the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$, as defined in Definition 2.17. Then*

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

The first inequality is called the easy direction, and the second inequality is called the hard direction which is the graph analog of Cheeger's original inequality for Riemannian manifolds. We will see that the easy direction corresponds to using the second eigenvalue as a "relaxation" for graph conductance, while the hard direction corresponds to "rounding" a fractional solution for graph conductance to an integral solution.

An important implication of Cheeger's inequality is that λ_2, the second smallest eigenvalue of the normalized Laplacian matrix, can be used to certify that a graph is an expander graph. We say that a graph G is an expander graph if $\phi(G) \geq c$ for some constant $0 < c < 1$. Sparse expander graphs are highly efficient combinatorial objects with numerous applications in theoretical computer science and mathematics [HLW06]. Cheeger's inequality implies that a graph is an expander if and only if λ_2 is a constant bounded away from zero. This provides an algebraic method for constructing expander graphs, bringing deep mathematical tools into their study, and leading to significant advances in the field.

3.2 Easy Direction: Continuous Relaxation

We prove the easy direction of Cheeger's inequality in this section. The key observation is that λ_2 and $\phi(G)$ can be written as optimization problems of the same form.

We start with the optimization formulation of the second eigenvalue of the normalized Laplacian matrix using the Rayleigh quotient in Definition A.11.

Lemma 3.3 (Optimization Formulation for λ_2)**.** *Let $G = (V = [n], E)$ be an undirected graph, and λ_2 be the second smallest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$. Then*

$$\lambda_2 = \min_{x \in \mathbb{R}^n} \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2} \qquad \text{subject to} \qquad \sum_{i \in V} \deg(i) \cdot x(i) = 0.$$

Proof. By the Rayleigh quotient characterization in Lemma A.13,

$$\lambda_2 = \min_{x \in \mathbb{R}^n : x \perp v_1} R_{\mathcal{L}}(x) = \min_{x \in \mathbb{R}^n : x \perp D^{\frac{1}{2}}\vec{1}} \frac{x^\top \mathcal{L} x}{x^\top x} = \min_{x \in \mathbb{R}^n : x \perp D^{\frac{1}{2}}\vec{1}} \frac{x^\top D^{-\frac{1}{2}} L D^{-\frac{1}{2}} x}{x^\top x},$$

where v_1 is the first eigenvector for the normalized Laplacian matrix, which is parallel to $D^{\frac{1}{2}}\vec{1}$ from Lemma 2.18. By a change of variable $x = D^{\frac{1}{2}} y$ for $y \in \mathbb{R}^n$, this can be rewritten as

$$\lambda_2 = \min_{y \in \mathbb{R}^n : D^{\frac{1}{2}} y \perp D^{\frac{1}{2}}\vec{1}} \frac{y^\top L y}{y^\top D y} = \min_{y \in \mathbb{R}^n : \sum_{i \in V} \deg(i) \cdot y(i) = 0} \frac{\sum_{ij \in E} \big(y(i) - y(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot y(i)^2},$$

where the last equality follows from the quadratic form of the Laplacian matrix in Lemma 2.14. □

Next, we observe that the graph conductance can also be written in the same form.

Lemma 3.4 (Optimization Formulation for Graph Conductance)**.** *Let $G = (V = [n], E)$ be an undirected graph. Then*

$$\phi(G) = \min_{x \in \{0,1\}^n} \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2} \qquad \text{subject to} \qquad \sum_{i \in V} \deg(i) \cdot x(i)^2 \le |E|.$$

Proof. For a set $S \subseteq V$, let $\chi_S \in \{0,1\}^n$ be the characteristic vector of S. Note that

$$\phi(S) = \frac{|\delta(S)|}{\mathrm{vol}(S)} = \frac{\sum_{ij \in \delta(S)} 1}{\sum_{i \in S} \deg(i)} = \frac{\sum_{ij \in E} |\chi_S(i) - \chi_S(j)|}{\sum_{i \in V} \deg(i) \cdot \chi_S(i)} = \frac{\sum_{ij \in E} (\chi_S(i) - \chi_S(j))^2}{\sum_{i \in V} \deg(i) \cdot \chi_S(i)^2} = \frac{\chi_S^\top L \chi_S}{\chi_S^\top D \chi_S}.$$

Each vector x in $\{0,1\}^n$ corresponds to the characteristic vector of the subset $S := \{i \mid x(i) = 1\}$. The graph conductance $\phi(G)$ minimizes over subsets with volume at most $|E|$, which corresponds to the constraint that $\sum_{i \in V} \deg(i) \cdot x(i)^2 \le |E|$. □

Intuition: Continuous Relaxation

There are two differences between these two formulations in Lemma 3.3 and Lemma 3.4: one involves the domain, and the other involves the constraint.

The major difference is that the former optimizes over the continuous domain $x \in \mathbb{R}^n$, while the latter optimizes over the discrete domain $x \in \{0,1\}^n$. A good way to think of the relationship between the two optimization problems is that the former problem is a relaxation of the latter. This is a common idea in the design of approximation algorithms. The latter problem is actually an NP-hard combinatorial optimization problem. The relaxation idea is to optimize over a larger continuous domain, so that the problem becomes solvable in polynomial time. Since we optimize over a larger domain, the objective value of the former problem can only be smaller than that of the latter, and so we expect that $\lambda_2 \leq \phi(G)$. This is the main intuition behind the easy direction.

In this course, we will see that this is a common theme in spectral graph theory, where the spectral quantities involve the continuous domain $x \in \mathbb{R}^n$ while the combinatorial properties involve discrete domains such as $x \in \{0,1\}^n$ or $x \in \{-1,1\}^n$.

For these two formulations, however, the constraints are also different. And it turns out that the inequality $\lambda_2 \leq \phi(G)$ does not hold, but the slightly weaker inequality $\lambda_2 \leq 2\phi(G)$ does.

Proof of the Easy Direction

To upper bound λ_2, we just need to find a vector x satisfying the constraint $\sum_{i \in V} \deg(i) \cdot x(i) = 0$, and compute its Rayleigh quotient $R_{\mathcal{L}}(x)$. Let $S \subseteq V$ be an optimal solution to graph conductance, with $\phi(S) = \phi(G)$ and $\mathrm{vol}(S) \leq |E|$. Consider the following binary-valued solution $z \in \mathbb{R}^n$ with

$$z(i) = \begin{cases} \frac{1}{\mathrm{vol}(S)} & \text{if } i \in S \\ -\frac{1}{\mathrm{vol}(V-S)} & \text{if } i \notin S \end{cases}.$$

By construction, $\sum_{i \in V} \deg(i) \cdot z(i) = \sum_{i \in S} \deg(i)/\mathrm{vol}(S) - \sum_{i \in V-S} \deg(i)/\mathrm{vol}(V-S) = 0$. Thus, z is a feasible solution to the optimization problem for λ_2 in Lemma 3.3, and it follows that

$$\lambda_2 \leq \frac{\sum_{ij \in E} \big(z(i) - z(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot z(i)^2} = \frac{|\delta(S)| \cdot \big(\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(V-S)}\big)^2}{\frac{1}{\mathrm{vol}(S)} + \frac{1}{\mathrm{vol}(V-S)}} = \frac{|\delta(S)| \cdot 2|E|}{\mathrm{vol}(S) \cdot \mathrm{vol}(V-S)} \leq 2\phi(S),$$

where the last inequality uses the assumption that $\mathrm{vol}(S) \leq |E|$ which implies that $|E| \leq \mathrm{vol}(V-S)$. This completes the proof of the easy direction.

3.3 Hard Direction: Rounding Algorithm

By optimizing over a larger domain, the objective value of the continuous problem will typically be smaller than that of the discrete problem. The hard direction is to prove that λ_2 cannot be much smaller than $\phi(G)$, ensuring that λ_2 is a good approximation to $\phi(G)$.

For graph conductance, the objective is to find an integral solution $z \in \{0,1\}^n$ that minimizes the ratio in Lemma 3.4. However, once the problem is relaxed to the continuous domain, the optimal solution $x \in \mathbb{R}^n$ to λ_2 may be very smooth and continuous. The task in the hard direction is to prove that there always exists an integral solution z whose objective value is not much worse than that of x. A common approach in approximation algorithms is to design a procedure to "round" the

continuous solution x to an integral solution z, while bounding the objective value of z in terms of the objective value of x. This is the approach taken by Trevisan [Tre08], providing a more intuitive proof of Cheeger's inequality.

Ideas and Overview

We can think of the optimizer $x \in \mathbb{R}^n$ to the optimization problem in Lemma 3.3 as an embedding of the vertices of the graph into the real line, such that the total squared edge length $\sum_{ij \in E}(x(i)-x(j))^2$ is small. To produce an integral solution z, a natural idea is to do a "threshold rounding", where we pick a threshold t and set $z(i) = 0$ if $x(i) < t$ and $z(i) = 1$ if $x(i) \geq t$. The intuition is that if most edges are short in this embedding, then there must exist a threshold with not many edges crossing. A simple analogy is that if the average number of nonzeros in the rows of a matrix is small, then there must exist a column with few nonzeros. This intuition can be made precise by introducing the intermediate optimization problem of the following ℓ_1-form:

$$\min_{y \in \mathbb{R}^n_+} \frac{\sum_{ij \in E} |y(i) - y(j)|}{\sum_{i \in V} \deg(i) \cdot y(i)}.$$

The proof of the hard direction consists of the following three steps:

1. Truncate an optimal solution to λ_2 in Lemma 3.3 to obtain a solution $x \in \mathbb{R}^n_+$ that satisfies

$$\frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2} \leq \lambda_2 \qquad \text{and} \qquad \mathrm{vol}(\mathrm{supp}(x)) \leq |E|, \tag{3.1}$$

   where $\mathrm{supp}(x) := \{i \in V \mid x(i) \neq 0\}$ is the support set of the vector x. This step is to ensure that the output of the rounding algorithm has volume at most $|E|$. The assumption that the optimal solution to λ_2 satisfies the constraint $\sum_{i \in V} \deg(i) \cdot x(i) = 0$ is only used here. This step can be thought of as bridging the gap between the constraints in Lemma 3.3 and Lemma 3.4.

2. Use the solution x to construct a solution $y \in \mathbb{R}^n_+$ that satisfies

$$\frac{\sum_{ij \in E} |y(i) - y(j)|}{\sum_{i \in V} \deg(i) \cdot y(i)} \leq \sqrt{2\lambda_2} \qquad \text{and} \qquad \mathrm{vol}(\mathrm{supp}(y)) \leq |E|. \tag{3.2}$$

   This step can be interpreted as embedding from ℓ_2^2 to ℓ_1, and it incurs the square root loss in Cheeger's inequality.

3. Apply the threshold rounding procedure described above to y to obtain a set S satisfying

$$\phi(S) \leq \sqrt{2\lambda_2} \qquad \text{and} \qquad \mathrm{vol}(S) \leq |E|.$$

   This step is lossless and relies on a simple probabilistic analysis.

With this overview in mind, we proceed to present the details in reverse order, as the main ideas are in the last two steps.

Threshold Rounding

In the threshold rounding step, we take a vector $y \in R_+^n$ from (3.2) and outputs a set $S \subseteq \operatorname{supp}(y)$ with the same objective value. Our analysis follows that of Trevisan [Tre08], whose idea is to choose a random $t > 0$ and consider the level set $S_t := \{i \in V \mid y(i) \geq t\}$. The conductance of S_t is then bounded by separately computing the expectation of the numerator $\mathbb{E}\left[|\delta(S_t)|\right]$ and the expectation of the denominator $\mathbb{E}\left[d|S_t|\right]$. The idea of choosing a random t is similar to randomized rounding in approximation algorithms, and Trevisan's approach of computing the expectations separately simplifies the analysis of the ratio.

Lemma 3.5 (Threshold Rounding)**.** *Let $G = (V = [n], E)$ be an undirected graph. Let $y \in \mathbb{R}_+^n$ be a non-zero vector with non-negative entries. There exists $t > 0$ such that the threshold set $S_t := \{i \in [n] \mid y(i) \geq t\}$ is nonempty and satisfies*

$$\phi(S_t) \leq \frac{\sum_{ij \in E} |y(i) - y(j)|}{\sum_{i \in V} \deg(i) \cdot y(i)}.$$

Proof. We scale y so that $\max_i y(i) = 1$. Let $t \in (0, 1]$ be chosen uniformly at random. Note that the threshold set $S_t := \{i \in V \mid y(i) \geq t\}$ is nonempty by construction. In the following, we compute separately the expected value of the numerator and of the denominator for S_t.

For an edge $ij \in E$, note that the probability that $ij \in \delta(S_t)$ is $|y(i) - y(j)|$, when the random threshold t falls between $y(i)$ and $y(j)$. By linearity of expectation,

$$\mathbb{E}_t\big[|\delta(S_t)|\big] = \sum_{ij \in E} \Pr_t(ij \in \delta(S_t)) = \sum_{ij \in E} \big|y(i) - y(j)\big|.$$

For a vertex $i \in V$, note that the probability that $i \in S_t$ is $y(i)$, when the random threshold t is at most $y(i)$. By linearity of expectation,

$$\mathbb{E}_t\big[\operatorname{vol}(S_t)\big] = \sum_{i \in V} \deg(i) \cdot \Pr_t(i \in S_t) = \sum_{i \in V} \deg(i) \cdot y(i).$$

It follows from Lemma 3.6 below that

$$\min_t \phi(S_t) = \min_t \frac{|\delta(S_t)|}{\operatorname{vol}(S_t)} \leq \frac{\mathbb{E}_t\big[|\delta(S_t)|\big]}{\mathbb{E}_t\big[\operatorname{vol}(S_t)\big]} \leq \frac{\sum_{ij \in E} |y(i) - y(j)|}{\sum_{i \in V} \deg(i) \cdot y(i)}.$$

□

Lemma 3.6 (Spielman's Favorite Inequality)**.** *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be positive numbers, and $p_1, \ldots, p_n$ be a probability distribution. Then*

$$\min_i \frac{a_i}{b_i} \leq \frac{\sum_{i=1}^n p_i a_i}{\sum_{i=1}^n p_i b_i} \leq \max_i \frac{a_i}{b_i}.$$

The proof of this inequality is left as an exercise to the reader.

Embedding from ℓ_2^2 to ℓ_1

In the embedding step, we construct an ℓ_1-solution y in (3.2) from an ℓ_2^2 solution x in (3.1). The most obvious mapping is $y(i) := x(i)^2$ so as to match the denominators, and it works.

Lemma 3.7 (Embedding Step). *Given an undirected graph $G = (V, E)$ and a vector $x \in \mathbb{R}^n_+$, there is a vector $y \in \mathbb{R}^n_+$ with $\operatorname{supp}(y) = \operatorname{supp}(x)$ such that*

$$\frac{\sum_{ij\in E} |y(i) - y(j)|}{\sum_{i\in V} \deg(i) \cdot y(i)} \le \sqrt{2 \cdot \frac{\sum_{ij\in E} \big(x(i) - x(j)\big)^2}{\sum_{i\in V} \deg(i) \cdot x(i)^2}}.$$

Proof. Set $y(i) = x(i)^2$ for all $i \in V$. By construction, the supports of x and y are the same. The main idea is to use the Cauchy-Schwarz inequality to bound the LHS by the RHS such that

$$\sum_{ij\in E} |y(i) - y(j)| = \sum_{ij\in E} |x(i) - x(j)| \cdot |x(i) + x(j)| \le \sqrt{\sum_{ij\in E} (x(i) - x(j))^2} \sqrt{\sum_{ij\in E} (x(i) + x(j))^2}.$$

Observe that

$$\sum_{ij\in E} (x(i) + x(j))^2 \le \sum_{ij\in E} 2\big(x(i)^2 + x(j)^2\big) = 2 \sum_{i\in V} \deg(i) \cdot x(i)^2.$$

Combining these inequalities, we conclude that

$$\frac{\sum_{ij\in E} |y(i) - y(j)|}{\sum_{i\in V} \deg(i) \cdot y(i)} \le \frac{\sqrt{\sum_{ij\in E} (x(i) - x(j))^2} \sqrt{2 \sum_{i\in V} \deg(i) \cdot x(i)^2}}{\sum_{i\in V} \deg(i) \cdot x(i)^2} = \sqrt{\frac{2 \sum_{ij\in E} \big(x(i) - x(j)\big)^2}{\sum_{i\in V} \deg(i) \cdot x(i)^2}}.$$

□

Truncation Step

Given an optimal solution x to λ_2, we first shift x to obtain $\bar{x}$ with at most the same objective value and the additional property that both the positive part of $\bar{x}$ and the negative part of $\bar{x}$ have volume at most $|E|$. The proof crucially relies on the assumption that $\sum_{i\in V} \deg(i) \cdot x(i) = 0$.

Lemma 3.8 (Shifting). *Let $x \in \mathbb{R}^n$ be an optimal solution to λ_2 in Lemma 3.3. There exists a vector $\bar{x} \in \mathbb{R}^n$ such that $\operatorname{vol}(\{i \mid \bar{x}(i) < 0\}) \le |E|$ and $\operatorname{vol}(\{i \mid \bar{x}(i) > 0\}) \le |E|$ and*

$$\frac{\sum_{ij\in E} \big(\bar{x}(i) - \bar{x}(j)\big)^2}{\sum_{i\in V} \deg(i) \cdot \bar{x}(i)^2} \le \frac{\sum_{ij\in E} \big(x(i) - x(j)\big)^2}{\sum_{i\in V} \deg(i) \cdot x(i)^2}.$$

Proof. Let $c \in \mathbb{R}$ be a median value such that $\operatorname{vol}(\{i \mid x(i) < c\}) \le |E|$ and $\operatorname{vol}(\{i \mid x(i) > c\}) \le |E|$. Set $\bar{x} := x - c\vec{1}$. By construction, $\operatorname{vol}(\{i \mid \bar{x}(i) < 0\}) \le |E|$ and $\operatorname{vol}(\{i \mid \bar{x}(i) > 0\}) \le |E|$.

For the ratio, observe that the numerator does not change by shifting, and the denominator cannot decrease because

$$\sum_{i\in V} \deg(i) \cdot \bar{x}(i)^2 = \sum_{i\in V} \deg(i) \cdot (x(i) - c)^2 = \sum_{i\in V} \deg(i) \cdot x(i)^2 + c^2 \sum_{i\in V} \deg(i) \ge \sum_{i\in V} \deg(i) \cdot x(i)^2,$$

where the last equality holds due to the assumption that $\sum_{i\in V} \deg(i) \cdot x(i) = 0$. This completes the proof. (Note that $\bar{x}$ may not satisfy the constraint $\sum_{i\in V} \deg(i) \cdot \bar{x}(i) = 0$, so it may not be a feasible solution to λ_2 in Lemma 3.3.) □

Next, we show that either the positive part of $\bar{x}$ or the negative part of $\bar{x}$ satisfies the requirements in (3.1).

Lemma 3.9 (Truncation). *Let $\bar{x} \in \mathbb{R}^n$ be a vector that satisfies the properties in Lemma 3.8. There exists a vector $\bar{x}_+ \in \mathbb{R}^n_+$ with $\mathrm{vol}(\mathrm{supp}(\bar{x})) \leq |E|$ and*

$$\frac{\sum_{ij \in E} \left(\bar{x}_+(i) - \bar{x}_+(j)\right)^2}{\sum_{i \in V} \deg(i) \cdot \bar{x}_+(i)^2} \leq \frac{\sum_{ij \in E} \left(\bar{x}(i) - \bar{x}(j)\right)^2}{\sum_{i \in V} \deg(i) \cdot \bar{x}(i)^2}$$

Proof. Let $\bar{x}_+ \in \mathbb{R}^n$ be the vector with $\bar{x}_+(i) := \max\{\bar{x}(i), 0\}$ for $1 \leq i \leq n$, and $\bar{x}_- \in \mathbb{R}^n$ be the vector with $\bar{x}_-(i) := \min\{\bar{x}(i), 0\}$ for $1 \leq i \leq n$. We argue that either $\bar{x}_+$ or $-\bar{x}_-$ would satisfy the requirements. By construction, both $\bar{x}_+$ and $-\bar{x}_-$ satisfy that the volume of the support is at most $|E|$. For the ratio, note that

$$\frac{\sum_{ij \in E} (\bar{x}_+(i) - \bar{x}_+(j))^2)^2 \ + \ \sum_{ij \in E} (\bar{x}_-(i) - \bar{x}_-(j))^2}{\sum_{i \in V} \deg(i) \cdot \bar{x}_+(i)^2 \ + \ \sum_{i \in V} \deg(i) \cdot \bar{x}_-(i)^2} \leq \frac{\sum_{ij \in E} \left(\bar{x}(i) - \bar{x}(j)\right)^2}{\sum_{i \in V} \deg(i) \cdot \bar{x}(i)^2},$$

where the denominators are equal, and the numerator on the LHS can only be smaller than that of the RHS by a simple case analysis. The conclusion then follows from Lemma 3.6. □

Proof of the Hard Direction

We summarize the proof of the hard direction. Let $v_2 \in \mathbb{R}^n$ be an eigenvector of $\mathcal{L}(G)$ with eigenvalue λ_2. First, we apply the transformation $u := D^{-\frac{1}{2}} v_2$ to obtain a vector u that satisfies the requirements in Lemma 3.3. Next, we apply the shifting and truncation steps in Lemma 3.8 and Lemma 3.9 on u to obtain a vector x that satisfies the requirements in (3.1). Then, we apply the embedding step in Lemma 3.7 on x to obtain a vector y that satisfies the requirements in (3.2). Finally, we apply the threshold rounding step in Lemma 3.5 on y to obtain a threshold set S_t with $\phi(S_t) \leq \sqrt{2\lambda_2}$ and $\mathrm{vol}(S_t) \leq |E|$. This completes the proof of the hard direction of Cheeger's inequality.

3.4 The Spectral Partitioning Algorithm

Finding a set of small conductance, also called a sparse cut, is an important algorithmic problem with numerous applications. It is useful in designing divide-and-conquer algorithms and has applications in image segmentation, data clustering, community detection, VLSI design, and more.

The problem of finding a sparsest cut is NP-hard. A popular heuristic for finding an approximate sparsest cut in practice is the following spectral partitioning algorithm.

Algorithm 1 The Spectral Partitioning Algorithm

Require: An undirected graph $G = (V, E)$ with $V = [n]$ and $m = |E|$.

1: Compute the second smallest eigenvalue λ_2 of $\mathcal{L}(G)$ and a corresponding eigenvector $x \in \mathbb{R}^n$.
2: Compute the vector $y := D^{-\frac{1}{2}} x$ and sort the vertices so that $y(1) \geq y(2) \geq \ldots \geq y(n)$.
3: For $1 \leq i \leq n-1$, let $S_i = [i]$ if $\mathrm{vol}_G\left([i]\right) \leq m$, and let $S_i = [n] \setminus [i]$ if $\mathrm{vol}_G\left([i]\right) > m$.
4: **return** $\min_{i:1 \leq i \leq n-1} \{\phi(S_i)\}$.

This algorithm is remarkably simple, with only a few lines of code when implemented in mathematical software such as MATLAB. This simplicity is one reason why this heuristic is popular.

There is a near-linear time randomized algorithm to compute an approximate eigenvector of the second eigenvalue, using the power method and a fast Laplacian solver. This makes the spectral partitioning algorithm both practically and theoretically efficient, which is another reason for its popularity.

The primary reason that it is popular is its excellent empirical performance in various applications, especially in image segmentation and clustering. The introduction of the concept of a normalized cut (closely related to a sparse cut) and the spectral partitioning algorithm by Shi and Malik [SM00] in the context of image segmentation were considered a breakthrough.

The proof of Cheeger's inequality provides a nontrivial performance guarantee of the spectral partitioning algorithm, that it will always output a set S with conductance $\phi(S) \leq \sqrt{2\lambda_2} \leq 2\sqrt{\phi(G)}$. This follows because the shifting, truncation, and embedding steps in the proof of the hard direction are only used for the analysis and do not change the ordering of the vertices. Therefore, the cuts considered in the threshold rounding step are also considered by the spectral partitioning algorithm.

The spectral partitioning algorithm is a constant factor approximation algorithm when $\phi(G)$ is a constant, providing an efficient way to certify that a graph is an expander as discussed earlier. However, its approximation ratio could be arbitrarily bad when $\phi(G)$ is small. For example, the approximation ratio is $\Theta(\sqrt{n})$ when $\phi(G) \leq 1/n$. Providing a theoretical explanation of the empirical success of the spectral partitioning algorithm remains an open problem. We will revisit this question when we study the improved Cheeger's inequality.

In the following, we discuss the theoretical performance of the spectral partitioning algorithm in more detail. We present examples where the easy direction is tight, where the hard direction is tight, and where the spectral partitioning algorithm is fooled.

Tight Example for the Hard Direction

Consider the cycle of $4n$ vertices. One can compute the second eigenvector of the cycle exactly (see Problem 2.19), but we do not need it here. Recall that $\lambda_2 = \min_{x \perp \vec{1}} x^T \mathcal{L} x / x^T x$, so to give an upper bound, we only need to demonstrate a solution with small objective value. Consider

$$x = \Big(1, 1 - \frac{1}{n}, 1 - \frac{2}{n}, \ldots, \frac{1}{n}, 0, -\frac{1}{n}, \ldots, -1 + \frac{1}{n}, -1, -1 + \frac{1}{n}, \ldots, -\frac{1}{n}, 0, \frac{1}{n}, \ldots, 1 - \frac{1}{n}\Big).$$

Then $x \perp \vec{1}$, and so

$$\lambda_2 \leq \frac{\sum_{ij \in E} \big(x(i) - x(j)\big)^2}{2 \sum_{i \in V} x(i)^2} = \Theta\Big(\frac{n\big(\frac{1}{n}\big)^2}{n}\Big) = \Theta\Big(\frac{1}{n^2}\Big).$$

On the other hand, it is easy to verify that the conductance of the cycle of $4n$ vertices is $\Theta(\frac{1}{n})$. This is an example where the hard direction $\phi(G) \leq \sqrt{2\lambda_2}$ is tight up to a constant factor.

In this example, λ_2 is not a good estimate of $\phi(G)$, but the spectral partitioning algorithm works perfectly to output a set S with $\phi(S) \approx \phi(G)$, as it outputs a set S with $\phi(S) \approx \sqrt{\lambda_2} \approx \phi(G)$. It is actually a general phenomenon that rounding algorithms work perfectly for the worst integrality gap examples.

Tight Example for the Easy Direction

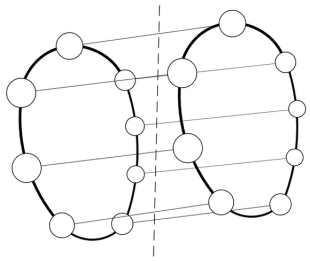
To find an example where the spectral partitioning algorithm performs poorly, we need to examine cases where the easy direction is tight, but the algorithm outputs a set S with $\phi(S) \approx \sqrt{\lambda_2} \approx \sqrt{\phi(G)}$.

For the easy direction, one can check that it is tight for the hypercubes; see Problem 2.20. Interestingly, for hypercubes, there are vectors in the second eigenspace where the spectral partitioning algorithm performs perfectly and performs poorly; see Problem 3.13. Since we do not have control over which eigenvector in the second eigenspace is returned, this provides an example where the spectral partitioning algorithm could perform poorly.

An Example Fooling the Spectral Partitioning Algorithm

The cycles and the hypercubes are the standard examples showing that both sides of Cheeger's inequality are tight. In the hypercube example, the spectral partitioning algorithm could only output a set S with $\phi(S) \approx \sqrt{\phi(G)}$. However, this example may not be fully satisfying, as the algorithm could still work perfectly. More importantly, we do not clearly see or gain intuition about how the spectral partitioning algorithm is fooled.

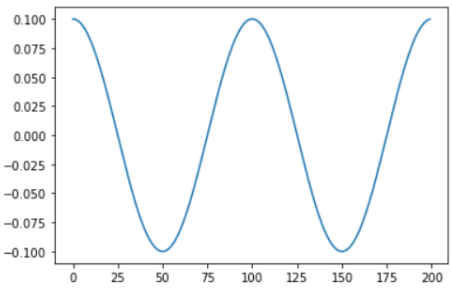
We construct such an example by tweaking the cycle example. Let G be the weighted graph with vertices $\{v_1, \ldots, v_n, v_{n+1}, \ldots, v_{2n}\}$, and two cycles $(v_1, v_2, \ldots, v_n)$ and $(v_{n+1}, v_{n+2}, \ldots, v_{2n})$ where every edge in these cycles is of weight one, and a "hidden" matching $\{v_1v_{n+1}, v_2v_{n+2}, \ldots, v_nv_{2n}\}$ where every edge in the matching has weight $100/n^2$. It is easy to see that the set of smallest



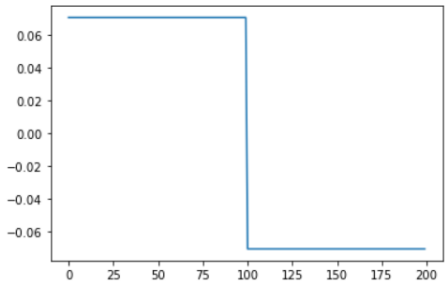
conductance is the set $S := \{v_1, \ldots, v_n\}$ with $\phi(S) = O(1/n^2)$. However, the edges in the hidden matching are just barely heavy enough that the spectral partitioning algorithm does not "feel" them, and still considers the smooth embedding of the cycle as the best way to map the vertices onto the real line. Indeed, one can verify that the second eigenvector x in this example is still the same as that in the cycle of n vertices, with $x(v_i) = x(v_{n+i})$ for $1 \leq i \leq n$. See Figure 3.1 for an illustration. Therefore, λ_2 is still $O(1/n^2)$ which is close to $\phi(G)$, but the cut of smallest conductance is completely lost in x and every threshold set has conductance $\Omega(1/n)$. This example provides a more insightful view into how the spectral partitioning algorithm is fooled. Althout this example is a weighted graph, one can modify it slightly to keep the same structure while making the graph unweighted.

3.5 Variants of Cheeger's Inequality

We discuss two variants of Cheeger's inequality, based on different definitions of the isoperimetric constant of a graph.



Matching edge weight 1/500



Matching edge weight 1/1000

Figure 3.1: The plots of the second eigenvector are generated for the graph with 200 vertices. Vertices 0 to 99 belong to the first cycle, and vertices 100 to 199 belong to the second cycle. A matching edge connects vertex i and vertex $i + 100$ for $0 \leq i \leq 99$. When the edge weight in the hidden matching is slightly heavier, the second eigenvector is the same as that of the cycle, with matched vertices having the same value. When the edge weight in the hidden matching is slightly lighter, the second eigenvalue becomes a binary vector, indicating the sparsest cut.

Edge Expansion

Cheeger's inequality is often stated using edge expansion rather than edge conductance.

Definition 3.10 (Edge Expansion)**.** *Let $G = (V, E)$ be an undirected graph. The edge expansion of a subset $S \subseteq V$ and the edge expansion of the graph G are defined as*

$$\Phi(S) := \frac{|\delta(S)|}{|S|} \qquad \text{and} \qquad \Phi(G) := \min_{S:|S|\leq|V|/2} \Phi(S).$$

Both edge expansion and edge conductance aim to identify the "bottleneck" in the graph. For d-regular graphs, the two definitions are essentially equivalent, with $\Phi(G) = d \cdot \phi(G)$. For non-regular graphs, the relationship between the edge conductance and the second smallest eigenvalue of the normalized Laplacian matrix is more elegant, without any dependency on the maximum degree of the graph.

Vertex Expansion

Another natural definition of an isopermetric constant of graphs is based on measuring the "volume" using the vertices of the graph.

Definition 3.11 (Vertex Expansion)**.** *Let $G = (V, E)$ be an undirected graph. The vertex expansion of a subset $S \subseteq V$ and the vertex expansion of the graph G are defined as*

$$\psi(S) := \frac{|\partial(S)|}{\min\{|S|, |V - S|\}} \qquad \text{and} \qquad \psi(G) := \min_{S\subseteq V} \psi(S),$$

where $\partial(S) := \{v \notin S \mid \exists u \in S \text{ with } uv \in E\}$ denotes the vertex boundary of S.

There is a Cheeger-type inequality relating vertex expansion and the second smallest eigenvalue of the Laplacian matrix.

Theorem 3.12 (Cheeger's Inequality for Vertex Expansion). *Let $G = (V, E)$ be an undirected graph with maximum degree d, and let λ_2' be the second smallest eigenvalue of its Laplacian matrix $L(G)$, as defined in Definition 2.11. Then*

$$\psi(G) \geq \frac{2\lambda_2'(G)}{d + 2\lambda_2'(G)} \quad \text{and} \quad \lambda_2'(G) \geq \frac{\psi(G)^2}{4 + 2\psi(G)^2},$$

The first inequality is the easy direction proved by Tanner [Tan84] and Alon and Milman [AM85]. The second inequality is the hard direction proved by Alon [Alo86]. Note that these imply that $\lambda_2'(G)$ can be used to provide an $O(\sqrt{d \cdot \psi(G)})$-approximation algorithm to $\psi(G)$.

Comparing to Cheeger's inequality for edge conductance, there is an extra factor d loss between the upper and lower bounds. In Chapter **??**, we will introduce a new Cheeger's inequality for vertex expansion using a concept called reweighted eigenvalues, which improves the dependency on the maximum degree from d to $\log d$.

3.6 Problems

Problem 3.13 (Spectral Partitioning for Hypercubes). *Let G be the hypercube of dimension d with 2^d vertices and $\mathcal{L}(G)$ be its normalized Laplacian matrix.*

(a) Show that there is an eigenvector $x \in \mathbb{R}^{2^d}$ of $\mathcal{L}(G)$ with eigenvalue λ_2 so that the spectral partitioning algorithm applied on x outputs a set S with $\phi(S) = R_{\mathcal{L}}(x) = \frac{1}{2}\lambda_2$.

(b) Show that there is an eigenvector $y \in \mathbb{R}^{2^d}$ of $\mathcal{L}(G)$ with eigenvalue λ_2 so that the spectral partitioning algorithm applied on y outputs a set S with $\phi(S) \approx \sqrt{R_{\mathcal{L}}(y)} = \sqrt{\lambda_2}$.

(Hint: Consider a convex combination of the good vectors in the previous part.)

Problem 3.14 (Houdré-Tetali Isoperimetric Constant). *Consider an isoperimetric constant of graphs introduced by Houdré and Tetali [HT04]. Assume the graph $G = (V = [n], E)$ is d-regular for simplicity. For a vertex i and a subset $S \subset V$, denote $d(i, \overline{S}) := |\{ij \in E \mid j \in V - S\}|$. For any $p \in [0, 1]$, the isoperimetric constant φ_p of a subset $S \subset V$ and of the graph are defined as*

$$\varphi_p(S) := \frac{1}{|S|} \sum_{i \in S} \left(\frac{d(i, \overline{S})}{d}\right)^p \quad \text{and} \quad \varphi_p(G) := \min_{S : |S| \leq n/2} \varphi_p(S).$$

(a) Verify that $\varphi_1(G)$ is equal to the edge conductance $\phi(G)$.

(b) Check that $\varphi_0(G)$ is equal to the inner vertex expansion $\psi_{in}(G)$ if we use the convention that $0^0 = 0$. Let $\partial_{in}(S) := |\{i \in S \mid d(i, \overline{S}) > 0\}|$ be the inner vertex boundary. Define $\psi_{in}(S) := |\partial_{in}(S)|/|S|$ and $\psi_{in}(G) := \min_{S : |S| \leq n/2} \psi_{in}(S)$.

(c) Show that $\varphi_{\frac{1}{2}}(G)^2 \leq \varphi_1(G) \cdot \varphi_0(G)$.

(d) Prove that $\varphi_{\frac{1}{2}}(G)^2 \lesssim \lambda_2 \cdot \log d$, where λ_2 is the second eigenvalue of the normalized Laplacian.

(e) Prove that $\varphi_p(G)^2 \lesssim \frac{1}{2p-1} \cdot \lambda_2$ for any $p \in (\frac{1}{2}, 1]$.

(Hint: A similar randomized rounding proof as in the hard direction of Cheeger's inequality would work to prove (d) and (e). See [LT24].)

Cheeger-Type Inequality for λ_n and Bipartiteness Ratio

Through his exposition of Cheeger's inequality using the intermediate ℓ_1-problem presented in this chapter, Trevisan [Tre09] discovered an analog of Cheeger's inequality for λ_n.

In this subsection, we follow his thought process to derive the result, which starts with a spectral characterization relating λ_n to the bipartiteness of the graph.

Exercise 3.15 (Spectral Characterization of Bipartiteness). *Let $G = (V, E)$ be an undirected graph and λ_n be the largest eigenvalue of its normalized Laplacian matrix $\mathcal{L}(G)$. Then $\lambda_n = 2$ if and only if G has a bipartite component, i.e., a connected component that is a bipartite graph.*

Trevisan [Tre09] proved a robust generalization that λ_n is close to 2 if and only if G is close to having a bipartite component, in the same style as in Cheeger's inequality in Theorem 3.2. To state his result, we write the optimization formulation for $2 - \lambda_n$ and then motivate the corresponding combinatorial property.

Exercise 3.16 (Optimization Formulation for $2 - \lambda_n$). *Let $G = (V, E)$ be an undirected graph and λ_n be the largest eigenvalue of $\mathcal{L}(G)$. Then*

$$2 - \lambda_n = \min_{x \in \mathbb{R}^n} \frac{\sum_{ij \in E} \big(x(i) + x(j)\big)^2}{\sum_{i \in V} \deg(i) \cdot x(i)^2}.$$

Trevisan *defined* the combinatorial property to measure the bipartiteness ratio of a subset of vertices using the ℓ_1-version of the optimization problem in Exercise 3.16.

Definition 3.17 (Bipartiteness Ratio). *Let $G = (V = [n], E)$ be an undirected graph. The bipartiteness ratio of a vector $x \in \{-1, 0, 1\}^n$ is defined as*

$$\beta(x) := \frac{\sum_{ij \in E} \big|x(i) + x(j)\big|}{\sum_{i \in V} \deg(i) \cdot |x(i)|}.$$

The bipartiteness ratio of a graph G is defined as

$$\beta(G) := \min_{x \in \{-1,0,1\}^n} \beta(x).$$

Given a subset S and a bipartition of S into (L, R), the corresponding vector $x \in \{-1, 0, 1\}$ is such that

$$x(i) = \begin{cases} +1 & \text{if } i \in L \\ -1 & \text{if } i \in R \\ 0 & \text{otherwise} \end{cases}.$$

Trevisan proved the following analog of Cheeger's inequality for $2 - \lambda_n$ and $\beta(G)$.

Problem 3.18 (Cheeger's Inequality for λ_n [Tre09]). *Let $G = (V, E)$ be an undirected graph and λ_n be the largest eigenvalue of $\mathcal{L}(G)$. Then*

$$\frac{1}{2}(2 - \lambda_n) \le \beta(G) \le \sqrt{2(2 - \lambda_n)}.$$

An interesting application of this inequality is the design of an approximation algorithm for the maximum cut problem using spectral techniques. An observation is that if λ_n is bounded away from 2, then the graph does not have a very large max-cut.

Problem 3.19. *Use the easy direction of Problem 3.18 to show that the trivial approximation algorithm of cutting* 50% *of edges is a* $1/\lambda_n$*-approximation algorithm for the maximum cut problem.*

On the other hand, if λ_n is close to 2, the hard direction of Problem 3.18 can be used to find a subset $S = (L, R)$ with small bipartiteness ratio. This ensures that more than 50% of the edges with an endpoint in S will be cut. We can then apply the same idea recursively on $V - S$ to obtain a better than 50% approximation algorithm for the maximum cut problem. See [Tre09] for details and [Sot15] for an improved analysis.

References

[Alo86] Noga Alon. Eigenvalues and expanders. *Comb.*, 6(2):83–96, 1986. 19, 20, 30

[AM85] Noga Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985. 19, 20, 30

[Che70] J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems Anal.*, page 195, 1970. 19, 20

[Chu97] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997. 14, 20

[Dod84] J. Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Trans. Amer. Math. Soc.*, 284:787–794, 1984. 19, 20

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 3, 19, 20, 54, 57, 61, 62, 66, 69, 70, 72

[HT04] Christian Houdré and Prasad Tetali. Isoperimetric invariants for product Markov chains and graph products. *Combinatorica*, 24:359–388, 2004. 30

[LT24] Lap Chi Lau and Dante Tjowasi. On the Houdré-Tetali conjecture about an isoperimetric constant of graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, volume 317 of *LIPIcs*, pages 36:1–36:18, 2024. 30

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Inf. Comput.*, 82(1):93–133, 1989. 19, 20

[SM00] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000. 4, 27

[Sot15] José A. Soto. Improved analysis of a max-cut algorithm based on spectral partitioning. *SIAM J. Discret. Math.*, 29(1):259–268, 2015. 32

[Tan84] Robert Michael Tanner. Explicit concentrators from generalized n-gons. *SIAM Journal on Algebraic and Discrete Methods*, 5:287–293, 1984. 30

[Tre08] Luca Trevisan. The spectral partitioning algorithm. *Blog post:* https://lucatrevisan.wordpress.com/2008/05/11/the-spectral-partitioning-algorithm/, 2008. 19, 23, 24

[Tre09] Luca Trevisan. Max cut and the smallest eigenvalue. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC*, pages 263–272. ACM, 2009. 31, 32

Chapter 4

Random Walks on Graphs

Given an undirected graph $G = (V, E)$, a random walk is a simple stochastic process where it starts from a vertex, and in each step the walk moves to a uniformly random neighbor of the current vertex. We are interested in understanding the long-term behavior of the random walk. Is there a limiting distribution on the vertices as the number of steps tends to infinity? If so, how many steps are needed to converge to the limiting distribution?

There are two main approaches to addressing these questions. One is probabilistic, based on the concept of "coupling" two random processes. The other is spectral, using the eigenvalues of the transition matrix. We study the spectral approach and refer the reader to [Häg02, LPW06] for expositions on the probabilistic approach.

In this chapter, we begin with the more general setting of a finite Markov chain and state the fundamental theorem. Next, we specialize the fundamental theorem to the case of random walks on undirected graphs and use spectral analysis to prove it. The spectral analysis builds naturally on the results in Chapter 2 and Chapter 3 and provides a useful upper bound on the mixing time. Finally, we discuss some interesting applications of random walks and mention some known results for random walks on directed graphs.

4.1 Markov Chains

A finite Markov chain is defined by a finite state space and a transition matrix.

Definition 4.1 (Transition Matrix)**.** *Let $[n]$ be the state space. A matrix $P \in \mathbb{R}^{n \times n}$ is a probability transition matrix if P is non-negative and $\sum_{j \in [n]} P_{i,j} = 1$ for each $i \in [n]$. For $1 \leq i, j \leq n$, the entry P_{ij} is the transition probability from state i to state j.*

Definition 4.2 (Markov Chain)**.** *A sequence of random variables $(X_0, X_1, \ldots)$ is a Markov chain with state space $[n]$ and transition matrix $P \in \mathbb{R}^{n \times n}$ if, for all $i, j \in [n]$ and $t \geq 1$,*

$$\Pr\left[X_{t+1} = j \mid X_t = i \cap X_{t-1} = i_{t-1} \cap \ldots \cap X_0 = i_0\right] = \Pr\left[X_{t+1} = j \mid X_t = i\right] = P_{i,j}.$$

This property, known as the Markov property, states that the transition probability from i to j depends only on the current states X_t, regardless of the states $X_0, \ldots, X_{t-1}$ that precedes it.

Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution over the states. Then, $\vec{p}_t := \vec{p}_0 P^t$ is the probability distribution on the states after t steps of random walks.

A Markov chain can be viewed as a random walk on a weighted directed graph $G = ([n], w)$, where the transition probability from state i to state j is proportional to the edge weight $w(i, j)$ such that $P_{i,j} = w(i, j)/\sum_{j\in[n]} w(i, j)$.

Irreducibility and Aperiodicity

Two key properties ensures the existence of a unique limiting distribution.

Definition 4.3 (Irreducibility). *A Markov chain defined by transition matrix $P \in \mathbb{R}^{n\times n}$ is called irreducible if, for any two states i, j, there exists an integer t such that $\Pr[X_t = j \mid X_0 = i] > 0$.*

Equivalently, the underlying directed graph $G = ([n], E)$, where $E(G) := \{ij \mid P_{i,j} > 0\}$ of P, is strongly connected.

This property is called irreducibility because, if it is not satisfied, then the Markov chain can be reduced to a smaller one for studying the limiting distribution. Specifically, the limiting distribution, if exists, will only have support on a strongly connected component of the underlying directed graph.

Definition 4.4 (Aperiodicity). *The period of a state i is defined as $\gcd\{t \mid \Pr[X_t = i \mid X_0 = i] > 0\}$, the greatest common divisor of the set of times when it is possible to return to the starting state i. A state i is aperiodic if its period is equal to 1. A Markov chain is aperiodic if all states are aperiodic; otherwise it is periodic.*

For example, random walks on an undirected bipartite graph are periodic, as every state has period 2. Similarly, random walks on a directed cycle of length $k > 1$ are periodic, with every state having a period of k. In general, a Markov chain does not have a limiting distribution if it is periodic.

Irreducibility and aperiodicity together imply the following property.

Proposition 4.5 (Reachability). *For any finite, irreducible, and aperiodic Markov chain, there exists an integer $\tau < \infty$ such that $\Pr[X_t = j \mid X_0 = i] > 0$ for all i, j and all $t \geq \tau$.*

The proof of reachability uses aperiodicity and a simple number-theoretic argument to establish the statement for all $i = j$, and then uses irreducibility to extend the statement for all $i \neq j$. We do not need this result for the spectral analysis in this chapter. Interested readers are referred to [Häg02, LPW06] for a detailed proof.

Stationary Distribution and Convergence

Informally, when a limiting distribution exists, it is a stationary distribution, defined as follows.

Definition 4.6 (Stationary Distribution). *For a Markov chain defined by a transition matrix $P \in \mathbb{R}^{n\times n}$, a probability distribution $\vec{\pi} \in \mathbb{R}^n$ is a stationary distribution if $\vec{\pi}P = \vec{\pi}$, where $\vec{\pi}$ is represented as a row vector.*

A stationary distribution $\vec{\pi}$ satisfies $\vec{\pi}P^t = \vec{\pi}$ for any $t \geq 1$. From a linear algebraic perspective, a stationary distribution is simply an eigenvector of $P^\top$ corresponding to the eigenvalue 1.

To define convergence, we need a measure of how close two probability distributions are. One commonly used measure is the total variation distance.

Definition 4.7 (Total Variation Distance)**.** *Given two probability distributions $\vec{p}, \vec{q} \in \mathbb{R}^n$, the total variation distance of $\vec{p}$ and $\vec{q}$ is defined as*

$$d_{\mathrm{TV}}(\vec{p}, \vec{q}) := \frac{1}{2} \sum_{i=1}^{n} |p(i) - q(i)| = \frac{1}{2} \|\vec{p} - \vec{q}\|_1.$$

We say that $\vec{p}_t$ converges to a probability distribution $\vec{q}$ as $t \to \infty$ if $\lim_{t\to\infty} d_{\mathrm{TV}}(\vec{p}_t, \vec{q}) = 0$.

Fundamental Theorem of Markov Chains

The fundamental theorem states that any finite, irreducible and aperiodic Markov chain has a unique limiting distribution. Moreover, this limiting distribution is independent of the initial distribution.

Theorem 4.8 (Fundamental Theorem of Markov Chains)**.** *Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n\times n}$. Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution on the states. Let $\vec{p}_t \in \mathbb{R}^n$ be the probability distribution after t steps, i.e., $\vec{p}_t := \vec{p}_0 P^t$ after t steps. If the Markov chain is finite, irreducible, and aperiodic, then the distribution $\vec{p}_t$ converges to a unique stationary distribution $\vec{\pi}$, regardless of the initial distribution $\vec{p}_0$.*

The intuition behind the probabilistic proof of the fundamental theorem is as follows: For any finite, irreducible, and aperiodic Markov chain defined by P, running the chain for a sufficiently long time ensures that it is possible to reach any state from any other state (Proposition 4.5). If two instances of the Markov chains, $(X_1, X_2, \ldots)$ and $(Y_1, Y_2, \ldots)$, meet at the same state at some time t (i.e., $X_t = Y_t$), then their future behavior becomes indistinguishable because Markov chains "forget" their history. By Proposition 4.5, any two instances of the Markov chains will eventually meet, and thus all distributions converge to the same limiting distribution as $t \to \infty$.

In the following sections, we specialize the fundamental theorem to the case of random walks on undirected graphs and use a spectral approach to prove it. The spectral analysis has the advantage that it also provides a useful upper bound on the mixing time.

For the general result, we refer the reader to [Häg02] for a probabilistic proof using coupling, [LPW06] for a probabilistic and algebraic proof, and [HJ13] for a purely algebraic proof related to the Perron-Frobenius Theorem A.18.

4.2 Random Walks on Undirected Graphs

We consider random walks on an unweighted undirected graph $G = (V, E)$, where in each step the walk moves to a uniformly random neighbor of the current vertex. The fundamental theorem becomes easier in this special case, as there are simple characterizations of irreducibility, aperiodicity, and the limiting distribution. We also consider lazy random walks at the end of this section.

Matrix Formulation: The transition probability P_{ij} from a vertex i to a vertex j is simply $1/\deg(i)$, and so the transition matrix is $P = D^{-1}A$ where D is the diagonal degree matrix in Definition 2.10 and A is the adjacency matrix in Definition 2.1. Let $\vec{p}_0 : V \to \mathbb{R}$ be an initial probability distribution, and $\vec{p}_t$ be the probability distribution after t steps of random walks. Then $\vec{p}_{t+1} = \vec{p}_t P^\top = \vec{p}_t D^{-1} A$, and by induction $\vec{p}_t = \vec{p}_0 (D^{-1}A)^t$.

Stationary Distribution: Recall that a probability distribution $\vec{\pi} : V \to \mathbb{R}$ is a stationary distribution of P if $\vec{\pi}P = \vec{\pi}$. It is equivalent to saying that $\vec{\pi}$ is a an eigenvector of $P^\top$ with eigenvalue 1. Given that $P = D^{-1}A$ for random walks on undirected graphs, it is not difficult to identify one such eigenvector with probabilities proportional to the degrees.

Lemma 4.9 (Stationary Distribution of Undirected Graphs)**.** *Let $G = (V, E)$ be an undirected graph and $P = D^{-1}A$ be its transition matrix. The distribution $\vec{\pi} : V \to \mathbb{R}$ with*

$$\vec{\pi}(i) = \frac{\deg(i)}{\sum_{j \in V} \deg(j)} = \frac{\deg(i)}{2|E|}$$

for all $i \in V$ is a stationary distribution of P.

Irreducibility: Is $\vec{\pi}$ in Lemma 4.9 the unique stationary distribution? Not necessarily. For example, if the graph is disconnected, the distribution after many steps depends on the initial distribution (e.g., which component does the starting vertex belongs to). This corresponds to the irreducibility condition in the fundamental theorem. For undirected graphs, the irreducibility condition is equivalent to the graph being connected.

Aperiodicity: Even if the graph is connected, a limiting distribution may not exist. For example, in a connected bipartite graph, if the initial distribution $\vec{p}_0$ is on a single vertex, then the distribution $\vec{p}_t$ depends on the parity of t, as the support of $\vec{p}_t$ oscillates between the two sides of the bipartite graph. This corresponds to the aperiodicity condition in the fundamental theorem. For connected undirected graphs, observe that the aperiodicity condition is equivalent to the condition that the graph is non-bipartite.

Fundamental Theorem: Given the simple characterizations of the conditions in the fundamental theorem, it reduces to the following statement for undirected graphs.

Theorem 4.10 (Fundamental Theorem for Undirected Graphs)**.** *Let G be a connected, non-bipartite undirected graph. Let $P = D^{-1}A$ be the transition matrix of random walks on G. The distribution $\vec{\pi}$ in Lemma 4.9 is the unique stationary distribution. Furthermore, $\vec{p}_t := \vec{p}_0 P^t$ converges to $\vec{\pi}$ as $t \to \infty$ regardless of the initial distribution $\vec{p}_0$.*

Lazy Random Walks

The non-bipartiteness condition ensures that the Markov chain is aperiodic. There is a simple modification of the random walks so that this assumption can be removed by adding self-loops in the graph.

Definition 4.11 (Lazy Random Walks)**.** *Let G be an undirected graph. The transition matrix W of the lazy random walks is defined as $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$. In other words, the lazy random walks stay at the current vertex with probability $\frac{1}{2}$ and moves to a uniform random neighbor of the current vertex with probability $\frac{1}{2}$.*

By performing lazy random walks, we make the graph non-bipartite and obtain the following corollary of Theorem 4.10.

Corollary 4.12 (Fundamental Theorem for Lazy Undirected Graphs)**.** *Let G be a connected undirected graph. Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. The distribution $\vec{\pi}$ in Lemma 4.9 is the unique stationary distribution. Furthermore, $\vec{p}_t := \vec{p}_0 W^t$ converges to $\vec{\pi}$ as $t \to \infty$ regardless of the initial distribution $\vec{p}_0$.*

It will be clear from the spectral analysis in the next section why the constant 1/2 is used.

4.3 Spectral Analysis of Mixing Time for Undirected Graphs

In this section, we will prove the fundamental theorem for undirected graphs in Theorem 4.10 using spectral analysis. Besides being elegant and insightful, spectral analysis can be used to analyze the mixing time, which is the rate of convergence to the unique stationary distribution.

We first assume that the undirected graph is d-regular. We then explain the modifications needed for non-regular undirected graphs.

Spectrum of the Transition Matrix for Regular Graphs

For a d-regular graph, the transition matrix P for random walks and the transition matrix W for lazy random walks are

$$P = D^{-1}A = \frac{1}{d}A = \mathcal{A} \quad \text{and} \quad W = \frac{1}{2}I + \frac{1}{2}\mathcal{A},$$

where $\mathcal{A}$ is the normalized adjacency matrix in Definition 2.17.

This is the main simplification from the d-regular assumption, as the matrices P and W are still real symmetric. Another simplification is that the stationary distribution $\vec{\pi}$ in Lemma 4.9 is simply the uniform distribution $\vec{1}/n$ for a d-regular graph.

Our goal is to prove that

$$\lim_{t\to\infty} P^t \vec{p}_0 = \frac{\vec{1}}{n} \quad \text{and} \quad \lim_{t\to\infty} W^t \vec{p}_0 = \frac{\vec{1}}{n},$$

regardless of the initial distribution $\vec{p}_0$, as long as the graph is connected and non-bipartite for random walks, and connected for lazy random walks.

For the spectral analysis, we write $\vec{p}_0$ and $\vec{p}_t$ as column vectors. To compute $P^t\vec{p}_0$ and $W^t\vec{p}_0$, a repeated application of the same operator, it is helpful to know the spectrum of the matrices P and W, as discussed in Appendix A.

Let $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ be the eigenvalues of $\mathcal{A}$ and $v_1, \ldots, v_n$ be the corresponding orthonormal eigenvectors. Recall that

1. $\alpha_1 = 1$ and $v_1 = \vec{1}/\sqrt{n}$ from Lemma 2.7,
2. $\alpha_2 < 1$ if and only if G is connected from Proposition 2.15,
3. $\alpha_n > -1$ if and only if G is non-bipartite from Problem 2.21.

For the lazy random walk matrix W, the spectrum is $\frac{1}{2}(1+\alpha_1) \geq \frac{1}{2}(1+\alpha_2) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$, which implies that the smallest eigenvalue is always at least 0. This is why the non-bipartitness assumption can be removed when we consider lazy random walks.

Limiting Distribution

After translating the combinatorial conditions in the fundamental theorem into spectral conditions, we can restate the fundamental theroem for d-regular undirected graphs in Theorem 4.10 as follows, and the proof becomes transparent.

Proposition 4.13 (Limiting Distribution for Regular Graphs)**.** *Let $G = (V = [n], E)$ be a d-regular undirected graph. Let $P = \mathcal{A}$ be the transition matrix of random walks on G and $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -1$ be its eigenvalues. If $\alpha_2 < 1$ and $\alpha_n > -1$, then*

$$\lim_{t\to\infty} P^t \vec{p}_0 = \frac{\vec{1}}{n}.$$

Proof. Let $v_1, v_2, \ldots, v_n$ be the orthonormal eigenvectors. For any initial distribution $\vec{p}_0$, we write $\vec{p}_0 = c_1 v_1 + \ldots + c_n v_n$ where $c_i = \langle p_0, v_i \rangle$ for $1 \leq i \leq n$. Then,

$$P^t \vec{p}_0 = \mathcal{A}^t \Big(\sum_{i=1}^{n} c_i v_i \Big) = \sum_{i=1}^{n} c_i \mathcal{A}^t v_i = \sum_{i=1}^{n} c_i \alpha_i^t v_i.$$

The assumptions $\alpha_2 < 1$ and $\alpha_n > -1$ imply that $|\alpha_i| < 1$ for $2 \leq i \leq n$. Hence,

$$\lim_{t\to\infty} P^t \vec{p}_0 = \lim_{t\to\infty} \sum_{i=1}^{n} c_i \alpha_i^t v_i = c_1 v_1,$$

as all but the first term go to zero as $t \to \infty$. In the d-regular case, $v_1 = \vec{1}/\sqrt{n}$ and thus $c_1 = \langle \vec{p}_0, \vec{1}/\sqrt{n} \rangle = 1/\sqrt{n}$ as $\vec{p}_0$ is a probability distribution. Therefore,

$$\lim_{t\to\infty} P^t \vec{p}_0 = c_1 v_1 = \frac{1}{\sqrt{n}} \cdot \frac{\vec{1}}{\sqrt{n}} = \frac{\vec{1}}{n}.$$

□

The proof shows that under $|\alpha_i| < 1$ for $2 \leq i \leq n$, the distribution $P^t \vec{p}_0$ converges to the first eigenvector, which is proportional to the all-one vector. Check that the same proof works for lazy random walks (Corollary 4.12) on d-regular graphs.

Mixing Time

The mixing time quantifies how fast $\vec{p}_t := P^t \vec{p}_0$ converges to the limiting distribution. The following definition applies to general Markov chains.

Definition 4.14 (Mixing Time)**.** *Consider the Markov chain defined by a transition matrix $P \in \mathbb{R}^{n\times n}$. Let $\vec{p}_0 \in \mathbb{R}^n$ be an initial probability distribution on the states, and let $\vec{p}_t \in \mathbb{R}^n$ be the probability distribution $\vec{p}_t := (P^\top)^t \vec{p}_0$ after t steps.*

Suppose the limiting distribution $\vec{\pi} = \lim_{t\to\infty} \vec{p}_t$ exists. For any $0 < \epsilon \leq 1$, the ϵ-mixing time $\tau_\epsilon(P)$ of P is defined as the smallest t such that $d_{\mathrm{TV}}(\vec{p}_t, \vec{\pi}) \leq \epsilon$ for any initial distribution $\vec{p}_0$, where d_{TV} is the total variation distance in Definition 4.7.

When ϵ is not specified, it is assumed to be a small constant such as $1/4$, and we simply say $\tau_{1/4}(P)$ the mixing time of the Markov chain P.

To bound the mixing time, we use the same approach as in Proposition 4.13, but assume that α_2 and $|\alpha_n|$ are bounded away from one, ensuring α_i^t converges to zero quickly for $2 \le i \le n$.

Theorem 4.15 (Bounding Mixing Time by Spectral Gap)**.** *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $P = \mathcal{A}$ be the transition matrix of random walks on G and $1 = \alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n \ge -1$ be its eigenvalues. Let $g := \min\{1-\alpha_2, 1-|\alpha_n|\}$ be the two-sided spectral gap. Then the ϵ-mixing time of P satisfies*

$$\tau_\epsilon(P) \lesssim \frac{1}{g} \ln\left(\frac{n}{\epsilon}\right).$$

Proof. Continuing from Proposition 4.13,

$$P^t \vec{p}_0 = \frac{\vec{1}}{n} + \sum_{i=2}^{n} c_i \alpha_i^t v_i,$$

where $v_1, \ldots, v_n$ are the orthonormal eigenvectors, $c_i = \langle \vec{p}_0, v_i \rangle$ for $2 \le i \le n$, and $\vec{\pi} = \vec{1}/n$ is the limiting distribution. Then,

$$d_{\mathrm{TV}}(\vec{p}_t, \vec{\pi}) = d_{\mathrm{TV}}(P^t \vec{p}_0, \vec{\pi}) = \frac{1}{2}\left\| P^t \vec{p}_0 - \frac{\vec{1}}{n} \right\|_1 = \frac{1}{2}\left\| \sum_{i=2}^{n} c_i \alpha_i^t v_i \right\|_1 \lesssim \sqrt{n} \left\| \sum_{i=2}^{n} c_i \alpha_i^t v_i \right\|_2,$$

where the last inequality follows from $\|v\|_1 \le \sqrt{n}\|v\|_2$ for $v \in \mathbb{R}^n$ (which can be derived from Cauchy-Schwarz). Since $v_1, \ldots, v_n$ are orthonormal,

$$\left\| \sum_{i=2}^{n} c_i \alpha_i^t v_i \right\|_2^2 = \sum_{i=2}^{n} c_i^2 \alpha_i^{2t} \le (1-g)^{2t} \sum_{i=2}^{n} c_i^2.$$

Note that $\sum_{i=2}^n c_i^2 \le \sum_{i=1}^n c_i^2 = \|\vec{p}_0\|_2^2 \le \|\vec{p}_0\|_1^2 = 1$. Therefore,

$$d_{\mathrm{TV}}(\vec{p}_t, \vec{\pi}) \lesssim \sqrt{n} \left\| \sum_{i=2}^{n} c_i \alpha_i^t v_i \right\|_2 \le \sqrt{n(1-g)^{2t} \sum_{i=2}^{n} c_i^2} \le \sqrt{n}(1-g)^t \le \sqrt{n} e^{-gt}.$$

Setting $t \gtrsim \frac{1}{g} \ln(\frac{n}{\epsilon})$ ensures $d_{\mathrm{TV}}(p_t, \vec{\pi}) \le \epsilon$ for any initial distribution p_0. □

For the lazy random walk matrix W, the smallest eigenvalue is at least 0, so the spectral gap for W is simply $g = \frac{1}{2}(1-\alpha_2) = \frac{1}{2}\lambda_2$, where λ_2 is the second smallest eigenvalue of the normalized Laplacian matrix. Cheeger's inequality in Theorem 3.2 then implies the following important consequence.

Theorem 4.16 (Bounding Mixing Time by Edge Conductance)**.** *Let $G = (V, E)$ be a d-regular undirected graph with $V = [n]$. Let $W = \frac{1}{2}I + \frac{1}{2}\mathcal{A}$ be the transition matrix of lazy random walks on G. Then,*

$$\tau_\epsilon(W) \lesssim \frac{1}{\phi(G)^2} \ln\left(\frac{n}{\epsilon}\right).$$

This result provides a combinatorial condition for fast mixing. For an expander graph with $\phi(G) = \Omega(1)$, the mixing time of lazy random walks is $O(\ln n)$. Establishing a polylogarithmic mixing time is crucial for many applications, as we will discuss with examples.

Theorem 4.16 is useful in designing random sampling algorithms. For the purpose of uniform random sampling, the analysis for regular graphs is usually sufficient, as we can set up the Markov chain (e.g., by adding self-loops) so that the underlying graph is regular.

Spectrum of the Transition Matrix for General Graphs

The random walk matrix for general graphs is $P = D^{-1}A$, and the lazy random walk matrix is $W = \frac{1}{2}I + \frac{1}{2}P$. The main difference from the d-regular case is that these matrices are in general not symmetric, and so the spectral theorem in Theorem A.5 cannot be directly applied to reason about their eigenvalues and eigenvectors.

A simple but important observation is that P and W are similar to a real symmetric matrix (see Definition A.3), and so the eignevalues of P and W are still all real numbers.

Lemma 4.17 (Spectrum of Random Walk Matrices)**.** *Let $G = (V, E)$ be a connected undirected graph with $V = [n]$, and $\mathcal{A}$ be its normalized adjacency matrix. Let the eigenvalues of $\mathcal{A}$ be $\alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n$ and let $v_1, v_2, \ldots, v_n$ be a corresponding orthonormal basis of eigenvectors.*

Then the eigenvalues of the random walk matrix $P = D^{-1}A$ are also $\alpha_1 > \alpha_2 \geq \ldots \geq \alpha_n$, and the corresponding eigenvectors of $P^\top = AD^{-1}$ are $D^{\frac{1}{2}}v_1, D^{\frac{1}{2}}v_2, \ldots, D^{\frac{1}{2}}v_n$.

The eigenvalues of the lazy random walk matrix $W = \frac{1}{2}I + \frac{1}{2}P$ are $\frac{1}{2}(1+\alpha_1) > \frac{1}{2}(1+\alpha_2) \geq \ldots \geq \frac{1}{2}(1+\alpha_n)$, and the corresponding eigenvectors of $W^\top$ are $D^{\frac{1}{2}}v_1, D^{\frac{1}{2}}v_2, \ldots, D^{\frac{1}{2}}v_n$.

Proof. Note that $P = D^{-1}A = D^{-\frac{1}{2}}(D^{-\frac{1}{2}}AD^{-\frac{1}{2}})D^{\frac{1}{2}} = D^{-\frac{1}{2}}\mathcal{A}D^{\frac{1}{2}}$, and so P is similar to $\mathcal{A}$ as D is non-singular when the graph is connected. By the same argument, W is similar to $\frac{1}{2}I + \frac{1}{2}\mathcal{A}$. By Fact A.4, P and $\mathcal{A}$ have the same spectrum, and W and $\frac{1}{2}I + \frac{1}{2}\mathcal{A}$ have the same spectrum.

Note that $D^{\frac{1}{2}}v_i$ is an eigenvector of $P^\top$ with eigenvalue α_i, as

$$P^\top\big(D^{\frac{1}{2}}v_i\big) = \big(D^{\frac{1}{2}}\mathcal{A}D^{-\frac{1}{2}}\big)\big(D^{\frac{1}{2}}v_i\big) = D^{\frac{1}{2}}\mathcal{A}v_i = \alpha_i\big(D^{\frac{1}{2}}v_i\big).$$

Similarly, $D^{\frac{1}{2}}v_i$ is an eigenvector of $W^\top$ with eigenvalue $\frac{1}{2}(1+\alpha_i)$. □

The vectors $D^{\frac{1}{2}}v_1, \ldots, D^{\frac{1}{2}}v_n$ are linearly independent because D is non-singular for a connected graph. Note that these vectors are in general not orthonormal with respect to the standard inner product, but they are orthonormal if we use the following weighted inner product:

$$\langle u, v\rangle_{D^{-1}} := u^\top D^{-1} v \quad \text{and} \quad \|v\|_{D^{-1}} := \sqrt{v^\top D^{-1} v}.$$

Spectral Analysis for General Undirected Graphs

Using this weighted inner product, we can generalize the spectral analysis in Proposition 4.13 and Theorem 4.15 to non-regular graphs. We describe the main modifications and leave the verification of the details to the reader.

To compute the limiting distribution $P^\top\vec{p}_0 = (D^{\frac{1}{2}}\mathcal{A}D^{-\frac{1}{2}})^t\vec{p}_0 = D^{\frac{1}{2}}\mathcal{A}^tD^{-\frac{1}{2}}\vec{p}_0$, we write the initial distribution $\vec{p}_0$ as $\sum_{i=1}^n c_i D^{\frac{1}{2}}v_i$ to take advantage of the orthonormality of $v_1, \ldots, v_n$, where $c_i = \langle\vec{p}_0, D^{\frac{1}{2}}v_i\rangle_{D^{-1}}$ for $1 \leq i \leq n$. We can then adapt the proof in Proposition 4.13 to prove the following equivalent form of the fundamental theorem for undirected graphs in Theorem 4.10.

Exercise 4.18 (Limiting Distribution for Undirected Graphs)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $P = D^{-1}A$ be the transition matrix of random walks on G, and let $1 = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -1$ be its eigenvalues. If $\alpha_2 < 1$ and $\alpha_n > -1$, then*

$$\lim_{t\to\infty}\big(P^\top\big)^t\vec{p}_0 = \frac{\vec{d}}{2|E|},$$

where $\vec{d}$ is the degree vector with $\vec{d}(i) = \deg(i)$ for $1 \le i \le n$.

To bound the mixing time, we adapt the proof in Theorem 4.15. The key steps are

$$\|\vec{p}_t - \vec{\pi}\|_1 \le \|\vec{1}\|_D \cdot \|\vec{p}_t - \vec{\pi}\|_{D^{-1}} \asymp \sqrt{|E|} \cdot \|\vec{p}_t - \vec{\pi}\|_{D^{-1}} \le (1-g)^t \sqrt{|E|} \cdot \|\vec{p}_0\|_{D^{-1}}$$

where the first inequality is by Cauchy-Schwarz and the second inequality is by an orthonormality argument as in Theorem 4.15. Then the same theorem as in the d-regular case can be proved.

Theorem 4.19 (Bounding Mixing Time by Spectral Gap and Edge Conductance)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $P = D^{-1}A$ be the transition matrix of random walks on G and $1 = \alpha_1 \ge \alpha_2 \ge \ldots \ge \alpha_n \ge -1$ be its eigenvalues. Let $g := \min\{1 - \alpha_2, 1 - |\alpha_n|\}$ be the two-sided spectral gap. Then the ϵ-mixing time of P satisfies*

$$\tau_\epsilon(P) \lesssim \frac{1}{g} \ln\left(\frac{n}{\epsilon}\right).$$

Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. Then

$$\tau_\epsilon(W) \lesssim \frac{1}{\phi(G)^2} \ln\left(\frac{n}{\epsilon}\right).$$

For weighted undirected graphs, the same arguments can be used to prove that

$$\tau_\epsilon(P) \lesssim \frac{1}{g} \ln\left(\frac{1}{\epsilon \cdot \pi_{\min}}\right) \quad \text{and} \quad \tau_\epsilon(W) \lesssim \frac{1}{\phi(G)^2} \ln\left(\frac{1}{\epsilon \cdot \pi_{\min}}\right), \tag{4.1}$$

where $\pi_{\min} := \min_i \vec{\pi}(i)$ is the minimum stationary probability of a vertex. We leave the verification of these bound to Problem 4.22.

Remark 4.20. *This spectral approach can be further extended to prove the fundamental theorem for directed graphs, but it is considerably more involved and requires the Perron-Frobenius theorem and the Jordan normal form (see [HJ13] for proofs).*

4.4 Applications of Random Walks

We briefly discuss two applications of random walks that will be studied in later chapters. In both cases, random walks are used to solve the problem while exploring only a small portion of the graph.

Random Sampling

An important application of random walks is in random sampling. As an example, consider the following algorithm for generating a random spanning tree of an undirected graph.

To analyze this algorithm, we interpret it as performing random walks on a large "spanning tree exchange graph" H. In H, each vertex represents a spanning tree of the original graph, and two vertices are connected if their corresponding spanning trees T and T' can be obtained through one step of the algorithm (i.e. $T' = T + e - f$ for some edges e, f in the input graph).

Note that the exchange graph H could have as many as $\Omega(n^{n-2})$ vertices when the original graph has n vertices. Therefore, to show that $\tau \lesssim \text{poly}(n)$ is sufficient to return an almost uniform random

Algorithm 2 Random Exchange Algorithm for Sampling Random Spanning Trees

Require: An undirected graph $G = (V, E)$.

1: Compute an arbitrary spanning tree T_0 of the graph.
2: **for** $1 \leq t \leq \tau$ **do**
3: Add a uniform random edge $e \in E - T_{t-1}$ to T_{t-1}.
4: Remove a uniform random edge f in the unique cycle formed in $T_{t-1} + e$.
5: Set $T_t := T_{t-1} + e - f$.
6: **end for**
7: **return** T_τ.

spanning tree, we must prove that the random walks on H mix in polylogarithmic time relative to its size. From a combinatorial perspective, this requires proving that the spanning tree exchange graph H is an expander graph, a task that is generally quite challenging.

There are different approaches to proving fast mixing of Markov chains. One is the coupling method, the most common and versatile probabilistic technique in bounding mixing time (see [LPW06]). Another is the canonical path method, which uses multicommodity flow to lower bound the graph conductance, so that Theorem 4.16 can be used to upper bound the mixing time. A famous application of the canonical path method is in approximating the permanent of a non-negative matrix [JSV04], which is equivalent to counting the number of perfect matchings in a bipartite graph.

These methods are beyond the scope of this course. Instead, we will analyze the random exchange algorithm for sampling random spanning trees using the new techniques derived from high dimensional expanders in Chapter **??**.

Local Graph Partitioning

Another useful application of random walks is in graph partitioning. This idea, originally proposed by Spielman and Teng [ST13], is to use the random walk distribution $W^t\chi_i$ from some starting vertex i to identify a small sparse cut of the graph. They proved that the performance of the random walk algorithm for graph partitioning is comparable to that of the spectral partitioning algorithm in Chapter 3. Furthermore, the random walk algorithm has the significant advantage that it can be implemented locally, such that the running time depends only on output size but not on the original graph size. This provides a sublinear time algorithm for graph partitioning in some situations. Local graph partitioning is an active research topic on its own, and there are several other algorithms such as using PageRank vector [ACL06] and evolving sets [AOPT16]. We will discuss these results in Chapter **??**.

4.5 Random Walks on Directed Graphs

For directed graphs, there are currently no direct relationships between the eigenvalues of their transition matrix and the mixing time of random walks.

In this section, we discuss some known results about the mixing time of random walks on directed graphs, using the second eigenvalue of symmetric matrices associated with directed graphs.

Stationary Flow Graph

Given a directed graph $G = (V, E)$ with an edge weight function $w : E \to \mathbb{R}_+$, let P be the transition matrix of the random walk, where $P_{i,j} = w(ij)/\sum_{j:ij\in E} w(ij)$. Assume P is irreducible and aperiodic. Let $\vec{\pi}$ be the unique stationary distribution of P, and define $\Pi := \mathrm{diag}(\vec{\pi})$.

To study the mixing time of the weighted directed graph $G = (V, E, w)$, it is helpful to consider the stationary flow graph $G_f = (V, E, f)$, where $f(i, j) = \vec{\pi}(i) \cdot P_{ij}$ is the probability flow on edge ij in the stationary distribution $\vec{\pi}$. Verify that the weighted directed graph $G_f = (V, E, f)$ is Eulerian, satisfying $\sum_{j:ji\in E} f(j, i) = \sum_{k:ik\in E} f(i, k)$ for all $i \in V$.

The adjacency matrix of G_f is denoted by $F := \Pi P$. The Eulerian property implies that the i-th row sum of F is equal to the i-th column sum of F for all i.

Symmetric Matrices for Directed Graphs

Fill [Fil91] defined the sum matrix as

$$\mathfrak{A} := \frac{1}{2}(P + \Pi^{-1}P^\top\Pi).$$

Chung [Chu05] defined the directed Laplacian matrix of G as

$$\mathfrak{L} = I - \Pi^{-\frac{1}{2}}\left(\frac{F + F^\top}{2}\right)\Pi^{-\frac{1}{2}}. \tag{4.2}$$

Observe that $\mathfrak{L}$ is the normalized Laplacian matrix of the symmetrized flow graph where the weight of edge ij is $\frac{1}{2}(f(i, j) + f(j, i))$, as the diagonal degree matrix of the symmetrized flow graph is still Π because of the Eulerian property.

Note that the spectra of $\mathfrak{A}$ and $\mathfrak{L}$ are essentially the same, as $\mathfrak{A}$ and $I - \mathfrak{L}$ are similar matrices.

Bounding Mixing Time by Spectral Gap of Symmetric Matrix

A main result from [Fil91, Chu05] uses the spectral gap of $\mathfrak{A}$ or $\mathfrak{L}$ to bound the mixing time of random walks on G.

Theorem 4.21 (Bounding Mixing Time by Second Eigenvalue of Directed Graphs [Fil91, Chu05])**.** *Let G be a strongly connected directed graph $G = (V, E)$ with a weight function $w : E \to \mathbb{R}_+$, and P be the transition matrix of the random walks on G with $P(i, j) = w(ij)/\sum_{i\in V} w(ij)$ for $ij \in E$. The ϵ-mixing time of the lazy random walks on G (with transition matrix $\frac{1}{2}(I+P)$) to the stationary distribution $\vec{\pi}$ satisfies*

$$\tau_\epsilon\Big(\frac{I + P}{2}\Big) \lesssim \frac{1}{\lambda_2(\mathfrak{L})} \cdot \log\Big(\frac{1}{\pi_{\min} \cdot \epsilon}\Big)$$

where $\lambda_2(\mathfrak{L})$ is the second smallest eigenvalue of $\mathfrak{L}$ in (4.2), and $\pi_{\min} = \min_{i\in V} \vec{\pi}(i)$.

Cheeger Constant of Directed Graphs

The Cheeger constant of a set and of a directed graph [Fil91, Chu05, LPW06] are defined as

$$h(S) := \frac{\sum_{i\in S, j\notin S} \pi(i)P(i, j)}{\pi(S)} = \frac{\sum_{i\in S, j\notin S} F(i, j)}{\pi(S)} \quad \text{and} \quad h(G) := \min_{S:\pi(S)\le\frac{1}{2}} h(S). \tag{4.3}$$

This is also known as the conductance or the bottleneck ratio in the literature [LPW06].

Since the flow graph F is Eulerian, the Cheeger constant of the flow graph is the same as that of the symmetrized flow graph $\frac{1}{2}(F + F^\top)$. So the following Cheeger's inequality by Chung [Chu05] for directed graphs is a direct consequence of the Cheeger's inequality for undirected graphs in Theorem 3.2:

$$\frac{1}{2}\lambda_2(\mathfrak{L}) \le h(G) \le \sqrt{2\lambda_2(\mathfrak{L})}.$$

A consequence of Theorem 4.21 is that

$$\tau_\epsilon\Big(\frac{I+P}{2}\Big) \lesssim \frac{1}{h(G)^2} \cdot \log\Big(\frac{1}{\pi_{\min}\cdot\epsilon}\Big).$$

In Chapter **??**, we will study a combinatorial method to directly prove this consequence.

4.6 Problems

Problem 4.22 (Weighted Undirected Graphs)**.** *Extend the proof of Theorem 4.19 to establish (4.1). Fill in the proof details for Exercise 4.18 and Theorem 4.19.*

Problem 4.23 (Upper Bound on Mixing Time and Initial Distribution)**.** *Suppose the initial distribution $\vec{p}$ satisfies $\vec{p}(i) \le 2\pi(i)$ for all i where π is the unique stationary distribution. Prove that*

$$\tau_\epsilon(P) \lesssim \frac{1}{g}\ln\Big(\frac{1}{\epsilon}\Big) \quad \text{and} \quad \tau_\epsilon(W) \lesssim \frac{1}{\phi(G)^2}\ln\Big(\frac{1}{\epsilon}\Big).$$

In other words, the factor $\log(n)$ in Theorem 4.19 is only needed to get away from distributions concentrated on a small set.

Problem 4.24 (Lower Bound on Mixing Time)**.** *Let $G = (V, E)$ be an undirected graph with $V = [n]$. Let $W = \frac{1}{2}I + \frac{1}{2}D^{-1}A$ be the transition matrix of lazy random walks on G. Prove that the ϵ-mixing time of W is*

$$\tau_\epsilon(W) \gtrsim \frac{1}{1-\alpha_2}\ln\Big(\frac{1}{\epsilon}\Big),$$

where α_2 is the second largest eigenvalue of the normalized adjacency matrix $\mathcal{A}(G)$. A simpler problem is to prove that

$$\tau_\epsilon(W) \gtrsim \frac{1}{\phi(G)}\ln\Big(\frac{1}{\epsilon}\Big),$$

where $\phi(G)$ is the edge conductance of G. You may also consider the special case when G is d-regular.

Problem 4.25 (Page Ranking)**.** *Suppose someone searches a keyword (e.g., "car"), and we want to identify the webpages that are the most relevant for this keyword and those that are the most reliable sources (a page is considered reliable if it points to many highly relevant pages).*

First, we identify the pages with this keyword and ignore all others. Then we run the following ranking algorithm on the remaining pages. Each vertex corresponds to a remaining page, and there is a directed edge from page i to page j if there is a link from page i to page j. Call this directed graph $G = (V, E)$.

For each vertex i, we have two values, $s(i)$ and $r(i)$, where $r(i)$ represents the relevance of the page and $s(i)$ represents its reliability as a source (larger values are better). We start with arbitrary initial values, such as $s(i) = 1/|V|$ for all i, as we have no prior information.

At each step, we update s and r (where s and r are vectors of $s(i)$ and $r(i)$ values) as follows:

1. *Update* $r(i) = \sum_{j:ji \in E} s(j)$ *for all* i*, as a page is more relevant if it is linked by many reliable sources.*

2. *Update* $s(i) = \sum_{j:ij \in E} r(j)$ *for all* i *(using the just updated values* $r(j)$*), as a page is a more reliable source if it points to many relevant pages.*

To keep the values bounded, let $R = \sum_{i=1}^{|V|} r(i)$ *and* $S = \sum_{i=1}^{|V|} s(i)$*, and normalize by dividing each* $s(i)$ *by* S *and divide each* $r(i)$ *by* R*. We repeat these steps multiple times to refine the values.*

Let $s, r \in \mathbb{R}^{|V|}$ *be the vectors of the* s *and* r *values. Provide a matrix formulation for computing* s *and* r*.*

Suppose G *is weakly connected (the underlying undirected graph is connected when ignoring edge directions) and has a self-loop at each vertex. Prove that there is a unique limiting* s *and a unique limiting* r *for any initial* s*, provided* $s \geq 0$ *and* $s \neq 0$*. You may use the Perror-Frobenius Theorem A.18 to solve this problem.*

References

[ACL06] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. Local graph partitioning using pagerank vectors. In *FOCS*, pages 475–486. IEEE Computer Society, 2006. 42

[AOPT16] Reid Andersen, Shayan Oveis Gharan, Yuval Peres, and Luca Trevisan. Almost optimal local graph clustering using evolving sets. *J. ACM*, 63(2):15:1–15:31, 2016. 42

[Chu05] Fan Chung. Laplacians and cheeger inequality for directed graphs. *Annals of Combinatorics*, 9:1–19, 2005. 43, 44

[Fil91] James Allen Fill. Eigenvalue bounds on convergence to stationary for nonreversible markov chains, with an application to the exclusion process. *The Annals of Applied Probability*, 1:62–87, 1991. 43

[Häg02] Olle Häggström. *Finite Markov Chains and Algorithmic Applications*. London Mathematical Society Student Texts. Cambridge University Press, 2002. 33, 34, 35

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edition, 2013. 35, 41, 94, 95

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004. 42

[LPW06] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006. 3, 33, 34, 35, 42, 43, 44

[ST13] Daniel A. Spielman and Shang-Hua Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM J. Comput.*, 42(1):1–26, 2013. 4, 42

Chapter 5

Expander Graphs: Properties

There are several possible ways to define expander graphs:

1. Algebraically, expander graphs are graphs with a large spectral gap.
2. Combinatorically, expander graphs are graphs with very good connectivity properties.
3. Probabilistically, expander graphs are graphs in which random walks mix rapidly.

From what we have learnt in Chapter 3 and Chapter 4, these three perspectives are closely related:

- Cheeger's inequality in Theorem 3.2 states that $\phi(G) = \Omega(1)$ if and only if $\lambda_2 = \Omega(1)$.
- The spectral analysis in Theorem 4.16 and Problem 4.24 show that the mixing time τ of the lazy random walks on $G = (V, E)$ satisfies

$$\frac{1}{\lambda_2} \lesssim \tau \lesssim \frac{1}{\lambda_2} \log |V|.$$

Complete graphs are the best expander graphs under all three definitions, but we are interested in sparse expander graphs with a linear number of edges. In constructions of expander graphs, the spectral definition is the most convenient. We use the following stronger spectral definition that also bounds the last eigenvalue.

Definition 5.1 (Two-Sided Spectral Expanders)**.** *Let G be a d-regular graph and let the spectrum of its adjacency matrix be*

$$d = \alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n \geq -d.$$

We say that G is an (n, d, α)-graph if G has n vertices, G is d-regular, and $\max\{\alpha_2, |\alpha_n|\} \leq \alpha$.

In this chapter, we first prove the expander mixing lemma and its converse, providing a combinatorial characterization of two-sided spectral expanders. Then we discuss the extremal question of how small can α be. Finally, we investigate the stronger combinatorial and probabilistic properties that an (n, d, α)-graph has when $\alpha = o(d)$, including small-set vertex expansion and constant mixing time.

5.1 Expander Mixing Lemma

A well-known and useful property of two-sided spectral expanders is that they behaves as random d-regular graphs. Consider the number of edges between two subsets S, T of vertices.

Definition 5.2 (Induced Edges)**.** *Given an undirected graph $G = (V, E)$ and $S, T \subseteq V$, define $\vec{E}(S,T) := \{(u,v) \mid u \in S, v \in T, uv \in E\}$ as the set of ordered pairs where $u \in S$ and $v \in T$. Note that an edge with $u \in S \cap T$ and $v \in S \cap T$ is counted twice, as both (u,v) and (v,u) are in $\vec{E}(S,T)$.*

In a random graph where every pair of vertices has an edge with probability d/n, the expected value of $|\vec{E}(S,T)|$ is $d|S||T|/n$. The expander mixing lemma by Alon and Chung [AC88] says that in a two-sided spectral expander, $|\vec{E}(S,T)|$ is close to this expected value for all $S, T \subseteq V$. This can be interpreted as a pseudorandom or discrepancy property of a two-sided spectral expander.

Theorem 5.3 (Expander Mixing Lemma [AC88])**.** *Let $G = (V, E)$ be an (n, d, α)-graph. Then, for every $S \subseteq V$ and $T \subseteq V$,*

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

Proof. First, we write $|\vec{E}(S,T)|$ as an algebraic expression. Let χ_S and χ_T be the characteristic vectors of S and T. Notice that $|\vec{E}(S,T)| = \chi_S^\top A \chi_T$, where A is the adjacency matrix of G.

Next, we use eigen-decompositions of χ_S and χ_T to relate $|\vec{E}(S,T)|$ to the eigenvalues of A. Let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors of A. Write $\chi_S = \sum_{i=1}^n a_i v_i$ and $\chi_T = \sum_{j=1}^n b_j v_j$, where $a_i = \langle \chi_S, v_i \rangle$ and $b_j = \langle \chi_T, v_j \rangle$. Recall that $\alpha_1 = d$ and $v_1 = \vec{1}/\sqrt{n}$, so $a_1 = |S|/\sqrt{n}$ and $b_1 = |T|/\sqrt{n}$. Then, by orthonormality of $v_1, \ldots, v_n$,

$$|\vec{E}(S,T)| = \chi_S^\top A \chi_T = \Big(\sum_{i=1}^n a_i v_i\Big)^\top A \Big(\sum_{j=1}^n b_j v_j\Big) = \sum_{i=1}^n \alpha_i a_i b_i = \frac{d|S||T|}{n} + \sum_{i=2}^n \alpha_i a_i b_i.$$

Therefore, by the definition of α and the Cauchy-Schwarz inequality,

$$\left| |\vec{E}(S,T)| - \frac{d|S||T|}{n} \right| = \left| \sum_{i=2}^n \alpha_i a_i b_i \right| \leq \alpha \sum_{i=2}^n |a_i||b_i| \leq \alpha \|\vec{a}\|_2 \|\vec{b}\|_2 = \alpha \|\chi_S\|_2 \|\chi_T\|_2 = \alpha \sqrt{|S||T|},$$

where $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$ with $\|\vec{a}\|_2 = \|\chi_S\|_2$ and $\|\vec{b}\|_2 = \|\chi_T\|_2$. □

See Problem 5.16 for a slightly better upper bound for the expander mixing lemma. The same proof can be extended to non-regular graphs using the normalized adjacency matrix.

Exercise 5.4 (Expander Mixing Lemma for Non-Regular Graphs)**.** *Let $G = (V, E)$ be an undirected graph with $|V| = n$. Let $\mathcal{A}$ be its normalized adjacency matrix with eigenvalues $1 = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq -1$ and let $\alpha := \max\{\alpha_2, |\alpha_n|\}$. Prove that, for every $S \subseteq V$ and $T \subseteq V$,*

$$\left| |\vec{E}(S,T)| - \frac{\mathrm{vol}(S) \cdot \mathrm{vol}(T)}{\mathrm{vol}(V)} \right| \leq \alpha \sqrt{\mathrm{vol}(S) \cdot \mathrm{vol}(T)}.$$

The following is an application of the expander mixing lemma.

Exercise 5.5 (Maximum Independent Set and Chromatic Number of Two-Sided Spectral Expanders)**.** *Let $G = (V, E)$ be an (n, d, α) graph. Show that the size of a maximum independent set is at most $\alpha n/d$. Conclude that an (n, d, ϵ)-graph has chromatic number at least d/α.*

5.2 Converse of Expander Mixing Lemma

Interestingly, Bilu and Linial [BL06] proved a converse of the expander mixing lemma, showing that it comes close to providing a combinatorial characterization of two-sided spectral expanders.

Theorem 5.6 (Converse of Expander Mixing Lemma [BL06]). *Let $G = (V = [n], E)$ be a d-regular graph. Suppose that*

$$\left| |\vec{E}(S,T)| - \frac{d|S||T|}{n} \right| \leq \alpha\sqrt{|S||T|} \quad \text{for all } S,T \subseteq V \text{ with } S \cap T = \emptyset. \tag{5.1}$$

Then all but the largest eigenvalue of $A(G)$ are bounded in absolute value by $O\big(\alpha(1+\log\frac{d}{\alpha})\big)$.

Compared to Cheeger's inequality in Theorem 3.2, this provides a tighter relationship between the spectral quantity and the combinatorial property, without a square root loss.

The proof of Theorem 5.6 is based on the following linear algebraic lemma, where the spectral radius of a matrix is defined as the maximum of the absolute values of its eigenvalues.

Lemma 5.7 (Bounding Spectral Radius [BL06]). *Let B be an $n \times n$ real symmetric matrix such that the ℓ_1-norm of each row of B is at most d, and all diagonal entries of B have absolute value $O\big(\alpha(1+\log\frac{d}{\alpha})\big)$ for some $1 \leq \alpha \leq d$. Suppose that*

$$\left|\chi_S^\top B \chi_T\right| \leq \alpha\sqrt{|S||T|} \quad \text{for all } S,T \subseteq [n] \text{ with } S \cap T = \emptyset. \tag{5.2}$$

Then the spectral radius of B is $O\big(\alpha(1+\log\frac{d}{\alpha})\big)$.

We first explain how the lemma implies the converse of the expander mixing lemma.

Proof of Theorem 5.6 Assuming Lemma 5.7. The idea is to apply Lemma 5.7 to the error matrix $B := A(G) - d \cdot v_1 v_1^\top$, where $v_1 = \vec{1}/\sqrt{n}$ is the first eigenvector of $A(G)$. Let the eigenvalues of A be $(d, \alpha_2, \ldots, \alpha_n)$. Then, the eigenvalues of B are $(0, \alpha_2, \ldots, \alpha_n)$, as A and B have the same eigenvectors. Therefore, bounding $\max_{2\leq i\leq n}|\alpha_i|$ is equivalent to bounding the spectral radius of B.

Note that the ℓ_1-norm of each row of B is at most $2d$, the diagonal entries have absolute value at most 1, and that the assumption in (5.1) and the assumption in (5.2) are equivalent. Therefore, we can apply Lemma 5.7 to obtain that the spectral radius of B is $O\big(\alpha(1+\log\frac{d}{\alpha})\big)$, and this implies that $\max_{2\leq i\leq n}|\alpha_i| \lesssim \alpha(1+\log\frac{d}{\alpha})$. □

The proof of Lemma 5.7 in [BL06] is by combining the constraints in (5.2) to establish that

$$\frac{|x^\top B x|}{x^\top x} \lesssim \alpha\left(1+\log\frac{d}{\alpha}\right) \quad \text{for all } x \in \mathbb{R}^n. \tag{5.3}$$

The combination of the constraints is guided by linear programming duality, and as a result the proof is not quite intuitive and it is not clear how the numbers were chosen.

We provide a different presentation of their proof following Trevisan's style. We prove the contrapositive that if (5.3) is violated then there must be a violating constraint in (5.2). To do so, we use a simple randomized rounding argument as in the proof of the hard direction of Cheeger's inequality in Section 3.3. We hope that this argument clarifies how the numbers in the proof are chosen.

Proof of a Weaker Version

To highlight the main idea, we first prove a weaker version of Lemma 5.7. We then explain the modifications needed to match Lemma 5.7 in the next subsection.

Contrapositive: In this subsection, our goal is to prove the weaker statement that

$$\exists x \in \mathbb{R}^n \text{ with } \frac{|x^\top B x|}{\|x\|_2^2} > 2(\log_2 n + 1)\alpha \quad \Longrightarrow \quad \exists y, z \in \{-1, 0, 1\}^n \text{ with } \frac{|y^\top B z|}{\|y\|_2 \|z\|_2} > \alpha.$$

Assumption: We assume that $\|x\|_2 = 1$ and each entry of x is a negative power of two. This is a natural first step to discretize x, and we will justify this assumption in the next subsection.

Notations: Let $S_k := \{i \in [n] \mid |x(i)| = 2^{-k}\}$ be the set of indices with absolute value 2^{-k}, and let $s_k := |S_k|$. Let χ_k be the signed pattern of x restricted to S_k, such that $\chi_k(i) = 1$ if $x(i) = 2^{-k}$, $\chi_k(i) = -1$ if $x(i) = -2^k$, and $\chi_k(i) = 0$ otherwise. Note that $x = \sum_k 2^{-k}\chi_k$.

Probability Distribution: We sample y and z independently from the same distribution, where

$$y, z := \chi_k \text{ with probability } 2^{-k}/c \text{ if } S_k \neq \emptyset,$$

and $c := \sum_{k:S_k \neq \emptyset} 2^{-k}$ is the normalizing constant to make this a probability distribution.

Probabilistic Arugment: As in the proof of Cheeger's inequality, we argue by Lemma 3.6 that there exist $y, z \in \{-1, 0, 1\}^n$ with

$$\frac{|y^\top B z|}{\|y\|_2 \cdot \|z\|_2} \geq \frac{\mathbb{E}\left[|y^\top B z|\right]}{\mathbb{E}\left[\|y\|_2 \cdot \|z\|_2\right]},$$

and so it remains to compute the expected values separately.

Expected Numerator: By the triangle inequality,

$$c^2 \cdot \mathbb{E}\left[|y^\top B z|\right] = \sum_k \sum_l 2^{-k} \cdot 2^{-l} \cdot |\chi_k^\top B \chi_l| \geq \left|\left(\sum_k 2^{-k}\chi_k\right)^\top B \left(\sum_l 2^{-l}\chi_l\right)\right| = |x^\top B x|. \quad (5.4)$$

This is the main motivation for defining the probability distribution in this way, so that the expected numerator can be easily compared to the numerator $|x^\top B x|$.

Expected Denominator: By independence,

$$c^2 \cdot \mathbb{E}\left[\|y\|_2 \cdot \|z\|_2\right] = c^2 \cdot \mathbb{E}\left[\|y\|_2\right]^2 = \left(\sum_k 2^{-k}\|\chi_k\|_2\right)^2 = \sum_k \sum_l 2^{-k} \cdot 2^{-l} \cdot \sqrt{s_k s_l}. \quad (5.5)$$

We would like to compare this to the denominator

$$x^\top x = \left(\sum_k 2^{-k}\chi_k\right)^\top \left(\sum_l 2^{-l}\chi_l\right) = \sum_k 2^{-2k} s_k. \quad (5.6)$$

To do so, we divide the right hand side of (5.5) into three terms:

$$\sum_k 2^{-2k} s_k \; + \; \sum_k \sum_{l:k<l\leq k+\log n} 2^{-k-l+1}\sqrt{s_k s_l} \; + \; \sum_k \sum_{l:l>k+\log n} 2^{-k-l+1}\sqrt{s_k s_l}.$$

Using the AM-GM inequality, the second term is bounded by

$$\sum_k \sum_{l:k<l\le k+\log n} 2^{-k-l+1}\sqrt{s_k s_l} \le \sum_k \sum_{l:k<l\le k+\log n} (2^{-2k}s_k + 2^{-2l}s_l) \le 2\log n \cdot \sum_k 2^{-2k}s_k. \tag{5.7}$$

The third term is bounded by

$$\sum_k \sum_{l:l>k+\log n} 2^{-2k-\log n}\sqrt{s_k s_l} = \frac{1}{n}\sum_k 2^{-2k}\sqrt{s_k}\sum_l \sqrt{s_l} \le \sum_k 2^{-2k}\sqrt{s_k} \le \sum_k 2^{-2k}s_k,$$

where the second last inequality uses $\sum_l \sqrt{s_l} \le \sqrt{n}\sqrt{\sum_l s_l} = n$ by Cauchy-Schwarz and $\sum_l s_l \le n$. Combining these inequalities,

$$c^2 \cdot \mathbb{E}\left[\|y\|_2 \cdot \|z\|_2\right] \le 2(\log n + 1)\sum_k 2^{-2k}s_k = 2(\log n+1)\cdot x^\top x.$$

Conclusion: Therefore, there exist $y, z \in \{-1,0,1\}^n$ such that

$$\frac{|y^\top Bz|}{\|y\|_2\|z\|_2} \ge \frac{\mathbb{E}\left[|y^\top Bz|\right]}{\mathbb{E}\left[\|y\|_2\cdot\|z\|_2\right]} \ge \frac{|x^\top Bx|}{2(\log n+1)\cdot\|x\|_2^2} > \alpha.$$

We would like to reduce the $\log n$ factor to $\alpha\log(d/\alpha)$. Observe that we did not use the assumption about the ℓ_1-norm of the rows in this proof. To exploit this assumption, we will modify the probability distribution to sample y and z.

Proof of Lemma 5.7

The proof has a similar structure to that in the previous subsection. We explain the modifications and the missing details here.

Zero Diagonal Entries: We assume that the diagonal entries of B are zero. See Exercise 5.17.

Contrapositive: To prove Lemma 5.7, we prove the contrapositive that

$$\exists x\in\mathbb{R}^n \text{ with } \frac{|x^\top Bx|}{\|x\|_2^2} > \alpha\Big(\log\frac{d}{\alpha}+1\Big) \quad\Longrightarrow\quad \exists y,z\in\{0,1\}^n, \langle y,z\rangle = 0 \text{ with } \frac{|y^\top Bz|}{\|y\|_2\|z\|_2} \gtrsim \alpha.$$

Negative Powers of Two: We rescale x to satisfy $\|x\|_2 = 1$. Then, using a simple rounding argument in Problem 5.18, we construct a vector $\tilde{x}\in\mathbb{R}^n$ such that each entry of $\tilde{x}$ is a negative power of two and

$$\frac{|\tilde{x}^\top B\tilde{x}|}{\|\tilde{x}\|_2^2} \ge \frac{1}{4}\cdot\frac{|x^\top Bx|}{\|x\|_2^2}.$$

We let $x := \tilde{x}$ in the following. We use the same notations $S_k \subseteq [n]$ and $\chi_k \in \{-1,0,1\}^n$ as before.

Probability Distribution: We sample $y, z \in \{-1,0,1\}^n$ jointly (not independently) where

$$(y,z) = (\chi_k,\chi_l) \text{ with probability } 2^{-k-l}/c \text{ if } S_k\neq\emptyset, S_l\neq\emptyset, \text{ and } |k-l|\le\gamma,$$

where c is the normalization constant that makes this a probability distribution, and γ is a parameter that we will choose to be $\log(d/\alpha)$.

This is the main modification, where pairs (χ_k,χ_l) with $|k-l|>\gamma$ are not sampled. The motivation is to avoid the $\log n$ loss in the denominator in the previous analysis. However, this will make the numerator smaller, and the choice of γ is to balance the denominator and the numerator.

Expected Numerator:

$$c \cdot \mathbb{E}\left[\left|y^\top B z\right|\right] = \sum_{k,l:|k-l|\le\gamma} 2^{-k-l} \cdot |\chi_k^\top B \chi_l| \ge |x^\top B x| - \sum_{k,l:|k-l|>\gamma} 2^{-k-l} |\chi_k^\top B \chi_l|,$$

where the inequality is from the previous analysis (5.4) that $\sum_{k,l} 2^{-k-l}|\chi_k^\top B\chi_l| \ge |x^\top Bx|$.

The key observation is that the second term can be bounded using the ℓ_1-norm assumption:

$$\sum_{k,l:|k-l|>\gamma} 2^{-k-l}|\chi_k^\top B\chi_l| = \sum_{k,l:l>k+\gamma} 2^{-k-l+1}|\chi_k^\top B\chi_l| \le 2^{-\gamma}\sum_k 2^{-2k} \sum_{l>k+\gamma} |\chi_k^\top B\chi_l| \le 2^{-\gamma} d \sum_k 2^{-2k} s_k,$$

where the last inequality holds as $\sum_l |\chi_k^\top B\chi_l| \le d \cdot s_k$ which follows from the ℓ_1-norm assumption. Therefore, by (5.6),

$$c \cdot \mathbb{E}\left[\left|y^\top Bz\right|\right] \ge |x^\top Bx| - 2^{-\gamma} d \sum_k 2^{-2k} s_k = |x^\top Bx| - 2^{-\gamma} d \cdot x^\top x.$$

Expected Denominator: Recall that $\|\chi_k\|_2 = \sqrt{s_k}$, so

$$c \cdot \mathbb{E}\left[\|y\|_2 \cdot \|z\|_2\right] = \sum_{k,l:|k-l|\le\gamma} 2^{-k-l}\sqrt{s_k s_l} = \sum_k 2^{-2k} s_k + \sum_k \sum_{l:k<l\le k+\gamma} 2^{-k-l+1}\sqrt{s_k s_l}.$$

Using the same calculation in (5.7), we obtain that

$$c \cdot \mathbb{E}\left[\|y\|_2 \cdot \|z\|_2\right] \le \sum_k 2^{-2k} s_k + 2\gamma \sum_k 2^{-2k} s_k = (2\gamma+1) x^\top x.$$

Good $\{-1,0,1\}^n$ Vectors: By Lemma 3.6, there exist $y, z \in \{-1,0,1\}^n$ such that

$$\frac{|y^\top Bz|}{\|y\|_2\|z\|_2} \ge \frac{\mathbb{E}\left[\left|y^\top Bz\right|\right]}{\mathbb{E}\left[\|y\|_2 \cdot \|z\|_2\right]} \ge \frac{|x^\top Bx| - 2^{-\gamma} d \cdot x^\top x}{(2\gamma+1)x^\top x} = \underbrace{\frac{|x^\top Bx|}{(2\gamma+1)x^\top x}}_{(*)} - \underbrace{\frac{2^{-\gamma} d}{2\gamma+1}}_{(**)} \gtrsim \alpha,$$

where the last inequality holds by choosing $\gamma = \log(d/\alpha)$ and the assumption in the contrapositive.

To see how to choose γ and set the assumption in the contrapositive, the idea is to ensure the $(*)$ is $\Omega(\alpha)$ and $(**)$ is $o(\alpha)$. Setting $\gamma = \log_2(d/\alpha)$ ensures that the second term is $o(\alpha)$, and setting $|x^\top Bx|/x^\top x \gtrsim \gamma \cdot \alpha$ in the assumption of the contrapositive ensures that the first term is $\Omega(\alpha)$.

Good $\{0,1\}^n$ Vectors: Write $y = y^+ - y^-$ and $z = z^+ - z^-$, where $y^+, y^-, z^+, z^- \in \{0,1\}^n$. Let $(\bar{y}, \bar{z})$ be one of the four options of $(y^\pm, z^\pm)$ that maximizes the absolute value of the quadratic form. Check that

$$\frac{|\bar{y}^\top B\bar{z}|}{\|\bar{y}\|_2\|\bar{z}\|_2} \ge \frac{1}{4} \cdot \frac{|y^\top Bz|}{\|y\|_2\|z\|_2}.$$

Disjoint Supports: If $\mathrm{supp}(\bar{y}) \cap \mathrm{supp}(\bar{z}) = \emptyset$, we are done. Otherwise, by our construction, $\mathrm{supp}(\bar{y}) = \mathrm{supp}(\bar{z})$. Let $Y := \{i \in [n] \mid \bar{y}(i) = 1\}$. Consider a random partition (S,T) of Y where each $i \in Y$ is put in S with probability $1/2$ and in T with probability $1/2$ independently. Use the assumption that the diagonal entries of B are zero to argue that there exists a partition $S \cup T = Y$ with

$$\frac{|\chi_S^\top B\chi_T|}{\|\chi_S\|_2\|\chi_T\|_2} \ge \frac{1}{2} \cdot \frac{|\bar{y}^\top B\bar{y}|}{\|\bar{y}\|_2^2}.$$

Conclusion: The proof of Lemma 5.7 follows by chaining together the inequalities:

$$\frac{|\chi_S^\top B \chi_T|}{\|\chi_S\|_2 \|\chi_T\|_2} \gtrsim \frac{|\bar{y}^\top B \bar{y}|}{\|\bar{y}\|_2^2} \gtrsim \frac{|y^\top B z|}{\|y\|_2 \|z\|_2} \geq \frac{|\tilde{x}^\top B \tilde{x}| - 2^{-\gamma} d \cdot \tilde{x}^\top \tilde{x}}{(2\gamma+1)\tilde{x}^\top \tilde{x}} \gtrsim \alpha,$$

where the last inequality is by $|\tilde{x}^\top B \tilde{x}|/\|\tilde{x}\|_2^2 \gtrsim |x^\top B x|/\|x\|_2^2 > \alpha(\log(d/\alpha)+1)$.

Tightness: Bilu and Linial [BL06] proved that Theorem 5.6 is tight that there are graphs satisfying the conditions but have spectral radius $\Omega(\alpha(\log(d/\alpha)+1)$.

5.3 Graphs with Large Spectral Gap

How large can the spectral gap be? Or, equivalently, how small can α be in an (n,d,α)-graph?

In this section, we present matching lower and upper bounds to this question, and discuss some strong properties of graphs with large spectral gap.

Lower Bounds

We begin with a simple proof that $\alpha \gtrsim \sqrt{d}$ using a trace argument.

Claim 5.8 (Easy Lower Bound for α)**.** *Let G be an (n,d,α)-graph. Then*

$$\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}.$$

Proof. Let A be the adjacency matrix of G with eigenvalues $\alpha_1 \geq \ldots \geq \alpha_n$. By Fact A.35,

$$\mathrm{Tr}(A^2) = \sum_{i=1}^n \alpha_i^2 \leq d^2 + (n-1)\alpha^2.$$

On the other hand, $\mathrm{Tr}(A^2) \geq nd$, as each edge uv contributes a length-two walk from u to u and a length-two walk from v to v. Combining the two inequalities establishes the claim. □

A higher-order trace argument can be used to prove a lower bound close to $2\sqrt{d-1}$.

Theorem 5.9 (Trace Lower Bound for α)**.** *Let G be an (n,d,α)-graph. Then*

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

Proof. Let A be the adjacency matrix of G with eigenvalues $\alpha_1 \geq \ldots \geq \alpha_n$. For any $k \in \mathbb{N}$,

$$\mathrm{Tr}(A^{2k}) = \sum_{i=1}^n \alpha_i^{2k} \leq d^{2k} + (n-1)\alpha^{2k}.$$

On the other hand, recall from Lemma 2.5 that $\mathrm{Tr}(A^{2k})$ is equal to the number of length-$2k$ walks in G. For each vertex v, the number of length-$2k$ walks from v to v is at least the number of such walks in an infinite d-regular tree. In an infinite d-regular tree, the number of such walks is at least $C_k \cdot (d-1)^k$, where C_k is the k-th Catalan number. This is because each such walk has k forward

steps and k backward steps, where every prefix of a walk has at least as many forward steps as backward steps, and there are at least $d-1$ options for each forward step. Thus,

$$\mathrm{Tr}(A^{2k}) \geq n \cdot C_k \cdot (d-1)^k = \frac{n}{k+1}\binom{2k}{k}(d-1)^k.$$

Combining the inequalities and using an estimate of Catalan numbers,

$$\alpha^{2k} \geq \frac{1}{k+1}\binom{2k}{k}(d-1)^k - \frac{d^{2k}}{n} \geq \frac{4^k(d-1)^k}{2(k+1)^{3/2}} - \frac{d^{2k}}{n}.$$

Therefore, by choosing $k \ll \log n / \log d$ so that $d^{2k}/n \ll 1$, but letting k goes to infinity as n grows,

$$\alpha \geq 2\sqrt{d-1}\Big(1 - \frac{O(\log k)}{k}\Big) - o_n(1) \geq 2\sqrt{d-1} - o_n(1).$$

□

The following well-known result by Alon and Boppana provides a tight lower bound on the second eigenvalue of the adjacency matrix of a d-regular graph.

Theorem 5.10 (Alon-Boppana Bound [Nil91])**.** *Let $G = (V, E)$ be a d-regular graph and α_2 be the second largest eigenvalue of its adjacency matrix. Then*

$$\alpha_2 \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{\lfloor \mathrm{diag}(G)/2 \rfloor},$$

where $\mathrm{diag}(G)$ *denotes the diameter of the graph G.*

The theorem implies that if we have an infinite family of d-regular graphs each has second eigenvalue at most α_2, then $\alpha_2 \geq 2\sqrt{d-1}$ as the diameter grows to infinity with the graph size.

The proof is by constructing a vector $x \perp \vec{1}$ with Rayleigh quotient $x^\top Ax / x^\top x \approx 2\sqrt{d-1}$. The vector x is similar to the first eigenvector of a d-regular tree (see Problem 2.22). We will not prove Theorem 5.10 and refer readers to [HLW06, Section 5.2] or Trevisan's exposition [Tre08].

Upper Bounds

A major discovery is that graphs with $\alpha \leq 2\sqrt{d-1}$ exist.

Theorem 5.11 (Lubotzky, Phillip, Sarnak [LPS88], Margulis [Mar88])**.** *For every prime p and every positive integer k, there exist infinitely many (n, d, α)-graphs with $\alpha \leq 2\sqrt{d-1}$ and $d = p^k + 1$.*

The graphs constructed in [LPS88] are the Cayley graph of certain groups. The proof relies on some deep results in number theory, specifically on proven cases of conjectures by Ramanujan, which is well beyond the scope of this course. That is the reason why they called graphs with $\alpha \leq 2\sqrt{d-1}$ "Ramanujan graphs".

This naturally leads to the question of whether there are combinatorial or probabilistic constructions of Ramanujan graphs. The simplest probabilistic construction is to generate a random d-regular graph. A famous result by Friedman shows that most d-regular graphs are nearly-Ramanujan.

Theorem 5.12 (Friedman [Fri08]). *Let G be a random d-regular graph on n vertices and let $\alpha := \max\{\alpha_2, |\alpha_n|\}$ where α_2 and α_n are the second and last eigenvalues of the adjacency matrix of G. Then, for every $\epsilon > 0$,*

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

It has been a long standing open question whether most d-regular graphs are Ramanujan. Recent progress has significantly advanced our understanding of this problem, including simpler proofs, new approaches, and sharper bounds. A remarkable new paper by Huang, McKenzie and Yau [HMY24] proves that a random d-regular graph is Ramanujan with a probability of approximately 69%. The proofs of these results are also well beyond the scope of this course.

It is still an open problem to design deterministic combinatorial constructions of Ramanujan graphs, although there are breakthroughs in such constructions for bipartite Ramanujan graphs [MSS15, MSS18] using the method of interlacing polynomials.

Properties

What additional properties do Ramanujan graphs possess that typical expander graphs do not? Typical d-regular expander graphs $G = ([n], E)$ satisfy the following properties:

1. Algebraic: $\lambda_2(\mathcal{L}(G)) = \Omega(1)$ where $\mathcal{L}$ is the normalized Laplacian matrix;
2. Combinatorial: $\phi(G) = \Omega(1)$ where $\phi(G)$ is the edge conductance of G;
3. Probabilistic: $\tau(G) = O(\log n)$ where $\tau(G)$ is the mixing time of lazy random walks on G.

Note that the mixing time bound is optimal when d is a constant as the graph has diameter $\Omega(\log n)$, but it does not improve even if we assume d is large.

Definition 5.13 (Graphs with Large Spectral Gap). *We say a (n, d, α)-graph G has a large spectral gap if $\alpha = O(d^{1-c})$ for some constant $0 < c \leq 1/2$, and say G is near-Ramanujan if $c = 1/2$.*

For near-Ramanujan graphs, the lower bound on edge conductance is $\frac{1}{2}(1 - O(\frac{1}{\sqrt{d}}))$. This is only slightly stronger than that of typical expander graphs, and does not quantify the additional expansion properties that near-Ramanujan graph possess. The right combinatorial parameter to measure the additional expansion properties of near-Ramanujan graphs is the small-set vertex expansion.

A d-regular graph $G = ([n], E)$ with a large spectral gap satisfies the following properties:

1. Algebraic: $\lambda_2(\mathcal{L}(G)) = 1 - d^{-c}$ where $\mathcal{L}$ is the normalized Laplacian matrix;
2. Combinatorial: $\psi(S) \gtrsim d^{2c}$ if $|S| \lesssim n/d^{2c}$, where $\psi(S)$ is the vertex expansion of a set S;
3. Probabilistic: $\tau(G) \lesssim \log n / \log d^c$, where $\tau(G)$ is the mixing time of random walks on G.

These are significant upgrades of the combinatorial and probabilistic properties over that of typical expander graphs: The second property implies that a near-Ramanujan graph has near-perfect small-set vertex expansion. The third property implies that a graph with a large spectral gap has constant mixing time when $d = n^\epsilon$ for some constant $\epsilon > 0$. We will explore these in the next sections.

5.4 Small-Set Vertex Expansion

For an (n, d, α)-graph, Tanner's theorem shows that sets of size up to $\alpha^2 n/d^2$ have vertex expansion d^2/α^2. Note that the bound becomes interesting when $\alpha \ll d$. In the extreme case of a near-Ramanujan graph, sets of size up to $\Omega(n/d)$ have near-perfect vertex expansion of $\Omega(d)$.

Theorem 5.14 (Tanner's Theorem)**.** *Let $G = (V, E)$ be an (n, d, α)-graph. For any $0 < \delta \leq 1/2$ and any subset $S \subseteq V$ with $|S| = \delta n$,*

$$\psi(S) \geq \left(\delta\Big(1 - \frac{\alpha^2}{d^2}\Big) + \frac{\alpha^2}{d^2}\right)^{-1} - 1.$$

Proof. The idea is to consider the quantity $\|A\chi_S\|_2^2$, where A is the adjacency matrix and χ_S is the characteristic vector of $S \subseteq V$. This quantity is bounded in two ways. One way is to show that if the closed vertex boundary $|\partial[S]|$ is small, then $\|A\chi_S\|_2^2$ is large, where $\partial[S] := \partial(S) \cup S$. The other way is to upper bound $\|A\chi_S\|_2^2$ using the spectral property, as $\|Ax\|_2^2 \leq \alpha^2\|x\|_2^2$ for $x \perp \vec{1}$.

For a vertex $v \in V$, let $\deg_S(v) := |\{u \in S \mid uv \in E\}|$ be the number of neighbors of v in S. Then,

$$\|A\chi_S\|_2^2 = \sum_{v \in V} \deg_S(v)^2 = \sum_{v \in \partial[S]} \deg_S(v)^2 \geq \frac{\big(\sum_{v \in \partial[S]} \deg_S(v)\big)^2}{|\partial[S]|} = \frac{\big(d|S|\big)^2}{|\partial[S]|},$$

where the inequality follows from Cauchy-Schwarz. This proves the lower bound.

For the upper bound, we write $\chi_S = \sum_{i=1}^n c_i v_i$ as a linear combination of the orthonormal eigenvectors of A, with $v_1 = \vec{1}/\sqrt{n}$ and $c_1 = \langle \chi_S, v_1 \rangle = |S|/\sqrt{n}$. Then

$$\|A\chi_S\|_2^2 = \Big\|\sum_{i=1}^n c_i \alpha_i v_i\Big\|_2^2 = \sum_{i=1}^n c_i^2 \alpha_i^2 \leq \frac{d^2|S|^2}{n} + \alpha^2\big(\|\chi_S\|^2 - c_1^2\big) = d^2\delta|S| + \alpha^2|S| - \alpha^2\delta|S|,$$

Combining the inequalities yields

$$\psi(S) + 1 = \frac{|\partial[S]|}{|S|} \geq \frac{d^2}{\delta(d^2 - \alpha^2) + \alpha^2} = \left(\delta\Big(1 - \frac{\alpha^2}{d^2}\Big) + \frac{\alpha^2}{d^2}\right)^{-1}.$$

□

For Ramanujan graphs, Tanner's theorem show that sets of size up to $n/(Cd)$ for a large constant C have vertex expansion close to $d/4$.

Kahale [Kah95] improved Tanner's theorem and showed that small linear-sized subsets in Ramanujan graphs have vertex expansion close to $d/2$. The same paper provided an example where $\alpha \leq 2\sqrt{d-1} + o(1)$, but the graph has a small set with vertex expansion at most $d/2$, proving that the $d/2$ bound is tight.

For some applications in constructing error correcting codes, graphs with small-set vertex expansion strictly greater than $d/2$ are required. We will revisit this question in the following chapters.

5.5 Random Walks on Expander Graphs

Mixing Time

We usually consider random walks on (n, d, α)-graphs when $d = \Theta(1)$. In this regime, as shown in Theorem 4.15, random walks converge to the uniform distribution in $O(\log n)$ steps as long as $\alpha \leq (1-\epsilon)d$ for some constant $\epsilon > 0$. This bound is optimal because a d-regular graph with $d = \Theta(1)$ has diameter $\Omega(\log n)$, so even if the graph is Ramanujan the mixing time remains $\Omega(\log n)$.

Now, consider the regime when $d = n^{\epsilon}$ for some constant $\epsilon > 0$, so that the diameter of the graph could be a constant. For typical expander graphs with $\alpha = \Theta(d)$, there still exist (n, d, α)-graphs that have mixing times of $\Omega(\log n)$. In contrast, for graphs with a large spectral gap, where $\alpha = O(d^{1-c})$ for some constant $c > 0$, every (n, d, α)-graph has constant mixing time.

The verification of these claims is left as an exercise in Problem 5.20. This demonstrates that graphs with large spectral gaps, as defined in Definition 5.13, exhibit significantly stronger randomness properties compared to typical expander graphs.

Concentration Property

Interestingly, random walks on expander graphs not only provide strong randomness properties for the final vertex in the walk, but also for the sequence of vertices traversed during the walk. In some applications, the sequence of vertices in a walk can effectively replace a sequence of independent uniform random variables.

The following result is not presented in its most general form, but will suffice for the application of probability amplification that we will see in Chapter **??**. For more general statements, the reader is referred to [HLW06, Vad12]. To develop intuition, it is useful to compare the probability bound stated below with the corresponding bound when each X_i is an independent uniform random sample.

Theorem 5.15 (Concentration Property of Random Walks on Two-Sided Spectral Expanders)**.** *Let $G = (V, E)$ be a (n, d, α)-graph with $\alpha \leq d/10$. Let $B \subseteq V$ with $|B| \leq \frac{1}{100}|V|$. Let X_0 be a uniform random vertex, and let $X_1, \ldots, X_t$ be the vertices produced by t steps of a random walk. Let $S = \{i \mid X_i \in B\}$ be the set of times when the random walk is in B. Then,*

$$\Pr\left(|S| > \frac{t}{2}\right) \leq \left(\frac{2}{\sqrt{5}}\right)^{t+1}.$$

Proof. We first set up the matrix formulation of the problem. The initial distribution of X_0 is $\vec{p}_0 = \vec{1}/n$. Let I_B be the diagonal matrix with a 1 in the i-th diagonal entry if $i \in B$ and zero otherwise, and similarly define $I_{\overline{B}}$ for $\overline{B} = V - B$. For a probability vector $\vec{p}$, $I_B \cdot \vec{p}$ restricts $\vec{p}$ to B. The probability that the random walk is in B precisely at the times in S is

$$p_S := \vec{1}^{\top}(I_{Z_t}\mathcal{A})(I_{Z_{t-1}}\mathcal{A})(I_{Z_{t-2}}\mathcal{A})\ldots(I_{Z_2}\mathcal{A})(I_{Z_1}\mathcal{A})\vec{p}_0,$$

where $Z_i = B$ if $i \in S$ and $Z_i = \overline{B}$ otherwise, and $\mathcal{A}$ is the normalized adjacency matrix (the transition matrix of the random walk).

We will prove that $p_S \leq (\frac{1}{5})^{|S|}$. The theorem will then follow by a union bound as

$$\Pr\left(|S| > \frac{t}{2}\right) \leq \sum_{S:|S|>t/2} p_S \leq \sum_{S:|S|>t/2} \left(\frac{1}{5}\right)^{|S|} \leq \sum_{S:|S|>t/2} \left(\frac{1}{5}\right)^{\frac{t+1}{2}} \leq 2^{t+1}\left(\frac{1}{5}\right)^{\frac{t+1}{2}} = \left(\frac{2}{\sqrt{5}}\right)^{t+1}.$$

To prove $p_S \le (\frac{1}{5})^{|S|}$, we use the operator norm $\|\cdot\|_{\text{op}}$ (see Definition A.19). Note that $\|I_B\|_{\text{op}} = \|I_{\overline{B}}\|_{\text{op}} = \|\mathcal{A}\|_{\text{op}} = 1$. We will prove $\|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5}$, which implies $p_S \le (\frac{1}{5})^{|S|}$ as follows:

$$\begin{aligned}
p_S &= \vec{1}^\top (I_{Z_t}\mathcal{A}) \dots (I_{Z_1}\mathcal{A})\vec{p}_0 \\
&\le \|\vec{1}\|_2 \cdot \big\|(I_{Z_t}\mathcal{A}) \dots (I_{Z_1}\mathcal{A})\vec{p}_0\big\|_2 && \text{(Cauchy-Schwarz)} \\
&\le \|\vec{1}\|_2 \cdot \Big(\prod_{i=1}^{t} \|I_{Z_i}\mathcal{A}\|_{\text{op}}\Big) \cdot \|\vec{p}_0\|_2 && \text{(operator norm properties in Fact A.21)} \\
&\le \|\vec{1}\|_2 \cdot \Big(\frac{1}{5}\Big)^{|S|} \cdot \|\vec{p}_0\|_2 && \text{(from } \|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5} \text{ and } \|I_{\overline{B}}\mathcal{A}\|_{\text{op}} \le 1) \\
&= \Big(\frac{1}{5}\Big)^{|S|}. && \text{(since } \|\vec{1}\|_2 = \sqrt{n} \text{ and } \|\vec{p}_0\|_2 = \frac{1}{\sqrt{n}})
\end{aligned}$$

It remains to prove that $\|I_B\mathcal{A}\|_{\text{op}} \le \frac{1}{5}$, which is equivalent to proving that $\|I_B\mathcal{A}x\|_2^2 \le \|x\|_2^2/25$ for any nonzero vector x. Write $x = c_1v_1 + \ldots + c_nv_n$, where $v_1, \ldots, v_n$ are the orthonormal eigenvectors of $\mathcal{A}$ with eigenvalues $1 = \alpha_1 \ge \ldots \ge \alpha_n \ge -1$. Since G is an (n, d, α)-graph, we have $\max_{2\le i\le n}\{|\alpha_i|\} \le \alpha/d$. It is then natural to decompose $\|I_B\mathcal{A}x\|_2^2$ into two terms:

$$\|I_B\mathcal{A}x\|_2^2 = \|I_B\mathcal{A}(c_1v_1 + \ldots + c_nv_n)\|_2^2 = \Big\|I_B\sum_{i=1}^{n} c_i\alpha_iv_i\Big\|_2^2 \le 2\|I_Bc_1\alpha_1v_1\|_2^2 + 2\Big\|I_B\sum_{i=2}^{n} c_i\alpha_iv_i\Big\|_2^2.$$

Using $c_1 = \langle x, v_1\rangle = \langle x, \frac{\vec{1}}{\sqrt{n}}\rangle = \frac{1}{\sqrt{n}} \cdot \sum_{i=1}^{n} x(i)$ and $|B| \le \frac{n}{100}$, the first term is

$$2\big\|I_Bc_1\alpha_1v_1\big\|_2^2 = 2\Big\|\frac{1}{n}\Big(\sum_{i=1}^{n} x(i)\Big)I_B\vec{1}\Big\|_2^2 = 2|B|\Big(\frac{\sum_{i=1}^{n} x(i)}{n}\Big)^2 \le 2|B| \cdot \frac{\|x\|_2^2}{n} \le \frac{1}{50}\|x\|_2^2,$$

where the first inequality is by Cauchy-Schwarz. Using $\alpha/d \le 1/10$, the second term is

$$2\Big\|I_B\sum_{i=2}^{n} c_i\alpha_iv_i\Big\|_2^2 \le 2\|I_B\|_{\text{op}}^2 \cdot \Big\|\sum_{i=2}^{n} c_i\alpha_iv_i\Big\|_2^2 = 2\sum_{i=2}^{n} c_i^2\alpha_i^2 \le 2\Big(\frac{\alpha}{d}\Big)^2\sum_{i=2}^{n} c_i^2 \le 2\Big(\frac{\alpha}{d}\Big)^2\|x\|_2^2 \le \frac{1}{50}\|x\|_2^2,$$

Adding the two terms finishes the proof. □

5.6 Problems

Problem 5.16 (Tighter Expander Mixing Lemma). *Let $G = (V, E)$ be an (n, d, α)-graph. Prove that, for every $S \subseteq V$ and $T \subseteq V$,*

$$\Big||E(S,T)| - \frac{d|S||T|}{n}\Big| \le \alpha\sqrt{|S||T|\Big(1 - \frac{|S|}{n}\Big)\Big(1 - \frac{|T|}{n}\Big)}.$$

Obtain the corresponding improvement in the non-regular case.

Exercise 5.17 (Zero Diagonal Entries). *Argue that if Lemma 5.7 is true for matrices with zero diagonal entries, then Lemma 5.7 is true.*

Problem 5.18 (Powers of Two). *Show that given any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$, there is a vector $\tilde{x} \in \mathbb{R}^n$ such that each entry of $\tilde{x}$ is a negative power of two and*

$$\frac{|\tilde{x}^\top B \tilde{x}|}{\|\tilde{x}\|_2^2} \geq \frac{1}{4} \cdot \frac{|x^\top B x|}{\|x\|_2^2}.$$

Hint: Design a rounding such that $\mathbb{E}[\tilde{x}(i)] = x(i)$, and use the assumption that the diagonal entries of B are zero to argue that $\mathbb{E}\left[\tilde{x}^\top B \tilde{x}\right] = x^\top B x$.

Problem 5.19 (Small-Set Vertex Expansion from Expander Mixing Lemma). *Use the expander mixing lemma in Theorem 5.3 to prove that an (n, d, α)-graph with a large spectral gap (as defined in Definition 5.13) has good small-set vertex expansion. Compare your bound to Tanner's bound in Theorem 5.14.*

Problem 5.20 (Mixing Time of Graphs with Large Spectral Gap). *Let G be an (n, d, α)-graph with $d = n^\epsilon$ for some constant $\epsilon > 0$.*

(a) Give an example with $\alpha = c \cdot d$ for some constant $0 < c < 1$ and mixing time $\Omega(\log n)$.

(b) Show that the mixing time is $O(1)$ when $\alpha = d^{1-c}$ for a constant $0 < c \leq 1/2$.

References

[AC88] Noga Alon and Fan R. K. Chung. Explicit construction of linear sized tolerant networks. *Discret. Math.*, 72(1-3):15–19, 1988. 48

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Comb.*, 26(5):495–519, 2006. 49, 53, 66

[Fri08] Joel Friedman. *A Proof of Alon's Second Eigenvalue Conjecture and Related Problems*. American Mathematical Society, 2008. 55, 61

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 3, 19, 20, 54, 57, 61, 62, 66, 69, 70, 72

[HMY24] Jiaoyang Huang, Theo Mckenzie, and Horng-Tzer Yau. Ramanujan property and edge universality of random regular graphs. *arXiv preprint arXiv:2412.20263*, 2024. 55, 61

[Kah95] Nabil Kahalé. Eigenvalues and expansion of regular graphs. *J. ACM*, 42(5):1091–1106, 1995. 56

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. 54, 62

[Mar88] Gregory A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988. 54, 62

[MSS15] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *Annals of Mathematics*, 182:307–325, 2015. 55, 66

[MSS18] Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families IV: bipartite ramanujan graphs of all sizes. *SIAM J. Comput.*, 47(6):2488–2509, 2018. 55

[Nil91] Alon Nilli. On the second eigenvalue of a graph. *Discret. Math.*, 91(2):207–210, 1991. 54

[Tre08] Luca Trevisan. The Alon-Boppana theorem. *Blog post: https://lucatrevisan.wordpress.com/2014/09/01/the-alon-boppana-theorem-2/*, 2008. 54

[Vad12] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. 57, 70, 72

Chapter 6

Expander Graphs: Constructions

In this chapter, we explore various constructions of expander graphs. We begin with a discussion of probabilistic and algebraic constructions, followed by an analysis of a combinatorial construction known as the zig-zag product, along with an overview of other combinatorial constructions. The content of this chapter is mostly based on [HLW06].

6.1 Probabilistic Constructions

Constructing expander graphs is generally considered a challenging task. Ironically, almost all graphs are expander graphs and even near-Ramanujan graphs!

For combinatorial properties, the probabilistic method can be used to prove that a random d-regular graph has constant edge conductance with probability tending to 1 as the graph size goes to infinity. Furthermore, it can be proved the vertex expansion of linear-sized subsets is close to $d-2$ [HLW06, Section 4.6], which goes beyond the $d/2$ bound achieved by Ramanujan graphs.

For spectral properties, as discussed in Section 5.3, Friedman [Fri08] proved that almost all d-regular graphs are near-Ramanujan. Moreover, the recent work by Huang, McKenzie, and Yau [HMY24] proved that a random d-regular graph is Ramanujan with a probability of approximately 69%.

For applications such as using expander graphs for designing randomized algorithms, these randomized constructions can be applied directly.

Generating Random d-Regular Graphs

A technical question is how to generate a random d-regular graph.

The configuration model is a commonly used method for constructing random graphs with a specified degree sequence. It works by creating "half-edges" for each vertex according to its degree and then randomly pairing these half-edges to form edges. The resulting graph may contain self-loops or multi-edges, but if we condition on simple graphs, the distribution becomes uniform.

In the permutation model, a $2d$-regular graph on n vertices is generated by independently choosing d random permutations $\pi_1, \ldots, \pi_d$ on $[n]$ and adding an edge $(v, \pi_i(v))$ for each $v \in [n]$ and $i \in [d]$. This does not yield the uniform distribution on $2d$-regular graphs, but it is known that a family of events has probability $1 - o(1)$ in the distribution induced by the permutation model if and only if it has probability $1 - o(1)$ in the uniform distribution [HLW06, JLR11].

For fast algorithms for generating random regular graphs, we refer the reader to [GW17].

6.2 Algebraic Constructions

For applications such as derandomization, the probabilistic constructions cannot be used. Moreover, for some applications, we cannot afford to generate the whole graph and need to explore the graph locally. Algebraic constructions are deterministic and are explicit, allowing efficient computation of neighbors for any vertex.

The first construction of expanders is due to Margulis, a family of 8-regular graphs G_m for every integer m. The vertex set is $V_m = \mathbb{Z}_m \times \mathbb{Z}_m$. The neighbors of the vertex (x, y) are $(x + y, y), (x - y, y), (x, y + x), (x, y - x), (x + y + 1, y), (x - y + 1, y), (x, y + x + 1), (x, y - x + 1)$, where all operations are modulo m. We recommend the proof by Lee and Trevisan [Tre17, Chapter 19].

Another interesting construction is a family of 3-regular p-vertex graphs for every prime p. The vertex set is $\mathbb{Z}_p$, and a vertex x is connected to $x + 1$, $x - 1$ and to its multiplicative inverse x^{-1} modulo p. The proof relies on a deep result in number theory.

The constructions of Ramanujan graphs by Lubotzky, Phillip, Sarnak [LPS88] and Margulis [Mar88] are Cayley graphs of specific groups. We refer the reader to [HLW06] for the construction and an exposition of Cayley expander graphs.

6.3 Combinatorial Constructions

Developing more elementary constructions and analysis of expander graphs remains an area of great interest. The general approach in combinatorial constructions is to build larger expander graphs from smaller ones. It turns out that this approach has significant applications in algorithms and complexity, as we will see in the next chapter.

Replacement Products

The replacement product is perhaps the most natural product to begin with. The base case could simply be a constant size complete graph. Let G be an $(n, k, \epsilon_1 k)$-graph and H be an $(k, d, \epsilon_2 d)$-graph. A replacement product of G and H is to replace each vertex v in G with a copy of H, so that each edge incident on v connects to a different vertex of H.

Definition 6.1 (Replacement Product)**.** *Let G be a k-regular graph on n vertices and H be a d-regular graph on k vertices. The replacement product $G \textcircled{r} H$ is a graph where the vertex set is the Cartesian product $[n] \times [k]$ of the vertex set of G and H. Two vertices (u, i) and (v, j) have an edge if and only if:*

1. *$u = v$ and $ij \in E(H)$, or*
2. *$vu \in E(G)$, v is the i-th neighbor of u in G, and u is the j-th neighbor of v in G.*

Intuitively, $G \textcircled{r} H$ is a combinatorial expander if G and H are combinatorial expanders. Consider a set $S \subseteq V(G \textcircled{r} H)$. If S has either a large of small intersection with each "cloud" (copy of H), then S should have large expansion due to the large expansion of G, as S essentially corresponds to a subset of vertices in G. If S has medium intersections with many clouds, then S should have large expansion because H has large expansion, and there are many crossing edges within each cloud.

However, it is unclear how to formalize this intuition, as there seems to be no clean way to decompose a subset's contribution into those from G and H. The spectral approach, which we will see shortly, can be thought of as a linear algebraic method to make this idea rigorous in a more general setting.

Zig-Zag Product

The actual construction by Reingold, Vadhan and Wigderson [RVW02] that we will analyze is slightly more complex.

Definition 6.2 (Zig-Zag Product)**.** *Let G be a k-regular graph on n vertices and H be a d-regular graph on k vertices. The zig-zag product $G \textcircled{z} H$ is a graph with the same vertex set $[n] \times [k]$ as the replacement product. Two vertices (u, i) and (v, j) are connected by an edge if and only if $u \neq v$ and there exists $a \in [k]$ such that (u, i)–(u, a), (u, a)–(v, b), and (v, b)–(v, j) are all edges in the replacement product $G \textcircled{r} H$, where (u, a)–(v, b) is the unique edge incident on (u, a) with $v \neq u$ (i.e., the unique edge incident on (u, a) that leaves the cloud of u in the replacement product).*

In words, each edge in the zig-zag product $G \textcircled{z} H$ corresponds to a length three walk in the replacement product $G \textcircled{r} H$, where the first step is within a cloud, the second step is the unique way to leave a cloud, and the third step is within the other cloud.

The intuition that the zig-zag product is a spectral expander comes from random walks. Edge edge in $G \textcircled{z} H$ corresponds to a random step in H, a deterministic step in G, and another random step in H. We should think of the first two steps as going to a random neighboring cloud, and the third step corresponds to moving to a random neighbor within that cloud. Since both G and H are spectral expanders with fast mixing properties, after a few steps, the walk loses information about both the cloud and the location within the cloud. Thus, $G \textcircled{z} H$ inherits the fast mixing property and is a spectral expander.

Theorem 6.3 (Zig-Zag Theorem [RVW02])**.** *Let G be an $(n, k, \epsilon_1 k)$-graph and H be an $(k, d, \epsilon_2 d)$-graph. Then $G \textcircled{z} H$ is an $(nk, d^2, (\epsilon_1 + \epsilon_2 + \epsilon_2^2)d^2)$-graph.*

We will prove the theorem in the next subsection. First, let us see how the zig-zag product can be used to construct larger and larger constant degree expander graphs. The idea is to combine with the following standard operation that increases the spectral gap.

Definition 6.4 (Graph Power)**.** *Let G be a graph with adjacency matrix A. The k-th power G^k is the graph with the same vertex set as G and adjacency matrix A^k.*

In words, the number of parallel edges between u and v in G^k is equal to the number of length k walks between u and v in G. Note that while the spectral gap of G^k improves significantly, its degree also increases significantly.

Exercise 6.5 (Spectral Gap of Power)**.** *G^k is an $(n, d^k, \epsilon^k d^k)$-graph if G is an $(n, d, \epsilon d)$-graph.*

The idea of the combinatorial construction is to use graph power to increase the spectral gap, and then use the zig-zag product to reducing the degree without significantly reducing the spectral gap.

Theorem 6.6 (Expanders from Zig-Zag Product)**.** *For every sufficiently large constant d, there exists an infinite family of $(n, d^2, d^2/4)$-graphs.*

Proof. Let H be a $(d^4, d, d/16)$-graph. Its existence can be shown using a probabilistic argument when d is sufficiently, and it can be found via exhaustive search in constant time.

Using H as the building block, we define G_i inductively by $G_1 = H^2$ and $G_{i+1} = G_i^2 \textcircled{z} H$. We claim that G_i is a $(d^{4i}, d^2, d^2/4)$-graph for all $i \geq 1$. The base case follows from Exercise 6.5. Assuming G_i is a $(d^{4i}, d^2, d^2/4)$-graph, then G_i^2 is a $(d^{4i}, d^4, d^4/16)$-graph by Exercise 6.5, and $G_i^2 \textcircled{z} H$ is a $(d^{4(i+1)}, d^2, d^2/4)$-graph by Theorem 6.3. □

Proof of the Zig-Zag Theorem

Check that $G \textcircled{z} H$ has nk vertices and is d^2-regular. We bound the spectral gap of $G \textcircled{z} H$.

Matrix Formulation: The first step is to write down the walk matrix Z of the zig-zag product $G \textcircled{z} H$. Let $W(H)$ be the $k \times k$ walk matrix of H, which is simply $\frac{1}{d}A(H)$ where $A(H)$ is the adjacency matrix of H. Let W be the $nk \times nk$ matrix with n copies of W_H on the diagonal. This represents the transition matrix for one step of the random walk within the clouds in $G \textcircled{z} H$. The steps between clouds are deterministic: the walk moves from a vertex (u, i) to a unique vertex (v, j) with $v \neq u$. The transition matrix for this deterministic step is thus a permutation matrix P where $P_{(u,i),(v,j)} = 1$ for each inter-cloud edge and zero otherwise. It follows from the definition of the zig-zag product that

$$Z = WPW.$$

Thus, the random walk matrix of $G \textcircled{z} H$ has a very clean form, which should be the reason for the definition of the zig-zag product in Definition 6.2.

The graph $G \textcircled{z} H$ is regular, so $\vec{1}_{nk}$ is an eigenvector of Z with eigenvalue 1. To prove the zig-zag product theorem, we will prove that for all $f \perp \vec{1}_{nk}$, the Rayleigh quotient satisfies

$$R_Z(f) = \frac{|f^T Z f|}{\|f\|_2^2} \leq \epsilon_1 + \epsilon_2 + \epsilon_2^2.$$

This implies that all but the largest eigenvalue of Z have absolute value at most $\epsilon_1 + \epsilon_2 + \epsilon_2^2$, and hence all but the largest eigenvalue of $A(G \textcircled{z} H)$ have absolute value at most $(\epsilon_1 + \epsilon_2 + \epsilon_2^2)d^2$.

Vector Decomposition: For any $f \perp \vec{1}_{nk}$, we decompose f into two vectors to apply the results in G and H. This step demonstrates the power of linear algebra, as in the larger domain $\mathbb{R}^{nk}$, there is a natural way to decompose a vector. In contrast, in the combinatorial setting, it is unclear how to decompose a set of vertices in $G \textcircled{z} H$ into subsets of G and H to utilize their expansion properties as previously discussed.

Define f_G as the average of f on clouds, such that $f_G(u, i) = \frac{1}{k}\sum_{j=1}^{k} f(u, j)$ for all $(u, i) \in V(G \textcircled{z} H)$, so that two vertices in the same cloud have the same value in f_G. Define $f_H = f - f_G$. Note that f_H sums to zero in each cloud, such that $\sum_{j=1}^{k} f_H(u, j) = 0$ for each $u \in G$. Using the triangle inequality,

$$|f^\top Z f| = |f^\top WPWf| = |(f_G + f_H)^\top WPW(f_G + f_H)| \leq |f_G^\top WPWf_G| + 2|f_G^\top WPWf_H| + |f_H^\top WPWf_H|.$$

Since $W(H) \cdot \vec{1}_k = \vec{1}_k$ as H is regular, it follows that $Wf_G = f_G$, as vertices in the same cloud have the same value in f_G. Thus,

$$|f^\top Z f| \leq |f_G^\top P f_G| + 2|f_G^\top PWf_H| + |f_H^\top WPWf_H|.$$

We will use the spectral properties of G and H to establish:

- $|f_G^\top P f_G| \le \epsilon_1 \|f_G\|_2^2$ (Claim 6.9),
- $|f_H^\top W P W f_H| \le \epsilon_2^2 \|f_H\|_2^2$ (Claim 6.7),
- $2|f_G^\top P W f_H| \le 2\epsilon_2 \|f_G\|_2 \|f_H\|_2$ (Claim 6.8).

Assuming these claims and using $\|f\|_2^2 = \|f_G\|_2^2 + \|f_H\|_2^2$ we conclude that

$$\begin{aligned} |f^\top Z f| &\le \epsilon_1 \|f_G\|_2^2 + 2\epsilon_2 \|f_G\|_2 \|f_H\|_2 + \epsilon_2^2 \|f_H\|_2^2 \\ &\le \epsilon_1 \|f_G\|_2^2 + \epsilon_2 \big(\|f_G\|_2^2 + \|f_H\|_2^2\big) + \epsilon_2^2 \|f_H\|_2^2 \\ &\le (\epsilon_1 + \epsilon_2 + \epsilon_2^2) \|f\|_2^2. \end{aligned}$$

This completes the proof of Theorem 6.3, leaving the three claims to be proven.

Spectral Bounds: The following claim uses the spectral property of H and the fact that f_H sums to zero in each cloud.

Claim 6.7 (Quadratic Term of H). $|f_H^\top W P W f_H| \le \epsilon_2^2 \|f_H\|_2^2$.

Proof. First, we claim that $\|W(H) \cdot x\|_2 \le \epsilon_2 \|x\|_2$ for any $x \perp \vec{1}_k$. To see this, let $x = \sum_{i=1}^k c_i v_i$, where $v_1, \ldots, v_n$ is an orthonormal basis of eigenvectors of $W(H)$ with eigenvalues $\alpha_1, \ldots, \alpha_k$. Note that $c_1 = 0$, as $v_1 = \vec{1}/\sqrt{k}$ and $x \perp \vec{1}$. Since H is an $(k, d, \epsilon_2 d)$-graph, $\alpha_i^2 \le \epsilon_2^2$ for $2 \le i \le d$. Thus,

$$\|W(H) \cdot x\|_2^2 = \left\| W(H) \cdot \left(\sum_{i=2}^k c_i v_i \right) \right\|_2^2 = \left\| \sum_{i=2}^k c_i \alpha_i v_i \right\|_2^2 = \sum_{i=2}^k c_i^2 \alpha_i^2 \le \epsilon_2^2 \sum_{i=2}^k c_i^2 \le \epsilon_2^2 \|x\|_2^2.$$

This implies that $\|W f_H\|_2 \le \epsilon_2 \|f_H\|$, as the sum of the entries in each cloud is zero in f_H. Therefore,

$$|f_H^\top W P W f_H| \le \|W f_H\|_2 \cdot \|P W f_H\|_2 = \|W f_H\|_2^2 \le \epsilon_2^2 \|f_H\|_2^2,$$

where the inequality is by Cauchy-Schwarz and the equality holds as P is a permutation matrix. □

The second claim is straightforward.

Claim 6.8 (Cross Term). $|f_G^\top P W f_H| \le \epsilon_2 \|f_G\|_2 \|f_H\|_2$.

Proof. By Cauchy-Schwarz and $\|W f_H\|_2 \le \epsilon_2 \|f_H\|_2$ established in Claim 6.7,

$$|f_G^\top P W f_H| \le \|f_G\|_2 \cdot \|P W f_H\|_2 = \|f_G\|_2 \cdot \|W f_H\|_2 \le \epsilon_2 \|f_G\|_2 \|f_H\|_2.$$

□

The final claim uses the spectral property of G and the fact that $f \perp \vec{1}_{nk}$.

Claim 6.9 (Quadratic Term of G). $|f_G^\top P f_G| \le \epsilon_1 \|f_G\|_2^2$.

Proof. The main point is that $f_G^\top P f_G$ is equal to a corresponding quadratic form of the walk matrix of G. To see this, we "contract" each cloud to a single vertex. Define $g : V(G) \to \mathbb{R}$ as $g(v) = \sqrt{k} \cdot f_G(v,1)$, so that $\|g\|_2^2 = \|f_G\|_2^2$. We claim that $f_G^\top P f_G = g^\top W(G) g$, where $W(G)$ is the random walk matrix of G. This is because each edge (u,i)-(v,j) in $G ⓩ H$ contributes $f_G(u,i) \cdot f_G(v,j)$ to $f_G^\top P f_G$, while the corresponding edge $uv \in G$ contributes $g(u) \cdot W(G)_{u,v} \cdot g(v) = \left(\sqrt{k} f_G(u,1)\right)\left(\frac{1}{k}\right)\left(\sqrt{k} f_G(v,1)\right) = f_G(u,i) \cdot f_G(v,j)$ to $g^\top W(G) g$.

Since $f \perp \vec{1}$, it follows that $f_G \perp \vec{1}$ and thus $g \perp \vec{1}$. Therefore,

$$\frac{f_G^\top P f_G}{\|f_G\|_2^2} = \frac{g^\top W(G) g}{\|g\|^2} \leq \epsilon_1,$$

where the inequality is because G is an $(n, k, \epsilon_1 k)$-graph, □

This concludes the proof of Theorem 6.3. The idea of decomposing a vector into different components is useful in many proofs. We will use it again when we study high dimensional expanders.

Lifts of Graphs

Another combinatorial approach to constructing expanders, first proposed by Friedman, is to "lift" a smaller Ramanujan graph into a larger one. A k-lift of an (n,d,α)-graph G is an (nk,d,α')-graph, where each vertex u of G is replaced by k vertices $u_1,\ldots,u_k$ in H, and each edge $uv \in G$ is replaced by a perfect matching between $u_1,\ldots,u_k$ and $v_1,\ldots,v_k$ in H.

There are several results on random k-lifts. The best known result is by Puder [Pud15] who proved that a random k-lift of a Ramanujan graph satisfies $\alpha' \leq 2\sqrt{d-1}+1$ with probability tending to 1 as $k \to \infty$.

Bilu and Linial [BL06] studied 2-lifts. They used the converse of expander mixing lemma in Theorem 5.6 and a derandomized probabilistic method to construct an infinite family of $\left(n,d,O(\sqrt{d\log^3 d})\right)$-graphs.

Marcus, Spielman, and Srivastava [MSS15] developed an original approach using interlacing polynomials to prove that every bipartite Ramanujan graph has a 2-lift which is bipartite Ramanujan (i.e., with $\max_{2\leq i\leq n-1}|\alpha_i| \leq 2\sqrt{d-1}$).

Lossless Expanders

A graph is called a lossless expander if every small linear-sized subsets has vertex expansion close to d. These graphs have applications in constructing error correcting codes, as we will see in the next chapter.

As discussed previously, a random d-regular graph has small-set vertex expansion close to $d-2$, while there are Ramanujan graphs with small-set vertex expansion at most $d/2$. It remains a major open problem to design deterministic constructions of lossless expanders.

Capalbo, Reingold, Vadhan, and Wigerson [CRVW02] constructed one-sided bipartite lossless expanders using an intricate zig-zag product on "conductors" (a randomness-enchancing object); see also [HLW06, Chapter 10].

Recently, Golowich [Gol24] presented a simple and improved constructions of one-sided bipartite lossless expanders, by composing a small lossless expander with a large two-sided bipartite spectral expander. A new paper [HLM+24] demonstrated an explicit construction of two-sided vertex expanders beyond the spectral barrier $d/2$.

References

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Comb.*, 26(5):495–519, 2006. 49, 53, 66

[CRVW02] Michael Capalbo, Omer Reingold, Salil Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 659–668, 2002. 66, 74

[Fri08] Joel Friedman. *A Proof of Alon's Second Eigenvalue Conjecture and Related Problems*. American Mathematical Society, 2008. 55, 61

[Gol24] Louis Golowich. New explicit constant-degree lossless expanders. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4963–4971. SIAM, 2024. 67

[GW17] Pu Gao and Nicholas Wormald. Uniform generation of random regular graphs. *SIAM Journal on Computing*, 46(4):1395–1427, 2017. 62

[HLM+24] Jun-Ting Hsieh, Ting-Chun Lin, Sidhanth Mohanty, Ryan O'Donnell, and Rachel Yun Zhang. Explicit two-sided vertex expanders beyond the spectral barrier. *arXiv preprint arXiv:2411.11627*, 2024. 67

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 3, 19, 20, 54, 57, 61, 62, 66, 69, 70, 72

[HMY24] Jiaoyang Huang, Theo Mckenzie, and Horng-Tzer Yau. Ramanujan property and edge universality of random regular graphs. *arXiv preprint arXiv:2412.20263*, 2024. 55, 61

[JLR11] Svante Janson, Tomasz Luczak, and Andrzej Rucinski. *Random graphs*. John Wiley & Sons, 2011. 61

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. 54, 62

[Mar88] Gregory A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988. 54, 62

[MSS15] Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *Annals of Mathematics*, 182:307–325, 2015. 55, 66

[Pud15] Doron Puder. Expansion of random graphs: New proofs, new results. *Inventiones mathematicae*, 201(3):845–908, 2015. 66

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. *Annals of Mathematics*, 155(1):pp. 157–187, 2002. 63

[Tre17] Luca Trevisan. *Lecture Notes on Graph Partitioning, Expanders and Spectral Methods*. 2017. 9, 62, 90

Chapter 7

Expander Graphs: Applications

This chapter highlights some key applications of expander graphs. More details are given on expander codes, which form the basis of recent breakthroughs in designing asymptotically good and locally testable codes [DEL+22, PK22]. The content of this chapter is mostly based on [HLW06].

7.1 Pseudorandomness

Expander graphs are pseudorandom objects, as suggested by the expander mixing lemma in Theorem 5.3. This pseudorandom property makes them useful for reducing or even eliminating randomness in certain settings.

Reducing Randomness in Probability Amplification

Suppose we have a randomized algorithm with error probability $1/100$ that requires n random bits. To decrease the failure probability, the standard approach is to run the randomized algorithm independently k times and take the majority answer as the output. By a standard Chernoff bound argument, this decreases the failure probability to δ^k for some small constant δ. However, this approach uses kn random bits.

We show how to achieve exponentially small error probability while using only $n + O(k)$ bits. To do so, let us reinterpret the above analysis from a random walk perspective.

Let V denote the set of all n-bit strings. The condition that the randomized algorithm has an error probability of at most $1/100$ is equivalent to saying that, among the 2^n n-bit strings, at most $2^n/100$ of them are "bad" strings. Let $B \subseteq V$ denote this set of bad strings. The majority-vote algorithm fails if more than $k/2$ strings sampled randomly are from B. Sampling k independent n-bit strings can be interpreted as performing a random walk of length k on the complete graph over V and use the corresponding bit strings of the vertices on this walk.

The key idea is to replace the random walk on the complete graph with a random walk on a constant degree expander graph over V. Let G be an $(2^n, d, d/10)$-graph, where d is a constant. Such graphs exist, for example, by taking a sufficiently large constant power of a Margulis expander as described in Section 6.2. In the first step of the random walk, we use an n-bit random string. In subsequent steps, instead of using n random bits to select the next n-bit string, we choose a random neighbor of the current string in G and use the corresponding string of this random neighbor. Since G is a d-regular graph, each subsequent step requires only $\log_2 d$ random bits to select a random neighbor.

Thus, the total number of bits used is $n+(k-1)\cdot\log_2 d = n+O(k)$. Importantly, the neighbors of a Margulis expander can be computed efficiently, allowing the corresponding strings to be computed quickly without constructing the entire graph G. This is where the explicitness of an algebraic construction becomes useful.

What about the error probability of this expander-walk algorithm? This is precisely what Theorem 5.15 is formulated for. It shows that the failure probability of taking the majority answer from a random walk of length k on a two-sided spectral expander with $\alpha \leq d/100$ is at most $(2/\sqrt{5})^k$.

Expander Random Walks

This is just one example of using expander graphs in derandomization; see [Vad12] for more. See [Gil98] for a well-known Chernoff bound for random walks on two-sided spectral expanders and [GLSS18] for a recent generalization to the matrix setting.

The above results can be interpreted as showing that the majority functions are fooled by expander random walks, meaning they cannot distinguish independent random samples from those produced by expander random walks. See [CPT21] for a recent paper demonstrating that many functions are similarly fooled by expander random walks.

7.2 Constructing Efficient Networks and Algorithms

A d-regular expander graph can be viewed as highly efficient, having only a linear number of edges while achieving very high connectivity. From this perspective, it is natural that expander graphs are used in constructing efficient networks.

Efficient Networks

One interesting example is the construction of superconcentrators, which are directed graphs with n input nodes and n output nodes (and possibly other notes) that satisfy the strong connectivity property: for any $k \leq n$, there exist k vertex disjoint paths between any k input nodes and any k output nodes. For example, the complete bipartite graph $K_{n,n}$ satisfies this property but requires $\Theta(n^2)$ edges. Valiant initially conjectured that no superconcentrator with $O(n)$ edges exists as part of an attempt to prove circuit lower bounds. Later, he developed a recursive construction of superconcentrator with $O(n)$ edges using expander graphs as building blocks; see [HLW06].

Another classical application of expander graphs is the construction of optimal sorting networks with $O(n \log n)$ edges and $O(\log n)$ depth [AKS83].

Efficient Algorithms

Superconcentrators and expander graphs can also be used to design efficient algorithms. A simple example is the design of fast algorithms for computing matrix rank [CKL13]. In this application, an expander graph or superconcentrator is used to "compress" a rectangular matrix $A \in \mathbb{F}^{m\times n}$ with $n \gg m$ into a square matrix $B \in \mathbb{F}^{m\times m}$ in linear time, such that $\text{rank}(A) = \text{rank}(B)$ with high probability. This leads to faster randomized algorithm for computing the rank of a rectangular matrix and finding linearly independent columns. For this application, probabilistic constructions of bipartite expander graphs are sufficient.

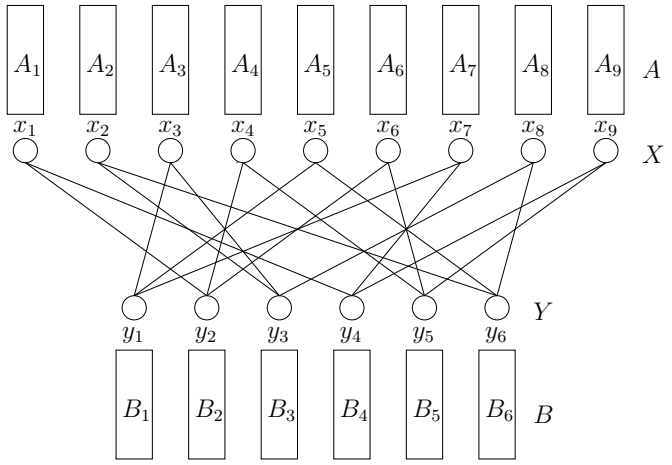


Figure 7.1: A bipartite expander graph $G = (X, Y; E)$ is used to compress the input matrix A into a smaller matrix B. Each column of B is a random linear combination of the columns of its neighbors; for example, B_3 is a random linear combination of A_2, A_3 and A_8.

Another example is to use superconcentrators to design a faster algorithm for computing edge connectivities in a directed acyclic graph [CLL13].

7.3 Complexity Theory

The construction of expanders using zig-zag products has inspired remarkable developments in complexity theory. In an interesting parallel, the construction of expanders using cut-matching games has been applied widely to designing efficient algorithms, as we will see later in this course. These results demonstrate the value of combinatorial constructions of expander graphs.

Graph Connectivity in Log-Space

A striking application of the zig-zag product in Definition 6.2 is solving the s-t connectivity problem in an undirected graph using logarithmic space.

If randomized algorithms are allowed, solving s-t connectivity in log-space is simple: just perform a random walk for $O(n^3)$ steps. This works because the expected cover time for any undirected graph is at most $O(n^3)$.

For deterministic algorithms, Savitch's algorithm solves the more general problem of s-t connectivity in directed graphs using $O(\log^2 n)$ space, by recursively guessing the midpoint of a directed s-t path. It remains an important open problem whether s-t connectivity in directed graphs can be solved in $O(\log n)$ space. Such an algorithm would imply $NL = L$, meaning the non-deterministic and deterministic log-space complexity classes are equal.

Reingold [Rei08] discovered a deterministic $O(\log n)$ space algorithm for s-t connectivity in undirected graphs using zig-zag products. If the input graph G is a d-regular expander graph with constant d, then G has a diameter of $O(\log n)$. Paths of length $O(\log n)$ in such a graph can be enumerated in $O(\log n)$ space, since each neighbor requires only $\log_2 d$ bits to describe. Thus, solving s-t connectivity in constant-degree expander graphs in $O(\log n)$ space is straightforward.

Reingold's approach is to transform any graph G into a regular expander graph G_k such that s, t are connected in G if and only if they are connected in G_k. First, G is converted into a D-regular graph G_1 with constant $D = d^{16}$ by replacing high degree vertices with constant-degree expander graphs and adding self-loops to low degree vertices, similar to the replacement product in Definition 6.1. The resulting G_1 is an $(n, d^{16}, \epsilon_1 d^{16})$-graph where $\epsilon_1 \leq 1 - 1/n^2$ since any connected undirected graph has a spectral gap of at least $1/n^2$.

To improve expansion, the idea is to construct $G_{i+1} := (G_i \textcircled{z} H)^8$, where H is a $(d^{16}, d, d/2)$-graph. Using a variant of the zig-zag theorem in Theorem 6.3, it can be shown that the spectral gap doubles with each iteration. More precisely, if G_i is an $(nd^{16i}, d^{16}, \epsilon_i d^{16})$-graph, then G_{i+1} is an $(nd^{16(i+1)}, d^{16}, \epsilon_{i+1} d^{16})$-graph where $\epsilon_{i+1} \leq \epsilon_i^2$. Therefore, repeating this construction for $k = O(\log n)$ times yields G_k, an $(nd^{16k}, d^{16}, 3/4)$-graph with constant spectral gap. Note that the size of G_k is at most a polynomial factor larger than G for $k = O(\log n)$, and s and t are connected in G if and only if they are connected in G_k.

A technical challenge in this approach is computing a neighbor of a vertex in G_k in log-space. The intuition is that there are only $O(\log n)$ recursion levels in the zig-zag construction, and each level requires only constant space since there are just three steps and the degree is constant. Reingold proved that this can indeed be achieved using a clever data structure; see [Rei08, Vad12] for details.

Hardness Amplification

Random walks on expander graphs can also be used for hardness amplifications, transforming instances that are hard to approximate into instances that are even harder to approximate. See [HLW06, Section 3.3] or [AB06, Chapter 22] for a simple application of expander random walks in proving hardness of approximating maximum independent sets.

Dinur [Din07] provided an elegant proof of the important PCP theorem using expander random walks. Her proof, inspired by Reingold's result, involves multiple iterations of "powering" and "degree reduction", which makes the underlying constraint satisfaction problem increasingly harder to approximate. See [AB06, Chapter 22] for a good exposition of the PCP theorem.

7.4 Error Correcting Codes

A major motivation behind the early development of expander graphs comes from coding theory, where small-set vertex expansion is the key combinatorial quantity.

A code $C \subseteq \{0,1\}^n$ of length n is a subset of n-bit strings, where each string in C is called a codeword. To design a good error correcting code, we aim to choose codewords that are far from each other so as to correct more errors, while also maximizing the number of codewords so as to achieve a high information rate. This can be viewed as a sphere packing problem, where the objective is to fit in as many disjoint spheres of a certain radius as possible within $\mathbb{F}_2^n$.

Definition 7.1 (Distance of Code). *Given $C \subseteq \{0,1\}^n$, the distance of C is defined as* $\text{dist}(C) := \min_{x \neq y \in C} d_H(x, y)$, *where $d_H(x, y)$ is the Hamming distance between two codewords x and y. The relative distance of C is defined as* $\text{dist}(C)/n$.

Definition 7.2 (Rate of Code). *Given $C \subseteq \{0,1\}^n$, the rate of C is defined as* $\log(|C|)/n$, *where* $\log |C|$ *can be thought of as the number of bits of information sent.*

Definition 7.3 (Asymptotically Good Code)**.** *A family $C_n \in \{0,1\}^n$ of codes is asymptotically good if there are constants $r > 0$ and $\delta > 0$ such that for all n, the relative distance of C_n is at least δ, and the rate of C_n is at least r.*

The existence of an asymptotically good code can be proved using a standard probabilistic method. However, for the codes to be practical, encoding and decoding should also be achievable in polynomial time in n (ideally linear time in n). This requirement makes designing good codes much more challenging.

A common class of codes is the class of linear codes, where C is a linear subspace of $\mathbb{F}_2^n$. Linear codes have the advantage that they can be described by a basis and so encoding can be done in $O(n^2)$ time. Additionally, a simple but useful property of linear codes is that the minimum distance of the code is equal to the minimum ℓ_1-norm of a non-zero codeword, because $d_H(x,y) = \|x-y\|_1$ and $x-y$ is a codeword. The natural decoding strategy is to find the nearest codeword of a received word, but this is an NP-complete problem even for linear codes.

Low Density Parity Check Codes

The idea of constructing codes from graphs was first suggested by Gallager, who uses sparse bipartite graphs to design low-density parity check codes (LDPC codes).

Let A be a parity check matrix for code C, such that $C = \{x \mid Ax = 0\}$ where $A \in \{0,1\}^{m\times n}$ with $m < n$. Each row i of A is a parity-check constraint, requiring $\sum_{j=1}^{n} A_{ij} \cdot x(j) = 0 \pmod 2$. Note that the rate of this code is $1 - m/n$. Thus, one objective is to minimize the number of constraints to ensure a lower bound on the rate.

The matrix A can be viewed as a bipartite graph $G = (L, R; E)$, where $L = [n]$ represents variables, $R = [m]$ represents constraints, and there is an edge between variable $i \in L$ and constraint $j \in R$ if $A_{ij} = 1$. The small-set vertex expansion of G is the key property in analyzing LPDC codes.

Definition 7.4 (One-Sided Small-Set Vertex Expansion)**.** *Let $G = (L, R; E)$ be a bipartite graph with $|L| = n$ and $|R| = m$ and $m < n$. For any $0 < \delta < 1$, define the left δ-small-set vertex expansion of G as*

$$\psi_\delta^L(G) := \min_{S\subseteq L : |S| \le \delta n} \frac{|\partial(S)|}{|S|},$$

where $\partial(S)$ is the vertex boundary as in Definition 3.11.

The following theorem relates the one-sided small-set vertex expansion of the graph to the minimum distance of the code. The proof uses the unique neighbor property guaranteed by strong enough one-sided small-set vertex expansion.

Theorem 7.5 (Distance of Expander Code [SS96])**.** *Let $G = (L, R; E)$ be a left d-regular bipartite graph with $\psi_\delta^L(G) > d/2$. Then the parity check code $C(G)$ defined by G has relative distance greater than δ.*

Proof. Let $S \subseteq L$ be a subset of left vertices with $|S| \le \delta n$. By the left small-set vertex expansion assumption of G, $|\partial(S)| > d|S|/2$. A simple counting argument shows that there exists a vertex $v \in \partial(S) \subseteq R$ with only one neighbor in S. Let us call such a vertex a unique neighbor of S.

To lower bound the minimum distance, recall that it is equivalent to lower bounding the ℓ_1-norm/support-size of a codeword $x \in \{0,1\}^n$. Let S be the support of x. If $|S| \le \delta n$, by the

above argument, there exists a unique neighbor $v \in R$ of S. This implies that the parity constraint on v is not satisfied by x, so x is not a codeword of the parity check code defined by G. Therefore, any codeword of this parity check code must have support size greater than δn, and thus the minimum distance of this code is greater than δn. □

For efficient decoding, a stronger requirement $\psi_\delta^L(G) \geq 3d/4$ is needed. As discussed in Section 6.3, this strong condition is satisfied with high probability in random d-left-regular bipartite graphs. For deterministic constructions, Capalbo, Reingold, Vadhan, Wigderson [CRVW02] used an intricate variant of the zig-zag product to construct one-sided small-set vertex expanders with $\psi_\delta^L(G) \geq 0.99d$ for some $\delta > 0$, while $m/n < 0.99$ so that the rate of the code is at least 0.01.

Fast Decoding Algorithm

A key feature of LPDC codes defined by expander graphs is that there exists a surprisingly simple and efficient decoding algorithm.

Algorithm 3 Flip Algorithm for Expander Code

Require: A parity check matrix $A \in \{0,1\}^{m \times n}$ and a bit string $x \in \{0,1\}^n$.

1: Let $x^{(0)} := x$ and $t = 0$.
2: **while** there is an unsatisfied parity check constraint **do**
3: Find a bit i such that flipping it decreases the number of unsatisfied parity constraints. That is, an $i \in [n]$ such that $\left\|A\left(x^{(t)} + \chi_i\right)\right\|_1 < \left\|Ax^{(t)}\right\|_1$, where χ_i is the characteristic vector of i and the addition is mod 2. Set $x^{(t+1)} := x^{(t)} + \chi_i$ and $t \leftarrow t + 1$.
4: **end while**
5: **return** $x^{(t)}$.

The analysis of the flip algorithm relies on a stronger assumption about the left small-set vertex expansion than that in Theorem 7.5.

Theorem 7.6 (Efficient Decoding of Expander Code [SS96])**.** *Let $G = (L, R; E)$ be a left d-regular bipartite graph with $L = [n]$ and $R = [m]$ and $\psi_\delta^L(G) > 3d/4$. Let x be an n-bit string whose distance from a codeword y is at most $\delta n/2$. Then Algorithm 3 will return y in at most m iterations.*

Proof. The plan is to argue that:

1. There exists a bit i such that flipping it decreases the number of unsatisfied constraints, as long as $\mathrm{dist}_H\left(x^{(t)}, y\right) \leq \delta n$;
2. $\mathrm{dist}_H\left(x^{(t)}, y\right) \leq \delta n$ for all t if $\mathrm{dist}_H\left(x^{(0)}, y\right) \leq \delta n/2$.

These imply that the number of unsatisfied constrained decreases with each iteration, so the algorithm must stop after at most $\tau \leq m$ iterations. At that point, $x^{(\tau)}$ is a codeword because all constraints are satisfied. Moreover, $x^{(\tau)}$ must be equal to y, as $\mathrm{dist}_H(x^{(\tau)}, y) \leq \delta n$, while the distance between y and other codewords is strictly bigger than δn by Theorem 7.5.

Let $\Delta := \{i \in [n] \mid x^{(t)}(i) \neq y(i)\}$ be the set of errors at the t-th iteration. Assume $0 < |\Delta| \leq \delta n$, we argue that a bit i exists such that flipping it decreases the number of unsatisfied constraints.

Partition $\partial(\Delta)$ into satisfied neighbors $\partial_+(\Delta)$ and unsatisfied neighbors $\partial_-(\Delta)$. Since $|\Delta| \leq \delta n$, by the left small-set vertex expansion assumption,

$$|\partial_+(\Delta)| + |\partial_-(\Delta)| = |\partial(\Delta)| > 3d|\Delta|/4.$$

Consider the $d|\Delta|$ number of edges between Δ and $\partial(\Delta)$. Each vertex in $\partial_+(\Delta)$ contributes at least two such edges while each vertex in $\partial_-(\Delta)$ contributes at least one such edge, and so

$$2|\partial_+(\Delta)| + |\partial_-(\Delta)| \leq d|\Delta|.$$

Combining these two inequalities shows $|\partial_-(\Delta)| > d|\Delta|/2$. Thus, there must exist a vertex $i \in \Delta$ with strictly more than $d/2$ unsatisfied neighbors, so flipping i decreases the number of unsatisfied constraints.

To complete the proof, we argue that $|\Delta| \leq \delta n$ in any iteration. Suppose this is not true. Since $|\Delta|$ changes by one in each iteration, there must be an (earliest) iteration such that $|\Delta| = \delta n$. At this point, using the argument in the previous paragraph, there are strictly more than $d|\Delta|/2 = d\delta n/2$ unsatisfied constraints. However, since the initial error is at most $\delta n/2$, the number of unsatisfied constraints in the beginning is at most $d\delta n/2$. This contradicts with the previous paragraph that the number of unsatisfied constraints is decreasing when $|\Delta| \leq \delta n$. □

Spielman [Spi96] further designed "superconcentrator codes" to construct asymptotically good codes that are linear-time encodable and decodable.

Tanner Codes

Tanner codes generalize LDPC codes by allowing the "base code" to be more general than a simple parity check. Let $C_0 \subseteq \{0,1\}^d$ be the base code, and let $G = (V, E)$ be a d-regular graph with $V = [n]$ and $E = [m]$. The Tanner code is defined as

$$C(G) := \{y \in \{0,1\}^m \mid y_{|\delta(i)} \in C_0 \ \forall i \in [n]\},$$

where $y_{|\delta(i)}$ is the vector y restricted to the d edges in $\delta(i)$ for a vertex $i \in V$. Each bit $y(j)$ of a codeword corresponds to an edge $j \in E$ of G, and a binary string y is a codeword if $y_{|\delta(i)}$ is a codeword of the base code C_0 for every vertex $i \in V$ of G.

The advantage of Tanner codes is that we could use a stronger base code with larger minimum distance, instead of a parity check code with minimum distance two. For a base code C_0 with minimum distance d_0, the vertex expansion requirement for G can be relaxed to d/d_0 to achieve the same distance as the corresponding LDPC code. In particular, by Tanner's theorem in Theorem 5.14, a spectral expander can be used as G to design asymptotically good codes that are linear time encodable and decodable, without requiring lossless expanders.

The decoding algorithm for Tanner codes is still an iterative "fixing" algorithm, where invalid codewords on a vertex are replaced by their nearest valid codewords in the base code. The analysis is similar to LDPC codes: if the decoding algorithm fails, one can argue that there must be a "denser" subgraph exists than what is allowed by the expander mixing lemma.

Recent breakthroughs [DEL+22, PK22] in designing asymptotically good codes that are also locally testable generalize Tanner codes to 2-dimensional expanders (where graphs are considered 1-dimensional expanders).

References

[AB06] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2006. 72

[AKS83] Miklós Ajtai, János Komlós, and Endre Szemerédi. An o(n log n) sorting network. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 1–9. ACM, 1983. 70

[CKL13] Ho Yee Cheung, Tsz Chiu Kwok, and Lap Chi Lau. Fast matrix rank algorithms and applications. *J. ACM*, 60(5):31:1–31:25, 2013. 70

[CLL13] Ho Yee Cheung, Lap Chi Lau, and Kai Man Leung. Graph connectivities, network coding, and expander graphs. *SIAM Journal on Computing*, 42(3):733–751, 2013. 71

[CPT21] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a Fourier-analytic approach. In *53rd Annual ACM SIGACT Symposium on Theory of Computing,*, pages 1643–1655. ACM, 2021. 70

[CRVW02] Michael Capalbo, Omer Reingold, Salil Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 659–668, 2002. 66, 74

[DEL+22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 357–374. ACM, 2022. 69, 75

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):12, 2007. 72

[Gil98] David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998. 70

[GLSS18] Ankit Garg, Yin Tat Lee, Zhao Song, and Nikhil Srivastava. A matrix expander Chernoff bound. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, (STOC)*, pages 1102–1114. ACM, 2018. 70

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, 2006. 3, 19, 20, 54, 57, 61, 62, 66, 69, 70, 72

[PK22] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In *54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 375–388. ACM, 2022. 69, 75

[Rei08] Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4):17:1–17:24, 2008. 71, 72

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inf. Theory*, 42(6):1723–1731, 1996. 75

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inf. Theory*, 42(6):1710–1722, 1996. 73, 74

[Vad12] Salil P. Vadhan. Pseudorandomness. *Found. Trends Theor. Comput. Sci.*, 7(1-3):1–336, 2012. 57, 70, 72

Topic I

Generalizations of Cheeger's Inequality

Chapter 8

Higher-Order Cheeger Inequality

Cheeger's inequality in Chapter 3 is a robust generalization of the basic fact that $\lambda_2(G) = 0$ if and only if G is disconnected. The higher-order Cheeger inequality in this chapter is a robust generalization of the basic fact that $\lambda_k(G) = 0$ if and only if G has k connected components.

Arora, Barak and Steurer [ABS10] were the first to establish such a generalization. Informally, they proved that if λ_k is small then there exists a set S with small edge conductance and $|S| \lesssim |V|/k$, generalizing the fact that if $\lambda_k = 0$ then there exists a set S with edge conductance 0 and $|S| \leq |V|/k$. This result was used in [ABS10] to design a subexponential-time algorithm for Unique Games. This work has inspired many subsequent studies that use higher eigenvalues to design and analyze approximation algorithms. We will present their result later in Chapter **??** when we study random-walk based techniques for graph partitioning.

The higher-order Cheeger inequality provides a conceptually stronger generalization, asserting that λ_k is small if and only if G has at least k disjoint subsets $S_1, \ldots, S_k$ each with small edge conductance.

Definition 8.1 (k-Way Edge Conductance)**.** *Let $G = (V, E)$ be a graph. The k-way edge conductance is defined as*

$$\phi_k(G) = \min_{S_1, S_2, \ldots, S_k \subseteq V} \max_{1 \leq i \leq k} \phi(S_i),$$

where the minimization is over pairwise disjoint subsets $S_1, \ldots, S_k$ of V.

The following results were obtained independently by two research groups.

Theorem 8.2 (Higher-Order Cheeger Inequalities [LOT14, LRTV12])**.** *Let $G = (V, E)$ be a graph and let λ_k denote the k-th smallest eigenvalue of its normalized Laplacian matrix. Then*

$$\lambda_k \lesssim \phi_k(G) \lesssim k^2 \cdot \sqrt{\lambda_k}.$$

Moreover, for fewer disjoint subsets, there is an improved dependency on k:

$$\phi_{k/2}(G) \lesssim \sqrt{\log k \cdot \lambda_k}.$$

The direction $\lambda_k \lesssim \phi_k(G)$ is called the easy direction, while the direction $\phi_k(G) \lesssim k^2\sqrt{\lambda_k}$ is called the hard direction. As in Cheeger's inequality, the easy direction shows that λ_k is a relaxation of the k-way edge-conductance problem, while the hard direction is proved using a rounding algorithm for this relaxation. We will prove the hard direction and leave the easy direction in Problem 8.3. We assume the graph is d-regular throughout this chapter.

Problem 8.3. *Prove the easy direction. Hint: Use the Courant-Fischer theorem in Theorem A.14.*

8.1 Cheeger Rounding and Spectral Embedding

We first revisit the Cheeger rounding algorithm from Chapter 3 and then motivate the use of the spectral embedding for the k-way edge-conductance problem.

Cheeger Rounding

The following rounding algorithm is a consequence of the threshold rounding step in Lemma 3.5 and the embedding step in Lemma 3.7. Note that it applies to all vectors, not just eigenvectors.

Lemma 8.4 (Cheeger Rounding)**.** *Given graph $G = (V = [n], E)$ and a vector $x \in \mathbb{R}^n$, there exists an efficient algorithm to find a subset $S \subseteq \mathrm{supp}(x)$ such that*

$$\phi(S) \leq \sqrt{2R(x)} \quad \textit{where} \quad R(x) = \frac{x^\top \mathcal{L} x}{x^\top x}$$

is the Rayleigh quotient of x, and $\mathrm{supp}(x) := \{i \in [n] \mid x(i) \neq 0\}$ is the support of vector x.

When λ_k is small, there are k orthogonal eigenvectors $v_1, \ldots, v_k$, each with a small Rayleigh quotient. Applying Lemma 8.4 to each v_i produces subsets $S_1, \ldots, S_k$, each with small edge conductance. Since the eigenvectors $v_1, \ldots, v_k$ are orthogonal, it is natural to expect that the sets $S_1, \ldots, S_k$ differ significantly. However, dealing with each vector separately does not provide a clear way to combine the resulting subsets $S_1, \ldots, S_k$ into k disjoint subsets with small edge conductance.

This motivates a more global view that considers all the k vectors simultaneously.

Spectral Embedding

An interesting idea in [LOT14, LRTV12] is to use the spectral embedding defined by the first k eigenvectors to find k disjoint subsets of small edge conductance (a set of small edge conductance is also called a sparse cut).

Definition 8.5 (Spectral Embedding)**.** *Let $G = (V = [n], E)$ be a graph. Let $\lambda_1 \leq \ldots \leq \lambda_k$ be the k smallest eigenvalues of $\mathcal{L}(G)$, and $v_1, \ldots, v_k \in \mathbb{R}^n$ be the corresponding orthonormal eigenvectors. Let $U \in \mathbb{R}^{n \times k}$ be the $n \times k$ matrix where the j-th column is v_j. The spectral embedding $u_i \in \mathbb{R}^k$ of a vertex i is defined as the i-th row of U.*

In the spectral embedding, each vertex i is mapped to a k-dimensional point $u_i \in \mathbb{R}^k$. This embedding is commonly used in practice to find disjoint sparse cuts. Heuristics often apply geometric clustering algorithms, such as the k-means algorithm, to partition the points into k clusters, which are then used to partition the graph.

These heuristics are reported to work well in applications such as image segmentation and data clustering, but no theoretical guarantees were known. The proof of the higher-order Cheeger inequality provides a rigorous analysis of certain variants of these methods [NJW01], justifying the use of spectral embedding for graph partitioning. Analyzing the k-means heuristic rigorously remains an open problem.

8.2 Isotropy Condition and Clustering by Directions

For the spectral embedding to provide useful information for finding k disjoint sparse cuts, it is necessary that the points are reasonably "well spread out". For example, if all vertices are mapped to only two points in $\mathbb{R}^k$, then there is no well-defined way to partition the points into k clusters.

As expected, such bad cases do not occur because the k eigenvectors are orthogonal. The question is how to use the orthogonality to derive conditions that facilitate clustering.

Isotropy Condition

As $v_1, \ldots, v_k$ are orthonormal vectors, the matrix U in Definition 8.5 satisfies $U^\top U = I_k$, which can be rewritten as $\sum_{i=1}^n u_i u_i^\top = I_k$. This implies that the spectral embedding satisfies the following isotropy condition.

Lemma 8.6 (Isotropy Condition)**.** *Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 8.5. For any $x \in \mathbb{R}^k$ with $\|x\|_2 = 1$,*

$$\sum_{i=1}^n \langle x, u_i \rangle^2 = 1.$$

Proof. The condition $U^\top U = I_k$ implies that $x^\top U^\top U x = x^\top x = 1$ for any $x \in \mathbb{R}^k$ with $\|x\|_2 = 1$. Writing $U^\top U = \sum_{i=1}^n u_i u_i^\top$ shows that

$$1 = x^\top \Big(\sum_{i=1}^n u_i u_i^\top \Big) x = \sum_{i=1}^n x^\top u_i u_i^\top x = \sum_{i=1}^n \langle x, u_i \rangle^2.$$

□

This lemma says that for any direction $x \in \mathbb{R}^k$, the sum of the square of the projection of u_i on x is equal to 1. To develop some intuition, suppose $u_1 = u_2 = \cdots = u_l = y \in \mathbb{R}^k$ (that the first l vertices all mapped to the same point y), then the lemma implies that

$$1 = \sum_{i=1}^n \Big\langle \frac{y}{\|y\|_2}, u_i \Big\rangle^2 \geq \sum_{i=1}^l \Big\langle \frac{u_i}{\|u_i\|_2}, u_i \Big\rangle^2 = \sum_{i=1}^l \|u_i\|^2.$$

On the other hand, as each eigenvector satisfies $\|v_i\|_2 = 1$,

$$\sum_{i=1}^n \|u_i\|_2^2 = \|U\|_F^2 = \sum_{i=1}^k \|v_i\|_2^2 = k. \tag{8.1}$$

Therefore, if we consider the "mass" of a point i as $\|u_i\|_2^2$, then the above calculation shows that at most $1/k$ fraction of the total mass can be mapped to the same point. This ensures that the spectral embedding maps to at least k distinct points, ruling out the bad cases mentioned earlier.

Clustering by Directions

The same reasoning shows that points with similar directions carry at most $\approx 1/k$ fraction of the total mass. This motivates clustering points by their directions, ensuring they are reasonably well spread out. The following distance measure and spreading property formalize this idea.

Definition 8.7 (Radial Projection Distance)**.** *Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 8.5. The radial projection distance between two vertices i and j is defined as*

$$d(i,j) = \left\| \frac{u_i}{\|u_i\|_2} - \frac{u_j}{\|u_j\|_2} \right\|_2$$

if $\|u_i\| > 0$ and $\|u_j\| > 0$. Otherwise, if $u_i = u_j = 0$ then $d(i,j) := 0$, else $d(i,j) := \infty$.

Lemma 8.8 (Spreading Property)**.** *Let $G = (V = [n], E)$ be a graph. Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 8.5. If $S \subseteq V$ satisfies $d(i,j) \leq \Delta$ for all $i, j \in S$, then*

$$\sum_{i \in S} \|u_i\|^2 \leq \frac{1}{1-\Delta^2}.$$

Proof. Choose an arbitrary vertex $j \in S$ and consider the unit vector $u_j/\|u_j\|_2$. By the isotropy condition in Lemma 8.6,

$$1 = \sum_{i=1}^{n} \left\langle \frac{u_j}{\|u_j\|_2}, u_i \right\rangle^2 \geq \sum_{i \in S} \|u_i\|_2^2 \cdot \left\langle \frac{u_j}{\|u_j\|_2}, \frac{u_i}{\|u_i\|_2} \right\rangle^2 = \sum_{i \in S} \|u_i\|_2^2 \cdot \left(1 - \frac{d^2(i,j)}{2}\right)^2,$$

where the last equality holds because $\langle u, v \rangle = 1 - \|u - v\|_2^2/2$ for any two unit vectors u, v. Using the assumption $d(i,j) \leq \Delta$, it follows that

$$1 \geq \sum_{i \in S} \|u_i\|_2^2 \cdot \left(1 - \frac{\Delta^2}{2}\right)^2 \geq \sum_{i \in S} \|u_i\|_2^2 \cdot (1 - \Delta^2).$$

Rearranging gives the lemma. □

To ensure mass balance, we will choose Δ such that $\frac{1}{1-\Delta^2} \leq 1 + \frac{1}{2k}$ (e.g., $\Delta = \frac{1}{2\sqrt{k}}$). By forming subsets with diameter Δ, the lemma ensures that each subset has mass at most $1 + \frac{1}{2k}$. Thus, after selecting $k-1$ subsets, the remaining mass is still at least $1/2$. This ensures that k groups, each containing $\Omega(1/k)$ fraction of the total mass, can be formed by clustering based on directions.

8.3 Ideal Case: Far Apart Clusters

The previous section rules out the bad cases for clustering based on directions. In this section, we analyze the ideal scenario when the spectral embedding provides exactly what we want: k clusters that are far apart from each other.

Suppose there exist k disjoint subsets $S_1, S_2, \ldots, S_k$ such that:

- Each subset S_i has mass 1, i.e., $\sum_{j \in S_i} \|u_j\|_2^2 = 1$;
- The clusters are well-separated, i.e., $d(S_i, S_j) \geq \delta$ for all $i \neq j$ where $d(S_i, S_j) := \min\{d(a,b) \mid a \in S_i, b \in S_j\}$.

Can we conclude that these k subsets correspond to k disjoint sparse cuts in the graph?

Rayleigh Quotient of the Spectral Embedding and Cheeger Rounding

To analyze the spectral embedding, we extend the Rayleigh quotient definition and the Cheeger rounding result in Lemma 8.4 to work with k-dimensional embeddings.

Recall that the Rayleigh quotient of a vector $x : V \to \mathbb{R}$ is defined as

$$R(x) = \frac{x^\top \mathcal{L} x}{x^\top x} = \frac{x^\top L x}{d x^\top x} = \frac{\sum_{ij\in E}(x(i) - x(j))^2}{d\sum_{i\in V} x(i)^2}.$$

For k-dimensional embeddings, the Rayleigh quotient is defined as follows.

Definition 8.9 (Rayleigh Quotient of a k-Dimensional Embedding)**.** *Let $\psi_1, \ldots, \psi_n \in \mathbb{R}^k$ and $\Psi = (\psi_1, \ldots, \psi_n)$ be a k-dimensional embedding of the vertices. The Rayleigh quotient of Ψ is defined as*

$$R(\Psi) := \frac{\sum_{ij\in E}\|\psi_i - \psi_j\|_2^2}{d\sum_{i\in V}\|\psi_i\|_2^2}.$$

From a k-dimensional embedding with a small Rayleigh quotient, we can apply Cheeger rounding to obtain a sparse cut within its support.

Lemma 8.10 (Cheeger Rounding for k-Dimensional Embedding)**.** *Given graph $G = (V = [n], E)$ and a k-dimensional embedding $\Psi = (\psi_1, \ldots, \psi_n)$ of the vertices, there exists an efficient algorithm to find a subset $S \subseteq \mathrm{supp}(\Psi)$ such that*

$$\phi(S) \le \sqrt{2R(\Psi)},$$

where $\mathrm{supp}(\Psi) := \{i \in [n] \mid \psi_i \neq \vec{0}\}$ is the support of Ψ.

Proof. Expanding the k-dimensional embedding coordinate-wise and applying Lemma 3.6,

$$R(\Psi) = \frac{\sum_{ij\in E}\sum_{l=1}^k(\psi_i(l) - \psi_j(l))^2}{d\sum_{i\in V}\sum_{l=1}^k \psi_i(l)^2} = \frac{\sum_{l=1}^k\sum_{ij\in E}(\psi_i(l) - \psi_j(l))^2}{d\sum_{l=1}^k\sum_{i\in V}\psi_i(l)^2} \ge \min_l \frac{\sum_{ij\in E}(\psi_i(l) - \psi_j(l))^2}{d\sum_{i\in V}\psi_i(l)^2}.$$

Let $x_l(i) := \psi_i(l)$ for $i \in [n]$ and $1 \le l \le k$. The above implies that there is a coordinate l such that $R(x_l) \le R(\Psi)$. Applying Cheeger rounding in Lemma 8.4 on x_l gives the lemma. □

The spectral embedding provides an initial k-dimensional embedding with a small Rayleigh quotient.

Lemma 8.11 (Rayleigh Quotient of the Spectral Embedding)**.** *Let $u_1, \ldots, u_n \in \mathbb{R}^k$ be the spectral embedding of the vertices in Definition 8.5. The Rayleigh quotient of the spectral embedding is*

$$R(U) := \frac{\sum_{ij\in E}\|u_i - u_j\|_2^2}{d\sum_{i\in V}\|u_i\|_2^2} = \frac{1}{k}\sum_{l=1}^k \lambda_l \le \lambda_k.$$

Proof. Since $u_i(l) = v_l(i)$ where v_l is the l-th eigenvector of $\mathcal{L}$,

$$R(U) = \frac{\sum_{ij\in E}\sum_{l=1}^k(u_i(l) - u_j(l))^2}{d\sum_{i\in V}\sum_{l=1}^k u_i(l)^2} = \frac{\sum_{l=1}^k\sum_{ij\in E}(v_l(i) - v_l(j))^2}{d\sum_{l=1}^k\sum_{i\in V}v_l(i)^2} = \frac{\sum_{l=1}^k \lambda_l d}{dk} = \frac{1}{k}\sum_{l=1}^k \lambda_l,$$

where we used $\sum_{i\in V} v_l(i)^2 = \|v_l\|_2^2 = 1$ and $\sum_{ij\in E}(v_l(i)^2 - v_l(j)^2) = d\cdot R(v_l)\cdot\sum_{i\in V} v_l(i)^2 = d\cdot\lambda_l$. □

The Basic Idea

We are now ready to analyze the ideal scenario. Starting from the spectral embedding U with a Rayleigh quotient at most λ_k, the plan is to construct k embeddings $\Psi_1, \ldots, \Psi_k$ such that $\mathrm{supp}(\Psi_l) \subseteq S_l$. If this can be done in a way that the Rayleigh quotient of Ψ_l is small for all $1 \le l \le k$, then Cheeger rounding in Lemma 8.10 can be applied to each Ψ_l to obtain a sparse cut supported on S_l.

The question is how to construct Ψ_l and upper bound its Rayleigh quotient. The most natural way to define Ψ_l is to zero out everything outside S_l:

$$\Psi_l = (\psi_{l,1}, \ldots, \psi_{l,n}) \quad \text{where} \quad \psi_{l,i} = \begin{cases} u_i & \text{if } i \in S_l \\ 0 & \text{otherwise.} \end{cases}$$

The Rayleigh quotient of Ψ_l is

$$R(\Psi_l) = \frac{\sum_{ij \in E} \|\psi_{l,i} - \psi_{l,j}\|_2^2}{d \sum_{i \in V} \|\psi_{l,i}\|_2^2} = \frac{\sum_{ij \in E, i \in S_l, j \in S_l} \|u_i - u_j\|^2 + \sum_{ij \in E, i \in S_l, j \notin S_l} \|u_i\|^2}{d \sum_{i \in S_l} \|u_i\|_2^2}$$

To bound $R(\Psi_l)$, we compare its numerator and denominator with $R(U)$ term-by-term.

For the denominator, $\sum_{i \in S_l} \|u_i\|^2 = 1$ by the assumption of the ideal scenario, and $\sum_{i \in V} \|u_i\|^2 = k$ by (8.1). Thus, the denominator of $R(\Psi_l)$ is $1/k$ times of that of $R(U)$.

For the numerator, edges $ij \in E$ with $i \in S_l$ and $j \in S_l$ contribute equally to $R(\Psi_l)$ and $R(U)$. For edges with $i \in S$ and $j \notin S$, the contribution to $R(\Psi_l)$ is $\|u_i\|^2$, while the contribution to $R(U)$ is $\|u_i - u_j\|^2$. By Claim 8.12, $\|u_i\|_2 \le 2\|u_i - u_j\|_2 / d(i,j) \le 2\|u_i - u_j\|_2 / \delta$.

Combining these bounds,

$$R(\Psi_l) = \frac{\sum_{ij \in E, i \in S_l, j \in S_l} \|u_i - u_j\|_2^2 + \sum_{ij \in E, i \in S_l, j \notin S_l} \|u_i\|_2^2}{d \sum_{i \in S_l} \|u_i\|_2^2} \lesssim \frac{k}{\delta^2} \cdot \frac{\sum_{ij \in E} \|u_i - u_j\|_2^2}{d \sum_{i \in V} \|u_i\|_2^2} = \frac{k}{\delta^2} \cdot R(U).$$

Applying Cheeger rounding in Lemma 8.10 on each Ψ_l gives a set $S_l' \subseteq S_l$ with $\phi(S_l') \lesssim \sqrt{k\lambda_k}/\delta$. Assuming δ is a constant, the sets $S_1', \ldots, S_l'$ are disjoint sparse cuts with edge conductance $O(\sqrt{k\lambda_k})$. This is the basic idea behind the hard direction of Theorem 8.2.

Claim 8.12. *For two vectors u_i, u_j, $\|u_i\|_2 \le 2\|u_i - u_j\|_2 / d(i,j)$ where $d(i,j)$ is the radial projection distance in Definition 8.7.*

Proof. By the triangle inequality,

$$d(i,j) = \left\| \frac{u_i}{\|u_i\|} - \frac{u_j}{\|u_j\|} \right\| \le \left\| \frac{u_i}{\|u_i\|} - \frac{u_j}{\|u_i\|} \right\| + \left\| \frac{u_j}{\|u_i\|} - \frac{u_j}{\|u_j\|} \right\|.$$

For the second term, again by the triangle inequality,

$$\left\| \frac{u_j}{\|u_i\|} - \frac{u_j}{\|u_j\|} \right\| = \frac{\big|\|u_j\| - \|u_i\|\big|}{\|u_i\|} \le \frac{\|u_i - u_j\|}{\|u_i\|}.$$

Therefore, $d(i,j) \le 2\|u_i - u_j\| / \|u_i\|$. Rearranging gives the claim. □

Improved Dependency on k for Fewer Disjoint Sparse Cuts: If we focus on finding only $k/2$ disjoint sparse cuts, we can achieve a better dependency on k. Sort the sets $S_1, \ldots, S_k$ so that

$$\sum_{ij \in E, i \in S_1} \|u_i - u_j\|_2^2 \leq \cdots \leq \sum_{ij \in E, i \in S_k} \|u_i - u_j\|_2^2.$$

Since each edge can contribute to at most two groups, an averaging argument implies that

$$\sum_{ij \in E, i \in S_{k/2}} \|u_i - u_j\|_2^2 \lesssim \frac{1}{k} \sum_{ij \in E} \|u_i - u_j\|_2^2.$$

This reduces the numerator by a factor of k and gives $R(\Psi_{k/2}) \lesssim R(U)/\delta^2$. Assuming δ is a constant, the sets $S_1', \ldots, S_{k/2}'$ are disjoint sparse cuts with edge conductance $O(\sqrt{\lambda_k})$.

8.4 General Case: Space Partitioning and Smooth Localization

We follow a similar strategy in the general case. We aim to find k disjoint subsets $S_1, \ldots, S_k \subseteq V$ such that:

- Each subset S_i has mass $\Omega(1)$, or equivalently $\sum_{j \in S_i} \|u_j\|_2^2 \gtrsim \frac{1}{k} \sum_{i=1}^n \|u_i\|_2^2$;
- The clusters are well-separated, i.e., $d(S_i, S_j) \geq 2\delta$ for all $i \neq j$.

We describe below how to partition the space to construct these disjoint subsets, and then how to use smooth localization to obtain disjoint sparse cuts from these subsets.

Space Partitioning

The difficult case is when the points are evenly distributed across the space, in which it is not clear how to identify the disjoint subsets $S_1, \ldots, S_k$ with the required properties.

A simple approach, presented in [Lee13], is to partition the directions in $\mathbb{S}^{k-1}$ (the k-dimensional sphere) into cubes with side length $L = \frac{1}{2k}$. All points $u_i \in \mathbb{R}^k$ with directions $u_i/\|u_i\|_2$ in the same cube Q are grouped into a block B. The diameter of each cube Q is $L\sqrt{k} = \frac{1}{2\sqrt{k}}$. By the spreading property in Lemma 8.8, the points in each block B has mass at most $1 + \frac{1}{2k}$.

To ensure that points in different cubes are sufficiently far apart, we define $\tilde{Q} \subseteq Q$ as the set of points on $\mathbb{S}^{k-1}$ that are at least $\frac{L}{4k^2}$ away (in Euclidean distance) from every side of Q. Similarly, $\tilde{B} \subseteq B$ denotes the points $u_i \in \mathbb{R}^k$ with directions $u_i/\|u_i\|_2$ in $\tilde{Q}$. By this construction,

$$\mathrm{vol}(\tilde{Q}) = \left(1 - \frac{1}{4k^2}\right)^k \cdot \mathrm{vol}(Q) \geq \left(1 - \frac{1}{4k}\right) \cdot \mathrm{vol}(Q).$$

Thus, if we choose a uniformly random axis-parallel translation of the cube partition, the expected total mass of points in the shrunk blocks is at least $k - \frac{1}{4}$.

To construct disjoint $S_1, \ldots, S_k \subseteq V$ where each S_i has mass at least $\frac{1}{4}$, we sort the shrunk blocks by non-increasing mass and greedily form $S_1, \ldots, S_k$ so that each S_i has mass at least $\frac{1}{4}$. This is always possible since no block has mass greater than $1 + \frac{1}{2k}$ (in the worst case, the last subset S_k has mass $(k - \frac{1}{4}) - (k-1)(1 + \frac{1}{2k}) \geq \frac{1}{4}$).

To summarize, the result achieved in this subsection is as follows.

Lemma 8.13 (Disjoint Subsets). *There are disjoint subsets $S_1, \ldots, S_k$ such that*

- *Each subset S_i has mass at least $\frac{1}{4}$, i.e., $\sum_{j \in S_i} \|u_j\|_2^2 \geq \frac{1}{4}$;*
- *The clusters are well-separated such that for all $i \neq j$,*

$$d(S_i, S_j) \geq \frac{L}{2k^2} = \frac{1}{4k^3}.$$

Smooth Localization

Given disjoint subsets $S_1, \ldots, S_k$ in Lemma 8.13, the goal is to construct k embeddings $\Psi_1, \ldots, \Psi_k : V \to \mathbb{R}^k$ with small Rayleigh quotients and $\mathrm{supp}(\Psi_i) \subseteq S_i$, as in the ideal scenario.

The difference from the ideal scenario is that there are points not in $S_1 \cup \cdots \cup S_k$. Suppose there is a point $j \notin S_1 \cdots \cup S_k$ but very close to some point $i \in S_l$. In this case, if Ψ_l is defined by zeroing out all points outside S_l, the length of the edge ij in Ψ_l would be $\|u_i\|_2$, which could be much larger than $\|u_i - u_j\|_2$. The ratio could be unbounded, and the term-by-term analysis would fail.

To handle this issue, we use the condition $d(S_i, S_j) \geq 2\delta$ to provide some room to "smoothly" decrease the length of the points close to S_l to zero.

Definition 8.14 (Smooth Localization). *Let δ be a parameter and $S_1, \ldots, S_k$ be disjoint subsets. For each $1 \leq l \leq k$ and each point j, let $d(j, S_l) = \min_{i \in S_l} d(i, j)$ and define*

$$c_j := \max\left\{1 - \frac{d(j, S_l)}{\delta}, 0\right\} \quad \text{and} \quad \psi_{l,j} := c_j u_j,$$

where $\Psi_l = (\psi_{l,1}, \ldots, \psi_{l,k})$ is the k-dimensional embedding that we construct and $U = (u_1, \ldots, u_n)$ is the spectral embedding in Definition 8.5.

Note that if $d(j, S_l) \geq \delta$, then $\psi_{l,j} = 0$, so the embeddings $\Psi_1, \ldots, \Psi_k$ are disjointly supported if $d(S_i, S_j) \geq 2\delta$ for all $i \neq j$.

On the other hand, if $d(j, S_l) \leq \delta$ then c_j decreases linearly with distance, with a slope of $1/\delta$. This smooth localization is designed to ensure that the term-by-term analysis works.

Lemma 8.15 (Distortion from Smooth Localization). *Using the same notations in Definition 8.14, for all $ij \in E$,*

$$\|\psi_{l,i} - \psi_{l,j}\|_2 \leq \left(1 + \frac{2}{\delta}\right) \cdot \|u_i - u_j\|_2.$$

Proof. Following the definitions in Definition 8.14,

$$\|\psi_{l,i} - \psi_{l,j}\| = \|c_i u_i - c_j u_j\| = \|c_i u_i - c_j u_i + c_j u_i - c_j u_j\| \leq |c_i - c_j| \cdot \|u_i\| + |c_j| \cdot \|u_i - u_j\|.$$

Note that $|c_i - c_j| \leq d(i,j)/\delta$, as $d(j, S) - d(i, S) \leq d(i, j)$ by the triangle inequality. Combining this with Claim 8.12, the first term is

$$|c_i - c_j| \cdot \|u_i\| \leq \frac{d(i,j)}{\delta} \cdot \frac{2\|u_i - u_j\|}{d(i,j)} = \frac{2}{\delta}\|u_i - u_j\|.$$

Since the second term is at most $\|u_i - u_j\|$, the proof is complete. □

Putting Together

We put the pieces together to prove a weaker version of the hard direction of Theorem 8.2.

First, compute the spectral embedding $U = (u_1, \ldots, u_n)$ as in Definition 8.5.

Next, apply the space partitioning scheme in Lemma 8.13 to obtain the disjoint subsets $S_1, \ldots, S_k$, each with mass at least $1/4$, and pairwise well-separated such that $d(S_i, S_j) \geq 1/4k^3$. Note that this step relies on the spreading property in Lemma 8.8.

Then, apply smooth localization from Definition 8.14 with $\delta := 1/8k^3$ to obtain k-dimensional embeddings $\Psi_1, \ldots, \Psi_k$. By the choice of δ, the embeddings $\Psi_1, \ldots, \Psi_k$ are disjointly supported. By construction, for each $1 \leq l \leq k$, $\psi_{l,i} = u_i$ for all $i \in S_l$, and so

$$\sum_{i \in V} \|\psi_{l,i}\|_2^2 \geq \sum_{i \in S_l} \|u_i\|_2^2 \geq \frac{1}{4} \quad \Longrightarrow \quad \sum_{i \in V} \|\psi_{l,i}\|_2^2 \gtrsim \frac{1}{k} \sum_{i \in V} \|u_i\|_2^2.$$

Combining with Lemma 8.15 and Lemma 8.11, for each $1 \leq l \leq k$,

$$R(\Psi_l) = \frac{\sum_{ij \in E} \|\psi_{l,i} - \psi_{l,j}\|_2^2}{d \sum_{i \in V} \|\psi_{l,i}\|_2^2} \lesssim \frac{k}{\delta^2} \cdot \frac{\sum_{ij \in E} \|u_i - u_j\|_2^2}{d \sum_{i \in V} \|u_i\|_2^2} \leq \frac{k}{\delta^2} \cdot \lambda_k \asymp k^7 \lambda_k.$$

Finally, applying Cheeger rounding from Lemma 8.10 to each Ψ_l yields disjoint subsets $S_1', \ldots, S_l'$, each with conductance at most $O(\sqrt{k^7 \lambda_k})$. This completes the proof.

Discussions

The tighter bound $O(k^2 \sqrt{\lambda_k})$ is obtained using a more sophisticated random partitioning technique developed in metric embedding, namely the padded decomposition.

The bound $\phi_{k/2}(G) \lesssim \mathrm{polylog}(k) \cdot \sqrt{\lambda_k}$ is achieved by applying a dimension reduction technique to reduce the spectral embedding from k-dimensional space to $O(\log k)$-dimensional space.

The noisy hypercube provides an example where $\phi_k(G) \gtrsim \sqrt{\log k \cdot \lambda_k}$. It remains an open question whether the $\mathrm{poly}(k)$ factor in Theorem 8.2 can be replaced by a $\mathrm{polylog}(k)$ factor.

Question 8.16. *Is it true that* $\phi_k(G) \lesssim \mathrm{polylog}(k) \cdot \sqrt{\lambda_k}$*?*

Although Theorem 8.2 provides a conceptually stronger generalization, its quantitative bounds are weaker than that in [ABS10]. Finding a common generalization of these two results is an interesting open problem for future research.

8.5 Alternative Randomized Rounding Algorithms

The algorithm in [LRTV12] is elegant and simple to describe.

It is proved in [LRTV12] that this algorithm will return $\Omega(k)$ subsets with constant probability. The proof is by computing the expectation and the variance of the numerator and the denominator, using properties of Gaussian random variables.

See also [MMSV24] for a recent generalization of the higher-order Cheeger inequality, obtained using the orthogonal separator technique.

Algorithm 4 Randomized Rounding on Spectral Embedding [LRTV12]

Require: An undirected graph $G = (V, E)$ with $V = [n]$ and $m = |E|$, and a parameter k.

1: Compute the spectral embedding $u_1, \ldots, u_n \in \mathbb{R}^k$ in Definition 8.5.

2: Pick k independent Gaussian vectors $g_1, \ldots, g_k \in N(0,1)^n$. Construct disjointly supported vectors $h_1, \ldots, h_k \in \mathbb{R}^n$ as follows:

$$h_i(j) = \begin{cases} \langle u_j, g_i \rangle & \text{if } i = \text{argmax}_{i \in [k]} \{\langle u_j, g_i \rangle\} \\ 0 & \text{otherwise.} \end{cases}$$

3: Apply Cheeger rounding from Lemma 3.5 to each h_i to obtain a set $S_i \subseteq \text{supp}(h_i)$ and $\phi(S_i) \leq \sqrt{2R_{\mathcal{L}}(h_i)}$.

4: **return** all S_i with $\phi(S_i) \lesssim \sqrt{\log k \cdot \lambda_k}$.

References

[ABS10] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 563–572. IEEE Computer Society, 2010. 4, 5, 79, 87

[Lee13] James R. Lee. No frills proof of higher-order cheeger inequality. 2013. 85

[LOT14] James R. Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *J. ACM*, 61(6):37:1–37:30, 2014. 4, 79, 80

[LRTV12] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh S. Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC*, pages 1131–1140. ACM, 2012. 4, 79, 80, 87, 88

[MMSV24] Konstantin Makarychev, Yury Makarychev, Liren Shan, and Aravindan Vijayaraghavan. Higher-order cheeger inequality for partitioning with buffers. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2236–2274. SIAM, 2024. 87

[NJW01] Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001. 80

Chapter A

Linear Algebra

A.1 Eigenvalues and Eigenvectors

Definition A.1 (Eigenvalues and Eigenvectors)**.** *Let A be an $n \times n$ matrix. A nonzero vector v is called an eigenvector of A if $Av = \lambda v$ for some scalar λ. A scalar λ is called an eigenvalue of A if there exists an eigenvector v with $Av = \lambda v$.*

The multi-set of eigenvalues of A is given by the roots of the characteristic polynomial. While this viewpoint is not often used in this course, it plays a central role in recent breakthroughs in spectral graph theory using interlacing polynomials (see [Spi19]).

Definition A.2 (Characteristic Polynomial)**.** *Let A be an $n \times n$ matrix. The characteristic polynomial of A is defined as $p_A(x) := \det(xI - A)$.*

Two matrices are said to be similar if one is obtained from another by a change of basis.

Definition A.3 (Similar Matrices)**.** *A matrix X is similar to another matrix Y if there exists a non-singular matrix B so that $X = BYB^{-1}$.*

It is well known that similar matrices have the same spectrum.

Fact A.4 (Spectrum of Similar Matrices)**.** *If X is similar to Y, then the multiset of eigenvalues of X and that of Y are the same.*

Proof. One way to see this is that they have the same characteristic polynomial:

$$p_X(x) = \det(xI - X) = \det(xI - BYB^{-1}) = \det(B(xI - Y)B^{-1}) = \det(xI - Y) = p_Y(x),$$

where the second-to-last equality follows from Fact A.27. □

Real Symmetric Matrices

In this course, we mostly work with real symmetric matrices, which have real eigenvalues and an orthonormal basis of eigenvectors.

Theorem A.5 (Spectral Theorem for Real Symmetric Matrices)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then all eigenvalues of A are real numbers, and there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.*

A proof of this fundamental theorem can be found in most linear algebra textbooks. We recommend the proofs in [GR, Tre17].

Remark A.6 (Undirected and Directed Graphs)**.** *The spectral theorem applies to the adjacency and Laplacian matrices of undirected graphs, but not to those of directed graphs. This is why the spectral theory for undirected graphs is much more devleoped. Developing a comparable spectral theory for directed graphs remains an open direction.*

Diagonalization: Using the spectral theorem, real symmetric matrices can be written in the following form. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be an orthonormal basis of eigenvectors guaranteed by Theorem A.5 with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Let V be the $n \times n$ matrix whose i-th column is v_i. Let D be the $n \times n$ diagonal matrix with $D_{i,i} = \lambda_i$. Then the conditions $Av_i = \lambda_i v_i$ for $1 \le i \le n$ can be compactly written as $AV = VD$. Since the columns of V form an orthonormal basis, it follows that $V^\top V = I$ and thus $V^{-1} = V^\top$. Hence, we can rewrite $AV = VD$ as

$$A = VDV^{-1} = VDV^\top.$$

Power of Matrices: For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, the diagonalized form $A = VDV^\top$ simplifies computations. To compute A^k, note that

$$A^k = (VDV^\top)^k = VD^kV^\top,$$

where D^k is computed by raising the diagonal entries of D to the k-th power.

This is particularly useful in analyzing random walks. For instance, P^t, the transition matrix of a random walk after t steps, can be expressed in terms of the eigenvalues of P to bound the mixing time.

Eigen-Decomposition and Pseudoinverse: If $v_1, \ldots, v_n$ form an orthonormal basis, then any $x \in \mathbb{R}^n$ can be written as a linear combination $c_1v_1 + \ldots + c_nv_n$. By orthonormality, for any $1 \le i \le n$,

$$\langle x, v_i\rangle = \langle c_1v_1 + \ldots + c_nv_n, v_i\rangle = \langle c_iv_i, v_i\rangle = c_i.$$

Therefore, for any $x \in \mathbb{R}^n$,

$$x = \langle x, v_1\rangle v_1 + \ldots + \langle x, v_n\rangle v_n = v_1v_1^\top x + \ldots v_nv_n^\top x = \big(v_1v_1^\top + \ldots + v_nv_n^\top\big)x.$$

Since this is true for all $x \in \mathbb{R}^n$, it follows that

$$v_1v_1^\top + \ldots + v_nv_n^\top = I_n.$$

Now, if $v_1, \ldots, v_n$ are also eigenvectors of a matrix $A \in \mathbb{R}^{n\times n}$, then for any $x \in \mathbb{R}^n$,

$$Ax = A(v_1v_1^\top + \ldots + v_nv_n^\top)x = \big(\lambda_1v_1v_1^\top + \ldots + \lambda_nv_nv_n^\top\big)x.$$

This implies that

$$A = \lambda_1 v_1 v_1^\top + \ldots + \lambda_n v_n v_n^\top.$$

Verify that we can also write the inverse using the eigen-decomposition as

$$A^{-1} = \frac{1}{\lambda_1} v_1 v_1^\top + \ldots + \frac{1}{\lambda_n} v_n v_n^\top.$$

This form will also be used to define the "pseudo-inverse" of a matrix A when A is not of full rank.

Definition A.7 (Moore-Penrose Pseudoinverse)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix with eigen-decomposition $A = \sum_{i=1}^n \lambda_i v_i v_i^\top$. The pseudoinverse of A, denoted by $A^\dagger$, is defined as*

$$A^\dagger := \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} v_i v_i^\top.$$

Check the following properties of pseudoinverse.

Fact A.8 (Properties)**.** *Let A be a real symmetric matrix and $A^\dagger$ be its pseudoinverse. Then*

$$AA^\dagger A = A \quad \text{and} \quad A^\dagger A A^\dagger = A^\dagger \quad \text{and} \quad \left(A^\dagger\right)^\dagger = A.$$

Positive Semidefinite Matrices

An important class of real symmetric matrices is the class of positive semidefinite matrices. A real symmetric matrix is called positive semidefinite if all of its eigenvalues are nonnegative. This can be seen as a matrix analog of a non-negative number. The following are some equivalent characterizations of a positive semidefinite matrix.

Fact A.9 (Positive Semidefinite Matrix)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix. The following statements are equivalent.*

1. *A is positive semidefinite, i.e., all eigenvalues of A are non-negative.*
2. *For any $x \in \mathbb{R}^n$, it holds that $x^\top A x \geq 0$, i.e., all quadratic forms are non-negative.*
3. *$A = B^\top B$ for some matrix $B \in \mathbb{R}^{m\times n}$ for some positive integer m.*

The notation $A \succcurlyeq 0$ is used to denote that A is a positive semidefinite matrix.

It is a good exercise to prove this fact. A matrix is called positive definite if all eigenvalues of A are positive. It is left as an exercise to find the equivalent characterizations for positive definite matrices as in Fact A.9.

Check that the set of positive semidefinite matrices forms a convex set. Optimizing a linear function over the set of positive semidefinite matrices with linear constraints is called semidefinite programming. This is an important class of convex optimization problems that can be solved in polynomial time and is a powerful tool in designing approximation algorithms. More background are provided in the relevant chapters.

Exercise A.10. *Prove that for any two positive semidefinite matrices $A, B \in \mathbb{R}^{n\times n}$,*

$$\langle A, B\rangle := \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij} \geq 0.$$

Optimization Formulation for Eigenvalues

The main reason why eigenvalues are related to optimization problems is through the following formulation, which is the quadratic form normalized by the vector length.

Definition A.11 (Rayleigh Quotient). *The Rayleigh quotient of a vector $x \in \mathbb{R}^n$ with respect to a matrix $A \in \mathbb{R}^{n \times n}$ is defined to be*

$$R_A(x) := \frac{x^\top A x}{x^\top x} = \frac{\sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j}{\sum_{i=1}^n x_i^2}.$$

The largest eigenvalue is the maximum value of the Rayleigh quotient, with the corresponding eigenvectors being an optimal solution.

Lemma A.12 (Optimization Formulation for α_1). *Suppose $A \in \mathbb{R}^{n \times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Then*

$$\alpha_1 = \max_{x \in \mathbb{R}^n} \frac{x^\top A x}{x^\top x}.$$

Proof. Let $v_1, v_2, \ldots, v_n$ be the corresponding orthonormal basis of eigenvectors guaranteed by Theorem A.5. As $v_1, \ldots, v_n$ forms a basis of $\mathbb{R}^n$, any vector $x \in \mathbb{R}^n$ can be written as a linear combination $x = c_1 v_1 + \cdots + c_n v_n$. Then, the numerator can be written as

$$x^\top A x = (c_1 v_1 + \cdots + c_n v_n)^\top A (c_1 v_1 + \cdots + c_n v_n) = (c_1 v_1 + \cdots + c_n v_n)^\top (c_1 \alpha_1 v_1 + \cdots + c_n \alpha_n v_n) = \sum_{i=1}^n c_i^2 \alpha_i,$$

where the second equality follows since $v_1, \cdots, v_n$ are eigenvectors and the last equality is because $v_1, \cdots, v_n$ are orthonormal. Similarly, the denominator can be written as

$$x^\top x = (c_1 v_1 + \cdots + c_n v_n)^\top (c_1 v_1 + \cdots + c_n v_n) = \sum_{i=1}^n c_i^2.$$

Thus, the Rayleigh quotient of x is

$$\frac{x^\top A x}{x^\top x} = \frac{\sum_{i=1}^n c_i^2 \alpha_i}{\sum_{i=1}^n c_i^2} \leq \frac{\alpha_1 \sum_{i=1}^n c_i^2}{\sum_{i=1}^n c_i^2} = \alpha_1.$$

On the other hand, note that v_1 attains the maximum, and the lemma follows. □

This result can be extended to characterize other eigenvalues. The following lemma is useful for characterizing the second largest eigenvalue of a matrix when the first eigenvector is known.

Lemma A.13 (Optimization Formulation for α_k). *Suppose $A \in \mathbb{R}^{n \times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and corresponding orthonormal eigenvectors $v_1, \ldots, v_n$. Let T_k be the set of vectors that are orthogonal to $v_1, v_2, \ldots, v_{k-1}$. Then*

$$\alpha_k = \max_{x \in T_k} \frac{x^\top A x}{x^\top x}.$$

Proof. Let $x \in T_k$. Write $x = c_1 v_1 + \cdots + c_n v_n$. Recall that $c_i = \langle x, v_i \rangle$ from the eigen-decomposition part. Since $x \in T_k$, it follows that $c_1 = c_2 = \cdots = c_{k-1} = 0$. Using the same calculation as in Lemma A.12,

$$\frac{x^\top A x}{x^\top x} = \frac{\sum_{i=k}^n c_i^2 \alpha_i}{\sum_{i=k}^n c_i^2} \leq \frac{\alpha_k \sum_{i=k}^n c_i^2}{\sum_{i=k}^n c_i^2} = \alpha_k.$$

On the other hand, $v_k \in T_k$ and $v_k^\top A v_k / v_k^\top v_k = \alpha_k$, so the lemma follows. □

The above result gives a characterization of α_k, but it requires knowledge of the previous eigenvectors. The Courant-Fischer theorem provides a characterization without requiring prior knowledge of the eigenvectors, and this is useful for proving upper and lower bounds on eigenvalues. In words, the Courant-Fischer theorem says that to prove a lower bound on α_k, one needs to exhibit a k-dimensional subspace in which every vector has large Rayleigh quotient, and the best k-dimensional subspace gives the tight lower bound. Similarly, to prove an upper bound on α_k, one needs to exhibit an $(n-k+1)$-dimensional subspace in which every vector has small Rayleigh quotient, and the best $(n-k+1)$-dimensional subspace gives the tight upper bound.

Theorem A.14 (Courant-Fischer Theorem). *Suppose $A \in \mathbb{R}^{n \times n}$ is a real symmetric matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$. Then*

$$\alpha_k = \max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^\top A x}{x^\top x} = \min_{S \subseteq \mathbb{R}^n : \dim(S) = n-k+1} \max_{x \in S} \frac{x^\top A x}{x^\top x}.$$

Proof. We prove the max-min equality. The min-max equality is similar and is left as an exercise.

Let S_k be the k-dimensional subspace spanned by the first k orthonormal eigenvectors $v_1, \ldots, v_k$, i.e. $S_k = \{x \mid x = c_1 v_1 + \cdots + c_k v_k \text{ for some } c_1, \ldots, c_k \in \mathbb{R}\}$. For any $x \in S_k$,

$$\frac{x^\top A x}{x^\top x} = \frac{(c_1 v_1 + \cdots + c_k v_k)^\top A (c_1 v_1 + \cdots + c_k v_k)}{(c_1 v_1 + \cdots + c_k v_k)^\top (c_1 v_1 + \cdots + c_k v_k)} = \frac{\sum_{i=1}^k c_i^2 \alpha_i}{\sum_{i=1}^k c_i^2} \geq \frac{\alpha_k \sum_{i=1}^k c_i^2}{\sum_{i=1}^k c_i^2} = \alpha_k.$$

Therefore,

$$\max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^\top A x}{x^\top x} \geq \min_{x \in S_k} \frac{x^\top A x}{x^\top x} \geq \alpha_k.$$

To prove that the maximum cannot exceed α_k, observe that any k-dimensional subspace must intersect the $(n-k+1)$-dimensional subspace T_k spanned by $\{v_k, v_{k+1}, \ldots, v_n\}$. For any $x \in T_k$,

$$\frac{x^\top A x}{x^\top x} = \frac{\sum_{i=k}^n c_i^2 \alpha_i}{\sum_{i=k}^n c_i^2} \leq \alpha_k.$$

Therefore,

$$\max_{S \subseteq \mathbb{R}^n : \dim(S) = k} \min_{x \in S} \frac{x^\top A x}{x^\top x} \leq \min_{x \in S \cap T_k} \frac{x^\top A x}{x^\top x} \leq \alpha_k.$$

□

One consequence of the Courant-Fischer theorem is the eigenvalue interlacing theorem.

Theorem A.15 (Cauchy's Interlacing Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and B be a $(n-1)\times(n-1)$ principal submatrix of A. Then*

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \ldots \geq \alpha_{n-1} \geq \beta_{n-1} \geq \alpha_n,$$

where $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{n-1}$ are the eigenvalues of A and B, respectively.

Proof. Assume without loss of generality that B is in the top left corner of A, i.e., the first $n-1$ coordinates. It should be clear that $\alpha_k \geq \beta_k$ because the search space for α_k is larger than that for β_k. More formally,

$$\alpha_k = \max_{S\subseteq\mathbb{R}^n:\dim(S)=k} \min_{x\in S} \frac{x^\top Ax}{x^\top x} \geq \max_{S\subseteq\mathbb{R}^{n-1}:\dim(S)=k} \min_{x\in S} \frac{x^\top Ax}{x^\top x} = \max_{S\subseteq\mathbb{R}^{n-1}:\dim(S)=k} \min_{x\in S} \frac{x^\top Bx}{x^\top x} = \beta_k.$$

Next, we show $\beta_k \geq \alpha_{k+1}$. For any $S \subseteq \mathbb{R}^n$ with $\dim(S) = k+1$, its restriction to the first $n-1$ coordinates (i.e., $S \cap \mathbb{R}^{n-1}$) has dimension at least k. Thus, if there is a good $(k+1)$-dimensional subspace for A, then there is a good k-dimensional subspace for B, so β_k can do as well as α_{k+1}. More formally, let S_{k+1} be the $(k+1)$-dimensional subspace that attains maximum for α_{k+1},

$$\alpha_{k+1} = \min_{x\in S_{k+1}} \frac{x^\top Ax}{x^\top x} \leq \min_{x\in S_{k+1}\cap\mathbb{R}^{n-1}} \frac{x^\top Ax}{x^\top x} \leq \max_{S\subseteq\mathbb{R}^{n-1}:\dim(S)=k} \min_{x\in S} \frac{x^\top Ax}{x^\top x} = \beta_k.$$

□

Perron-Frobenius Theorem

The Perron-Frobenius theorem is an important result about the largest eigenvalue and its corresponding eigenvectors of non-negative matrices. To state it, we need the definitions of an irreducible matrix and the spectral radius of a matrix.

Definition A.16 (Irreducible Matrix)**.** *A matrix $A \in \mathbb{R}^{n\times n}$ is irreducible if its underlying directed graph $G = (V, E)$ is strongly connected, where the vertex set of G is $V = [n]$ and the edge set of G is $E = \{ij \mid A_{ij} \neq 0\}$.*

The spectral radius of a real symmetric matrix is simply the eigenvalue with largest absolute value. The following is the more general definition for matrices with complex eigenvalues.

Definition A.17 (Spectral Radius)**.** *The spectral radius $\rho(A)$ of a matrix A is the maximum of the moduli of its eigenvalues.*

The Perron-Frobenius theorem is particularly useful in the study of random walks. See Chapter 8.8 in [GR] and Chapter 8.4 in [HJ13] for more details and proofs.

Theorem A.18 (Perron-Frobenius Theorem)**.** *Let $A \in \mathbb{R}^{n\times n}$ be a non-negative irreducible matrix.*

1. *The spectral radius $\rho(A)$ is an eigenvalue of A with multiplicity one. In particular, if A is also a real symmetric matrix, then the largest eigenvalue is of multiplicity one, and its absolute value is the largest.*
2. *If v is an eigenvector with eigenvalue $\rho(A)$, then all the entries of v are nonzero and have the same sign.*

Matrix Norms

Definition A.19 (Operator Norm)**.** *Let A be an $m \times n$ matrix. The operator norm $\|A\|_{\text{op}}$ of A is defined as*

$$\|A\|_{\text{op}} := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

This norm is also denoted as $\|A\|_2$ to indicate that it relates to the 2-norm of vectors. However, this can sometimes be confused with the Schatten 2-norm of A, also denoted as $\|A\|_2$, so we use the notation $\|A\|_{\text{op}}$ which is also commonly used.

Exercise A.20 (Operator Norm of a Real Symmetric Matrix)**.** *Show that $\|A\|_{\text{op}}$ is equal to the largest eigenvalue of A when A is a real symmetric matrix.*

The following are some simple properties that will be useful.

Fact A.21 (Properties of Operator Norm)**.** *Let $A \in \mathbb{R}^{m \times n}$.*

1. *$\|A\|_{\text{op}} \geq 0$ and $\|A\|_{\text{op}} = 0$ if and only if $A = 0$.*
2. *$\|cA\|_{\text{op}} = |c| \|A\|_{\text{op}}$ for every scalar c.*
3. *$\|A + B\|_{\text{op}} \leq \|A\|_{\text{op}} + \|B\|_{\text{op}}$.*
4. *$\|Ax\|_2 \leq \|A\|_{\text{op}} \|x\|_2$ for every $x \in \mathbb{R}^n$.*
5. *$\|BA\|_{\text{op}} \leq \|B\|_{\text{op}} \|A\|_{\text{op}}$*

A.2 Formulas and Inequalities

In this section, we record some useful formulas and inequalities in this section. A general reference is the book by Horn and Johnson [HJ13].

Inverse

The following formulas are for updating the inverse of a matrix. Proofs can be found online, such as the Wikipedia.

Fact A.22 (Sherman-Morrison Formula)**.** *Suppose $A \in \mathbb{R}^{n \times n}$ is an invertible square matrix and $u, v \in \mathbb{R}^n$ are column vectors. Then $A + uv^\top$ is invertible if and only if $1 + v^\top A^{-1} u \neq 0$, and*

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1} u v^\top A^{-1}}{1 + v^\top A^{-1} u}$$

Fact A.23 (Woodbury Formula)**.** *Let A be a square invertible $n \times n$ matrix, U an $n \times k$ matrix, and V a $k \times n$ matrix. Assuming $(I_k + VA^{-1}U)$ is invertible, then*

$$(A + UV)^{-1} = A^{-1} - A^{-1} U \big(I_k + VA^{-1}U\big)^{-1} VA^{-1}.$$

The following formula is for inverting a block matrix, where the Schur complement is a useful definition.

Fact A.24 (Block Matrix Inversion). *Let A and D be square matrices and*

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

If A and the Schur complement $D - CA^{-1}B$ are invertible, then

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B\big(D - CA^{-1}B\big)^{-1}CA^{-1} & -A^{-1}B\big(D - CA^{-1}B\big)^{-1} \\ -\big(D - CA^{-1}B\big)^{-1}CA^{-1} & \big(D - CA^{-1}B\big)^{-1} \end{pmatrix}.$$

Determinant

Fact A.25 (Laplace Co-Factor Expansion). *Let A be a $n \times n$ matrix. For every $1 \le i \le n$,*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{i,j} \det\left(A_{[n]\setminus i,[n]\setminus j}\right),$$

where $A_{S,T}$ is the submatrix with rows in $S \subseteq [n]$ and columns in $T \subseteq [n]$.

Applying Laplace expansion recursively gives the Leibniz formula.

Fact A.26 (Leibniz Formula). *Let A be a $n \times n$ matrix. Then*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} A_{\sigma(i),i},$$

where S_n is the set of permutations of the set $[n] = \{1, \ldots, n\}$ and $\operatorname{sgn}(\sigma)$ is the sign function of permutation σ, which returns $+1$ for even permutations and -1 for odd permutations.

The following is a simple fact.

Fact A.27 (Product Rule for Determinants).

$$\det(AB) = \det(A)\det(B).$$

The following result, sometimes known as Sylvester's determinant identity, can be used to deduce that the nonzero eigenvalues of AB and BA are the same (with multiplicity).

Fact A.28 (Weinstein-Aronszajn Identity).

$$\det(I + AB) = \det(I + BA).$$

The matrix determinant formula tracks how the determinant changes after a rank-one update.

Fact A.29 (Matrix Determinant Formula).

$$\det\left(M - uv^T\right) = \det(M) \cdot \left(1 - v^T M^{-1} u\right).$$

The Cauchy-Binet formula is a useful tool with applications such as computing the number of spanning trees of a graph.

Fact A.30 (Cauchy-Binet Formula). *Let A be an $m \times n$ matrix and B be an $n \times m$ matrix. Then*

$$\det(AB) = \sum_{S \in \binom{[n]}{m}} \det\left(A_{[m],S}\right) \det\left(B_{S,[m]}\right)$$

The following formula provides the coefficients of the characteristic polynomials.

Fact A.31 (Characteristic Polynomial). *Let A be an $n \times n$ matrix.*

$$\det(\lambda I_n - A) = \sum_{k=0}^{n} \lambda^{n-k} (-1)^k \sum_{S \in \binom{[n]}{k}} \det(A_{S,S}).$$

Using Fact A.31 and Cauchy-Binet formula in Fact A.30, we obtain the following identity for the characteristic polynomial of a sum of outer products.

Fact A.32 (Characteristic Polynomial of Sum of Outer Products). *Let $u_1, \ldots, u_m \in \mathbb{R}^n$.*

$$\det\left(xI - \sum_{i=1}^{m} u_i u_i^\top\right) = \sum_{k=0}^{n} \lambda^{n-k} (-1)^k \sum_{S \in \binom{[m]}{k}} \det_k\left(\sum_{i \in S} u_i u_i^\top\right),$$

where $\det_k(A) = \sum_{S \in \binom{[n]}{k}} \det(A_{S,S})$.

Trace

Definition A.33 (Trace). *The trace of a matrix $A \in \mathbb{R}^{n \times n}$, denoted by $\mathrm{Tr}(A)$, is the sum of the diagonal entries of A.*

Fact A.34 (Cyclic Property of Trace). *For two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$,*

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA).$$

By examining the coefficient of x^{n-1} in the characteristic polynomial $\det(xI - A)$ of $A \in \mathbb{R}^{n \times n}$ in two ways, one can derive the following result.

Fact A.35 (Trace is the Sum of Eigenvalues). *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A \in \mathbb{R}^{n \times n}$. Then*

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i.$$

The following are two advanced results about traces; see [Bha97].

Fact A.36 (Golden-Thompson Inequality).

$$\mathrm{Tr}(e^{A+B}) \leq \mathrm{Tr}(e^A) \cdot \mathrm{Tr}(e^B)$$

Fact A.37 (Lieb-Thirring Inequality). *Let A and B be positive definite matrices and $q \geq 1$. Then*

$$\mathrm{Tr}\left((BAB)^q\right) \leq \mathrm{Tr}(B^q A^q B^q).$$

Matrix Calculus

The formula for differenting the inverse is derived by differentiating the identity $AA^{-1} = I$.

Fact A.38 (Inverse)**.**

$$d(A^{-1}) = -A^{-1}(dA)A^{-1}.$$

Jacobi's formula is obtained by differentiating the cofactor expansion in Fact A.25.

Fact A.39 (Jacobi's Formula)**.**

$$\frac{d}{dt}\det A(t) = \mathrm{Tr}\left(\mathrm{adj}(A(t)) \cdot \frac{dA(t)}{dt}\right) = \left(\det A(t)\right) \cdot \mathrm{Tr}\left(A(t)^{-1} \cdot \frac{dA(t)}{dt}\right).$$

References

[Bha97] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997. 97

[GR] Christopher D. Godsil and Gordon F. Royle. *Algebraic Graph Theory*. 90, 94

[HJ13] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge; New York, 2nd edition, 2013. 35, 41, 94, 95

[Spi19] Daniel A. Spielman. *Spectral and Algebraic Graph Theory*. 2019. 9, 89

[Tre17] Luca Trevisan. *Lecture Notes on Graph Partitioning, Expanders and Spectral Methods*. 2017. 9, 62, 90

Chapter B

Notations

$[n]$: the set of integers $\{1, \ldots, n\}$.
$\mathbb{R}_+$: the set of non-negative real number.
$a \lesssim b$: denotes $a = O(b)$.
$a \gtrsim b$: denotes $a = \Omega(b)$.
$a \asymp b$: denotes $a = \Theta(b)$.

$v \in \mathbb{R}^n$: an n-dimensional column vector v.
$v(i)$: the i-th entry of a vector v.
$\vec{1}$: the all-ones vector.
χ_S: the characteristic vector of a subset S with $\chi_S(i) = 1$ if $i \in S$ and $\chi_S(i) = 0$ otherwise.
$\|v\|_1$: the ℓ_1-norm of v defined as $\sum_i |v(i)|$.
$\|v\|_2$: the ℓ_2-norm of v defined as $\sqrt{\sum_i v(i)^2}$.

$A_{i,j}$: the (i, j)-th entry of a matrix A.
$A(i, j)$: the (i, j)-th entry of a matrix A.
I: the identity matrix.
J: the all-ones matrix.
$M^\top$: the transpose of the matrix M.
$\mathrm{Tr}(M)$: the trace of the matrix M in Definition A.33.
$\|M\|_{\mathrm{op}}$: the operator norm of M.
$\|M\|_F$: the Frobenius norm of M.

$A(G)$: the adjacency matrix of graph G in Definition 2.1.
$\mathcal{A}(G)$: the normalized adjacency matrix of G in Definition 2.17.
α_i: the i-th largest eigenvalue of the adjacency matrix or the normalized adjacency matrix.
$L(G)$: the Laplacian matrix of a graph G in Definition 2.11.
$\mathcal{A}(G)$: the normalized Laplacian matrix of G in Definition 2.17.
λ_i: the i-th smallest eigenvalue of the Laplacian matrix or the normalized Laplacian matrix.

$\deg(v)$: the degree of vertex v.
$\delta(S)$: the set of edges with one endpoint in S and the other endpoint not in S.
$\phi(S)$: the edge conductance of a subset S in Definition 3.1.
$\phi(G)$: the edge conductance of a graph G in Definition 3.1.

$\text{vol}(S)$: the volume of a subset S in Definition 3.1.
$\Phi(S)$: the edge expansion of a subset S in Definition 3.10.
$\Phi(G)$: the edge expansion of a graph G in Definition 3.10.
$\psi(S)$: the vertex expansion of a subset S in Definition 3.11.
$\psi(G)$: the vertex expansion of a graph G in Definition 3.11.
$\partial(S)$: the vertex boundary of a subset S in Definition 3.11.
$E(S,T)$: the set of edges with one endpoint in S and the other endpoint in T.

$d_{TV}(\vec{p},\vec{q})$: the total variation distance of two distributions $\vec{p}$ and $\vec{q}$ in Definition 4.7.
$\tau_\epsilon(P)$: the ϵ-mixing time of a Markov chain P in Definition 4.14.
$\tau(P)$: the mixing time of a Markov chain P in Definition 4.14.