Graph Expansions and Applications

KWOK, Tsz Chiu

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Computer Science and Engineering

The Chinese University of Hong Kong
April 2015

Thesis Assessment Committee

Professor Shengyu ZHANG (Chair)
Professor Lap Chi LAU (Thesis Supervisor)
Professor Andrej BOGDANOV (Committee Member)
Professor Nisheeth K. VISHNOI (External Examiner)

Abstract

We study problems related to graph expansions, which measure how well a graph is connected. Various practical problems can be modelled as finding a small non-expanding subset of vertices, and this is also closely related to the unique games conjecture in the theory of approximation algorithms. Hence our study of graph expansions is both practically and theoretically motivated.

We present our new results about graph expansions. The results include the design and analysis of algorithms for finding non-expanding subsets, hardness of approximation results and algorithmic applications of expanders:

- We prove a generalization of Cheeger's inequality using higher eigenvalues, providing a better analysis of the spectral graph partitioning algorithm.
- We design a local graph partitioning algorithm using random walks on graph, matching the performance guaranteed by Cheeger's inequality.
- We give a tight lower bound for the expansion of graph powers and use it to prove hardness results for small set expansions.
- We use expanders to design fast and simple algorithms for computing matrix ranks, significantly improving over previous works.

摘要

我們研究與圖的擴張值相關的問題。圖的擴張值反映了一個圖的連通程度。很多不同的應用問題都可以轉換成在圖中找一個不擴張的點集，而這亦跟近似理論中的唯一遊戲猜想 (unique games conjecture) 有密切的關係。故此，圖的擴張值在應用和理論上均有研究動機。

我們介紹有關圖的擴張值的新發現。這些發現包括設計和分析找不擴張的子集的算法、近似算法的困難結論，以及擴張圖的算法應用：

- 我們用更多的特徵值去證明一個 Cheeger 不等式的推廣，從而較好地分析譜分割算法 (spectral graph partitioning algorithm)。
- 我們分析圖上的隨機漫走來設計一個局部的圖分割算法。此算法跟 Cheeger 不等式的保證相合。
- 我們給出圖冪的擴張值的緊下界，並用以得到一個關於找小的不擴張子集的困難結論 (hardness result)。
- 我們利用擴張圖來設計一個快而簡單的算法來計算矩陣的秩，新算法顯著地改進了以往的算法。

Acknowledgements

Foremost, I would like to express my deepest gratitude to my advisor, Lap Chi Lau, who has introduced me to researching in theoretical computer science. Lap Chi has been supporting me throughout my PhD study, and from whom I learn the attitude of a good researcher, teacher and advisor. I would also like to thank Andrej Bogdanov, Shengyu Zhang and Nisheeth Vishnoi for forming my thesis committee.

I thank all my friends in the theory group and ACM programming team in the Chinese University of Hong Kong, especially my 117-mates, for the enjoyable life in these years. I also thank all my coauthours; I have learnt a lot from them.

I want to thank James Lee and the Simons Institute for the Theory of Computing for giving me the opportunity to join the algorithmic spectral graph theory workshop and meet top researchers in the area.

Finally, I thank my family for their unconditional support. A special thank goes to my beloved Sin Man Choi, for her love and support.

Contents

Chapter 1

Introduction

The expansion of a graph is a robust measure on how well the graph is connected. Roughly speaking, the expansion of a subset of vertices is the ratio between the size of the boundary and the size of the subset, and the expansion of a graph is the minimum expansion over all small subsets. Hence, if the expansion of the graph is large, then every small subset of vertices expands well. Various definitions of graph expansions arise with different measures on the sizes, and most of them are related to each other. In this thesis, the notion of expansion we study is mainly the graph conductance. It is a well studied notion. For simplicity, in this chapter, we consider only d-regular unweighted simple graphs, of which the conductance is defined by

$$\phi(G) = \min_{S \subseteq V, |S| \leq |V|/2} \frac{|\delta(S)|}{d\,|S|},$$

where $\delta(S)$ is the set of edges crossing S and $V - S$. The graph conductance is an important measure on graphs in many aspects.

Clustering: If we define a graph by putting an edge between each pair of similar objects, then the sets of small conductance are intuitively corresponding to clusters. Finding clusters is a fundamental problem with applications in various areas, including image segmentation [65, 71], data clustering [40, 54, 75], community detection [47] and VLSI design [30].

Mixing time: The mixing time of a graph is the number of steps needed for a random walk starting from an arbitrary vertex to generate a distribution close to the uniform distribution. This is useful in analyzing random sampling algorithms for combinatorial objects, like computing the permanent and the volume of a convex

object [66, 26]. It is proven that the mixing time is nearly inverse polynomial to the conductance [38]. Therefore, one way to upper bound the mixing time is by lower bounding the conductance.

Expanders: Expanders are graphs where every subset of vertices has large expansion. They are useful objects with applications in derandomization, error correcting codes, extractors and pseudorandom generators [34]. Sparse expanders with linear number of edges can be efficiently constructed, and some of the fastest combinatorial algorithms use sparse expanders as gadgets. For example, those expanders can be used to replace high degree vertices in a graph so as to reduce the maximum degree while maintaining the graph connectivities [18].

Small set expansion: Raghavendra and Steurer [58] proposed the small set expansion (SSE) conjecture, which roughly states that it is NP-hard to determine whether all small subsets of a graph are expanding well or there is a small subset that is barely connected to the rest of the graph. They showed that the SSE conjecture implies the unique games conjecture (UGC) and is implied by the UGC with mild assumptions. Since the UGC is an important conjecture in the theory of approximation algorithms, this connection raises theoretical interest in studying the complexity of approximating graph expansions.

1.1 Main results

In this thesis, we present our new results about graph expansions. We prove a generalization of Cheeger's inequality using higher eigenvalues, providing a better analysis of the spectral graph partitioning algorithm. We study random walks using the approach by Lovász and Simonovits to design and analyze a local graph partitioning algorithm. We also use their approach to answer a basic graph theoretical question, which can be applied to obtain a hardness result for approximating graph expansions. As an application, we use expanders to design fast and simple algorithms for computing matrix ranks.

Chapter 3 [43]: Finding a subset of vertices attaining the conductance of the graph is in general computationally intractable, so we aim at finding a good approximation S such that $|S| \leq |V|/2$ and $\phi(S) := |\delta(S)|/(d\,|S|) \approx \phi(G)$ in polynomial time. The spectral graph partitioning algorithm, Algorithm 1, is the first and most commonly used heuristic in finding such a set. Its performance is guaranteed by

Cheeger's inequality [14, 4] (see Section 2.2), which states that

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

where λ_2 is the second smallest eigenvalue of the Laplacian matrix of G, and in the proof it shows that the spectral partitioning algorithm output a set of conductance at most $\sqrt{2\lambda_2}$. This gives an $O(1/\sqrt{\phi})$ approximation algorithm for finding the set of smallest conductance. Since $1/\lambda_2$ and the mixing time of the graph only differ by at most a multiplicative logarithmic factor, Cheeger's inequality also relates the conductance and the mixing time (see Subsection 2.3.1).

In this work, we prove that for any graph G and any $k \geq 2$,

$$\frac{\lambda_2}{2} \leq \phi(G) = O\left(\frac{k\lambda_2}{\sqrt{\lambda_k}}\right).$$

The original Cheeger's inequality only guarantees constant approximation when the conductance of the graph is constant, and is unable to explain the empirical performance of the spectral partitioning algorithm. We show that the spectral partitioning algorithm works well when λ_k is large for a small k. This gives a theoretical justification for the good performance in practice. For example, in the problem of image segmentation, if an image contains only k interesting objects, then λ_{k+1} would be large and our result proves that the spectral partitioning algorithm performs well.

Chapter 4 [41, 42]: Lovász and Simonovits [51] introduced a combinatorial approach to study the relation between the conductance and the mixing time. For a probability distribution $p : V \to \mathbb{R}_{\geq 0}$ over the vertices, let $C(p, x)$ be the cumulative sum of decreasingly sorted probabilities. More precisely, suppose we label the vertices such that $p(u_1) \geq p(u_2) \geq \cdots \geq p(u_n)$. Then we define

$$C(p, x) = \sum_{i=1}^{x} p(u_i),$$

for integral x. Note that when the distribution p is uniform, then the function $C(p, x) = x/n$. Lovász and Simonovits give an upper bound on C after the t-steps lazy random walks. They prove that for any initial disbution p, the function satisfies

$$\frac{x}{n} \leq C(pW^t, x) \leq \frac{x}{n} + \sqrt{x}\left(1 - \frac{\phi(G)^2}{8}\right)^t,$$

where W is the lazy random walk matrix. This shows that the distribution converges to uniform after lazy random walks, and the rate of convergence depends on the conductance.

Local graph partitioning: We present a random walk based local graph partitioning algorithm. Given a graph G and a parameter k, suppose there is a subset $S \subseteq V$ of size at most k with $\phi(S) \leq \phi$, then we can find a set S' with conductance $O(\sqrt{\phi/\epsilon})$ and of size $O(k^{1+\epsilon})$. The algorithm can be implemented locally in the sense that given a good starting point, we are able to find S' in $O(k^{1+\epsilon}/\phi(G)^2)$ time. This runtime does not depend on the size of the original graph and this property is desirable when the graph is massive. The performance guarantee of our algorithm matches Cheeger's inequality and we restrict the size of the output set in addition. We note that if we can output a set of size $O(k)$ with the same guarantee, then we disprove the small set expansion conjecture.

We analyze the random walks using the function introduced by Lovász and Simonovits. Compared to the spectral algorithms, this combinatorial approach is better suited here, since local information of the graph is enough for this approach to give bounds.

Hardness amplification: Assume G is lazy and regular, and let G^t be the graph represented by the transition matrix of the t-steps random walk on G. It is a natural graph theoretical question to ask for the relation between the conductance of G^t and that of G. By Cheeger's inequality, it is easy to see that $\phi(G^t) = \Omega(t\phi(G)^2)$. We prove that $\phi(G^t) = \Omega(\sqrt{t}\phi(G))$, and show that similar relation holds for the small set expansions. This work is motivated by the gap amplification result for small set expansion by Raghavendra and Schramm [57]. They show that if it is hard to distinguish whether a graph has a small set of conductance ϵ or all small sets are of conductance at least $\omega(\sqrt{\epsilon})$, then it is also hard to distinguish whether a graph has a small set of conductance ϵ or all small sets are of conductance $1/2$.

We use the combinatorial approach introduced by Lovász and Simonovits as our main tool.

Chapter 5 [17] The rank of a matrix is the maximum number of independent columns of the matrix and it is a basic measure of the complexity of a matrix. Finding the rank is a fundamental problem in the field of computational linear algebra. It also has applications in graph algorithms and combinatorial optimizations: some of the

fastest algorithms for graph matchings [53, 32], graph connectivities [16, 61, 18] and matroid optimization problems [32, 19] are based on computing the matrix rank and finding the corresponding linear independent columns.

For a matrix $M \in \mathbb{F}^{n \times n}$, we give a randomized algorithm to compute the rank $r = \text{rank}(M)$ of M as well as a set of r linear independent columns in $\tilde{O}(\text{nnz}(M) + r^\omega)$ field operations, where $\text{nnz}(M)$ is the number of non-zeros in M and $\omega < 2.373$ is the matrix multiplication exponent. This improves the previous best known bound of $O(n^2 r^{\omega-2})$ [70]. Surprisingly, our main tool for this linear algebraic result is a sparse probabilistic vertex expander which can be constructed in linear time. We use the expander to efficiently compress the $n \times n$ matrix into $O(r) \times O(r)$ matrix, showing that the rank preserves with high probability because of the expanding property.

Chapter 2

Background

In this chapter, we present background materials for the thesis. We first give notation and definitions for graphs, review some basic linear algebra, and present basic spectral graph theory which relates combinatorial graph properties to the eigenvalues, in Section 2.1. Then we present Cheeger's inequality, which provides a fundamental connection between graph expansion and the eigenvalues, in Section 2.2. After that, we analyze random walks on the graph, and bound the mixing time by the eigenvalues and the conductance in Section 2.3. Finally, we talk about some previous works on spectral graph theory with higher eigenvalues and algorithms for graph partitioning problems, in Section 2.4.

2.1 Spectral graph theory

In this section, we first introduce notation and definitions for graphs in Subsection 2.1.1 and review some basic linear algebra in Subsection 2.1.2. Then we present basic spectral graph theory in Subsection 2.1.3. Finally, we set up some notation conventions for the whole thesis in Subsection 2.1.4.

2.1.1 Graphs

A weighted graph $G = (V, w)$ consists of an n-element set of vertices $V = V(G)$, and a weight function $w : V \times V \to \mathbb{R}_{\geq 0}$, which assigns a non-negative real value to each pair of vertices. If the weight function w satisfies $w(u, v) = w(v, u)$ for any $u \in V, v \in V$, then we call this graph undirected. In this thesis, we only consider undirected graphs. An edge $e = \{u, v\}$ is an unordered pair of vertices such that $w(u, v) > 0$. We say that $e = \{u, v\}$ is incident to the vertices u and v, and call u and v the end-vertices of

e. We also say that e connects the vertices u and v, and that u is adjacent to v. We use $E = E(G) := \{\{u, v\} \mid u \in V, v \in V, w(u, v) > 0\}$ to denote the set of all edges and $m = |E|$ to denote the number of edges[1]. A path is a finite sequence of vertices $(u_0, u_1, \ldots, u_k)$ such that $\{u_{i-1}, u_i\}$ is an edge for any $i = 1, 2, \ldots, k$, and the length of this path is k. We say that this path connects u_0 and u_k. A graph is called connected if for any two vertices u and v, there exists a path connecting u and v. A graph is called disconnected if the graph is not connected. A self loop is an edge that is incident to only one vertex. If a graph does not contain any self loops and $w(u, v) = 1$ for all edges $e = \{u, v\}$, then we call it a simple unweighted graph. The degree $\deg(u)$ of a vertex u, defined as $\deg(u) := \sum_{v \in V} w(u, v)$, is the sum of the weights of the edges incident to u. A graph is called regular if the degrees of all the vertices are the same, and is called d-regular if the degrees all equal to d. The volume $\mathrm{vol}(S)$ of a subset of vertices S, defined as $\mathrm{vol}(S) := \sum_{u \in S} \deg(u)$, is the total degree of the vertices inside S.

The set of cut edges $\delta(S)$ of a subset S, defined as $\delta(S) := \{\{u, v\} : u \in S, v \notin S\}$, is the set of edges crossing S and $V - S$. The cut value of a subset S is denoted as $w(\delta(S)) := \sum_{e \in \delta(S)} w(e)$, which is the total weight of the cut edges. Note that a graph is disconnected if and only if there exists a non-empty proper subset $S \subset V$ such that $w(\delta(S)) = 0$. The neighbor set $N(u)$ of a vertex u is the set of vertices adjacent to u, and the neighbor set $N(S)$ of a subset of vertices $S \subseteq V$ is the set of vertices adjacent to at least one vertex in S.

Graph expansions

Graph expansions are robust measures of how well a graph is connected. There are different notions of graph expansions.

The edge expansion of a subset of vertices $S \subseteq V$, defined as $h_{\mathrm{edge}}(S) := |\delta(S)|/|S|$, is the ratio of number of cut edges to the size of the set. The edge expansion of the graph is defined as

$$h_{\mathrm{edge}}(G) := \min_{S \subseteq V,\, 1 \le |S| \le n/2} h_{\mathrm{edge}}(S).$$

The sparsity of a subset of vertices $S \subseteq V$, defined as $\mathrm{sp}(S) := |\delta(S)|/(|S||V - S|)$, is the ratio of the number of cut edges to the maximum number of possible cut edges of

[1] It is more common to define the edge set E first and assign weights only on the edge set. However, in our study, we mostly only consider the weights and the edge set is less important, so we prefer to define in this way.

S. The sparsity of the graph is defined as

$$\mathrm{sp}(G) := \min_{S \subseteq V,\, 1 \le |S| \le n-1} \mathrm{sp}(S) = \min_{S \subseteq V,\, 1 \le |S| \le n/2} \mathrm{sp}(S),$$

where the last equality holds because if S attains the minimum with $|S| > n/2$, then $V - S$ also attains the minimum since $\delta(S) = \delta(V - S)$. Note that n times the sparsity of a graph and the edge expansion of the graph are within a multiplicative factor of 2. This is due to any for $|S| \le n/2$,

$$\frac{|\delta(S)|}{|S|} \le n \frac{|\delta(S)|}{|S||V - S|} \le 2 \frac{|\delta(S)|}{|S|}.$$

For general weighted graphs, it is more natural to include the weights in the definition of the expansion. The conductance of a subset of vertices $S \subseteq V$, defined as $\phi(S) := w(\delta(S))/\operatorname{vol}(S)$, is the ratio of the total weight of cut edges to the volume of the set. The conductance of the graph is defined as

$$\phi(G) = \min_{S \subseteq V,\, 0 < \operatorname{vol}(S) \le \operatorname{vol}(V)/2} \phi(S).$$

Note that for unweighted d-regular graphs, the notions of the edge expansion and the conductance are essentially the same, differing by exactly a factor of d.

In this thesis, we will mainly study the conductance and its generalizations. One generalization is about the volume of the sets we consider. For $\delta \in (0, 1/2]$, the small set conductance is defined as

$$\phi_\delta(G) = \min_{S \subseteq V,\, 0 < \operatorname{vol}(S) \le \delta \operatorname{vol}(V)} \phi(S).$$

Note that $\phi(G) = \phi_{1/2}(G)$ by the definition.

Another generalization is about the number of parts we partition into. For $1 \le k \le n$ and the k-way conductance is defined as

$$\phi_k(G) = \min_{S_1, S_2, \ldots, S_k} \max_i \phi(S_i),$$

where the minimum is over all non-empty disjoint subsets of vertices $S_1, S_2, \ldots, S_k$. Note that $\phi(G) = \phi_2(G)$ by the definition. We can also define the expansion by considering only the partitions, which means that we also require $\cup_{i=1}^k S_i = V$. The expansion under this definition differs from the k-way conductance by at most a factor of k [45].

There are also definitions for the expansion based on vertices. The vertex expansion

of a subset of vertices $S \subseteq V$, defined as $h_{\text{vertex}}(S) := |N(S) - S|/|S|$, is the ratio of new adjacent vertices to the size of the set. The vertex expansion of the graph is defined as

$$h_{\text{vertex}}(G) := \min_{S \subseteq V,\, 1 \leq |S| \leq n/2} h_{\text{vertex}}(S).$$

2.1.2 Eigenvectors and eigenvalues

In this subsection, we give definitions regarding spectrum of matrices and present some of its basic properties. Unless otherwise specified, all vectors we consider are column vectors.

For a real matrix $M \in \mathbb{R}^{n \times n}$, the eigenvalues and their corresponding eigenvectors are the pairs $\lambda_i \in \mathbb{C}$ and $v_i \in \mathbb{C}^n$ that satisfy $v_i \neq 0$ and $Mv_i = \lambda_i v_i$. In this thesis, we will focus on real symmetric matrices. For those matrices, we can choose n eigenvectors v_i such that they are real, orthogonal to each other, and with unit norm [35]. Their corresponding eigenvalues λ_i are also real. The multiset of these eigenvalues is called the spectrum of M. Throughout this thesis, when we say that v_i and λ_i are the eigenvectors and eigenvalues of a real symmetric matrix, we are referring to a set of n unit real orthogonal eigenvectors. We also call this set of eigenvectors the eigenbasis, and we can write any vector $x \in \mathbb{R}^n$ into eigenbasis representation $x = \sum_{i=1}^{n} c_i v_i$. The matrix M can be written as

$$M = V^T \Lambda V = \sum_{i=1}^{n} \lambda_i v_i v_i^T,$$

where

$$V = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{pmatrix}, \text{ and } \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

The decomposition $M = V^T \Lambda V$ is called the eigendecomposition. Note that V is orthogonal, which means $VV^T = V^T V = I$, and thus $M^k = V^T \Lambda^k V$ for any positive integer k. The equation $M^k = V^T \Lambda^k V$ is naturally generalized to any real k where Λ^k is defined to be $\operatorname{diag}(\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k)$. This is well defined for any real k when $\lambda_i > 0$ for all i.

The Rayleigh quotient of a non-zero vector $x \in \mathbb{R}^n$ is defined as

$$\mathcal{R}_M(x) = \frac{x^T M x}{x^T x}.$$

This concept is important in the study of the spectrum of the matrix. Suppose the eigenvalues are sorted in ascending order: $\lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$. Then

$$\lambda_1 = \min_{x \in \mathbb{R}^n,\, x \neq 0} \mathcal{R}_M(x),$$

and

$$\lambda_n = \max_{x \in \mathbb{R}^n,\, x \neq 0} \mathcal{R}_M(x).$$

The optimums are attained when x is the corresponding eigenvector. The reason is simple. We write $x = \sum_{i=1}^n c_i v_i$ into eigenbasis representation. Then $x^T M x = \sum_{i=1}^n c_i^2 \lambda_i$ and $x^T x = \sum_{i=1}^n c_i^2$. Now

$$\lambda_1 = \frac{\sum_{i=1}^n c_i^2 \lambda_1}{\sum_{i=1}^n c_i^2} \leq \frac{\sum_{i=1}^n c_i^2 \lambda_i}{\sum_{i=1}^n c_i^2} \leq \frac{\sum_{i=1}^n c_i^2 \lambda_n}{\sum_{i=1}^n c_i^2} = \lambda_n.$$

Therefore $\lambda_1 \leq \min_{x \in \mathbb{R}^n,\, x \neq 0} \mathcal{R}_M(x)$ and $\lambda_n \geq \max_{x \in \mathbb{R}^n,\, x \neq 0} \mathcal{R}_M(x)$. Since optimums are attainable by the corresponding eigenvectors, the equalities hold. The equalities can be generalized to the k-th smallest or largest eigenvalues by Courant-Fischer formula:

$$\lambda_k = \min_{\dim(W)=k} \max_{x \in W,\, x \neq 0} \mathcal{R}_M(x),$$

and

$$\lambda_{n-k+1} = \max_{\dim(W)=k} \min_{x \in W,\, x \neq 0} \mathcal{R}_M(x),$$

where the optimums are over all k dimensional subspaces of $\mathbb{R}^n$. The optimums are attained by $W = \text{span}\{v_1, v_2, \ldots, v_k\}$ and $W = \text{span}\{v_n, v_{n-1}, \ldots, v_{n-k+1}\}$ respectively. Another commonly used generalization is the following:

$$\lambda_k = \min_{x \in \text{span}\{v_1, v_2, \ldots, v_{k-1}\}^\perp,\, x \neq 0} \mathcal{R}_M(x), \tag{2.1}$$

and

$$\lambda_{n-k+1} = \max_{x \in \text{span}\{v_n, v_{n-1}, \ldots, v_{n-k+2}\}^\perp,\, x \neq 0} \mathcal{R}_M(x). \tag{2.2}$$

Here $W^\perp$ denote the subspace $\{x \mid \langle x, v \rangle = 0 \text{ for any } v \in W\}$. These equalities can be verified by considering the eigenbasis, using a similar approach as in the case $k = 1$. The optimums are attained by the k-th smallest or largest eigenvectors.

A matrix M is called positive definite if $\lambda_1 > 0$, or equivalently $x^T M x > 0$ for all $x \neq 0$. It is called positive semidefinite if $\lambda_1 \geq 0$, or equivalently $x^T M x \geq 0$ for all x.

Note that a positive definite matrix is invertible and its inverse is $M^{-1} = V^T \Lambda^{-1} V = \sum_{i=1}^{n} (1/\lambda_i) v_i v_i^T$. A positive definite matrix naturally induces a Hilbert space over $\mathbb{R}^n$ with the inner product $\langle x, y \rangle_M = x^T M y$, and hence an induced norm $\|x\|_M = \sqrt{x^T M x}$.

The generalized Rayleigh quotient $\mathcal{R}_{M,N}(x)$ of a vector x with respect to a pair of matrices M and N is defined to be $x^T Mx/(x^T Nx)$. When N is positive definite, we can get similar equalities for generalized Rayleigh quotient as those for Rayleigh quotient. Let $\mathcal{M} = N^{-1/2} M N^{-1/2}$, and let λ_i and $\tilde{v}_i$ be the eigenvalues and eigenvectors of $\mathcal{M}$. We have the Courant-Fischer formula for generalized Rayleigh quotient:

$$\begin{aligned}
\lambda_k &= \min_{\dim(\tilde{W})=k} \max_{\tilde{x} \in \tilde{W}, \tilde{x} \neq 0} \mathcal{R}_{\mathcal{M}}(\tilde{x}) \\
&= \min_{\dim(\tilde{W})=k} \max_{\tilde{x} \in \tilde{W}, \tilde{x} \neq 0} \frac{\tilde{x}^T N^{-1/2} M N^{-1/2} \tilde{x}}{\tilde{x}^T \tilde{x}}.
\end{aligned}$$

By substituting $x = N^{-1/2}\tilde{x}$ and $W = N^{-1/2}\tilde{W}$, we have

$$\begin{aligned}
\lambda_k &= \min_{\dim(W)=k} \max_{x \in W, x \neq 0} \frac{x^T M x}{(N^{1/2}x)^T N^{1/2} x} \\
&= \min_{\dim(W)=k} \max_{x \in W, x \neq 0} \frac{x^T M x}{x^T N x} \\
&= \min_{\dim(W)=k} \max_{x \in W, x \neq 0} \mathcal{R}_{M,N}(x),
\end{aligned}$$

and similarly,

$$\lambda_{n-k+1} = \max_{\dim(W)=k} \max_{x \in W, x \neq 0} \mathcal{R}_{M,N}(x).$$

Let $v_i = N^{-1/2}\tilde{v}_i$. We call these λ_i and v_i the generalized eigenvalues and eigenvectors of M with respect to N[2]. Note that v_i are also the (right) eigenvectors of $N^{-1}M$ with

[2]Do not confuse with the usual definition of generalized eigenvalues and eigenvectors of a single matrix M, which satisfy the equation $(M - \lambda_i \cdot I_{n\times n})^k v_i = 0$ for some integer $k > 0$.

eigenvalue λ_i. Equation 2.1 and Equation 2.2 can also be generalized as follows:

$$\begin{aligned}
\lambda_k &= \min_{\tilde{x}\in\text{span}\{\tilde{v}_1,\tilde{v}_2,\ldots,\tilde{v}_{k-1}\}^\perp,\,\tilde{x}\neq 0} \mathcal{R}_{\mathcal{M}}(\tilde{x}) \\
&= \min_{\tilde{x}\in\mathbb{R}^n,\,\langle\tilde{x},\tilde{v}_1\rangle=\langle\tilde{x},\tilde{v}_2\rangle=\cdots=\langle\tilde{x},\tilde{v}_{k-1}\rangle=0,\,\tilde{x}\neq 0} \frac{\tilde{x}^T N^{-1/2} M N^{-1/2}\tilde{x}}{\tilde{x}^T\tilde{x}} \\
&= \min_{x=N^{-1/2}\tilde{x}\in\mathbb{R}^n,\,\langle x,v_1\rangle_N=\langle x,v_2\rangle_N=\cdots=\langle x,v_{k-1}\rangle_N=0,\,x\neq 0} \frac{x^T M x}{x^T N x} \\
&= \min_{x\in\mathbb{R}^n,\,\langle x,v_1\rangle_N=\langle x,v_2\rangle_N=\cdots=\langle x,v_{k-1}\rangle_N=0,\,x\neq 0} \mathcal{R}_{M,N}(x), \qquad (2.3)
\end{aligned}$$

and similarly,

$$\lambda_{n-k+1} = \max_{x\in\mathbb{R}^n,\,\langle x,v_n\rangle_N=\langle x,v_{n-1}\rangle_N=\cdots=\langle x,v_{n-k+2}\rangle_N=0,\,x\neq 0} \mathcal{R}_{M,N}(x). \qquad (2.4)$$

The optimum can be attained by v_k and v_{n-k+1} respectively.

2.1.3 Graph expansions and eigenvalues

Spectral graph theory connects combinatorial graph properties and the spectrum of matrices naturally associated with the graph. In this subsection, we give definitions and present basic spectral graph theory.

Definitions

Eigenvectors and eigenvalues of graphs: We label the vertices from 1 to n, or $V = [n]$. The adjacency matrix $A = A(G) \in \mathbb{R}^{n\times n}$ of a graph G is defined such that $A(i,j) = w(i,j)$ for all $i \in V$, $j \in V$. The degree matrix $D = D(G) \in \mathbb{R}^{n\times n}$ of a graph G is a diagonal matrix with $A(i,i) = \deg(i)$. The Laplacian matrix $L = L(G)$ is defined as $L = D - A$. The normalized adjacency matrix and normalized Laplacian matrix are defined as $\mathcal{A} = D^{-1/2}AD^{-1/2}$ and $\mathcal{L} = D^{-1/2}LD^{-1/2} = I_{n\times n} - \mathcal{A}$ respectively. Note that both $\mathcal{A}$ and $\mathcal{L}$ are real symmetric, and so they have real eigenvalues and eigenvectors. Let α_i and $\tilde{v}_i$ be the eigenvalues and eigenvectors of $\mathcal{A}$. We assume α_i are sorted in decreasing order, or $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$. Note that $\tilde{v}_i$ are also eigenvectors of $\mathcal{L}$ with corresponding eigenvalues $1 - \alpha_i$. We let $\lambda_i = 1 - \alpha_i$ be these corresponding eigenvalues and hence they are sorted in increasing order. The spectrum of a graph is defined as the multiset $\{\lambda_i\}_{i=1}^n$. We let $v_i = D^{-1/2}\tilde{v}_i$ be the generalized eigenvectors of L with respect to D and call them the eigenvectors of the graph G. Note that $\langle v_i, v_j\rangle_D = \langle\tilde{v}_i, \tilde{v}_j\rangle$ and hence $\|v_i\|_D = 1$. These eigenvectors of the graph will be used frequently throughout

the thesis. We remark that v_i are (right) eigenvectors of $D^{-1}L$, satisfying

$$D^{-1}Lv_i = D^{-1}LD^{-1/2}\tilde{v}_i = D^{-1/2}\lambda_i\tilde{v}_i = \lambda_i v_i.$$

Similarly, $(Dv_i)^T$ are left eigenvectors of $D^{-1}L$, satisfying $(Dv_i)^T D^{-1}L = \lambda_i (Dv_i)^T$. Also since $D^{-1}L\chi_V = 0$, we can always assume $v_1 = \chi_V/\sqrt{\mathrm{vol}(V)}$.

Sweep: The support $\mathrm{supp}(x)$ of a vector $x \in \mathbb{R}^n$, defined as $\mathrm{supp}(x) = \{i \mid x(i) \neq 0\}$, is the set of positions of non-zeros in x. The level sets of a vector $x \in \mathbb{R}^n$ are the subsets of vertices in the form of $S_{>t} = \{i \mid x(i) > t\}$ or $S_{<t} = \{i \mid x(i) < t\}$ for some $t \in \mathbb{R}$. Note that for any vector x, there are at most $2n$ level sets. The sweep conductance $\phi_{\mathrm{sweep}}(x)$ of a vector x is define as

$$\phi_{\mathrm{sweep}}(x) = \min\left\{ \min_{t\in\mathbb{R},\, \mathrm{vol}(S_{>t})\leq \mathrm{vol}(V)/2} \phi(S_{>t}),\ \min_{t\in\mathbb{R}, \mathrm{vol}(S_{<t})\leq \mathrm{vol}(V)/2} \phi(S_{<t}) \right\}.$$

It is the minimum conductance over all level sets with volume not more than half the total volume. We use $\phi_{\mathrm{sweep}}(G) = \phi_{\mathrm{sweep}}(v_2(G))$ to denote the minimum conductance over all level sets of the second eigenvector of G.

Energy: The energy of a vector $x \in \mathbb{R}^n$, defined as $\mathcal{E}(x) = \sum_{i\in V, j\in V, i<j} w(i,j)(x(i) - x(j))^2$, is the weighted squared sum of the difference of their pointwise values. For $1 \leq i < j \leq n$, let $L_{i,j} \in \mathbb{R}^{n\times n}$ be the Laplacian matrix of a single edge $\{i,j\}$, which means that $L_{i,j}(i,i) = L_{i,j}(j,j) = 1$, $L_{i,j}(i,j) = L_{i,j}(j,i) = -1$, and all the other entries are zeros. Note that $L = \sum_{i\in V, j\in V, i<j} w(i,j)L_{i,j}$ and $x^T L_{i,j} x = (x(i) - x(j))^2$. Hence the energy can be written as

$$\mathcal{E}(x) = \sum_{i\in V, j\in V, i<j} w(i,j)(x(i) - x(j))^2 = \sum_{i\in V, j\in V, i<j} w(i,j) x^T L_{i,j} x = x^T L x. \tag{2.5}$$

Also, this shows that L is positive semidefinite as $w(i,j) \geq 0$.

Rayleigh quotient of graphs: For a weighted graph $G = (V, w)$ with $\deg(i) > 0$ for all $i \in V$, the Rayleigh quotient $\mathcal{R}_G(x)$ of a vector $x \in \mathbb{R}^V$ with respect to the graph G is defined as

$$\mathcal{R}_G(x) := \mathcal{R}_{L,D}(x) = \frac{x^T L x}{x^T D x} = \frac{\mathcal{E}(x)}{\|x\|_D^2}.$$

The characteristic vector $\chi_S \in \{0,1\}^V$ of the subset $S \subseteq V$ is defined by $\chi_S(i) = 1$ if and only if $i \in S$. The Rayleigh quotient of the characteristic vector is equal to the

conductance of S:

$$\mathcal{R}_G(\chi_S) = \frac{\mathcal{E}(\chi_S)}{\|\chi_S\|_D^2} = \frac{w(\delta(S))}{\sum_{i\in S}\deg(i)} = \frac{w(\delta(S))}{\mathrm{vol}(S)} = \phi(S).$$

Here $\mathcal{E}(\chi_S) = w(\delta(S))$ since the edges contributes to the energy are exactly those across S and $V - S$.

Basic results

We end this subsection with several basic properties on the graph spectrum. We assume the graphs have positive degrees so that the Rayleigh quotients are well defined. First we bound the range of the spectrum.

Fact 2.1 ([22])**.** The spectrum of the graph satisfies $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le 2$.

Proof. By Equation 2.3, we have

$$\lambda_1 = \min_{x\in\mathbb{R}^V} \mathcal{R}_{L,D}(x) = \min_{x\in\mathbb{R}^V} \frac{\mathcal{E}(x)}{\|x\|_D^2} \ge 0,$$

since both $\mathcal{E}(x)$ and $\|x\|_D^2$ are non-negative. Moreover $\mathcal{E}(x) = 0$ can be attained by the non-zero constant vector χ_V. Therefore $\lambda_1 = 0$. For the upperbound $\lambda_n \le 2$, we consider the matrix $\mathcal{N} = I_{n\times n} + \mathcal{A}$ instead. Let $\beta_1 \ge \beta_2 \ge \cdots \ge \beta_n$ be the eigenvalues of $\mathcal{N}$. Note that $\beta_i = 2 - \lambda_i$ for all i since $\mathcal{N} = 2I_{n\times n} - \mathcal{L}$. Since $\mathcal{N} = D^{-1/2}(D+A)D^{-1/2}$, by Equation 2.4, we have

$$\begin{aligned}
\beta_n &= \min_{x\in\mathbb{R}^V} \mathcal{R}_{D+A,D}(x) \\
&= \min_{x\in\mathbb{R}^V} \frac{x^T(D+A)x}{x^TDx} \\
&= \min_{x\in\mathbb{R}^V} \frac{\sum_{i\in V, j\in V, i<j} w(i,j)(x(i)+x(j))^2}{x^TDx} \\
&\ge 0,
\end{aligned}$$

where the last equality holds by similar argument as Equation 2.5. Hence we have $\lambda_n = 2 - \beta_n \le 2$. □

We can check the connectivity of a graph by examining its spectrum.

Fact 2.2 ([22])**.** A graph G is disconnected if and only if $\lambda_2 = 0$.

Proof. We first show that G is disconnected implies $\lambda_2 = 0$. If G is disconnected, then there exists a non-empty proper subset $S \subset V$ such that $w(\delta(S)) = 0$. Consider the vector space $W^* = \text{span}\{\chi_S, \chi_{V-S}\}$. Since S is a non-empty proper subset, $\dim(W) = 2$. For any vector $x \in W^*$,

$$\begin{aligned}\mathcal{E}(x) &= \sum_{i\in V, j\in V, i<j} w(i,j)(x(i)-x(j))^2 \\ &= \sum_{i\in S, j\in S, i<j} w(i,j)(x(i)-x(j))^2 + \sum_{i\in V-S, j\in V-S, i<j} w(i,j)(x(i)-x(j))^2 \\ &= 0,\end{aligned}$$

where the second equality holds since $w(i,j) = 0$ if i and j are on different sides, and the last equality holds since $x(i) = x(j)$ if i and j are on the same side. By the Courant-Fischer formula,

$$\lambda_2 = \min_{\dim(W)=2} \max_{x\in W, x\neq 0} \mathcal{R}_{L,D}(x) \leq \max_{x\in W^*, x\neq 0} \frac{\mathcal{E}(x)}{\|x\|_D^2} = 0.$$

Since $\lambda_2 \geq \lambda_1 = 0$, we have $\lambda_2 = 0$.

Next we show that $\lambda_2 = 0$ implies G is disconnected. Suppose to the contrary that G is connected. Then we will see that $\mathcal{E}(x) = 0$ implies $x(i) = x(j)$ for any $i \in V$, $j \in V$. Otherwise, we consider any path $(u_0 = i, u_1, \ldots, u_k = j)$ and find the first t such that $x(u_{t-1}) \neq x(u_t)$. The edge $\{u_{t-1}, u_t\}$ contributes strictly positive energy to $\mathcal{E}(x)$ and violates our assumption that $\mathcal{E}(x) = 0$. Hence $\mathcal{E}(x) = 0$ if and only if $x = c\chi_V$ for some constant c. Let W^* be the subspace attaining the minimum in the Courant-Fischer formula $\lambda_2 = \min_{\dim(W)=2} \max_{x\in W, x\neq 0} \mathcal{R}_{L,D}(x)$. Since $\dim(W^*) = 2$ and hence W^* contains a vector $x \neq c\chi_V$, we conclude that

$$\lambda_2 = \max_{x\in W^*, x\neq 0} \frac{\mathcal{E}(x)}{\|x\|_D^2} > 0.$$

□

We remark that the same argument can be extended to show that G has at least k connected components if and only if $\lambda_k = 0$.

Next we present a lower bound on the conductance by the second smallest eigenvalue of G. This is the easy side of Cheeger's inequality.

Fact 2.3 ([4])**.** We have $\lambda_2/2 \leq \phi(G)$.

Proof. Let v_i be the eigenvectors of G. Note that v_1 can be chosen as $\chi_V/\sqrt{\mathrm{vol}(V)}$ with $\lambda_1 = 0$, since $\mathcal{E}(\chi_V) = 0$. Here the $\sqrt{\mathrm{vol}(V)}$ is a normalizing term so that $\|v_1\|_D = 1$. By Equation 2.3, we have

$$\lambda_2 = \min_{x\in\mathbb{R}^n, \langle x, v_1\rangle_D, x\neq 0} \frac{x^T L x}{x^T D x} = \min_{x\in\mathbb{R}^n, \sum_{i\in V}\deg(i)x(i)=0, x\neq 0} \frac{x^T L x}{x^T D x}.$$

Now let S be the set such that $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and $\phi(S) = \phi(G)$, and let

$$x = \chi_S - \langle \chi_S, v_1\rangle_D v_1 = \chi_S - \frac{\langle \chi_S, \chi_V\rangle_D}{\mathrm{vol}(V)}\chi_V = \chi_S - \frac{\mathrm{vol}(S)}{\mathrm{vol}(V)}\chi_V.$$

Then $\langle x, v_1\rangle_D = 0$ and hence

$$\lambda_2 \leq \frac{x^T L x}{x^T D x}.$$

Note that x only has two values: $x(i) = 1 - \mathrm{vol}(S)/\mathrm{vol}(V)$ when $i \in S$ and $x(i) = -\mathrm{vol}(S)/\mathrm{vol}(V)$ otherwise. Since $x(i) - x(j) = 0$ if $\{i, j\} \notin \delta(S)$ and $x(i) - x(j) = \pm 1$ otherwise, we have

$$x^T L x = \sum_{i\in V, j\in V, i<j} w(i,j)(x(i) - x(j))^2 = \sum_{e\in\delta(S)} w(e) = w(\delta(S)),$$

On the other hand,

$$\begin{aligned}
x^T D x &= \sum_{i\in V} \deg(i) x(i)^2 \\
&= \sum_{i\in S} \deg(i) \left(1 - \frac{\mathrm{vol}(S)}{\mathrm{vol}(V)}\right)^2 + \sum_{i\in V-S} \deg(i) \left(\frac{\mathrm{vol}(S)}{\mathrm{vol}(V)}\right)^2 \\
&= \frac{\mathrm{vol}(S)(\mathrm{vol}(V) - \mathrm{vol}(S))^2 + \mathrm{vol}(V-S)\,\mathrm{vol}(S)^2}{\mathrm{vol}(V)^2} \\
&= \frac{\mathrm{vol}(S)\,\mathrm{vol}(V-S)}{\mathrm{vol}(V)} \\
&\geq \frac{\mathrm{vol}(S)}{2},
\end{aligned}$$

where the last inequality holds since $\mathrm{vol}(V - S) = \mathrm{vol}(V) - \mathrm{vol}(S) \geq \mathrm{vol}(V)/2$. Combining the inequalities, we have

$$\lambda_2 \leq \frac{x^T L x}{x^T D x} \leq \frac{w(\delta(S))}{\mathrm{vol}(S)/2} = 2\phi(S).$$

This completes the proof. □

Finally, we show that λ_k have to be small if we have k disjointly supported vectors, each with small Rayleigh quotient.

Lemma 2.1 ([43]). *Suppose $x_1, x_2, \ldots, x_k$ are vectors in $\mathbb{R}^V$ such that $\mathrm{supp}(x_i) \cap \mathrm{supp}(x_j) = \emptyset$ for any $i \neq j$, then*

$$\lambda_k \leq 2 \max_{i \in [k]} \mathcal{R}_G(x_i).$$

Proof. By the Courant-Fischer formula,

$$\lambda_k = \min_{\dim(W)=k} \max_{x \in W, x \neq 0} \mathcal{R}_G(x) \leq \max_{x \in \mathrm{span}\{x_1, x_2, \ldots, x_k\}, x \neq 0} \mathcal{R}_G(x).$$

We will show that for any $x = \sum_{i=1}^k c_i x_i \in \mathrm{span}\{x_1, x_2, \ldots, x_k\}$, $\mathcal{R}_G(x) \leq 2 \max_{i \in [k]} \mathcal{R}_G(x_i)$. This would complete the proof by the formula.

First we show that for any pair of vertices u and v,

$$(x(u) - x(v))^2 \leq 2 \sum_{i=1}^k (c_i(x_i(u) - x_i(v)))^2$$

Note that $x_i(u) \neq 0$ for at most one i since x_i are disjointly supported. Suppose $u \in \mathrm{supp}(x_i)$ and $v \in \mathrm{supp}(x_j)$ for some $i \neq j$, then

$$\begin{aligned}
(x(u) - x(v))^2 &= (c_i x_i(u) - c_j x_j(v))^2 \\
&\leq 2(c_i x_i(u))^2 + 2(c_j x_j(v))^2 \\
&= 2(c_i(x_i(u) - x_i(v)))^2 + 2(c_j(x_j(u) - x_j(v)))^2 \\
&= 2 \sum_{i=1}^k (c_i(x_i(u) - x_i(v)))^2.
\end{aligned}$$

The remaining case is when there is an i such that $x_j(u) = x_j(v) = 0$ for any $j \neq i$. Therefore,

$$(x(u) - x(v))^2 = (c_i(x_i(u) - x_i(v)))^2 \leq 2 \sum_{i=1}^k (c_i(x_i(u) - x_i(v)))^2.$$

Now we have

$$\begin{aligned}\mathcal{E}(x) &= \sum_{u\in V, v\in V, u<v} w(u,v)(x(u)-x(v))^2 \\ &\le 2 \sum_{u\in V, v\in V, u<v} w(u,v) \sum_{i=1}^{k} (c_i(x_i(u)-x_i(v)))^2 \\ &= 2\sum_{i=1}^{k} c_i^2 \mathcal{E}(x_i).\end{aligned}$$

Also

$$\|x\|_D^2 = \sum_{u\in V} \deg(u) x(u)^2 = \sum_{i=1}^{k} c_i^2 \sum_{u\in V} \deg(u) x_i(u)^2 = \sum_{i=1}^{k} c_i^2 \|x_i\|_D^2,$$

where the second last equality holds since x_i are disjointly supported. Therefore,

$$\mathcal{R}_G(x) = \frac{\mathcal{E}(x)}{\|x\|_D^2} \le 2\frac{\sum_{i=1}^{k} c_i^2 \mathcal{E}(x_i)}{\sum_{i=1}^{k} c_i^2 \|x_i\|_D^2} \le 2 \max_{i\in[k]} \frac{\mathcal{E}(x_i)}{\|x_i\|_D^2} = 2 \max_{i\in[k]} \mathcal{R}(x_i).$$

The last inequality holds by an averaging argument: for any $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}_{>0}$ and $p_i \in \mathbb{R}_{\ge 0}$,

$$\frac{\sum_i p_i a_i}{\sum_i p_i b_i} \le \max_i \frac{a_i}{b_i}.$$

This is because

$$(\sum_i p_i b_i)\left(\max_i \frac{a_i}{b_i}\right) \ge \sum_i p_i b_i \frac{a_i}{b_i} = \sum_i p_i a_i.$$

□

We remark that Lemma 2.1 actually deduces Fact 2.3 easily. For any S with $\mathrm{vol}(S) \le \mathrm{vol}(V)/2$, since χ_S and χ_{V-S} are disjointly supported, we have

$$\lambda_2 \le 2\max\{\mathcal{R}(\chi_S), \mathcal{R}(\chi_{V-S})\} = 2\max\{\phi(S), \phi(V-S)\} = 2\phi(S).$$

Fact 2.3 is then obtained by minimizing over all sets S with $\mathrm{vol}(S) \le \mathrm{vol}(V)/2$. By the same argument we can prove the easy side of higher order Cheeger's inequality, which states that

$$\frac{\lambda_k}{2} \le \phi_k(G).$$

Suppose ϕ_k is attained by the disjoint subsets $S_1, S_2, \ldots, S_k$, since χ_{S_i} are disjointly

supported, we have

$$\lambda_k \leq 2 \max_{i=1}^{k} \{\mathcal{R}(\chi_{S_i})\} \leq 2 \max_{i=1}^{k} \phi(S_i) = 2\phi_k(G).$$

2.1.4 Notation conventions

In this subsection, we introduce our notation conventions in this thesis.

Graphs: The graphs we consider are weighted undirected graphs and the degree of every vertex is strictly positive. Positive degrees are required so that the generalized Rayleigh quotient is well defined. We use G to denote a graph, V and E to denote a vertex set and an edge set respectively, and w to denote a weight function. We use $n = |V|$ and $m = |E|$ to denote the numbers of vertices and edges respectively.

Matrices: We use D, A, L, $\mathcal{A}$ and $\mathcal{L}$ to denote a degree matrix, an adjacency matrix, a Laplacian matrix, a normalized adjacency matrix and a normalized Laplacian matrix respectively. When dealing with multiple graphs, we use A_G or $A(G)$ to specify which graph we are referring to. The same convention apply to all notations that depends on a graph. If the graph we are referring to is clear from context, we will omit the symbol G.

Eigenvectors and eigenvalues: We use α_i and λ_i to denote the eigenvalues of $\mathcal{A}$ and $\mathcal{L}$ respectively. α_i is always sorted in decreasing order and λ_i is always sorted in increasing order. We use $\tilde{v}_i$ to denote the common eigenvectors of $\mathcal{A}$ and $\mathcal{L}$, and $v_i = D^{-1/2}\tilde{v}_i$ to denote the common eigenvectors of $D^{-1}A$ and $D^{-1}L$.

Vectors: We use x, y, z to denote column vectors in $\mathbb{R}^V$, and f, g to denote functions with continuous domain. We use $x(i)$ to denote the i-th element in the vector x. Note that x_i would be the i-th vector in a collection of vectors instead.

Norms: When dealing with generalized Rayleigh quotient, it is common that the vectors are naturally adapted with either the inner product $\langle \cdot, \cdot \rangle_2$ or the inner product $\langle \cdot, \cdot \rangle_D$. We use $\tilde{x}$ to denote a vector which is naturally adapted with $\langle \cdot, \cdot \rangle_2$ and x to denote a vector which is adapted with $\langle \cdot, \cdot \rangle_D$.

2.2 Cheeger's inequality

Fact 2.2 states that a graph is disconnected if and only if $\lambda_2 = 0$. Cheeger's inequality gives a robust version of this statement, stating that the conductance of a graph is small if and only if λ_2 is small.

Theorem 2.1 ([4]). *For any graphs G, we have*

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

The left inequality is usually called the easy direction of Cheeger's inequality, and we have already proven it in Fact 2.3. The right inequality is usually called the hard direction. It can be proved by explicitly constructing a set S with $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and $\phi(S) \leq \sqrt{2\lambda_2}$. One simple way to construct such a set is by the spectral partitioning algorithm.

In this section, we first describe the spectral partitioning algorithm in Subsection 2.2.1. Then we prove Cheeger's inequality in Subsection 2.2.2. Finally we show some examples to demonstrate when Cheeger's inequality gives a tight bound in Subsection 2.2.3.

2.2.1 Spectral partitioning algorithm

In this subsection, we discuss the spectral partitioning algorithm. The algorithm is stated in Algorithm 1.

Algorithm 1: Spectral partitioning algorithm

Input: A graph $G = (V, w)$ and the second eigenvector v_2 of G
Output: A subset of vertices $S \subseteq V$ with $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ and $\phi(S) \leq \sqrt{2\lambda_2}$.

1 Sort the vertices so that $v_2(u_1) \leq v_2(u_2) \leq \cdots \leq v_2(u_n)$;
2 For $k = 1, 2, \ldots, n$, set $S_k = \{u_1, u_2, \ldots u_k\}$ and $T_k = \{u_k, u_{k+1}, \ldots, u_n\}$;
3 Return $\mathrm{argmin}_{S=S_k \text{ or } S=T_k, \mathrm{vol}(S) \leq \mathrm{vol}(V)/2} \phi(S)$;

The algorithm clearly returns a set S with $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$, and we shall prove in Subsection 2.2.2 that $\phi(S) \leq \sqrt{2\lambda_2}$. Now we analyze the runtime of the algorithm. Suppose we are given the second eigenvector v_2. Sorting the vertices costs $O(n \log n)$ operations. Note that $w(\delta(S_i))$ can be computed incrementally. Computing $w(\delta(S_i))$ from $w(\delta(S_{i-1}))$ only needs the information on the edges incident at u_i, and thus costs $O(|\{u_j \mid w(u_i, u_j) > 0\}|)$ if the graph is input by the adjacency list. So the total number of operations to compute $w(\delta(S_i))$ for all i is $O(m)$. Similar argument holds for computing $w(\delta(T_i))$, $\mathrm{vol}(S_i)$ and $\mathrm{vol}(T_i)$. Therefore the total runtime after the computation of the second eigenvector is $O(m + n \log n) = \tilde{O}(m)$.

In general we are not able to compute the exact second eigenvector v_2, as the entries may contain irrational numbers. Instead we compute an approximation x in the sense that $\langle x, v_1 \rangle_D = 0$ and $\mathcal{R}_G(x) \approx \lambda_2$. The performance guarantee of the algorithm also

becomes $\phi(S) \leq \sqrt{2\mathcal{R}_G(x)}$ instead of $\phi(S) \leq \sqrt{2\lambda_2}$. Note that even if we cannot compute the exact eigenvector v_2, the inequality $\phi(G) \leq \sqrt{2\lambda_2}$ still holds, as we know the existence of v_2.

In the following, we demonstrate the power method which gives a $1-\epsilon$ approximation in finding the maximum eigenvalue of a positive semidefinite matrix. Then we will use it to find an approximate second eigenvector of $\mathcal{L}$.

Proposition 2.1. *Let $M \in \mathbb{R}^{n\times n}$ be a positive semidefinite matrix. Let $\mathrm{nnz}(M)$ be the number of non-zeros in a matrix M and let $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq 0$ be the eigenvalues of M. There exists an $O(\mathrm{nnz}(M)\log n/\epsilon)$ randomized algorithm that takes M and an error parameter ϵ as input and with probability $\Omega(1)$ outputs a vector x such that $\mathcal{R}_M(x) \geq (1-\epsilon)\alpha_1$.*

Proof. The algorithm starts with a uniform random vector $y \in \mathbb{S}^{n-1}$, which means that $y \in \mathbb{R}^n$ and $\|y\| = 1$. This can be done by choosing each coordinate of y independently with the Gaussian distribution and then rescaling y to unit length. Then we simply output $x = M^t y$, where $t = \lceil \log n/(2\epsilon)\rceil$. Note that Mx can be computed in $O(\mathrm{nnz}(M))$ operations, and hence the runtime of the above process is $O(\mathrm{nnz}(M)\log n/\epsilon)$.

Suppose $y = \sum_{i=1}^n c_i u_i$ where u_i are the eigenvectors of M. It is known that with probability at least $\Omega(1)$, we have $c_1^2 \geq 1/n$. We shall show that whenever this happens, the output vector x satisfies $\mathcal{R}_M(x) \geq (1-\epsilon)\alpha_1$. Note that

$$\mathcal{R}_M(x) = \frac{x^T M x}{x^T x} = \frac{y^T M^{2t+1} y}{y^T M^{2t} y} = \frac{\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}}{\sum_{i=1}^n c_i^2 \alpha_i^{2t}}.$$

Since $\sum_{i=1}^n c_i^2 = \|y\|^2 = 1$, by the power mean inequality[3], we have

$$\left(\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}\right)^{\frac{1}{2t+1}} \geq \left(\sum_{i=1}^n c_i^2 \alpha_i^{2t}\right)^{\frac{1}{2t}}.$$

Therefore

$$\frac{\sum_{i=1}^n c_i^2 \alpha_i^{2t+1}}{\sum_{i=1}^n c_i^2 \alpha_i^{2t}} \geq \left(\sum_{i=1}^n c_i^2 \alpha_i^{2t}\right)^{\frac{1}{2t}} \geq |c_1|^{1/t}\alpha_1.$$

Since $c_1^2 \geq 1/n$ and $t \geq \log n/(2\epsilon)$, we have

$$|c_1|^{1/t} \geq n^{-1/(2t)} \geq \exp(-\log n \frac{\epsilon}{\log n}) = \exp(-\epsilon) \geq 1-\epsilon.$$

[3]The power mean inequality states that for any non-negative random variable X, $\mathbb{E}[X^p]^{1/p} \geq \mathbb{E}[X^q]^{1/q}$ for any $p \geq q$.

This completes the proof. □

We remark that when the largest k eigenvectors are known, the power method can be applied to find a vector x which is orthogonal to $v_1, \cdots, v_k$ and $\mathcal{R}_M(x) \geq (1-\epsilon)\alpha_{k+1}$ in $O(kn + \text{nnz}(M)\log n/\epsilon)$ time. The proof is the same as Proposition 2.1 except that we first remove the components of the first k largest eigenvectors in the starting vector y in $O(kn)$ time.

There is an easy way to use power method to additively approximate the second eigenvector of a graph G. Since the smallest eigenvector of $\mathcal{L}(G)$ is known to be $\tilde{v}_1 = D^{1/2}\chi_V/\sqrt{\text{vol}(V)}$, we can apply Proposition 2.1 to compute the second largest eigenvector of the matrix $2I_{n\times n} - \mathcal{L}(G)$ and get an approximate vector $\tilde{x} \perp \tilde{v}_1$ with $\mathcal{R}_{2I_{n\times n}-\mathcal{L}(G)}(\tilde{x}) \geq (1-\epsilon)(2-\lambda_2(G)) \geq 2-\lambda_2(G)-2\epsilon$. Therefore, $\mathcal{R}_{\mathcal{L}(G)}(\tilde{x}) \leq \lambda_2(G)+2\epsilon$. Let $x = D^{-1/2}\tilde{x}$, we have $\langle x, v_1\rangle_D = 0$ and $\mathcal{R}_G(x) = \mathcal{R}_{\mathcal{L}(G)}(\tilde{x}) \leq \lambda_2(G) + 2\epsilon$. The total runtime of this process is $O(m\log n/\epsilon)$. This approach gives an additive approximation and is good when λ_2 is large.

To get a multiplicative $(1+\epsilon)$-approximation, we need to set the additive error to be $\epsilon\lambda_2$. Hence the runtime would become $O(m\log n/(\epsilon\lambda_2))$, having an undesirable dependence on λ_2. It is possible to get rid of the dependence of λ_2, and get a $1+\epsilon$ approximation in nearly linear time, using the Laplacian solvers. We consider the pseudo-inverse $\mathcal{L}^+ := \sum_{i:\lambda_i\neq 0}(1/\lambda_i)\tilde{v}_i\tilde{v}_i^T$ instead of $2I_{n\times n} - \mathcal{L}$. The eigenbasis of $\mathcal{L}^+$ is the same as that of $\mathcal{L}$ and the corresponding eigenvalue of v_2 becomes $1/\lambda_2$, which is the largest among the spectrum of $\mathcal{L}^+$. Now the power method requires us to compute $\mathcal{L}^+\tilde{b}$, which is equivalent to solving the linear system $\mathcal{L}\tilde{x} = \tilde{b}$. Spielman and Teng solved this problem approximately in [68] and obtained the following result.

Theorem 2.2 ([68]). *For any Laplacian matrix $L \in \mathbb{R}^{n\times n}$, there is a matrix $Z \in \mathbb{R}^{n\times n}$ such that*

$$(1-\epsilon)Z^+ \preceq L \preceq (1+\epsilon)Z^+.$$

Moreover for any vector b we can compute Zb in expected time $O(m(\log n)^c \log(1/\epsilon))$ for some absolute constant c.

We should view the matrix Z as an approximation of the pseudo-inverse L^+. Combining this theorem and the power method, Spielman and Teng are able to find a good approximation of the second eigenvector of G quickly. In the following, we give an alternate simpler proof of the following theorem.

Theorem 2.3 ([68]). *Given a graph G and an error parameter $\epsilon < 1/4$, we can output with probability $\Omega(1)$ an approximate second vector x such that $\langle x, v_1\rangle_D = 0$ and $\mathcal{R}_G(x) \leq (1+4\epsilon)\lambda_2$ in time $O(m(\log n)^c \log(1/\epsilon)/\epsilon)$ for some absolute constant c.*

Proof. The result is trivial when G is disconnected. So in the following we assume G is connected and thus $\lambda_2 > 0$.

Let Z be the matrix obtained in Theorem 2.2 and $\mathcal{Z} = D^{1/2}ZD^{1/2}$. Since $\mathcal{L} = D^{-1/2}LD^{-1/2}$ and $\mathcal{Z}^+ = D^{-1/2}Z^+D^{-1/2}$, we have

$$(1-\epsilon)\mathcal{Z}^+ \preceq \mathcal{L} \preceq (1+\epsilon)\mathcal{Z}^+.$$

Clearly $\mathcal{Z}b$ can also be computed in $O(m(\log n)^c \log(1/\epsilon))$ expected time for some absolute constant c.

Let α_i and u_i be the eigenvalues and eigenvectors of $\mathcal{Z}$ with $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n = 0$. Note that we have $(1-\epsilon)\alpha_1^{-1} \leq \lambda_2$ since $(1-\epsilon)\mathcal{Z}^+ \preceq \mathcal{L}$. Similar to the argument in the proof of Proposition 2.1, we choose a uniformly random unit vector y with $y \perp u_n = y \perp \tilde{v}_1 = 0$ and let $\tilde{x} = \mathcal{Z}^t y$, where $t = \lceil \log n/(2\epsilon) \rceil$. Suppose $y = \sum_{i=1}^{n-1} c_i u_i$. With probability $\Omega(1)$ we have $c_1^2 \geq 1/n$ and we assume this happens in the following analysis. Note that

$$\mathcal{R}_{\mathcal{Z}^+}(\tilde{x}) = \frac{\tilde{x}^T\mathcal{Z}^+\tilde{x}}{\tilde{x}^T\tilde{x}} = \frac{y^T\mathcal{Z}^t\mathcal{Z}^+\mathcal{Z}^t y}{y^T\mathcal{Z}^{2t}y} = \frac{y^T\mathcal{Z}^{2t-1}y}{y^T\mathcal{Z}^{2t}y} = \frac{\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t-1}}{\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t}}.$$

By the power mean inequality, we have

$$\left(\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t-1}\right)^{\frac{1}{2t-1}} \leq \left(\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t}\right)^{\frac{1}{2t}}.$$

Therefore

$$\frac{\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t-1}}{\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t}} \leq \left(\sum_{i=1}^{n-1} c_i^2\alpha_i^{2t}\right)^{-\frac{1}{2t}} \leq |c_1|^{-\frac{1}{t}}\alpha_1^{-1}.$$

Since $c_1^2 \geq 1/n$ and $t \geq \log n/(2\epsilon)$, we have

$$|c_1|^{-1/t} \leq n^{1/(2t)} \leq \exp(\log n\frac{\epsilon}{\log n}) = \frac{1}{\exp(-\epsilon)} \leq \frac{1}{1-\epsilon}.$$

Also $\alpha_1^{-1} \leq \lambda_2/(1-\epsilon)$. So we have $\mathcal{R}_{\mathcal{Z}^+}(\tilde{x}) \leq \lambda_2/(1-\epsilon)^2$. Finally, since $\mathcal{L} \preceq (1+\epsilon)\mathcal{Z}^+$, we have

$$\mathcal{R}_{\mathcal{L}}(\tilde{x}) \leq (1+\epsilon)\mathcal{R}_{\mathcal{Z}^+}(\tilde{x}) \leq (1+\epsilon)\lambda_2/(1-\epsilon)^2 \leq (1+4\epsilon)\lambda_2.$$

Let $x = D^{-1/2}\tilde{x}$ and we have $\langle x, v_1\rangle_D = 0$ and $\mathcal{R}_G(x) = \mathcal{R}_{\mathcal{L}}(\tilde{x}) \leq (1+4\epsilon)\lambda_2$. This completes the proof. □

2.2.2 Proof of Cheeger's inequality

In this subsection, we prove the performance of Algorithm 1. The proof consists of two parts. We first show that given an approximation of the second eigenvector, we can construct a vector x such that $\mathrm{vol}(\mathrm{supp}(x)) \leq \mathrm{vol}(V)/2$ without increasing the Rayleigh quotient, and every level set of x is also a level set of the approximate second eigenvector. Then we show that one of the level sets of x has small conductance.

Constructing vectors with small support size

We can construct a vector with small support by simply truncating the vector.

Lemma 2.2 ([34])**.** *For a graph G and its second eigenvector v_2, let $v_2^+ = \max\{v_2, 0\}$ and $v_2^- = \max\{-v_2, 0\}$. We have $\mathcal{R}_G(v_2^+) \leq \lambda_2$ and $\mathcal{R}_G(v_2^-) \leq \lambda_2$. Therefore either $x = v_2^+$ or $x = v_2^-$ satisfies $\mathrm{vol}(\mathrm{supp}(x)) \leq \mathrm{vol}(V)/2$ and $\mathcal{R}_G(x) \leq \lambda_2$.*

Proof. We would like to show that

$$\frac{x^T L x}{x^T D x} \leq \lambda_2.$$

Note that for any vertex i,

$$(Lv_2^+)(i) = \sum_{j\in V} w(i,j)(1 - v_2^+(j)) \leq \sum_{j\in V} w(i,j)(1 - v_2(j)) = (Lv_2)(i).$$

Since v_2 is a (right) eigenvector of $D^{-1}L$ with eigenvalue λ_2, we have $Lv_2 = \lambda_2 D v_2$, and thus

$$(Lv_2)(i) = \deg(i)\lambda_2 v_2(i).$$

Therefore

$$\begin{aligned} {v_2^+}^T L v_2^+ &= \sum_{i \in V} v_2^+(i)(Lv_2^+)(i) \\ &\leq \sum_{i \in V, v_2(i)>0} v_2(i)(Lv_2)(i) \\ &= \sum_{i \in V, v_2(i)>0} \deg(i)\lambda_2 v_2(i)^2 \\ &= \lambda_2 {v_2^+}^T D v_2^+. \end{aligned}$$

This shows that $\mathcal{R}_G(v_2^+) \leq \lambda_2$. Applying the same argument to $-v_2$, we have $\mathcal{R}_G(v_2^-) \leq \lambda_2$. □

The above argument only works when we are considering the exact second eigenvector. In the following, we present a more general approach to truncate a vector to obtain a vector with small Rayleigh quotient and small support.

Lemma 2.3. *For a graph G and a non-zero vector x such that $\langle x, v_1 \rangle_D = 0$, let $x_t = x - t\chi_V$, $x_t^+ = \max(x_t, 0)$ and $x_t^- = \max(-x_t, 0)$ for any $t \in \mathbb{R}$. Then we have either $\mathcal{R}_G(x_t^+) \leq \mathcal{R}_G(x)$ or $\mathcal{R}_G(x_t^-) \leq \mathcal{R}_G(x)$. Therefore, by choosing t such that $\mathrm{supp}(x_t^+) \leq \mathrm{vol}(V)/2$ and $\mathrm{supp}(x_t^-) \leq \mathrm{vol}(V)/2$, we have either $y = x_t^+$ or $y = x_t^-$ satisfying $\mathrm{vol}(\mathrm{supp}(y)) \leq \mathrm{vol}(V)/2$ and $\mathcal{R}_G(y) \leq \mathcal{R}_G(x)$.*

Proof. First note that

$$x_t^T L x_t = (x - t\chi_V)^T L (x - t\chi_V) = x^T L x,$$

and

$$\begin{aligned} x_t^T D x_t &= (x - t\chi_V)^T D (x - t\chi_V) \\ &= x^T D x - 2t\langle x, \chi_V \rangle_D + t^2 \|\chi_V\|_D^2 \\ &= x^T D x + t^2 \|\chi_V\|_D^2 \\ &\geq x^T D x, \end{aligned}$$

where $\langle x, \chi_V \rangle_D = 0$ since $v_1 = c\chi_V$ for some non-zero constant c and we assume that $\langle x, \chi_V \rangle_D = 0$. Hence we have $\mathcal{R}_G(x_t) \leq \mathcal{R}_G(x)$.

For any edge $\{i, j\}$, if $x_t(i)$ and $x_t(j)$ are of the same sign, then

$$(x_t(i) - x_t(j))^2 = (x_t^+(i) - x_t^+(j))^2 + (x_t^-(i) - x_t^-(j))^2.$$

Otherwise if $x_t(i)$ and $x_t(j)$ are of different sign, then

$$(x_t(i) - x_t(j))^2 \geq x_t(i)^2 + x_t(j)^2 = (x_t^+(i) - x_t^+(j))^2 + (x_t^-(i) - x_t^-(j))^2.$$

In any case we have $x_t^T L x_t \geq {x_t^+}^T L x_t^+ + {x_t^-}^T L x_t^-$. On the other hand $x_t^T D x_t = \sum_{i \in V} \deg(i) x_t(i)^2 = {x_t^+}^T D x_t^+ + {x_t^-}^T D x_t^-$. Hence we have

$$\mathcal{R}_G(x) \geq \mathcal{R}_G(x_t) = \frac{x_t^T L x_t}{x_t^T D x_t} \geq \frac{{x_t^+}^T L x_t^+ + {x_t^-}^T L x_t^-}{{x_t^+}^T D x_t^+ + {x_t^-}^T D x_t^-} \geq \min\{\mathcal{R}_G(x_t^+), \mathcal{R}_G(x_t^-)\}.$$

We almost complete the proof, except for a subtle issue. One of the x_t^+ or x_t^- may be a zero vector, where the Rayleigh quotient is not defined. If $x_t^- = 0$, then $x_t^+ = x_t$ and thus $\mathcal{R}_G(x) \geq \mathcal{R}_G(x_t^+)$. Similarly if $x_t^+ = 0$, then $\mathcal{R}_G(x) \geq \mathcal{R}_G(x_t^-)$. This completes the proof. □

We remark that every level set of x_t^+ or x_t^- is also a level set of x. Therefore in order to show that Algorithm 1 outputs a set of small conductance, it suffices to show that one of the level sets of x_t^+ or x_t^- has small conductance.

Random thresholding argument

There are multiple ways to prove Cheeger's inequality. Here we present a proof based on a random thresholding argument by Trevisan.

Theorem 2.4 ([74]). *Suppose x is a non-negative non-zero vector, then for some positive $t \in \mathbb{R}_{>0}$, the level set $S_t = \{i \mid x(i) \geq t\}$ of x satisfies $\phi(S_t) \leq \sqrt{2\mathcal{R}_G(x)}$. Since $t > 0$, we also have $S_t \subseteq \mathrm{supp}(x)$.*

Proof. We choose the level set S_t by randomly choosing a value t with probability proportional to its distance to 0. Let $M = \max_{i \in V} x(i)$ and $C = \int_0^M t dt = M^2/2$. Let $\mu : [0, M] \to \mathbb{R}$ satisfying $\mu(t) = t/C$ be a probability distribution over $[0, M]$. We analyze the conductance of the set S_t when t is randomly chosen following this distribution.

The expected cut value of S_t is

$$\begin{aligned}
\mathbb{E}_{t\sim\mu}[w(\delta(S_t))] &= \sum_{i\in V, j\in V} w(i,j)\mathbb{P}_{t\sim\mu}[i\in S_t, j\notin S_t] \\
&= \sum_{i\in V, j\in V, x(i)>x(j)} w(i,j)\int_{x(j)}^{x(i)} \mu(t)dt \\
&= \frac{1}{2C}\sum_{i\in V, j\in V, x(i)<x(j)} w(i,j)(x(i)^2 - x(j)^2) \\
&= \frac{1}{2C}\sum_{i\in V, j\in V, i<j} w(i,j)|x(i)^2 - x(j)^2| \\
&\le \frac{1}{2C}\sqrt{\sum_{i\in V, j\in V, i<j} w(i,j)(x(i)-x(j))^2 \sum_{i\in V, j\in V, i<j} w(i,j)(x(i)+x(j))^2} \\
&\le \frac{1}{2C}\sqrt{\mathcal{E}(x)\sum_{i\in V, j\in V, i<j} w(i,j)2(x(i)^2+x(j)^2)} \\
&= \frac{\sqrt{2}}{2C}\sqrt{\mathcal{E}(x)\sum_{i\in V}\deg(i)x(i)^2} \\
&= \frac{\sqrt{2}\|x\|_D}{2C}\sqrt{\mathcal{E}(x)},
\end{aligned}$$

where the first inequality is due to Cauchy-Schwarz inequality.

On the other hand, the expected volume of S_t is

$$\begin{aligned}
\mathbb{E}_{t\sim\mu}[\mathrm{vol}(S_t)] &= \sum_{i\in V}\deg(i)\mathbb{P}_{t\sim\mu}[i\in S_t] \\
&= \sum_{i\in V}\deg(i)\int_0^{x(i)}\mu(t)dt \\
&= \frac{1}{2C}\sum_{i\in V}\deg(i)x(i)^2 \\
&= \frac{\|x\|_D^2}{2C}.
\end{aligned}$$

Now suppose to the contrary that every level set S_t satisfies $\phi(S) = \delta(S)/\mathrm{vol}(S) > \sqrt{2\mathcal{R}_G(x)}$. Then $\mathbb{E}_{t\sim\mu}[w(\delta(S)) - \sqrt{2\mathcal{R}_G(x)}\,\mathrm{vol}(S)] > 0$. However we have shown that

$$\mathbb{E}_{t\sim\mu}[w(\delta(S)) - \sqrt{2\mathcal{R}_G(x)}\,\mathrm{vol}(S)] \le \frac{\sqrt{2}\|x\|_D}{2C}\sqrt{\mathcal{E}(x)} - \sqrt{2\mathcal{R}_G(x)}\frac{\|x\|_D^2}{2C} = 0.$$

This concludes that there exists a level set S_t satisfying $\phi(S) \leq \sqrt{2\mathcal{R}_G(x)}$. □

Finally we are ready to prove Cheeger's inequality.

Proof of Theorem 2.1. The easy direction is proven in Fact 2.3. By Lemma 2.3 and Theorem 2.4, when given the second eigenvector, Algorithm 1 outputs a set S such that $\phi(S) \leq \sqrt{2\lambda_2}$. Therefore $\phi(G) \leq \phi(S) \leq \sqrt{2\lambda_2}$. This completes the proof. □

2.2.3 Examples

In this subsection, we give several types of graphs which demonstrate when the spectral partitioning algorithm performs well or badly.

Cycles

We use the cycles to demonstrate that the hard side of Cheeger's inequality is tight. An n-cycle $C_n = (V, E)$ is an unweighted graph with n vertices such that $V = [n]$ and $\{i, j\} \in E$ if and only if $i - j \equiv \pm 1 \pmod{n}$. For any non-empty proper subset $S \subset V$, $w(\delta(S)) \geq 2$. Therefore

$$\phi(G) = \min_{S \subseteq V, \mathrm{vol}(S) \leq \mathrm{vol}(V)/2} \frac{w(\delta(S))}{\mathrm{vol}(S)} \geq \frac{4}{\mathrm{vol}(V)} = \frac{2}{n}.$$

On the other hand, the eigenvalues of $\mathcal{L}(G)$ are $1 - \cos(2k\pi/n)$ with corresponding (non-unit) eigenvectors $(1, \cos(2k\pi/n), \cos(4k\pi/n), \ldots, \cos(2(n-1)k\pi/n))$. Note that $1 - \cos(2k\pi/n) = \Theta(k^2/n^2)$ as $n \to \infty$ and thus $\lambda_2 = \Theta(1/n^2)$. So we have

$$\phi(G) \geq \frac{2}{n} = \Theta(\sqrt{\lambda_2})$$

This shows that the hard side of Cheeger's inequality is tight up to a constant factor.

Note that although this analysis shows that λ_2 is a bad approximation of ϕ, the spectral partitioning algorithm actually works well in this example. The cut given by the algorithm has conductance $O(1/n)$, which matches the conductance of the graph.

Hypercubes

We use the hypercubes to demonstrate that the easy side of Cheeger's inequality is tight. For $k \in \mathbb{N}$, the k-dimensional hypercube $H_k = (V, E)$ is an unweighted graph with $n = 2^k$ vertices such that $V = \{0, 1\}^k$ and $\{u, v\} \in E$ if and only if u and v differ in exactly one coordinate. The dimension cuts $S_i = \{u \in V \mid u(i) = 0\}$ satisfies

$w(\delta(S_i)) = 2^{k-1}$ and $\mathrm{vol}(S_i) = k2^{k-1}$, and hence $\phi(S_i) = 1/k$. We will see later that $\phi(H_k) = \phi(S_i)$. Now we construct an eigenbasis of H_k. For any k-bit string $a \in \{0,1\}^k$, the (non-unit) vector $v_a \in \mathbb{R}^V$ is defined by $v_a(u) = (-1)^{\sum_{i=1}^k a(i)u(i)}$. These vectors are orthogonal to each other and satisfies $\mathcal{L}v_a = \lambda_a v_a$ where $\lambda_a = (2/k)\sum_{i=1}^k a(i)$. Therefore $\lambda_2 = 2/k$ with corresponding (non-unit) eigenvectors v_a where $a(i) = 1$ for exactly one index i. By the easy side of Cheeger's inequality we have $\lambda_2/2 \le \phi(H_k)$. Therefore $\phi(H_k) \ge 1/k = \phi(S_i) \ge \phi(H_k)$ and all the inequalities must be equalities. So $\phi(H_k) = \phi(S_i)$ and the easy side of Cheeger's inequality is tight.

Note that in this example while λ_2 gives a good approximation of ϕ, the spectral partitioning algorithm can give a bad cut. The vector $v = \sum_{i=1}^k v_{e_i}$ is an second eigenvector satisfying $\mathcal{L}v = \sum_{i=1}^k \mathcal{L}v_{e_i} = (2/k)\sum_{i=1}^k v_{e_i} = (2/k)v$. Suppose in the spectral partitioning algorithm we are given v and we consider the level sets of v. Since $v(u) = \sum_{i=1}^k v_{e_i}(u) = \sum_{i=1}^k (-1)^{u(i)}$, the level sets are of the form $\{u \in V \mid \sum_{i=1}^k u(i) \le t\}$. Among these sets, the set S of smallest conductance is when $t = \lfloor k/2 \rfloor$, with $w(\delta(S)) = \binom{k}{t}(k-t) = \Theta(\sqrt{k}2^k)$ and $\mathrm{vol}(S) = k\sum_{i=1}^t \binom{k}{t} = \Theta(k2^k)$. Hence $\phi(S) = \Theta(1/\sqrt{k}) = \Theta(\sqrt{\lambda_2})$.

Ladder graphs

We use the ladder graphs to demonstrate that the spectral partitioning algorithm can give a set of conductance $\Omega(n)$ larger than the optimal set. For $k \in \mathbb{N}$ and $\epsilon \in (0,1)$, the (k,ϵ) ladder graph $G = (V,w)$ is an weighted graph with $n = 2k$ vertices. First we construct two unweighted k-cycles C_k and C'_k with vertex sets $V(C_k) = \{1,2,\ldots,k\}$ and $V(C'_k) = \{1',2',\ldots,k'\}$ respectively. Then we connect the two sets of vertices with a matching of weight ϵ, by letting $w(i,i') = \epsilon$ for all $i = 1,2,\ldots,k$. This is the construction of the (k,ϵ) ladder graph. In the following, we set $\epsilon = 100/k^2$ and k to be sufficiently large.

For any non-empty proper subset $S \subset V$, $\delta(S) \ge \min(1, \epsilon k) = 100/k$. This is because if we do not cut through any edges in the cycle, which have unit weight, then we have to cut through all the k edges in the ϵ matching. Therefore

$$\phi(G) \ge \frac{100/k}{\mathrm{vol}(V)/2} = \Theta(1/k^2),$$

and this can be attained by $S = V(C_k)$.

On the other hand, we want to show that $\phi_{\mathrm{sweep}} = \Omega(1/k)$ and conclude that the spectral partitioning algorithm can give a set of conductance $\Omega(n)$ larger. We will see

that there is a second eigenvector such that $v_2(i) = v_2(i')$ for any $i = 1, \ldots, k$, and hence the spectral partitioning algorithm does not separate i and i'. Suppose x is an eigenvector of $\mathcal{L}(C_k)$ with eigenvalue λ, then we can use it to construct two eigenvectors of G. The vector $y \in \mathbb{R}^V$ defined by $y(i) = y(i') = x(i)$ satisfies

$$\begin{aligned}
(\mathcal{L}(G)y)(i) &= \frac{1}{2+\epsilon}(L(G)y)(i) \\
&= \frac{1}{2+\epsilon}(\deg(i)y(i) - y(i-1) - y(i+1) - \epsilon y(i')) \\
&= \frac{1}{2+\epsilon}((2+\epsilon)x(i) - x(i-1) - x(i+1) - \epsilon x(i)) \\
&= \frac{1}{2+\epsilon}(L(C_k)x)(i) \\
&= \frac{2}{2+\epsilon}(\mathcal{L}(C_k)x)(i) \\
&= \frac{2}{2+\epsilon}\lambda x(i) \\
&= \frac{2}{2+\epsilon}\lambda y(i),
\end{aligned}$$

where we denote $x(0) = x(n)$ and $y(0) = y(n)$. By the same argument the equation also holds for i'. Therefore y is an eigenvector of $\mathcal{L}(G)$ with eigenvalue $2\lambda/(2+\epsilon)$. Similarly the vector $z \in \mathbb{R}^V$ defined by $z(i) = -z(i') = x(i)$ satisfies

$$(\mathcal{L}(G)z)(i) = \frac{1}{2+\epsilon}((2+\epsilon)x(i) - x(i-1) - x(i+1) + \epsilon x(i)) = (\frac{2}{2+\epsilon}\lambda + \frac{2\epsilon}{2+\epsilon})z(i),$$

and z is an eigenvector of $\mathcal{L}(G)$ with eigenvalue $2(\lambda+\epsilon)/(2+\epsilon)$. Each v_i in the eigenbasis of $\mathcal{L}(C_k)$ produces two eigenvectors y_i and z_i of $\mathcal{L}(G)$ and they are all orthogonal to each other. These vectors y_i and z_i form an eigenbasis of $\mathcal{L}(G)$. Now the second eigenvector of G is produced by either $v_1(C_k)$ or $v_2(C_k)$. Therefore $\lambda_2(G) = 2\lambda_2(C_k)/(2+\epsilon)$ or $2\epsilon/(2+\epsilon)$, whichever is smaller. Since $\lambda_2(C_k) = 1 - \cos(2\pi/k) \leq 2\pi^2/k^2$ and $\epsilon = 100/k^2$, we conclude that $\lambda_2(G) = 2\lambda_2(C_k)/(2+\epsilon)$ with corresponding eigenvector $y \in \mathbb{R}^V$ satisfying $y(i) = y(i') = v_2(C_k)(i)$. This shows that in the spectral partitioning algorithm, i and i' cannot be separated. Thus none of the ϵ edges are cut and the set S output satisfies $w(\delta(S)) \geq 1$. So we conclude $\phi_{\text{sweep}} = \Omega(1/k)$.

Geometric weighted paths[4]

We use the geometric weighted paths to demonstrate that the constant in the hard side of Cheeger's inequality is tight. For $k \in \mathbb{N}$ and $r \in (0,1)$, the (k,r) weighted path is a weighted graph with $n = 2k+1$ vertices. We label the vertices by $-k, -k+1, \ldots, k$. For $i = 0, 1, \ldots, k-1$, we add edges with weight r^i to the pairs of vertices $\{i, i+1\}$ and $\{-i, -(i+1)\}$. We also add self loops with weight $2r^k/(1-r) = 2(r^k + r^{k+1} + \ldots)$ to the vertices k and $-k$. For $i \geq 0$, let $S_i = \{i+1, \ldots, k\}$. The self loops are added such that whenever we cut an edge, the conductance is always the same:

$$\phi(S_i) = \frac{r^i}{r^i + 2r^{i+1} + \cdots + 2r^{k-1} + 2r^k/(1-r)} = \frac{1}{1 + 2r/(1-r)} = \frac{1-r}{1+r}.$$

These cuts are the ones attaining minimum conductance. First we note that an optimal cut S of the graph can be assumed to be connected, since otherwise the best connected component has conductance not larger than S. Therefore by considering $V - S$ if necessary, we conclude that the optimal conductance is attained by S_i.

Now we shall upper bound λ_2 in order to show that the hard side of Cheeger's inequality is tight. To do so we only need to construct a vector $x \in \mathbb{R}^V$ such that $\langle x, v_1 \rangle_D = 0$ and $\mathcal{R}_G(x)$ is small. Let x be defined by

$$x(i) = \begin{cases} 0 & \text{if } i = 0, \\ r^{(k-i)/2} & \text{if } i > 0, \\ -r^{(k+i)/2} & \text{if } i < 0. \end{cases}$$

By symmetry we have $\langle x, v_1 \rangle_D = 0$. The energy is

$$\begin{aligned} \mathcal{E}(x) &= 2\sum_{i=0}^{k-1} w(i,i+1)(x(i) - x(i+1))^2 \\ &= 2\left(r^{k-1} + \sum_{i=1}^{k-1} r^i (r^{(k-i)/2} - r^{(k-i-1)/2})^2\right) \\ &= 2\left(r^{k-1} + \sum_{i=1}^{k-1} r^i r^{k-i} (1 - r^{-1/2})^2\right) \\ &= 2\left(r^{k-1} + (k-1) r^k (1 - r^{-1/2})^2\right), \end{aligned}$$

[4]These graphs are provided by Yin Tat Lee, who views them as the discrete versions of the tight examples in the manifold settings.

and the squared D-norm is

$$\begin{aligned}\|x\|_D^2 &= \sum_{i=-k}^{k} \deg(i)x(i)^2 \\ &= 2\left(\sum_{i=1}^{k-1}(r^{i-1}+r^i)r^{k-i} + \left(r^{k-1}+\frac{2r^k}{1-r}\right)\right) \\ &= 2\left((k-1)r^k(r^{-1}+1) + r^{k-1} + \frac{2r^k}{1-r}\right).\end{aligned}$$

Hence the Rayleigh quotient is

$$\begin{aligned}\mathcal{R}_G(x) &= \frac{\mathcal{E}(x)}{\|x\|_D^2} \\ &= \frac{2(r^{k-1}+(k-1)r^k(1-r^{-1/2})^2)}{2((k-1)r^k(r^{-1}+1)+r^{k-1}+2r^k/(1-r))} \\ &= \frac{r^{k-1}/(k-1)+r^k(1-r^{-1/2})^2}{r^k(r^{-1}+1)+(r^{k-1}+2r^k/(1-r))/(k-1)} \\ &\to \frac{(1-r^{-1/2})^2}{r^{-1}+1} \quad \text{as } k\to\infty \\ &= \frac{(1-r^{1/2})^2}{1+r}\end{aligned}$$

Therefore for any fixed $r \in (0,1)$ and $\epsilon > 0$, we can find large enough k such that when G is the (k,r) geometric weighted path,

$$\begin{aligned}\frac{\phi(G)}{\sqrt{\lambda_2}} &\geq \frac{\phi(G)}{\sqrt{\mathcal{R}(x)}} \\ &\geq \frac{1-r}{1+r}\sqrt{\frac{1+r}{(1-r^{1/2})^2}}(1-\epsilon) \\ &= \frac{1+r^{1/2}}{\sqrt{1+r}}(1-\epsilon) \\ &\to \sqrt{2}(1-\epsilon) \quad \text{as } r\to 1^-.\end{aligned}$$

Since ϵ is arbitrary and r can be arbitrarily close to 1, we conclude that for any $C < \sqrt{2}$, we can construct a graph G such that $\phi(G) > C\sqrt{\lambda_2(G)}$.

2.3 Random walks

A random walk on a graph G starting with an initial probability distribution p is a sequence of random vertices $u_0 = u, u_1, u_2, \ldots$, such that $\mathbb{P}[u_0 = u] = p(u)$ and $\mathbb{P}[u_{t+1} = v | u_t = u] = w(u, v)/\deg(u)$. This means that the next vertex would be a neighbor of the current vertex with probability proportional to the edge weights. Instead of a particular instance of the random walks, we often study the probability distributions $p_t \in \mathbb{R}^n$, where $p_t(v) = \mathbb{P}[u_t = v]$. We view the probability distributions as row vectors. By the definition, it is easy to see that $p_0 = p$ and $p_{t+1} = p_t D^{-1} A$. We define the random walk matrix to be $W = D^{-1}A$, and hence $p_{t+1} = p_t W$. If the initial distribution $p = \chi_u^T$ only has probability on a single vertex u, we also call p_t as the probability distribution after t steps of random walks starting at the vertex u.

A distribution π is called stationary if the distribution does not change after one step of the random walk, that means $\pi = \pi W$. The random walk process is a Markov chain as the next vertex only depends on the current vertex but not the past. A Markov chain has a unique stationary distribution π if and only if it is irreducible and aperiodic [31], which, in this case, is equivalent to G being connected and non-bipartite. Also, when the Markov chain has a unique stationary distribution π, we have $p_t \to \pi$ no matter what the initial distribution is. It is possible to remove the non-bipartite requirement by considering the lazy random walk instead. In this setting, we have probability $1/2$ staying at the current vertex and probability $1/2$ moving to a neighbor with probability proportional to the edge weights. Since the lazy random walk is equivalent to the random walk on the graph G with an additional self loop of weight $\deg(u)$ on each vertex u, and the new graph is non-bipartite by construction, the lazy random walk has a unique stationary distribution if and only if G is connected. The probability distribution after t steps lazy random walk starting from the initial distribution p is then pW'^t, where $W' = (I_{n\times n} + W)/2$ is the lazy random walk matrix. In the remainder of this thesis, we use W to denote the lazy random walk matrix $(I_{n\times n} + D^{-1}A)/2$. We can check that the stationary distribution is $\pi = (D\chi_V)^T/\operatorname{vol}(V)$.

The mixing time of a graph G measures how fast the random walks starting from an arbitrary vertex converges to the stationery distribution π where $\pi(u) = \deg(u)/\operatorname{vol}(V)$. The most common definition for the mixing time is

$$\tau(\epsilon) = \min\{t \mid \max_{u \in V} \|\chi_u^T W^t - \pi\|_1 \leq \epsilon\},$$

and $\tau = \tau(1/4)$. The parameter τ is useful in analyzing random sampling algorithms

for combinatorial objects, like computing the permanent and the volume of a convex object [66, 26]. In this section we discuss two ways to analyze the mixing time of a graph, one through spectral arguments and the other through combinatorial arguments by Lovász and Simonovits. From the spectral approach, we get an upper bound on the mixing time by the second eigenvalue λ_2. From the combinatorial approach, we get an upper bound on the mixing time by the conductance $\phi(G)$ of the graph. We remark that by comparing the two bounds, we can obtain Cheeger's inequality with slightly worse constant.

2.3.1 Spectral approach

We first construct the left eigenvectors of the lazy random walk matrix W so that we have a good eigenbasis to work with. Suppose $v_1, \ldots, v_n$ are the eigenvectors of G, which means that $Lv_i = \lambda_i D v_i$. Then we have $v_i' = (Dv_i)^T$ are the left eigenvectors of W, since

$$\begin{aligned} v_i' W &= \frac{1}{2}\left(v_i' + v_i' D^{-1} A\right) \\ &= \frac{1}{2}\left(2v_i' - v_i' D^{-1} L\right) \\ &= \frac{1}{2}\left(2v_i' - v_i^T L\right) \\ &= \frac{1}{2}\left(2v_i' - \lambda_i v_i^T D\right) \\ &= \left(1 - \frac{\lambda_i}{2}\right) v_i'. \end{aligned}$$

This set of eigenvectors also satisfies $\langle v_i', v_j' \rangle_{D^{-1}} = v_i^T D D^{-1} D v_j = \langle v_i, v_j \rangle_D$. Hence $\|v_i'\|_{D^{-1}} = 1$ and $\langle v_i', v_j' \rangle_{D^{-1}} = 0$ for $i \neq j$.

In the following, we present a theorem that upper bounds the mixing time by the second eigenvalue of the graph G.

Theorem 2.5 ([48]). *Let u^* be a vertex with minimum degree. Then the mixing time satisfies*

$$\tau = O(\frac{\log(\pi(u^*)^{-1})}{\lambda_2}).$$

Proof. To bound the mixing time by t, we only need to argue that for any starting point u, $\|\chi_u^T W^t - \pi\|_1 \leq 1/4$. We write $\chi_u^T = \sum_{i=1}^n c_i v_i'$ into the left eigenbasis of W, where

$$c_i = \langle \chi_u^T, v_i' \rangle_{D^{-1}} = \frac{v_i'(u)}{\deg(u)} = v_i(u).$$

Then $c_1 = v_1(u) = 1/\sqrt{\mathrm{vol}(V)}$, and hence for any vertex v,

$$c_1 v_1'(v) = \frac{1}{\sqrt{\mathrm{vol}(V)}} \frac{\deg(v)}{\sqrt{\mathrm{vol}(V)}} = \frac{\deg(v)}{\mathrm{vol}(V)} = \pi(v).$$

This concludes $c_1 v_1' = \pi$. Note that $\lambda_1 = 0$. So we have

$$\begin{aligned}
\|\chi_u^T W^t - \pi\|_1 &= \|\sum_{i=1}^n c_i v_i' W^t - \pi\|_1 \\
&= \|\sum_{i=1}^n c_i \left(1 - \frac{\lambda_i}{2}\right)^t v_i' - \pi\|_1 \\
&= \|\sum_{i=2}^n c_i \left(1 - \frac{\lambda_i}{2}\right)^t v_i'\|_1 \\
&\leq \sum_{i=2}^n \left(1 - \frac{\lambda_i}{2}\right)^t \|c_i v_i'\|_1.
\end{aligned}$$

Here we used the fact that $\lambda_1 = 0$ and $c_1 v_1' = \pi$, and the last inequality holds by the triangle inequality. Now we need to upper bound $\|c_i v_i'\|_1$. Since

$$|c_i| = |v_i(u)| \leq \sqrt{\frac{\sum_{v \in V} v_i(v)^2 \deg(v)}{\deg(u)}} = \sqrt{\frac{1}{\deg(u)}},$$

we have

$$\begin{aligned}
\|c_i v_i'\|_1 &\leq \sqrt{\frac{1}{\deg(u)}} |\sum_{v \in V} v_i(v) \deg(v)| \\
&\leq \sqrt{\frac{1}{\deg(u)}} \sqrt{\left(\sum_{v \in V} v_i(u)^2 \deg(v)\right)\left(\sum_{v \in V} \deg(v)\right)} \\
&= \sqrt{\frac{1}{\deg(u)}} \sqrt{\mathrm{vol}(V)} \\
&= \sqrt{\frac{1}{\pi(u)}},
\end{aligned}$$

where the second inequality follows from Cauchy-Schwartz inequality. Therefore when

$t > 2\log(4n/\sqrt{\pi(u)})/\lambda_2$, we have

$$\begin{aligned}
\|\chi_u^T W^t - \pi\|_1 &\leq \sum_{i=2}^{n} \left(1 - \frac{\lambda_i}{2}\right)^t \|c_i v_i'\|_1 \\
&\leq \sum_{i=2}^{n} \exp(-\frac{\lambda_2}{2}t)\|c_i v_i'\|_1 \\
&= \sum_{i=2}^{n} \frac{\sqrt{\pi(u)}}{4n}\|c_i v_i'\|_1 \\
&\leq \sum_{i=2}^{n} \frac{1}{4n} \\
&\leq \frac{1}{4}.
\end{aligned}$$

Let u^* be the vertex that have minimum degree, that means $\pi(u^*) \leq \pi(u)$ for any vertex u. Then $\pi(u^*) \leq 1/n$ and for $t \geq 2\log(4\pi(u^*)^{-3/2})/\lambda_2$, we have $\|\chi_u^T W^t - \pi\|_1 \leq 1/4$ for any u. This concludes $\tau = O(\log(\pi(u^*)^{-1})/\lambda_2)$. □

2.3.2 Combinatorial approach

Given an undirected weighted graph $G = (V, w)$, for each probability distribution $p : V \to \mathbb{R}_{\geq 0}$ over the vertices, we define the following function for $x \in [0, \mathrm{vol}(V)]$:

$$C(p, x) = \max_{\sum_{i=1}^{n} \delta_i \deg(u_i) = x, 0 \leq \delta_i \leq 1} \sum_{i=1}^{n} \delta_i p(i).$$

The function can be interpreted as the total probability mass by picking the largest x fractional vertices. Suppose $p(u_1)/\deg(u_1) \geq p(u_2)/\deg(u_2) \geq \cdots \geq p(u_n)/\deg(u_n)$, then we can simply write

$$C\left(p, \sum_{i=1}^{k} \deg(u_i)\right) = \sum_{i=1}^{k} p(u_i).$$

We call $\sum_{i=1}^{k} \deg(u_i)$ the extreme points. Note that C is piecewise linear between the extreme points. When $p = \pi$ is the stationary distribution, then $p(u_i)/\deg(u_i)$ is constant, and the function C satisfies $C(p, x) = x/\mathrm{vol}(V)$ and becomes linear.

Another interpretation is viewing the graph as directed by replacing each undirected edge by two directed edges, and the function as the total probability mass by picking the largest x fractional directed edges (instead of vertices). It turns out that

this interpretation is more useful in our later analysis. We assign the probability mass $p(i,j) = p(i)w(i,j)/\deg(i)$ to each of i's outgoing directed edges (i,j). Then we have

$$C(p,x) = \max_{\sum_{i\in V}\sum_{j\in V}\delta_{i,j}w(i,j)=x, 0\le\delta_{i,j}\le 1} \sum_{i\in V}\sum_{j\in V}\delta_{i,j}p(i,j),$$

and suppose $p(e_1)/w(e_1) \ge p(e_2)/w(e_2) \ge \cdots \ge p(e_m)/w(e_m)$, we have

$$C\left(p, \sum_{i=1}^{k} w(e_i)\right) = \sum_{i=1}^{k} p(e_i).$$

Clearly for any probabilistic distribution p, the function is concave and thus $C(p,x) \ge x/\operatorname{vol}(V)$. Hence the function is always larger than the one corresponding to the stationary distribution. The following lemma shows that the function drops after one step of the lazy random walk, and the speed depends on the conductance of the graph.

Lemma 2.4 ([51])**.** *For a graph $G = (V,w)$ and a probability distribution $p : V \to \mathbb{R}$, let $p' = pW$ be the probability distribution after one step of the lazy random walk in G. Then for any extreme point x of p', we have*

$$C(p',x) \le \frac{1}{2}(C(p, x+\phi\overline{x}) + C(p, x-\phi\overline{x})),$$

where

$$\overline{x} = \min(x, \operatorname{vol}(V) - x)$$

is the distance of x to the boundaries.

Proof. Let x be an extreme point of p'. Suppose $C(p',x)$ is attained by picking the subset of vertices $S \subseteq V$, which means that $x = \operatorname{vol}(S)$ and $C(p',x) = \sum_{i\in S} p'(i)$. For each vertex $i \in V$, we have

$$\begin{aligned} p'(i) &= \frac{1}{2}\sum_{j\in V}\frac{w(j,i)}{\deg(j)}p(j) + \frac{1}{2}p(i) \\ &= \frac{1}{2}\sum_{j\in V}\frac{w(j,i)}{\deg(j)}p(j) + \frac{1}{2}\sum_{j\in V}\frac{w(i,j)}{\deg(i)}p(i) \\ &= \frac{1}{2}\sum_{j\in V}p(j,i) + \frac{1}{2}\sum_{j\in V}p(i,j). \end{aligned}$$

Therefore

$$\begin{aligned}
C(p', x) &= \sum_{i \in S} p'(i) \\
&= \frac{1}{2} \sum_{i \in S} \sum_{j \in V} p(j, i) + \frac{1}{2} \sum_{i \in S} \sum_{j \in V} p(i, j) \\
&= \frac{1}{2} \sum_{i \in S} \sum_{j \in S} p(i, j) + \frac{1}{2} \left(\sum_{i \in V - S} \sum_{j \in S} p(i, j) + \sum_{i \in S} \sum_{j \in V} p(i, j) \right) \\
&\leq \frac{1}{2} C(p, \mathrm{vol}(S) - w(\delta(S))) + \frac{1}{2} C(p, \mathrm{vol}(S) + w(\delta(S))),
\end{aligned}$$

where the last inequality is due to

$$\sum_{i \in S} \sum_{j \in S} w(i, j) = \sum_{i \in S} \sum_{j \in V} w(i, j) - \sum_{i \in S} \sum_{j \in V - S} w(i, j) = \mathrm{vol}(S) - w(\delta(S)),$$

$$\sum_{i \in V - S} \sum_{j \in S} w(i, j) + \sum_{i \in S} \sum_{j \in V} w(i, j) = w(\delta(S)) + \mathrm{vol}(S),$$

and the directed edges from $V - S$ to S and those from S to V are disjoint. Finally, since $w(\delta(S)) \leq \phi \min(\mathrm{vol}(S), \mathrm{vol}(V) - \mathrm{vol}(S)) = \phi \overline{x}$, by concavity of C we have

$$\begin{aligned}
C(p', x) &\leq \frac{1}{2} C(p, \mathrm{vol}(S) - w(\delta(S))) + \frac{1}{2} C(p, \mathrm{vol}(S) + w(\delta(S))) \\
&\leq \frac{1}{2} C(p, x - \phi \overline{x}) + \frac{1}{2} C(p, x + \phi \overline{x}).
\end{aligned}$$

□

We view the probability distribution p over the vertices V as a row vector indexed by the vertices. We use $C^{(t)}(x)$ to denote $C(pW^t, x)$ when p is clear from the context. The following theorem provides an upper bound on $C^{(t)}(x)$ and shows that for any starting distribution p, the function approaches to the linear function after a number of steps of the lazy random walk and the convergence rate depends of $\phi(G)$.

Theorem 2.6 ([51]). *For any graph G, any probability distribution $p \in \mathbb{R}^V_{\geq 0}$ over the vertices, any $t \geq 0$ and any $x \in [0, \mathrm{vol}(V)]$, we have*

$$C^{(t)}(x) \leq \frac{x}{\mathrm{vol}(V)} + \sqrt{\frac{\overline{x}}{\min_{i \in V} \deg(i)}} \left(1 - \frac{\phi^2}{8}\right)^t.$$

Proof. We shall only prove the statement for $x \leq \mathrm{vol}(V)/2$ where $\overline{x} = x$, and the rest

follows by symmetry. Note that the function on the right hand side is concave and $C^{(t)}$ is linear between the extreme points, it remains to show that the inequality holds for the extreme points. We shall prove the statement by induction. When $t = 0$, $C^{(t)}(x) = C(p, x)$. Since the first non-zero extreme point is at least $\min_{i\in V} \deg(i)$, the inequality holds when $t = 0$. When $t > 0$ and x is an extreme point, by Lemma 2.4 and the induction hypothesis, we have

$$\begin{aligned} C^{(t)}(x) &\le \frac{1}{2}(C^{(t-1)}(x - \phi x) + C^{(t-1)}(x + \phi x)) \\ &\le \frac{x}{\mathrm{vol}(V)} + \sqrt{\frac{x}{\min_{i\in V} \deg(i)}} \left(1 - \frac{\phi^2}{8}\right)^{t-1} \left(\frac{1}{2}(\sqrt{\frac{\overline{x - \phi x}}{x}} + \sqrt{\frac{\overline{x + \phi x}}{x}})\right) \end{aligned}$$

Since $\overline{x - \phi x} = x(1 - \phi)$ and $\overline{x + \phi x} \le x(1 + \phi)$, we have

$$\begin{aligned} C^{(t)}(x) &\le \frac{x}{\mathrm{vol}(V)} + \sqrt{\frac{x}{\min_{i\in V} \deg(i)}} \left(1 - \frac{\phi^2}{8}\right)^{t-1} \left(\frac{1}{2}(\sqrt{1 - \phi} + \sqrt{1 + \phi})\right) \\ &\le \frac{x}{\mathrm{vol}(V)} + \sqrt{\frac{x}{\min_{i\in V} \deg(i)}} \left(1 - \frac{\phi^2}{8}\right)^{t}, \end{aligned}$$

where the last inequality follows from the Taylor expansion of $\sqrt{1 \pm \phi}$ at $\phi = 0$. □

Note that for any probability vector p, $\|pW^t - \pi\|_1 = 2 \max_x (C^{(t)}(x) - x/\mathrm{vol}(V))$. Hence suppose u^* is the vertex with minimum degree and $t \ge 8 \log(8\pi(u^*)^{-1/2})/\phi^2)$, we have $\|pW^t - \pi\|_1 \le 1/4$ for any p. This concludes $\tau = O(\log(\pi(u^*)^{-1})/\phi^2)$.

Finally, we remark that Theorem 2.6 can give another proof of Cheeger's inequality with worse constant. Let v'_i be the left eigenbasis of W, as defined in Subsection 2.3.1. Let $u \in V$ be any vertex such that $v'_2(u) \neq 0$, and let $p = \chi_u^T = \sum_{i\in V} c_i v'_i$, where $c_i = \langle \chi_u, v'_i \rangle_{D^{-1}} = v'_i(u)/\deg(u) = v_i(u)$. Then $c_1 = 1/\sqrt{\mathrm{vol}(V)}$ and by our assumption $c_2 \neq 0$. Also

$$pW^t = \sum_{i\in V} c_i (1 - \frac{\lambda_i}{2})^t v'_i.$$

Thus we have

$$pW^t D^{-1} p^T = \langle pW^t, p \rangle_{D^{-1}} = \sum_{i\in V} c_i^2 (1 - \frac{\lambda_i}{2})^t \ge \frac{1}{\mathrm{vol}(V)} + c_2^2 (1 - \frac{\lambda_2}{2})^t.$$

On the other hand, note that $pW^t D^{-1} p^T$ can be viewed as starting from the distribution $p = \chi_u$, after t steps of the random walk, we pick $1/\deg(u)$ fraction of the probability

mass remaining in the vertex u. Hence, by Theorem 2.6, we have

$$pW^tD^{-1}p^T \leq C^{(t)}(1) \leq \frac{1}{\mathrm{vol}(V)} + \sqrt{\frac{1}{\min_{i\in V}\deg(i)}}\left(1-\frac{\phi^2}{8}\right)^t.$$

This shows that

$$\sqrt{\frac{1}{\min_{i\in V}\deg(i)}}\left(1-\frac{\phi^2}{8}\right)^t \geq c_2^2(1-\frac{\lambda_2}{2})^t.$$

Since t is arbitrary and $c_2 \neq 0$, we must have $1-\phi^2/8 \geq 1-\lambda_2/2$, and this implies that $\phi \leq 2\sqrt{\lambda_2}$.

In this argument, we only use the information of the conductances of the level sets of p_t. Hence by examining all the level sets of all the random walk vectors p_t, one of them would have conductance not greater than $2\sqrt{\lambda_2}$.

2.4 Previous works

In this section, we will discuss some previous results about generalizations of Cheeger's inequality using higher eigenvalues, and about approximating the conductance and the small set expansion.

2.4.1 Higher eigenvalues

We will discuss several recent results showing connections between the conductance profile of a graph and the higher eigenvalues of its normalized Laplacian matrix. The first result in this direction is about the small set expansion problem. Arora, Barak and Steurer show that we can obtain small sparse cuts efficiently if there are many small eigenvalues.

Theorem 2.7 ([7]). *Let G be a regular graph on n vertices such that $\lambda_k \leq \eta$, where $k = n^{100\eta/\gamma}$. Then there is an efficient algorithm to find a vertex set S of size at most $n^{1-\eta/\gamma}$ that satisfies $\phi(S) \leq \sqrt{\gamma}$.*

In particular, if $k = n^\epsilon$, by setting $\eta = \lambda_k$ and $\gamma = 100\lambda_k/\epsilon$, the graph has a sparse cut S with $\phi(S) \leq O(\sqrt{\lambda_k/\epsilon})$ and $|S| = O(n/k^{1/100})$. This concludes that

$$\phi_{O(k^{-1/100})} = O(\sqrt{\lambda_k \log_k n}).$$

This can be seen as a generalization of Cheeger's inequality to the small set expansion problem. This result is then improved in several papers with various approaches [69, 55,

56]. All of them are able to improve the size approximate from $1/100$ to $1-\epsilon$. They prove that

$$\phi_{O(k^{-1+\epsilon})} = O(\sqrt{\frac{\lambda_k \log_k n}{\epsilon}}),$$

and there is also an efficient algorithm computing a small set S achieving $|S| = O(n/k^{1-\epsilon})$ and $\phi(S) = O(\sqrt{\lambda_k \log_k n/\epsilon})$.

Cheeger's inequality for graph partitioning can also be extended to higher-order Cheeger's inequality for k-way graph partitioning [49, 45]: If there are k small eigenvalues, then there are k disjoint sparse cuts. This is a robust generalization of the fact that $\lambda_k = 0$ if and only if G has at least k connected components.

Theorem 2.8 ([45]). *There is an efficient algorithm to find k non-empty disjoint subsets $S_1, \ldots, S_k$, $S_i \subseteq V$, such that $\phi(S_i) \leq O(k^2)\sqrt{\lambda_k}$. Hence*

$$\frac{1}{2}\lambda_k \leq \phi_k(G) \leq O(k^2)\sqrt{\lambda_k}.$$

If only ck sets are needed for some constant c, the inequality can be further improved.

Theorem 2.9 ([49, 45]). *There exists a constant $c > 0$ such that we can efficiently find ck non-empty disjoint subsets $S_1, \ldots, S_{ck}$, $S_i \subseteq V$, such that $\phi(S_i) = O(\sqrt{\lambda_k \log k})$. This also implies that*

$$\phi_{O(k^{-1})}(G) = O(\sqrt{\lambda_k \log k}).$$

Lee et.al. [45] show that c can be chosen to be $1-\epsilon$ for any $\epsilon > 0$, while Louis et.al [49] can only show that the statement is true for some $c > 0$.

2.4.2 Sparsest cut

We will discuss about algorithms for the sparsest cut problem in this subsection. Given a graph G, the sparsest cut problem is to determine the conductance

$$\phi(G) = \min_{S \subseteq V,\, \mathrm{vol}(S) \leq \mathrm{vol}(V)/2} \phi(S),$$

and find a subset of vertices $S \subseteq V$ with $\mathrm{vol}(S) \leq \mathrm{vol}(V)/2$ that minimizes the conductance[5]. It is known that the sparsest cut problem is NP-hard [10], and approximating

[5]The more common definition of the sparsest cut problem is to find a subset of vertices $S \subseteq V$ that minimizes the sparsity $\mathrm{sp}(S)$ instead. Since the sparsity $\mathrm{sp}(S)$ and the conductance $\phi(S)$ are equivalent up to a constant factor (see Subsection 2.1.1), which we do not care about in this subsection, we will study the conductance instead here.

the conductance within constant factor is NP-hard assuming the unique games conjecture [13].

The spectral partitioning algorithm is the most commonly used heuristic for finding sparse cuts in practice because of its ease of implementation and efficiency (see Subsection 2.2.1). Its performance is guaranteed by Cheeger's inequality, which states that the output set S satisfies

$$\Omega(\lambda_2) \leq \phi(S) \leq O(\sqrt{\lambda_2}).$$

This inequality is tight up to constant. In the worst case, the spectral relaxation λ_2 only gives an $O(1/\sqrt{\lambda_2})$ approximation for the conductance, and this can be as bad as $\Theta(n)$ in a cycle graph (see Subsection 2.2.3). On the other hand, λ_2 is a constant factor approximation when G is an expander. Although this spectral relaxation only guarantees good approximation when the graph has large conductance, the spectral partitioning algorithm often performs well in practice, and this cannot be explained by merely Cheeger's inequality. We will have more discussion about this in Section 2.2.

While Cheeger's inequality gives an approximation with performance depending on the optimal value, there are other approaches that guarantees approximation ratio only depend on n, the size of the graph. The linear programming rounding algorithm by Leighton and Rao [46] finds a set of conductance $O(\phi(G) \log n)$. The authors uses the all pairs multicommodity flow to give a lower bound on the conductance, and this method is now commonly used to certify lower bounds of expansion and hence proves upper bounds on mixing time. The linear programming relaxation gives a good approximate when the graph can be well embedded into L^1 metric, for example, when the graph is a path or a cycle. On the other hand, there are examples where the relaxation only gives an $\Omega(\log n)$ approximation when the graph is a sparse expander. It is interesting to note that the bad examples for the spectral relaxation is good for linear programming relaxation and vice versa.

Arora, Rao and Vazirani [8] uses a semidefinite programming relaxation as a commonly generalization of the spectral approach and the linear programming approach, and finds a set of conductance $O(\phi(G)\sqrt{\log n})$.

Guruswami and Sinop [29] shows that we have better approximation using semidefinite programming if λ_r is large. They present an $O(2^{r/(\epsilon\delta)})\,\mathrm{poly}(n)$ algorithm to find a cut with conductance at most $(1+\epsilon)\phi(G)/\delta$ when $\lambda_r > \phi(G)/(1-\delta)$.

2.4.3 Small sparse cuts

We will discuss about the small sparse cuts problem in this subsection. Given a graph G and a size parameter $\delta \in (0, 1/2]$, the small sparsest cut problem is to determine the small set conductance

$$\phi_\delta(G) = \min_{\mathrm{vol}(S) \leq \delta \,\mathrm{vol}(V)} \phi(S),$$

and find a subset of vertices $S \subseteq V$ with $\mathrm{vol}(S) \leq \delta \,\mathrm{vol}(V)$ that minimizes the conductance. This is also known as the expansion profile of the graph [50, 59]. Raghavendra and Steurer [58] propose the following small set expansion conjecture about the hardness of approximating small sparsest cuts.

Conjecture 2.1 ([58])**.** For every constant $\epsilon > 0$, there exists a constant $\delta > 0$, such that it is NP-hard to distinguish the two cases: (1) $\phi_\delta(G) \leq \epsilon$; (2) $\phi_\delta(G) \geq 1 - \epsilon$.

They show that the small set expansion conjecture implies the unique games conjecture and is implied by the unique games conjecture with some mild assumptions, and so it is of interest to understand what algorithmic techniques can be used to estimate $\phi_\delta(G)$. Raghavendra and Schramm [57] prove the following gap amplification result about the hardness of approximating small sparsest cuts.

Theorem 2.10 ([57])**.** *Let f be a function such that $f(x) = \omega(\sqrt{x})$ as $x \to 0^+$. If for all $\epsilon > 0$, there exists $\delta > 0$ such that it is NP-hard to distinguish $\phi_\delta(G) \leq \epsilon$ or $\phi_\delta(G) \geq f(\epsilon)$, then for all $\epsilon > 0$, there exists $\delta > 0$ such that it is NP-hard to distinguish $\phi_\delta \leq \epsilon$ or $\phi_{\delta/8} \geq 1/2$.*

A bicriteria approximation algorithm for the small sparsest cut problem is to find a subset of vertices S with volume $\mathrm{vol}(S) \lesssim \delta \,\mathrm{vol}(V)$ and conductance $\phi(S) \lesssim \phi_\delta(G)$. There are bicriteria approximation algorithms for this problem using semidefinite programming relaxations: Raghavendra, Steurer and Tetali [59] obtain an algorithm that finds a set S with $\mathrm{vol}(S) \leq O(\delta \,\mathrm{vol}(V))$ and $\phi(S) \leq O(\sqrt{\phi_\delta(G) \log(\delta^{-1})})$, and Bansal et.al. [9] obtain an algorithm that finds a set S with $\mathrm{vol}(S) \leq (1 + \epsilon)\delta \,\mathrm{vol}(V)$ and $\phi(S) \leq O(f(\epsilon)\phi_\delta(G)\sqrt{\log n \log(\delta^{-1})})$ for any $\epsilon > 0$ where $f(\epsilon)$ is a function depends only on ϵ.

2.4.4 Local graph partitioning

We will discuss about local graph partitioning algorithms in this subsection. In some situations, we have a massive graph $G = (V, E)$ and a vertex $v \in V$, and we would like

to identify a small set S with small conductance that contains v. The graph may be too big that it is not feasible to read the whole graph and run some nontrivial approximation algorithms. So it would be desirable to have a local algorithm that only explores a small part of the graph, and outputs a set S with small conductance that contains v, and the running time of the algorithm depends only on $\mathrm{vol}(S)$ and $\mathrm{polylog}(n)$.

All existing local graph partitioning algorithms are based on some random walk type processes. The efficiency of the algorithm is measured by the work/volume ratio, which is defined as the ratio of the running time and the volume of the output set. Suppose S^* is our target sparse cut. Spielman and Teng [67] proposed the first local graph partitioning algorithm using truncated random walk, that returns a set S with $\phi(S) = O(\phi(S^*)^{1/2}(\log n)^{3/2})$ if the initial vertex is a random vertex in S^*, and the work/volume ratio of the algorithm is $O(\phi(S^*)^{-2}\,\mathrm{polylog}(n))$. Andersen, Chung, Lang [5] used local pagerank vectors to find a set S with $\phi(S) = O(\sqrt{\phi(S^*)\log k})$ and work/volume ratio $O(\phi(S^*)^{-1}\,\mathrm{polylog}(n))$, if the initial vertex is a random vertex in a set S^* with $\mathrm{vol}(S^*) = k$. Andersen and Peres [6] used the volume-biased evolving set process to obtain a local graph partitioning algorithm with work/volume ratio $O(\phi(S^*)^{-1/2}\,\mathrm{polylog}(n))$ and a similar conductance guarantee as in [5]. Note that the running time of these algorithms would be sublinear if the volume of the output set is small, which is the case of interest in massive graphs.

Chapter 3

Improved Cheeger's inequality

In this chapter, we present a generalization of Cheeger's inequality using higher eigenvalues. The results presented in this chapter are mainly based on joint work with Yin Tat Lee, Lap Chi Lau, Shayan Oveis Gharan and Luca Trevisan [43].

Finding a set of small conductance is a fundamental problem which comes up in different areas of computer science. The spectral graph partitioning algorithm is a common heuristic to this problem used in practice because of its ease of implementation and efficiency. The classical Cheeger's inequality provides a performance guarantee for this algorithm:

$$\frac{1}{2}\lambda_2 \leq \phi(G) \leq \phi_{\text{sweep}} \leq \sqrt{2\lambda_2},$$

where ϕ_{sweep} is the conductance of the set returned by the spectral graph partitioning algorithm. However, the inequality can be as bad as a $\Theta(n)$ approximation in the worst case (see Subsection 2.4.2), and gives a constant approximation only when the graph is an expander. On the contrary, the spectral graph partitioning algorithm often performs well in practice. Therefore Cheeger's inequality cannot give a satisfactory explanation.

3.1 Main result

Our main result is a generalization of Cheeger's inequality using higher eigenvalues.

Theorem 3.1. *For any undirected graph G and integer $k \geq 2$, we have*

$$\phi(G) \leq \phi_{sweep} \leq O(k)\frac{\lambda_2}{\sqrt{\lambda_k}},$$

where ϕ_{sweep} is the conductance of the set returned by the spectral partitioning algorithm.

This shows that the spectral partitioning algorithm is an $O(k/\sqrt{\lambda_k})$-approximation algorithm for the sparsest cut problem. In particular, the spectral partitioning algorithm performs well when λ_k is large for a small k, which happens when the graph has a clear k-partition.

The rest of this chapter is organized as follows. We first give a proof overview in Section 3.2, and prove our main result in Section 3.3. Then we improve the hidden constant in the big-Oh notation in Section 3.4, and show that $\phi(G) \leq 2\sqrt{2}k\lambda_2/\sqrt{\lambda_{k+1}}$. Finally we see some related problems where our techniques can be used to get better bounds in Section 3.5. For the maximum cut problem, suppose the optimal cut fraction is $1-\epsilon$, our techniques gives a cut with fraction $1-\tilde{O}(k\epsilon/(2-\lambda_{n-k}))$.

3.2 Proof overview

We first describe an informal intuition to the inequality when $k=3$ and see how it can be generalized afterwards. Recall that the Rayleigh quotient of a vector $x \in \mathbb{R}^V$ is

$$\mathcal{R}_G(x) = \frac{\sum_{uv\in E} w(u,v)(x(u)-x(v))^2}{\sum_{u\in V}\deg(u)x(u)^2}. \tag{3.1}$$

Let us assume that λ_2 is small and λ_3 is large for a graph G. Since λ_2 is small, G contains a set of small conductance. On the other hand, when λ_3 is large, by the higher order Cheeger's inequality (Theorem 2.8), $\phi_3(G)$ is large, and hence G cannot be partitioned into 3 sparse cuts. Intuitively, these together mean that the vertices of G can be partitioned into 2 sets, each with large internal conductance. Now the Rayleigh quotient of the second eigenvector v_2 is small implies the values $v_2(u)$ inside each subset have to be similar, otherwise the numerator in Equation 3.1 would be large. Hence v_2 is close to a 2-valued vector. However, if v_2 is 2-valued, then it is a normalized characteristic vector and therefore is at most factor 2 away from the conductance of a set. Therefore $\phi(G) = O(\lambda_2)$.

For general k, our proof consists of two main steps. First, we show that when λ_k is large, v_2 is close to a $O(k)$-valued vector in the sense that $\|v_2 - y\|_D$ is small for some $O(k)$-valued vector y. In general cases it is not easy to argue that G can be partitioned into $O(k)$ expanders, and conclude that v_2 is close to k-valued. Instead, we will show the contrapositive. We will prove that if v_2 is not close to any $O(k)$-valued vectors, then we can construct k disjointly supported vectors x_i with small Rayleigh quotients, and this contradicts to λ_k is large by Lemma 2.1. The remaining step is to show that when v_2 is close to a $O(k)$-valued vector, then ϕ_{sweep} is small. More precisely, we prove

that $\phi_{\text{sweep}} = O(k\lambda_2 + k\|v_2 - y\|_D\sqrt{\lambda_2})$. This can be done by a random thresholding argument similar to Theorem 2.4, by defining a probability distribution such that we are more likely to choose thresholds far away from the $O(k)$ values of y.

We have a second proof of the improved Cheeger's inequality [43], which uses the technique of smoothing the functions with well sperated disjoint support. This method comes from [45] for proving the higher order Cheeger's inequality. We will not discuss this second proof in this thesis.

3.3 Proof of improved Cheeger's inequality

We first state a slightly stronger statement that we shall prove later. This statement shows that we can find a good cut whenever we have a vector x perpendicular to v_1 with small Rayleigh quotient, instead of having to consider the second eigenvector v_2.

Theorem 3.2. *For any undirected graph G, integer $k \geq 2$ and non-zero vector $x \in \mathbb{R}^V$ with $\langle x, v_1\rangle_D = 0$, we have*

$$\phi_{sweep}(x) = O(k)\frac{\mathcal{R}(x)}{\sqrt{\lambda_k}}.$$

Recall that $\phi_{sweep}(x)$ is the minimum conductance over the level sets of x with volume not greater than half of the total volume.

From Theorem 3.2 it is easy to deduce Theorem 3.1.

Proof of Theorem 3.1. Let $x = v_2$, the second eigenvector of G. Then we have $\phi_{\text{sweep}}(x) = \phi_{\text{sweep}}$ and $\mathcal{R}(x) = \lambda_2$. Therefore we get

$$\phi_{\text{sweep}} = O(k)\frac{\lambda_2}{\sqrt{\lambda_k}}.$$

□

The proof consists of two steps. Before proving the theorem, let us first state and prove the two main lemmas first.

3.3.1 $O(k)$-valued approximation

Given a vector $x \in \mathbb{R}^V$, we say that $y \in \mathbb{R}^V$ is a k-valued approximation of x if y only contains k distinct elements, and $\|x - y\|_D$ is small. The following lemma argues that when λ_k is large, there exists a $O(k)$-valued approximation.

Lemma 3.1. *For any undirected graph G, integer $k \geq 2$ and non-constant vector $x \in \mathbb{R}^V$ with $\|x\|_D = 1$, when $\lambda_k > 0$, there exists a $(2k-1)$-valued approximation y such that*

$$\|x - y\|_D^2 \leq \frac{4\mathcal{R}(x)}{\lambda_k}$$

Proof of Lemma 3.1. We first show how we determine an approximate vector y by its values, and then choose the right values to achieve the bound. Given $2k-1$ values $t_1 \leq t_2 \leq \cdots \leq t_{2k-1}$, for any vertex $u \in V$, let $y(u) = t_k$ be such that $|x(u) - t_k| = \min_i |x(u) - t_i|$. In other words, $y(u)$ is the closest value of $x(u)$ in the set $\{t_1, \ldots, t_{2k-1}\}$. If there are multiple values satisfying the equality, then we can just choose an arbitrary value. Note that this assignment minimizes $\|x - y\|_D$ subject to the values of y are in the set $\{t_1, \ldots, t_{2k-1}\}$.

Now we use the values t_i to define disjointly supported vectors. Let $t_0 = -\infty$ and $t_{2k} = \infty$. Then these values divide the vertex set into $2k$ parts $V_i = \{u \in V \mid t_{i-1} \leq x(u) < t_i\}$. We shall define $2k$ vectors, one for each of these $2k$ parts, and then apply Lemma 2.1 to relate their norms to λ_k. Given two values $a \leq b$, define the vector $x_{a,b} \in \mathbb{R}^V$ by

$$x_{a,b}(u) = \begin{cases} \min(x(u) - a, b - x(u)) & \text{if } x(u) \in [a, b], \\ 0 & \text{otherwise.} \end{cases}$$

Let $x_i = x_{t_{i-1},t_i}$ for $1 \leq i \leq 2k$ be the $2k$ disjointly supported vectors. Note that $\|x - y\|_D^2 = \sum \|x_i\|_D^2$ by our definition, since both $|x(u) - y(u)|$ and $|x_i(u)|$ are the distance of $x(u)$ to the closest value in the set $\{t_1, \ldots, t_{2k-1}\}$, when $u \in [t_{i-1}, t_i]$. Also by our definition all the $2k$ vectors are 1-Lipschitz with respect to x, which means that $|x_i(u) - x_i(v)| \leq |x(u) - x(v)|$ for any pairs of vertices u, v.

Finally we choose the values t_i such that $\|x_i\|_D^2$ are all the same, and equal to $C = 2\mathcal{R}(x)/(k\lambda_k)$. This value C is chosen such that it leads to a contradiction when we can successfully define the $2k$ disjointly supported vectors. We start with $t_0 = -\infty$. After $t_0, t_1, \ldots, t_{i-1}$ are determined, we choose t_i such that $\|x_i\|_D^2 = C$. Note that $\|x_i\|_D^2 = \|x_{t_{i-1},t_i}\|_D^2$ is continuous and increases with t_i, therefore we can find such a t_i whenever $\|x_{t_{i-1},\infty}\|_D^2 \geq C$. Now we have two cases. Either there exists an $i < 2k$ such that $\|x_{t_{i-1},\infty}\|_D^2 < C$, or all t_i are determined successfully for all $i = 0, 1, \ldots, 2k$. In the second case, we shall redefine $t_{2k} = \infty$, and still have $\|x_{2k}\|_D^2 = \|x_{t_{2k-1},\infty}\|_D^2 \geq C$.

In the first case, we set $t_i = t_{i+1} = \cdots = t_{2k} = \infty$. Then

$$\|x - y\|_D^2 = \sum_{i=1}^{2k} \|x_i\|_D^2 \leq 2kC = \frac{4\mathcal{R}(x)}{\lambda_k},$$

and we are done.

In the second case, we obtain $2k$ disjointly supported vectors, each with the squared D-norm at least C. We would like to prove by contradiction using Lemma 2.1, which gives an upper bound of λ_k by the Rayleigh quotients of disjointly supported vectors. Hence we need to upper bound the Rayleigh quotients of the vectors. Since the norm known to be at least C, it remains to bound the energies of the vectors. For any pairs of vertices u, v, we will show that

$$\sum_{i=1}^{2k} (x_i(u) - x_i(v))^2 \leq (x(u) - x(v))^2. \tag{3.2}$$

Suppose $u \in V_p$ and $v \in V_q$ and assume without loss that $1 \leq p \leq q \leq 2k$. If $p = q$, then

$$\sum_{i=1}^{2k} (x_i(u) - x_i(v))^2 = (x_p(u) - x_p(v))^2 \leq (x(u) - x(v))^2,$$

where the inequality holds since x_p is 1-Lipschitz with respect to x. Otherwise $p < q$, and

$$\begin{aligned}
\sum_{i=1}^{2k} (x_i(u) - x_i(v))^2 &= (x_p(u) - x_p(v))^2 + (x_q(u) - x_q(v))^2 \\
&= x_p(u)^2 + x_q(v)^2 \\
&\leq (x(u) - t_p)^2 + (t_{q-1} - x(v))^2 \\
&\leq (x(u) - x(v))^2,
\end{aligned}$$

where the last inequality holds since $x(u) \leq t_p \leq t_{q-1} \leq x(v)$. Therefore Equation 3.2

holds in any case. Summing up these equations for all edges $\{u, v\}$, we have

$$\begin{aligned}
\sum_{i=1}^{2k} \mathcal{R}(x_i) &= \frac{1}{C} \sum_{i=1}^{2k} \sum_{uv \in E} w(u,v)(x_i(u) - x_i(v))^2 \\
&= \frac{1}{C} \sum_{uv \in E} w(u,v) \sum_{i=1}^{2k} (x_i(u) - x_i(v))^2 \\
&\leq \frac{1}{C} \sum_{uv \in E} w(u,v)(x(u) - x(v))^2 \\
&= \frac{k\lambda_k}{2\mathcal{R}(x)} \mathcal{R}(x) \\
&= \frac{k\lambda_k}{2}.
\end{aligned}$$

On the other hand, let x'_i be a permutation of x_i such that $\mathcal{R}(x'_1) \leq \mathcal{R}(x'_2) \leq \cdots \leq \mathcal{R}(x'_{2k})$. Then Lemma 2.1 states that $\mathcal{R}(x'_k) \geq \lambda_k/2$. Therefore

$$\sum_{i=1}^{2k} \mathcal{R}(x_i) = \sum_{i=1}^{2k} \mathcal{R}(x'_i) \geq \sum_{i=k}^{2k} \mathcal{R}(x'_i) \geq \frac{(k+1)\lambda_k}{2}.$$

Since $\lambda_k > 0$, this leads to a contradiction. Thus the second case cannot happen and we are done. □

3.3.2 Upper bound ϕ_{sweep} with distance to approximation

The next lemma states that when a non-negative vector x is close to having only a few values, then the spectral partitioning algorithm actually works better.

Lemma 3.2. *For any non-negative vector $x \in \mathbb{R}^V$ with $\|x\|_D = 1$ and any $(2k-1)$-valued vector y, we have*

$$\phi_{sweep}(x) \leq 2k\mathcal{R}(x) + 4\sqrt{2k}\|x - y\|_D \sqrt{\mathcal{R}(x)}.$$

Proof. We use the randomized thresholding argument to analyze the performance of the sweeping algorithm. The proof is a generalization of the original randomized thresholding proof for Cheeger's inequality. Let $M = \max_{u \in V} x(u)$ and $0 \leq t_1 \leq t_2 \leq \cdots \leq t_{2k-1}$ be the values of the vector y. Without loss we assume the vector y satisfies

$$|x(u) - y(u)| = \min_{1 \leq i \leq 2k-1} |x(u) - t_i|.$$

Otherwise we can redefine y such that the above equality holds. Then the new y still has the same values while $\|x-y\|_D$ does not increase. Let $d : [0,M] \to \mathbb{R}_{\geq 0}$ be the distance function to the closest value, defined as

$$d(t) = \min_{1 \leq i \leq 2k-1} |t - t_i|.$$

We abuse the notation and also use $d(u)$ to denote $d(x(u))$. Let $C = \int_0^M d(t)dt$. We use $\mu = d/C$ as the probability density function to choose our random threshold. We choose a value $t \in [0,M]$ following the distribution μ and analyze the level set $S_t = \{u \mid x(u) \geq t\}$. The expected volume of the set is

$$\begin{aligned}
\mathbb{E}_{t\sim\mu}[S_t] &= \sum_{u\in V} \deg(u) \mathbb{P}_{t\sim\mu}[u \in S_t] \\
&= \sum_{u\in V} \deg(u) \int_0^{x(u)} \mu(t)dt \\
&= \frac{1}{C} \sum_{u\in V} \deg(u) \int_0^{x(u)} d(t)dt.
\end{aligned}$$

Let $t_0 = -\infty$ and $t_{2k} = \infty$. Note that if $x(u) \in [t_{i-1}, t_i)$ for some $1 \leq i \leq 2k$, then

$$\begin{aligned}
\int_0^{x(u)} d(t)dt &= \int_0^{t_1} d(t)dt + \left(\sum_{j=2}^{i-1} \int_{t_{j-1}}^{t_j} d(t)dt \right) + \int_{t_{i-1}}^{x(u)} d(t)dt \\
&\geq \frac{t_1^2}{4} + \left(\sum_{j=2}^{i-1} \frac{(t_j - t_{j-1})^2}{4} \right) + \frac{(x(u) - t_{i-1})^2}{4} \\
&\geq \frac{1}{4} \left(\frac{\left(t_1 + \left(\sum_{j=2}^{i-1} (t_j - t_{j-1}) \right) + (x(u) - t_{i-1}) \right)^2}{i} \right) \\
&\geq \frac{x(u)^2}{8k},
\end{aligned}$$

where the second last inequality follows from Cauchy-Schwarz inequality. Therefore the expected volume of the set is

$$\mathbb{E}_{t\sim\mu}[\mathrm{vol}(S_t)] \geq \frac{1}{8kC} \sum_{u\in V} \deg(u) x(u)^2 = \frac{\|x\|_D^2}{8kC}.$$

On the other hand, the expected total weight of edges cut by the set is

$$\mathbb{E}_{t\sim\mu}[w(\delta(S_t))] = \sum_{uv\in E} w(u,v)\mathbb{P}_{t\sim\mu}[uv \in \delta(S_t)].$$

For any edge uv with $x(u) \le x(v)$, we have

$$\mathbb{P}_{t\sim\mu}[uv \in \delta(S_t)] = \mathbb{P}_{t\sim\mu}[x(u) \le t \le x(v)] = \int_{x(u)}^{x(v)} \mu(t)dt = \frac{1}{C}\int_{x(u)}^{x(v)} d(t)dt.$$

Note that for any $t \in [x(u), x(v)]$, $d(t) \le d(u) + (t - x(u))$ and $d(t) \le d(v) + (x(v) - t)$. So for any $c \in [x(u), x(v)]$, we have

$$\int_{x(u)}^{c} d(t)dt \le \int_{x(u)}^{c} d(u) + (t - x(u))dt = (c - x(u))d(u) + \frac{(c - x(u))^2}{2},$$

and

$$\int_{c}^{x(v)} d(t)dt \le \int_{c}^{x(v)} d(v) + (x(v) - t)dt = (x(v) - c)d(v) + \frac{(x(v) - c)^2}{2}.$$

We choose $c = (x(u) + x(v) - d(u) + d(v))/2$ to optimize the bound and get

$$\begin{aligned}
\int_{x(u)}^{x(v)} d(t)dt &= \int_{x(u)}^{c} d(t)dt + \int_{c}^{x(v)} d(t)dt \\
&\le (c - x(u))d(u) + \frac{(c - x(u))^2}{2} + (x(v) - c)d(v) + \frac{(x(v) - c)^2}{2} \\
&= \frac{(x(v) - x(u))^2}{4} + \frac{(x(v) - x(u))(d(u) + d(v))}{2} - \frac{(d(u) - d(v))^2}{4} \\
&\le \frac{(x(v) - x(u))^2}{4} + \frac{(x(v) - x(u))(d(u) + d(v))}{2},
\end{aligned}$$

where the last equality is obtained by expanding and simplifying the terms. Combining

the inequalities, we have

$$\begin{aligned}
\mathbb{E}_{t\sim\mu}[w(\delta(S_t))] &= \sum_{uv\in E} w(u,v)\mathbb{P}_{t\sim\mu}[uv \in \delta(S_t)] \\
&= \sum_{uv\in E} w(u,v)\frac{1}{C}\int_{x(u)}^{x(v)} d(t)dt \\
&\le \frac{1}{C}\sum_{uv\in E} w(u,v)\left(\frac{(x(v)-x(u))^2}{4} + \frac{(x(v)-x(u))(d(u)+d(v))}{2}\right) \\
&= \frac{1}{4C}\left(\mathcal{E}(x) + 2\sum_{uv\in E} w(u,v)(x(v)-x(u))(d(u)+d(v))\right) \\
&\le \frac{1}{4C}\left(\mathcal{E}(x) + 2\sqrt{\mathcal{E}(x)\left(\sum_{uv\in E} w(u,v)(d(u)+d(v))^2\right)}\right) \\
&\le \frac{1}{4C}\left(\mathcal{E}(x) + 2\sqrt{\mathcal{E}(x)\left(\sum_{uv\in E} w(u,v)2(d(u)^2+d(v)^2)\right)}\right) \\
&= \frac{1}{4C}\left(\mathcal{E}(x) + 2\sqrt{2\mathcal{E}(x)\left(\sum_{u\in V} \deg(u)d(u)^2\right)}\right) \\
&= \frac{1}{4C}\left(\mathcal{E}(x) + 2\sqrt{2}\sqrt{\mathcal{E}(x)}\|x-y\|_D\right),
\end{aligned}$$

where the second inequality is due to Cauchy-Schwarz inequality and the third inequality is due to the fact that $(a+b)^2 \le 2(a^2+b^2)$ for any $a, b \in \mathbb{R}$.

Finally, there exists t such that

$$\phi(S_t) = \frac{w(\delta(S_t))}{\mathrm{vol}(S_t)} \le \frac{\mathbb{E}_{t\sim\mu}[w(\delta(S_t))]}{\mathbb{E}_{t\sim\mu}[\mathrm{vol}(S_t)]} = 2\mathcal{E}(x) + 4\sqrt{2}\sqrt{\mathcal{E}(x)}\|x-y\|_D.$$

This completes the proof. □

3.3.3 Combining the lemmas

Now we are ready to prove the main theorem in this chapter.

Proof of Theorem 3.2. First we can assume x is non-negative with support size no more than half of the total volume. By Lemma 2.3, we can get $x' \ge 0$ with $\mathrm{supp}(x') \le \mathrm{vol}(V)/2$, $\mathcal{R}(x') \le \mathcal{R}(x)$, and the level sets of x' is a subset of level sets of x.

Consider the level sets $S_t = \{u \in V \mid x'(u) \ge t\}$. By combining Lemma 3.1 and

Lemma 3.2, we can find a probability density function μ such that

$$\begin{aligned}\frac{\mathbb{E}_{t\sim\mu}[w(\delta(S_t))]}{\mathbb{E}_{t\sim\mu}[\mathrm{vol}(S_t)]} &\le 2k\left(\mathcal{R}(x') + 2\sqrt{2}\frac{4\mathcal{R}(x')}{\sqrt{\lambda_k}}\right)\\ &\le 2k\mathcal{R}(x') + 8\sqrt{2}k\frac{\mathcal{R}(x')}{\sqrt{\lambda_k}}.\end{aligned}$$

Therefore there exists a t such that

$$\begin{aligned}\phi(S_t) &= \min_t \frac{w(\delta(S_t))}{\mathrm{vol}(S_t)}\\ &\le \frac{\mathbb{E}_{t\sim\mu}[w(\delta(S_t))]}{\mathbb{E}_{t\sim\mu}[\mathrm{vol}(S_t)]}\\ &\le 2k\mathcal{R}(x') + 8\sqrt{2}k\frac{\mathcal{R}(x')}{\sqrt{\lambda_k}},\end{aligned}$$

where the first inequality follows from an averaging argument similar to the argument in the proof of Theorem 2.4. This completes the proof. □

3.3.4 Tight example

We show that for any k we can find a graph G such that $\phi(G) = \Theta(k\lambda_2/\sqrt{\lambda_k})$. Our tight example is simply an n-cycle. Note that the eigenvalues of the Laplacian of an n-cycle is $1-\cos(2\pi j/n)$ where $j = 0, 1, \ldots, n-1$ (see Subsection 2.2.3). Therefore the k-th eigenvalue of the Laplacian of an n-cycle is $\Theta(k^2/n^2)$ as $n \to \infty$. The upper bound of the conductance given by our result is $O(k\lambda_2/\sqrt{\lambda_k}) = O(1/n)$, which matches the conductance of an n-cycle.

3.4 Improving the constant

In this section, we aim at improving the constant in the improved Cheeger's inequality. We shall prove that

$$\phi_{\text{sweep}} \le 2\sqrt{2}\frac{k\lambda_2}{\sqrt{\lambda_{k+1}}}.$$

This bound has tightened the previous bound in two ways. We improve the constant from $10\sqrt{2}$ to $2\sqrt{2}$ and also change λ_k to λ_{k+1}. The improvement is mainly based on two observations. First, we find a way to construct k-valued (instead of $(2k-1)$-valued) function which is close to the second eigenvector. Second, originally we only consider the distances from the values, and now we also use the signs to get a tighter bound.

We call a function $p : \mathbb{R} \to \mathbb{R}$ k-flipping if p is continuous with $k-1$ turning points, and between the turning points, p is linear with slope ± 1, which means $p(t)$ is in the form of $t + c$ or $-t + c$ for some c. For example, $p(t) = |t|$ is 2-flipping. Given a vector $x \in \mathbb{R}^V$, we will denote by $p \circ x$ the vector that satisfies $(p \circ x)(u) = p(x(u))$ for all $u \in V$. Note that $p \circ x$ is 1-Lipschitz with respect to x. The following is our main lemma.

Lemma 3.3. *Given a connected graph G and any vector $x \in \mathbb{R}^V$, there exists a k-flipping function p such that $\langle p \circ x, v_i \rangle_D = 0$ for $i = 1, 2, \ldots, k$.*

The intuition is that we have k degrees of freedom in choosing the k-flipping function. They include the positions of the $k-1$ turning points and one constant shift. So the freedom should allow us to satisfy k linear equations.

This intuition can be easily verified when $k = 2$. When $k = 2$, consider moving the turning point from negative infinity to positive infinity and set the constant shift so that $\langle p \circ x, v_1 \rangle = 0$. Then $a = p \circ x$ changes from x to $-x$ and thus $\langle a, v_2 \rangle_D$ changes sign. So at some point the inner product is zero.

We will assume the lemma for now and prove it at the end of this subsection.

Theorem 3.3. *For any graph G, we have*

$$\phi_{sweep} \leq 2\sqrt{2}\frac{k\lambda_2}{\sqrt{\lambda_{k+1}}}.$$

Proof. Let x be a non-negative vector with $\mathcal{R}(x) \leq \lambda_2$, $\mathrm{supp}(x) \leq \mathrm{vol}(V)/2$ and $\|x\|_D = 1$. Let p be the k-flipping function obtained from Lemma 3.3, and $a = p \circ x$. So we have $\mathcal{R}(a) \geq \lambda_{k+1}$ since $\langle a, v_i \rangle_D = 0$ for $i = 1, \ldots, k$. We will use $|p|$ as the probability distribution for the random threshold and show that one of the level sets has small conductance. Let $h(i) = \int_0^{x(i)} |p(t)|dt$ and $C = \int_0^{\max_u x(u)} |p(t)|dt$. Consider the randomized cut $S_t = \{u \mid x(u) > t\}$ where $t \in [0, \max_u x(u)]$ is chosen following the distribution $\mu = |p|/C$. Then

$$\begin{aligned}
\mathbb{E}_{t\sim\mu}[|S_t|] &= \sum_{u\in V} \deg(u)\mathbb{P}_{t\sim\mu}[u \in S_t] \\
&= \sum_{u\in V} \deg(u) \int_0^{x(u)} \frac{|p(t)|}{C} dt \\
&= \frac{1}{C}\sum_{u\in V} \deg(u)h(u),
\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbb{E}_{t\sim\mu}[w(\delta(S_t))] &= \sum_{uv\in E} w(u,v)\mathbb{P}_{t\sim\mu}[uv\in\delta(S_t)]\\ &= \sum_{uv\in E} w(u,v)\left|\int_{x(u)}^{x(v)} \frac{|p(t)|}{C}\right|\\ &= \frac{1}{C}\sum_{uv\in E} w(u,v)|h(u)-h(v)|.\end{aligned}$$

Therefore

$$\phi_{\text{sweep}} \le \frac{\mathbb{E}[w(\delta(S_t))]}{\mathbb{E}[|S_t|]} = \frac{\sum_{uv\in E} w(u,v)|h(u)-h(v)|}{\sum_{u\in V}\deg(u)h(u)}.$$

Now we will give an upper bound to $|h(u)-h(v)|$. You may refer to Figure 3.1.

If $a(u)$ and $a(v)$ are of the same sign, then

$$\left|\int_{x(u)}^{x(v)} |p(t)|dt\right| \le \frac{1}{4}(x(u)-x(v))^2 + \frac{1}{2}|x(u)-x(v)||a(u)+a(v)| - \frac{1}{4}(a(u)-a(v))^2.$$

Otherwise $a(u)$ and $a(v)$ are of different sign. Suppose $|a(u)| > |a(v)|$, then $|x(u)-x(v)| \ge |a(u)-a(v)| \ge |a(v)|$, and

$$\begin{aligned}\int_{x(u)}^{x(v)} |p(t)|dt &\le \frac{1}{4}(x(u)-x(v))^2 + \frac{1}{2}|x(u)-x(v)||a(u)-a(v)| - \frac{1}{4}(a(u)-a(v))^2\\ &\qquad + a(v)^2 - |a(v)||x(u)-x(v)|\\ &\le \frac{1}{4}(x(u)-x(v))^2 + \frac{1}{2}|x(u)-x(v)||a(u)-a(v)| - \frac{1}{4}(a(u)-a(v))^2.\end{aligned}$$

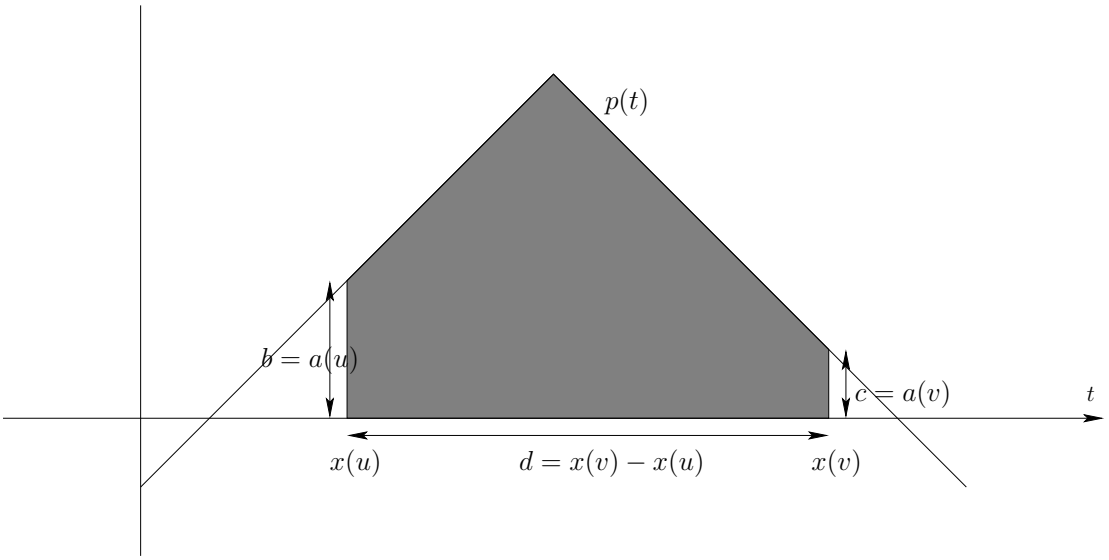
So in either case we have

$$|h(u)-h(v)| \le \frac{1}{4}(x(u)-x(v))^2 + \frac{1}{2}|x(u)-x(v)|(|a(u)|+|a(v)|) - \frac{1}{4}(a(u)-a(v))^2.$$

Now

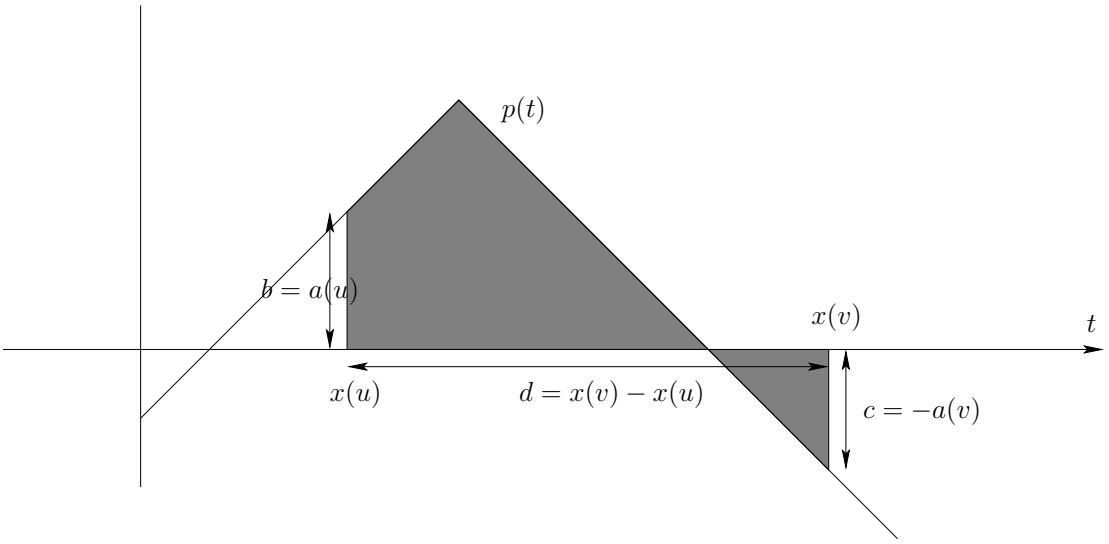
$$\begin{aligned}\sum_{uv\in E} w(u,v)|h(u)-h(v)| &\le \frac{1}{4}\mathcal{R}(x) + \frac{1}{2}\sqrt{\mathcal{R}(x)}\sqrt{2\sum_{u\in V}\deg(u)a(u)^2} - \frac{1}{4}\mathcal{R}(a)\|a\|_D^2\\ &\le \frac{1}{4}\lambda_2 + \frac{\sqrt{2}}{2}\sqrt{\lambda_2}\|a\|_D - \frac{1}{4}\lambda_{k+1}\|a\|_D^2,\end{aligned}$$

where the first inequality is by Cauchy-Schwarz inequality and the second inequality is by $\mathcal{R}(x)\le\lambda_2$ and $\mathcal{R}(a)\ge\lambda_{k+1}$. Note that $\|a\|_D^2 = \mathcal{E}(a)/\mathcal{R}(a) \le \mathcal{E}(x)/\mathcal{R}(a) \le \lambda_2/\lambda_{k+1}$.



(a) When $a(u)$ and $a(v)$ are of the same sign, the maximum attainable area is the area of the large triangle minus the areas of the smaller triangles:

$$\frac{(b+d+c)^2}{4} - \frac{b^2}{2} - \frac{c^2}{2} = \frac{1}{4}d^2 + \frac{1}{2}d(b+c) - \frac{1}{4}(b-c)^2.$$



(b) When $a(u)$ and $a(u)$ are of the different signs, the maximum attainable area when $|a(u)| > |a(v)|$ is the area of the large triangle minus the area of the left triangle plus the area of the right triangle:

$$\frac{(b+d-c)^2}{4} - \frac{b^2}{2} + \frac{c^2}{2} = \frac{1}{4}d^2 + \frac{1}{2}d(b+c) - \frac{1}{4}(b+c)^2 + c^2 - cd.$$

Figure 3.1: Maximum attainable areas

The quadratic equation

$$\frac{1}{4}\lambda_2 + \frac{\sqrt{2}}{2}\sqrt{\lambda_2}t - \frac{1}{4}\lambda_{k+1}t^2$$

is increasing when $t \le \sqrt{2\lambda_2/\lambda_{k+1}}$. Hence in the range $t \in [0, \sqrt{\lambda_2/\lambda_{k+1}}]$, the quadratic equation attains its maximum $(\sqrt{2}/2)\lambda_2/\sqrt{\lambda_{k+1}}$ at $t = \sqrt{\lambda_2/\lambda_{k+1}}$. This concludes that

$$\sum_{uv\in E} w(u,v)|h(u) - h(v)| \le \frac{\sqrt{2}}{2}\frac{\lambda_2}{\sqrt{\lambda_{k+1}}}.$$

On the other hand, suppose the $k-1$ turning points are $0 \le t_1 \le t_2 \le \cdots \le t_{k-1}$ and $t_i \le x(u) < t_{i+1}$, then

$$\begin{aligned}
h(u) &= \int_0^{x(u)} |p(t)|dt \\
&= \int_0^{t_1} |p(t)|dt + \int_{t_1}^{t_2} |p(t)|dt + \cdots + \int_{t_i}^{x(u)} |p(t)|dt \\
&\ge \frac{t_1^2}{4} + \frac{(t_2 - t_1)^2}{4} + \cdots + \frac{(x(u) - t_i)^2}{4} \\
&\ge \frac{x(u)^2}{4(i+1)} \\
&\ge \frac{x(u)^2}{4k},
\end{aligned}$$

where the first inequality is due to $p(t)$ is linear with slope ± 1 in each of the integrals, and the second inequality is by Cauchy-Schwarz inequality and the fact that there are $i+1$ terms. Therefore we have

$$\sum_{u\in V} \deg(u)h(u) \ge \frac{1}{4k}.$$

Combining the bounds, we conclude

$$\phi_{\text{sweep}} \le \frac{\sum_{uv\in E} w(u,v)|h(u) - h(v)|}{\sum_{u\in V} \deg(u)h(u)} \le \frac{2\sqrt{2}}{k\lambda_2}\sqrt{\lambda_{k+1}}.$$

This completes the proof. □

Finally we will prove Lemma 3.3.

Proof of Lemma 3.3. First recall that v_1 is parallel to χ_V and $\langle v_1, v_j\rangle_D = 0$ for $2 \le j \le$

k. So for any function p that satisfies $\langle p \circ x, v_j \rangle_D = 0$ for $2 \leq j \leq k$, we can choose

$$p' = p - \frac{\langle p \circ x, v_1 \rangle_D}{\langle \chi_V, v_1 \rangle_D},$$

so that

$$p' \circ x = p \circ x - \frac{\langle p \circ x, v_1 \rangle_D}{\langle \chi_V, v_1 \rangle_D} \chi_V,$$

and hence $\langle p' \circ x, v_j \rangle_D = 0$ for all j. Hence without loss we only need to satisfy to the later $k-1$ equations.

For any $c \in \mathbb{R}$, we define the function $p_c(t) = \max(t - c, 0)$. Then any k-flipping function p can be written as $p(t) = t - 2p_{c_1}(t) + 2p_{c_2}(t) - \cdots + (-1)^{k-1} 2p_{c_{k-1}}(t)$ for some real numbers $c_1 \leq \cdots \leq c_{k-1}$, where c_i are the turning points.

We define a function $g : \mathbb{R} \to \mathbb{R}^{\{2,\ldots,k\}}$ such that $g(c)_j = \langle p_c \circ x, v_j \rangle_D$ for $2 \leq j \leq k$. This function is continuous. For any $c \leq c_{\min} = \min_{i \in V} x(i)$, $g(c) = g(c_{\min})$ since $\langle \chi_V, v_j \rangle_D = 0$ for any $j \geq 2$. For any $c \geq c_{\max} = \max_{i \in V} x(i)$, $g(c) = 0$. Therefore $\langle p \circ x, v_j \rangle_D = \langle x, v_j \rangle_D - 2\langle p_{c_1}, v_j \rangle_D + 2\langle p_{c_2}, v_j \rangle_D - \cdots + (-1)^{k-1} 2\langle p_{c_{k-1}}, v_j \rangle_D = 0$ for $2 \leq j \leq k$ if and only if

$$g(c_{\min}) - 2g(c_1) + 2g(c_2) - \cdots + (-1)^{k-1} 2g(c_{k-1}) = 0.$$

This can be rewritten as

$$\sum_i (g(t_i) - g(s_i)) = \frac{1}{2}(0 - g(c_{\min})),$$

where (s_i, t_i) are disjoint intervals, and $g(t_1) = 0$ if k is even.

Now we make use of the following lemma[1].

Lemma 3.4 ([12]). *Let $r : [a, b] \to \mathbb{R}^n$ be a continuous curve. Then there exists a family of no more than $\lfloor (n+1)/2 \rfloor$ disjoint intervals $[s_i, t_i]$ such that*

$$\sum_i (r(t_i) - r(s_i)) = \frac{1}{2}(r(b) - r(a)).$$

Moreover when n is odd, we can choose $t_1 = b$.

Applying this lemma to g with $n = k - 1$ and interval $[c_{min}, c_{max}]$, we get no more

[1] We thank Sergei Ivanov for pointing us to this reference at mathoverflow.

than $\lfloor k/2 \rfloor$ disjoint intervals $[s_i, t_i]$ such that

$$\sum_i (g(t_i) - g(s_i)) = \frac{1}{2}(0 - g(c_{min})).$$

We set c_i such that $[c_{2i-1}, c_{2i}]$ are the intervals $[s_i, t_i]$. When k is odd, we are done. Otherwise $c_k = c_{max}$ and $g(c_k) = 0$. So we can omit that term. This completes the proof.

□

For the sake of completeness, we include the proof of Lemma 3.4 in the following.

Proof of Lemma 3.4. Let $(x_1, x_2, \ldots, x_{n+1})$ be a point on the n-sphere. For $0 \leq i \leq n+1$, let

$$y_i = (\sum_{j=1}^{i} x_j^2)(b - a) + a.$$

Now define $f : \mathbb{S}^n \to \mathbb{R}^n$ by

$$f(x_1, x_2, \ldots, x_{n+1}) = \sum_{i=1}^{n+1} \mathrm{sgn}(x_i)(r(y_i) - r(y_{i-1})).$$

Note that f is continuous and $f(-x) = -f(x)$. By the Borsuk-Ulam theorem[2], there exists a point $x \in \mathbb{S}^n$ such that $f(x) = 0$. Suppose $S_1 = \{i \mid x_i > 0\}$ and $S_2 = \{i \mid x_i < 0\}$. Let

$$A = \sum_{i \in S_1} (r(y_i) - r(y_{i-1})) \quad \text{and} \quad B = \sum_{i \in S_2} (r(y_i) - r(y_{i-1})).$$

Then $A = B$. Since $A + B = r(b) - r(a)$, we have $A = B = \frac{1}{2}(r(b) - r(a))$. Now we take the smaller index set S_1 or S_2, and when tie, take the one including $n + 1$. After combining the adjacent intervals if necessary, we obtain the result. □

3.5 Related problems and extensions

Our techniques can be extended to prove some results in related problems including k-way partitioning, balanced separator and maximum cut. In this section we will state some of these results [43] without proofs.

[2]The Borsuk-Ulam Theorem states that if $f : \mathbb{S}^n \to \mathbb{R}^n$ satisfies $f(-x) = -f(x)$, then there exists $x \in \mathbb{S}^n$ such that $f(x) = 0$.

Our results can be combined with several results in [45] to obtain the following multiway graph partitioning results. The proof idea is to apply the improved Cheeger's inequality on the disjointly supported function obtained by [45].

Corollary 3.1. *For any undirected graph G and integers $l > k \geq 2$, we have*

1. $$\phi_k(G) \leq O(lk^6)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

2. *For any $\delta \in (0,1)$,*
$$\phi_{(1-\delta)k}(G) \leq O\left(\frac{l(\log k)^2}{\delta^8 k}\right)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

3. *If G excludes K_h as a minor, then for any $\delta \in (0,1)$,*
$$\phi_{(1-\delta)k}(G) \leq O\left(\frac{h^4 l}{\delta^5 k}\right)\frac{\lambda_k}{\sqrt{\lambda_l}}.$$

We can use Theorem 3.1 repeatedly to prove a better bound for the balanced separator problem.

Theorem 3.4. *Let $\epsilon = \min_{|S|=|V|/2}\phi(S)$. There exists a polynomial time algorithm that finds a set S such that $|V|/5 \leq |S| \leq 4|V|/5$ and $\phi(S) \leq O(k/\lambda_k)\epsilon$.*

Similar techniques can be applied to improve the results in [73] for the maximum cut problem. The idea is to repeatedly find a near bipartite subgraph.

Theorem 3.5. *There exists a polynomial time algorithm that finds a cut $(S, \overline{S})$ such that if the optimal solution cuts $1-\epsilon$ fraction of edges, then $(S, \overline{S})$ cuts*

$$1 - O(k)\log\left(\frac{2-\lambda_{n-k}}{k\epsilon}\right)\frac{\epsilon}{2-\lambda_{n-k}}$$

fraction of edges.

It shows that we can obtain a better approximation for the maximum cut problem when there is a gap in the higher end of the spectrum, that is when $2-\lambda_{n-k}$ is large.

In the following we would like to remark that there is a relation between the spectral gap in the lower end and that in the higher end. We show that expander graphs implies such a gap in the higher end. This shows that if we can prove some properties of a graph assuming the spectral gap in the higher end, those properties also holds for expander graphs. Fact 3.1 and Theorem 3.5 implies the max cut problem is easy in expanders.

Fact 3.1. For any graph G, we have

$$\lambda_2 = O(2 - \lambda_{n-1}).$$

Proof. Let v_i be the eigenbasis of G. For each $\theta \in [0, 2\pi]$, let $v_\theta = (\cos\theta)v_{n-1} + (\sin\theta)v_n$ be a vector in the space spanned by v_{n-1} and v_n. Note that $\|v_\theta\|_D^2 = \cos^2\theta\|v_{n-1}\|_D^2 + \sin^2\theta\|v_n\|_D^2 = 1$. After that we define vectors z_θ such that

$$z_\theta(u) = |v_\theta(u)| - \langle |v_\theta|, v_1\rangle_D v_1(u).$$

We use the absolute value in the first term to guarantee the energy of z_θ is small, and we use the second term to normalize the vector so that $\langle z_\theta, v_1\rangle_D = 0$.

We will show that there exists θ such that the following two properties holds:

1. $z_\theta^T L z_\theta \le 2 - \lambda_{n-1}$.
2. $z_\theta^T D z_\theta = \Omega(1)$.

These properties together with $\langle z_\theta, v_1\rangle_D = 0$ by construction, we have

$$\lambda_2 \le \mathcal{R}(z_\theta) = \frac{z_\theta^T \mathcal{L} z_\theta}{z_\theta^T D z_\theta} = O(2 - \lambda_{n-1}).$$

The first property holds for all θ. Note that the Rayleigh quotient of v_θ satisfies

$$\mathcal{R}(v_\theta) = (\cos^2\theta)\lambda_{n-1} + (\sin^2\theta)\lambda_n \ge \lambda_{n-1}.$$

Therefore

$$\begin{aligned}
z_\theta^T L z_\theta &= \sum_{uv\in E} w(u,v)(z_\theta(u) - z_\theta(v))^2 \\
&= \sum_{uv\in E} w(u,v)(|v_\theta(u)| - |v_\theta(v)|)^2 \\
&\le \sum_{uv\in E} w(u,v)(v_\theta(u) + v_\theta(v))^2 \\
&= 2\sum_{u\in V} \deg(u) v_\theta(u)^2 - \sum_{uv\in E} w(u,v)(v_\theta(u) - v_\theta(v))^2 \\
&\le 2 - \lambda_{n-1}
\end{aligned}$$

Next we will show the second property holds for some θ. We will show that for a uniform random $\theta \in [0, 2\pi]$, in expectation $\mathbb{E}_\theta\left[\sqrt{1 - \|z_\theta\|_D^2}\right] \le \sqrt{8}/\pi < 1$, and therefore

there exists θ such that $\|z_\theta\|_D^2 \geq 1 - 8/\pi^2$. Since $z_\theta = |v_\theta| - \langle |v_\theta|, v_1\rangle_D \cdot v_1$, and $\langle z_\theta, v_1\rangle_D = 0$, we have

$$\|z_\theta\|_D^2 + \langle |v_\theta|, v_1\rangle_D^2 = \||v_\theta|\|_D^2 = 1.$$

So we have

$$\sqrt{1 - \|z_\theta\|_D^2} = |\langle |v_\theta|, v_2\rangle_D| = \left|\sum_{u\in V} \deg(u) v_1(u) |v_\theta(u)|\right|.$$

Without loss we assume v_1 is non-negative. Then we have

$$\begin{aligned}
\mathbb{E}_\theta\left[\sqrt{1 - \|z_\theta\|_D^2}\right] &= \mathbb{E}_\theta\left[\sum_{u\in V} \deg(u) v_1(u) |v_\theta(u)|\right] \\
&= \sum_{u\in V} \deg(u) v_1(u) \mathbb{E}_\theta[|v_\theta(u)|] \\
&= \sum_{u\in V} \deg(u) v_1(u) \frac{1}{2\pi} \int_0^{2\pi} |(\cos\theta) v_{n-1}(u) + (\sin\theta) v_n(u)| d\theta \\
&= \sum_{u\in V} \deg(u) v_1(u) \frac{1}{2\pi} \int_0^{2\pi} \sqrt{v_{n-1}(u)^2 + v_n(u)^2} \cdot |\cos(\theta - \alpha_n(u))| d\theta,
\end{aligned}$$

where $\alpha_n(u)$ is a number in $[0, 2\pi)$ such that $\cos\alpha_n(u) = v_{n-1}(u)/\sqrt{v_{n-1}(u)^2 + v_n(u)^2}$ and $\sin\alpha_n(u) = v_n(u)/\sqrt{v_{n-1}(u)^2 + v_n(u)^2}$. By the fact that $\int_0^{2\pi} |\cos\theta| d\theta = 4\int_0^{\pi/2} \cos\theta d\theta = 4$, we have

$$\begin{aligned}
\mathbb{E}_\theta\left[\sqrt{1 - \|z_\theta\|_D^2}\right] &= \sum_{u\in V} \deg(u) v_1(u) \frac{4}{2\pi} \sqrt{v_{n-1}(u)^2 + v_n(u)^2} \\
&\leq \frac{2}{\pi} \sqrt{\sum_{u\in V} \deg(u) v_1(u)^2 \cdot \sum_{u\in V} \deg(u) \left(v_{n-1}(u)^2 + v_n(u)^2\right)} \\
&= \frac{2}{\pi} \sqrt{\|v_1\|_D^2 (\|v_{n-1}\|_D^2 + \|v_n\|_D^2)} \\
&= \frac{\sqrt{8}}{\pi}.
\end{aligned}$$

This completes the proof. □

Remark 3.1. The hidden constant in the big-Oh notation is $1/(1-8/\pi^2) < 5.279$. This is the best we can do by this method (looking at the top 2-dimensional eigenspace and consider the absolute values). In the case of large odd cycles, $\lambda_2/(2-\lambda_{n-1}) = 4+O(1/n)$ and for any θ, $\mathcal{R}(z_\theta)/(2-\lambda_{n-1}) = 1/(1-8/\pi^2) + O(1/n)$. We believe 4 is the tight constant for this bound, however we do not have enough techniques to prove this now.

3.5.1 Local improved Cheeger's inequality

In this subsection, we give an extension of the improved Cheeger's inequality in the setting that we only have local information of the graph. Suppose there is a massive underlying graph G, and we only know the information over a small subset of vertices S. Then we can still try to relate the best conductance over this subset and the spectrum of some local matrix associated with S. More precisely, let A_S, D_S and L_S be the restrictions of $A(G)$, $D(G)$ and $L(G)$ over the subset S respectively, and let $\phi_S = \min_{T\subseteq S}\phi(T)$. Let $\mathcal{L}_S = D_S^{-1/2}L_S D_S^{-1/2} = I_{|S|\times|S|} - D_S^{-1/2}A_S D_S^{-1/2}$. Suppose $\tilde{v}_{S,1}, \tilde{v}_{S,2}, \ldots, \tilde{v}_{S,|S|}$ are the eigenvectors of $\mathcal{L}_S$ with corresponding eigenvalues $0 \le \lambda_{S,1} \le \lambda_{S,2} \le \ldots \lambda_{S,|S|}$, and as usual, $v_{S,i} = D_S^{-1/2}\tilde{v}_{S,i}$ are the (right) eigenvectors of $D_S^{-1}L_S$. Then Chung [20] proved the following theorem.

Theorem 3.6 ([20])**.** *For any graph G and any subset of vertices $S \subseteq V$, we have*

$$\lambda_{S,1} \le \phi_S \le \sqrt{2\lambda_{S,1}}.$$

Proof. We denote

$$\mathcal{R}_S(x) = \mathcal{R}_{L_S,D_S}(x) = \frac{x^T L_S x}{x^T D_S x}.$$

Suppose the minimum conductance ϕ_S is attained by $T \subseteq S$, then

$$\mathcal{R}_S(\chi_T) = \frac{\chi_T^T L_S \chi_T}{\chi_T^T D_S \chi_T} = \frac{w(\delta(T))}{\mathrm{vol}(T)} = \phi(T).$$

Therefore by the Courant-Fischer formula, $\phi(T) \ge \lambda_{S,1}$.

For the other side, we note that by applying Theorem 2.4 on the first local eigenvector $v_{S,1}$, we can get a subset $T \subseteq S$ such that $\phi(T) \le \sqrt{2\lambda_{S,1}}$. Therefore $\phi_S \le \phi(T) \le \sqrt{2\lambda_{S,1}}$. This completes the proof. □

We extend our main result in this setting that shows better bounds when the local higher eigenvalues are large.

Theorem 3.7. *For any graph G, any subset of vertices $S \subseteq V$ and any $k \ge 1$, we have*

$$\phi_S \le O(k)\frac{\lambda_{S,1}}{\sqrt{\lambda_{S,k}}}.$$

Proof. We first construct an auxiliary graph $G_M = (V', w_M)$ by collapsing all the vertices

other than those in S. More precisely, we let $V' = S \cup \{u^*\}$ and w_M satisfies

$$w_M(u,v) = \begin{cases} w(u,v) & \text{if } u \in S, v \in S. \\ \sum_{v \notin S} w(u,v) = \deg(u) - \sum_{v \in S} w(u,v) & \text{if } u \in S, v = u^*. \\ \sum_{u \notin S} w(u,v) = \deg(v) - \sum_{v \in S} w(u,v) & \text{if } u = u^*, v \in S. \\ M & \text{if } u = v = u^*. \end{cases}$$

We analyze the spectrum of G_M as $M \to \infty$. Note that

$$\mathcal{L}(G_M) = \left(\begin{array}{c|c} \mathcal{L}_S & b \\ \hline b^T & c \end{array} \right),$$

where $b \to 0$ and $c \to 0$ as $M \to \infty$, since $b(u) = w(u,u^*)/\sqrt{M}$ and $c = w(\delta(S))/M$. Since eigenvalues are continuous to the matrix, we have $\lambda_1(G_M) = 0$ with the corresponding eigenvector $(0,\dots,0,1)$ and $\lambda_{k+1}(G_M) \to \lambda_k(\mathcal{L}_S) = \lambda_{S,k}$ with the corresponding eigenvectors $(\tilde{v}_{S,k}, 0)$.

Now by applying Theorem 3.2 on the first local eigenvector $v_{S,1}$, we have for any $k \geq 1$,

$$\phi_{\text{sweep}}(v_{S,1}) \leq O(k) \frac{\mathcal{R}_{G_M}(v_{S,1})}{\sqrt{\lambda_{k+1}(G_M)}}.$$

Note that $\mathcal{R}_{G_M}(v_{S,1}) = \lambda_{S,1}$ regardless of M, and $\lambda_{k+1}(G_M) \to \lambda_{S,k}$. Hence by letting $M \to \infty$, we have for any $k \geq 2$,

$$\phi_{\text{sweep}}(v_{S,1}) \leq O(k) \frac{\lambda_{S,1}}{\sqrt{\lambda_{S,k}}}.$$

As $\text{supp}(v_{S,1}) \subseteq S$, $\phi_S \leq \phi_{\text{sweep}}(v_{S,1})$. This completes the proof. □

Chapter 4

Random walks

In this chapter we obtain two results using random walks. In Section 4.1, we show that by computing an approximate random walk vector, we can find a small sparse cut locally. This algorithm has almost the same performance guarantee as the spectral partitioning algorithm, while giving a non-trivial bound on the size of the output set.

In Section 4.2, we prove a tight lower bound on the expansions of graph powers. This result can be used to amplify the hardness result for small set expansion problems, and shows that some Cheeger-type inequalities can be reduced to the case where the parameters are constant. The results presented in this chapter are mainly based on joint work with Lap Chi Lau [41, 42].

4.1 Finding small sparse cuts

Throughout this section, we consider unweighted simple graphs, and hence $\mathrm{vol}(V) = 2m$. Recall that the small sparsest cut problem with size parameter δ is to determine $\phi_\delta(G) = \min_{\mathrm{vol}(S)\leq\delta\,\mathrm{vol}(V)} \phi(S)$ and find a subset S with $\mathrm{vol}(S) \leq \delta\,\mathrm{vol}(V)$ achieving the minimum; see Subsection 2.4.3.

4.1.1 Our results

We show that the techniques developed in local graph partitioning algorithms [67, 21] can be used to obtain bicriteria approximation algorithms for the small sparsest cut problem. We note that the algorithm in Theorem 4.1 is the same as the algorithm of Arora, Barak and Steurer [7], but we adapt the analysis in local graph partitioning algorithms to prove a tradeoff between the conductance guarantee and the volume of the output set.

Theorem 4.1. *Given an undirected graph $G = (V, w)$ and a volume parameter k, let $\varphi = \phi_{k/vol(V)}$ be the minimum conductance over sets with volume not larger than k. There is a polynomial time algorithm with the following guarantees:*

1. *Find a set S with $\phi(S) = O(\sqrt{\varphi/\epsilon})$ and $\mathrm{vol}(S) \leq k^{1+\epsilon}$ for any $\epsilon > 1/k$.*
2. *Find a set S with $\phi(S) = O(\sqrt{\varphi \log k/\epsilon})$ and $\mathrm{vol}(S) \leq (1+\epsilon)k$ for any $\epsilon > 2\log k/k$.*

For the small sparsest cut problem, when k is sublinear ($k = O(m^c)$ for some $c < 1$), the performance guarantee of the bicriteria approximation algorithm in Theorem 4.1(2) is similar to that of Raghavendra, Steurer and Tetali [59]. Also, when k is sublinear, the conductance guarantee of Theorem 4.1(1) is independent of n, which matches the performance of spectral partitioning while having a bound on the volume of the output set. These show that random walk algorithms can also be used to give nontrivial bicriteria approximations for the small sparsest cut problem. We note that the result of Andersen and Peres [6] implies a similar statement to Theorem 4.1(2), with the same conductance guarantee and $\mathrm{vol}(S) = O(k)$. The algorithms in Theorem 4.1 can also be implemented locally by using the truncated random walk algorithm.

Theorem 4.2. *For an undirected graph $G = (V, w)$ and a set $S^* \subseteq V$, given $\varphi \geq \phi(S^*)$ and $k \geq \mathrm{vol}(S^*)$, there exists an initial vertex $v \in S^*$ such that the truncated random walk algorithm can find a set S with $\phi(S) \leq O(\sqrt{\varphi/\epsilon})$ and $\mathrm{vol}(S) \leq O(k^{1+\epsilon})$ for any $\epsilon > 2/k$. The runtime of the algorithm is $\tilde{O}(k^{1+2\epsilon}\varphi^{-2})$.*

When k is sublinear, the conductance guarantee of Theorem 4.2 matches that of spectral partitioning, improving on the conductance guarantees in previous local graph partitioning algorithms. However, we note that our notion of a local graph partitioning algorithm is weaker than previous work [67, 5, 6], as they proved that a random initial vertex v will work with a constant probability, while we could only prove that there exists an initial vertex that will work and unable to prove the high probability statement.

In Subsection 4.1.4, we discuss a connection to the small set expansion conjecture.

Independent work

Oveis Gharan and Trevisan [55] prove Theorem 4.1 independently. They also prove a stronger version of Theorem 4.2, with a faster running time ($\tilde{O}(k^{1+2\epsilon}\phi^{-1/2})$) and also the algorithm works for a random initial vertex in S with constant probability. They use the evolving set process instead of the random walk process we considered here.

We note that their result implies that our truncated random walk algorithm will also succeed with constant probability if we start from a random initial vertex in S.

4.1.2 Finding small sparse cuts

The organization of this subsection is as follows. First we present our algorithm in Theorem 4.1 and the proof outline. Then we present the analysis and complete the proof of Theorem 4.1.

Algorithm

Our algorithm is simple and is the same as in Arora, Barak and Steurer [7]. For each vertex v, we use it as the initial vertex of the random walk, and compute the probability (row) vectors $p_t := \chi_v^T W^t = \chi_v^T (I_{n\times n} + D^{-1}A)/2$ for $1 \leq t \leq O(n^2 \log n)$. Then we output the set of smallest conductance among all level sets $S_{t,i} := \{u \mid p_t(u) \geq p_t(i)\}$ (of all initial vertices) of volume at most ck, where in Theorem 4.1(1) we set $c = k^\epsilon$ and in Theorem 4.1(2) we set $c = 1 + \epsilon$. Clearly this is a polynomial time algorithm.

Analysis

The techniques are from the work of Spielman and Teng [67] and Chung [21]. Our goal in Theorem 4.1(1) is equivalent to distinguishing the following two cases: (a) there is a set S^* with $\mathrm{vol}(S^*) \leq k$ and $\phi(S^*) \leq \varphi$, or (b) the conductance of every set of volume at most ck is at least $\Omega(\sqrt{\varphi})$ for some small $c > 1$ which may depend on k. As in [67], we use the method of Lovász and Simonovits [51] that considers the total probability of the k edges with largest probability after t steps of random walks, denoted by $C^{(t)}(k)$; see Subsection 2.3.2 for the definition. In case (a), we use the idea of Chung [21] that uses the local eigenvector of S^* of the Laplacian matrix to show that there exists an initial vertex such that $C^{(t)}(k) \geq (1 - \varphi/2)^t$. In case (b), we use Lemma 2.4 to prove that $C^{(t)}(k) \leq 1/c + \sqrt{k}(1 - M\varphi)^t$ for a small enough positive constant M, no matter what is the initial vertex of the random walk. Hence, say when $c \geq k^{0.01}$, by setting $t = \Theta(\log k/\varphi)$, we expect that $C^{(t)}(k)$ is significantly greater than $1/c$ in case (a) but at most $1/c$ plus a negligible term in case (b), and so we can distinguish the two cases. Theorem 4.1(2) is a corollary of Theorem 4.1(1). To prove Theorem 4.2, we use the truncated random walk algorithm as in [67] to give a bound on the runtime.

Upper bound

We prove the upper bound using Lemma 2.4. We note that the following statement is true for any initial probability distribution, in particular when $p = \chi_v$ for any v. The proof is very similar to that of Theorem 2.6. The difference is that we only consider the conductances of small sets here.

Theorem 4.3. *Suppose for all $t' \le t$ and $i \in [n]$, we have $\phi(S_{t',i}) \ge \phi_1$ whenever $\mathrm{vol}(S_{t',i}) \le l \le m$. Then the curve satisfies for all $x \le \mathrm{vol}(V)$,*

$$C^{(t)}(x) \le \frac{x}{l} + \sqrt{x}(1 - \frac{\phi_1^2}{8})^t.$$

Proof. Let the extreme points x_i satisfy $0 = x_0 \le x_1 \le x_2 \le \cdots \le x_i \le l < x_{i+1}$. Note that the function on the right hand side is concave and $C^{(t)}$ is linear between extreme points and between x_i and l. So we only need to show the inequality for extreme points and for $x \ge l$. When $x \ge l$, the inequality always hold as for any t,

$$\frac{x}{l} + \sqrt{x}(1 - \frac{\phi_1^2}{8})^t \ge 1 \ge C^{(t)}(x)$$

Now we would prove by induction. When $t = 0$ the inequality is trivial for any $x \ge 1$,

$$\frac{x}{l} + \sqrt{x}(1 - \frac{\phi_1^2}{8})^0 \ge 1 \ge C^{(0)}(x).$$

When $t > 0$ and x is an extreme point, by Lemma 2.4 and the induction hypothesis we have

$$\begin{aligned} C^{(t)}(x) &\le \frac{1}{2}(C_{t-1}(x - \phi_1 x) + C_{t-1}(x + \phi_1 x)) \\ &= \frac{x}{l} + \frac{1}{2}\sqrt{x}(1 - \frac{\phi_1^2}{8})^{t-1}(\sqrt{1-\phi_1} + \sqrt{1+\phi_1}) \\ &\le \frac{x}{l} + \sqrt{x}(1 - \frac{\phi_1^2}{8})^t, \end{aligned}$$

where the last inequality follows from Taylor expansions of $\sqrt{1 \pm \phi_1}$. □

Lower bound

For a probability vector p over V and any subset $S \subseteq V$, we denote $p(S) := \sum_{u \in S} p(u)$ as the probability mass in S. Our idea is to use the local eigenvector of S of the normalized Laplacian matrix to show that there is an initial distribution such that

$p_t(S) \geq (1 - \phi(S)/2)^t$.

Theorem 4.4. *Assume $S \subseteq V$ where $\mathrm{vol}(S) \leq m$ and $\phi(S) \leq \phi_2$. Then there exists a vertex v such that if $p = \chi_v^T$, then*

$$p_t(S) \geq (1 - \frac{\phi_2}{2})^t.$$

Proof. We will show that if the probability distribution p_0 is proportional to the smallest local eigenvector, then $p_t(S) = p_0 W^t \chi_S \geq (1 - \phi_2/2)^t$. The existence of the good starting vertex follows from the linearity of the operator $W^t \chi_S$ and the fact that p is a convex combination of χ_v where $v \in S$.

Let λ_S ans $\tilde{v}_S$ be the smallest eigenvector and the corresponding eigenvector of $\mathcal{L}_S$, which is the restriction of $\mathcal{L}$ over the vertex set S. By Theorem 3.6, we have $\lambda_S \leq \phi(S) \leq \phi_2$. Also, by the Perron-Frobenius theorem on $\mathcal{A}_T$ for every connected component T in S, the eigenvector $\tilde{v}_S$ can be assumed to be non-negative. Let p_S denote the restriction of p on S, and $p_{t,S}$ denote the restriction of p_t on S. We set the initial distribution p such that $p_S = (D_S^{-1/2} \tilde{v}_S)^T$ and $p_{V-S} = 0$, and we can rescale $\tilde{v}_S$ such that p_S is a probability distribution. Note that p_S is a left eigenvector of $D_S^{-1} A_S$ with eigenvalue $1 - \lambda_S$. We would show that $p_{t,S} \geq (1 - \lambda_S/2)^t p_S$ by induction. Clearly, the statement is true when $t = 0$. For $t > 0$, we have $p_{t,S} = (p_{t-1} W)_S \geq p_{t-1,S} W_S$ since p and W are non-negative. Therefore

$$\begin{aligned}
p_{t,S} &\geq p_{t-1,S} W_S \\
&= p_{t-1,S} \cdot \frac{1}{2}(I_S + D_S^{-1} A_S) \\
&\geq (1 - \frac{\lambda_S}{2})^{t-1} p_S \cdot \frac{1}{2}(I_S + D_S^{-1} A_S) \\
&= (1 - \frac{\lambda_S}{2})^t p_S,
\end{aligned}$$

where the second inequality follows from the induction hypothesis and the last equality holds since p_S is a left eigenvector of W_S. Finally,

$$p_t(S) = p_{t,S}(S) \geq (1 - \frac{\lambda_S}{2})^t p_S(S) \geq (1 - \frac{\phi_2}{2})^t.$$

This completes the proof. □

Remark 4.1. In the independent work [55], Oveis Gharan and Trevisan proves a stronger statement of Theorem 4.4. They show that there exists a subset $S' \subseteq S$ with

$\mathrm{vol}(S') \geq \mathrm{vol}(S)/2$, such that if the starting vertex is from S', then

$$p_t(S) = \Omega\left(\left(1 - \frac{3\phi_2}{2}\right)^t\right).$$

Combining the bounds

We combine the upper bound and the lower bound to prove Theorem 4.1.

Proof of Theorem 4.1. Note that Theorem 4.1(1) is trivial if $\varphi = \phi_{k/\mathrm{vol}(V)} \geq \epsilon$, and so we assume $\varphi < \epsilon$. We also assume $\epsilon \leq 0.01$, as otherwise we reset $\epsilon = 0.01$ and lose only a constant factor.

The algorithm is simple. Set $T = \epsilon k^2 \log k/4$. For each vertex v, set $p = \chi_v$ and compute $S_{t,i}$ for all $t \leq T$ and $i \in [n]$. Denote these sets by $S_{t,i,v}$ to specify the starting vertex u. Output a set $S = S_{t,i,v}$ that achieves the minimum in $\min_{\mathrm{vol}(S_{t,i,v}) \leq k^{1+\epsilon}} \phi(S_{t,i,v})$. Clearly, the algorithm runs in polynomial time.

We claim that $\phi(S) \leq 4\sqrt{\varphi/\epsilon}$. Suppose to the contrary that the algorithm does not return such a set. Consider $t = \epsilon \log k/(2\varphi)$. Note that $t \leq T$ as $\varphi \geq 1/k^2$ for a simple unweighted graph. Applying Theorem 4.3 with $l = k^{1+\epsilon}$, for any starting vertex v, we have

$$\begin{aligned} C^{(t)}(k) &\leq \frac{k}{k^{1+\epsilon}} + \sqrt{k}(1 - \frac{2\varphi}{\epsilon})^t \\ &\leq k^{-\epsilon} + \sqrt{k}\exp(-\frac{2\varphi}{\epsilon}\frac{\epsilon \log k}{2\varphi}) \\ &= k^{-\epsilon} + \sqrt{k}\exp(-\log k) \\ &= k^{-\epsilon} + k^{-1/2}. \end{aligned}$$

On the other hand, suppose S^* is a set with $\mathrm{vol}(S^*) \leq k$ and $\phi(S^*) = \varphi$. Then Theorem 4.4 says that there exists a starting vertex $v^* \in S^*$ such that

$$\begin{aligned} p_t(S^*) &\geq (1 - \frac{\varphi}{2})^t \\ &\geq \exp(-\varphi t) \\ &= \exp(-\frac{1}{2}\epsilon \log k) \\ &= k^{-\epsilon/2} \\ &> k^{-\epsilon} + k^{-1/2}, \end{aligned}$$

where the second inequality holds when $\varphi < 0.01$ and the last inequality holds for $k \geq 1/\epsilon$ and $\epsilon < 0.01$. This is contradicting since $C^{(t)}(k) \geq p_t(S^*)$ for that starting vertex, completing the proof of Theorem 4.1(1).

Now we obtain Theorem 4.1(2) as a corollary of Theorem 4.1(1). Set $\epsilon' = \frac{\epsilon}{2 \log k}$. Then $k^{1+\epsilon'} \leq (1+\epsilon)k$. By using Theorem 4.1(1) with ϵ', we have Theorem 4.1(2). □

4.1.3 Local graph partitioning

To implement the algorithm locally, we use truncated random walk as in [67]. Let $q_0 = p = \chi_v$. For each $t \geq 0$, we define p'_t by setting $p'_t(u) = 0$ if $q_t(u) < \epsilon \deg(u)$ and setting $p'_t(u) = q_t(u)$ if $q_t(u) \geq \epsilon \deg(v)$, and we define $q_{t+1} = p'_t W$. Then, we just use p'_t to replace p_t in the algorithm in Subsection 4.1.2. To prove that the truncated random walk algorithm works, we first show that p'_t is a good approximation of p_t and can be computed locally. Then we show that the curve defined by p'_t satisfies the upper bound in Theorem 4.3, and it almost satisfies the lower bound in Theorem 4.4. Finally we combine the upper bound and the lower bound to prove Theorem 4.2.

Computing truncated distributions

Lemma 4.1. *There is an algorithm that computes p'_t such that $p'_t \leq p_t \leq p'_t + \epsilon t d$ for every $0 \leq t \leq T$, with time complexity $O(T/\epsilon)$, where d is the degree (row) vector.*

Proof. First we prove the approximation guarantee. By induction, we have the upper bound

$$p'_t \leq q_t = p'_{t-1} W \leq p_{t-1} W = p_t.$$

Also, by induction, we have the lower bound

$$p_t = p_{t-1} W \leq (p'_{t-1} + \epsilon(t-1)d) W = q_t + \epsilon(t-1)d \leq p'_t + \epsilon t d.$$

Next we bound the computation time. Let S_t be the support of p'_t. In order to compute q_{t+1} from p'_t, we need to update each vertex $u \in S_t$ and its neighbors. Using a perfect hash function, the neighbors of a vertex u can be updated in $O(\deg(u))$ steps, and thus q_{t+1} and p'_{t+1} can be computed in $O(\mathrm{vol}(S_t))$ steps. Since each vertex $u \in S_t$ satisfies $p'_t \geq \epsilon \deg(u)$, we have $\mathrm{vol}(S_t) = \sum_{u \in S_t} d(u) \leq p_t(S_t)/\epsilon \leq 1/\epsilon$, and this completes the proof. □

Approximate upper bound

We use the truncated probability distributions to define the curve $C'^{(t)}$, and will prove an upper bound for $C'^{(t)}$. Since truncation only decreases the probabilities, intuitively the same upper bound applying to $C^{(t)}$ also applies to $C'^{(t)}$. Note that p'_t may not be a probability distribution and $C'^{(t)}(\mathrm{vol}(V))$ may be less than one. And we define the level sets $S'_{t,i} = \{u_1, u_2, \ldots, u_i\}$ when we order the vertices such that

$$\frac{p'_t(u_1)}{\deg(u_1)} \geq \frac{p'_t(u_2)}{\deg(u_2)} \geq \cdots \geq \frac{p'_t(u_n)}{\deg(u_n)}.$$

We show that C'_t would satisfy the same upper bound as in Theorem 4.3.

Lemma 4.2. *Suppose for all $t \leq T$ and $i \in [n]$, we have $\phi(S'_{t,i}) \geq \phi_1$ whenever $\mathrm{vol}(S'_{t,i}) \leq l \leq m$. Then for all $x \leq \mathrm{vol}(V)$, we have*

$$C'^{(t)}(x) \leq f_t(x) := \frac{x}{l} + \sqrt{x}(1 - \frac{\phi_1^2}{8})^t.$$

Proof. Let $x'_i = \sum_{u \in S'_{t,i}} \deg(u)$ be the extreme points defined by p'_t. By the same proof as in Theorem 4.3. It suffices to prove that Lemma 2.4 still holds after replacing p_t by p'_t. It means that we need to show if $x = x'_i \leq l$ is an extreme point (at time t), $S = S'_{t,j}$ is the corresponding set of vertices and $\mathrm{vol}(S) \geq \phi$, then

$$C'^{(t)}(x) \leq \frac{1}{2}(C'^{(t-1)}(x - \phi x) + C'^{(t-1)}(x + \phi x)).$$

Since $p'_t \leq q_t$, we have

$$C'^{(t)}(x) = C(p'_t, x) \leq C(q_t, x).$$

By Lemma 2.4,

$$C(q_t, x) = C(p'_{t-1}W, x) \leq \frac{1}{2}(C'^{(t-1)}(x - \phi x) + C'^{(t-1)}(x + \phi x)).$$

This completes the proof by Theorem 4.3. □

Approximate lower bound

By Lemma 4.1 we can easily get a lower bound on $p'(S)$ with a good initial vertex.

Lemma 4.3. *Assume $S \subseteq V$ where $\mathrm{vol}(S) \leq m$ and $\phi(S) \leq \phi_2$. Then there exists a*

vertex v such that if $p = \chi_v$, then

$$p'_t(S) \geq (1 - \frac{\phi_2}{2})^t - \epsilon t \operatorname{vol}(S).$$

Proof. By Theorem 4.4, we can get a vertex v such that if $p = \chi_v$, then $p_t(S) \geq (1-\phi_2/2)^t$. By Lemma 4.1, for any vertex u we have $p'_t(u) \geq p_t(u) - \epsilon t \deg(u)$. Therefore we have

$$p'_t(S) \geq \sum_{u \in S} p_t(u) - \epsilon t \sum_{u \in S} \deg(u) = p_t(S) - \epsilon t \operatorname{vol}(S) \geq (1 - \frac{\phi_2}{2})^t - \epsilon t \operatorname{vol}(S).$$

□

Combining the bounds

We combine the approximate bounds to prove Theorem 4.2.

Proof of Theorem 4.2. We would prove a slighter stronger statement. We show that whenever $v \in S^*$ satisfies $p_t(S^*) \geq (1 - \varphi/2)^t/4$ when $p = \chi_v$, then the truncated random walk algorithm will output a set S with $\operatorname{vol}(S) \leq O(k^{1+\epsilon})$ and $\phi(S) \leq 4\sqrt{\varphi/\epsilon}$. The running time of the algorithm is $O(\epsilon^2 k^{1+2\epsilon}(\log k)^3/\varphi^2)$. Note that Theorem 4.4 states the existence of such vertex. We relax the requirement $p_t(S^*) \geq (1 - \varphi/2)^t$ a bit so as to allow for more good initial vertices. The constant 4 in $p_t(S^*) \geq (1 - \varphi/2)^t/4$ is not important and can be replaced by any constant $c \geq 1$.

Set $T = \epsilon \log k/(2\varphi)$ and $\epsilon' = k^{-1-\epsilon}/(20T)$. Applying Lemma 4.1 with T and ϵ', we can compute all p'_t and thus $S'_{t,i}$ for all $t \leq T$ and $i \in [5k^{1+\epsilon}]$ in $O(T \log k/\epsilon') = O(\epsilon^2 k^{1+\epsilon}(\log k)^3/\varphi^2)$ steps (with an additional $\log k$ factor for sorting). By Lemma 4.3, a good starting vertex v will give

$$p'_T(S^*) \geq p_T(S^*) - \epsilon' T \operatorname{vol}(S^*) \geq \frac{1}{4}(1 - \varphi/2)^T - \epsilon' T \operatorname{vol}(S^*).$$

We claim that one of the set $S = S'_{t,i}$ must satisfy $\operatorname{vol}(S) \leq 5k^{1+\epsilon}$ and $\phi(S) \leq 4\sqrt{\varphi/\epsilon}$.

Otherwise, setting $\phi_1 \geq 4\sqrt{\varphi/\epsilon}$, we have

$$\begin{aligned} p'_T(S^*) &\geq \frac{1}{4}(1-\frac{\varphi}{2})^T - \epsilon' T \operatorname{vol}(S^*) \\ &\geq \frac{1}{4}\exp(-\varphi T) - \frac{k^{-\epsilon}}{20} \\ &= \frac{k^{-\epsilon/2}}{4} - \frac{k^{-\epsilon}}{20} \\ &> \frac{k^{-\epsilon}}{5} + k^{-1/2} \\ &\geq \frac{k}{5k^{1+\epsilon}} + \sqrt{k}(1-\frac{\phi_1^2}{8})^T \\ &\geq C'^{(T)}(k). \end{aligned}$$

Here the second inequality holds since $\varphi < 0.01$, and the third inequality holds using the fact that for $k \geq 1/\epsilon$ and $\epsilon \leq 0.01$,

$$k^{-\epsilon/2} > k^{-\epsilon} + 4k^{-\frac{1}{2}}.$$

This contradicts to $C'^{(T)}(k) \geq p'_T(S^*)$, completing the proof of Theorem 4.2. □

4.1.4 Concluding remarks

Small set expansion conjecture

We presented a bicriteria approximation algorithm for the small sparsest cut problem with conductance guarantee independent of n, but the volume of the output set is $k^{1+\epsilon}$. We note that if one can also guarantee that the volume of the output set is at most Mk for an absolute constant M, then one can disprove the small set expansion conjecture, which states that for any constant ϵ there exists a constant δ such that distinguishing $\phi_\delta(G) < \epsilon$ and $\phi_\delta(G) > 1-\epsilon$ is NP-hard. This can be viewed as an evidence that our analysis is almost tight, or an evidence that the small set expansion problem is not NP-hard. We note that this is also observed by Raghavendra, Steurer and Tulsiani [60].

More formally, suppose there is a polynomial time algorithm with the following guarantee: given a graph G, always output a set S with $\phi(S) = f(\phi_{k/\operatorname{vol}(V)})$ and $\operatorname{vol}(S) = Mk$ where $f(x)$ is a function that tends to zero when x tends to zero (e.g. $f(x) = x^{1/100}$) and M is an absolute constant. Then we claim that there is a (small) constant ϵ such that whenever $\phi_{k/\operatorname{vol}(V)} < \epsilon$ there is a polynomial time algorithm to return a set S with $\phi(S) < 1-\epsilon$ and $\operatorname{vol}(S) \leq k$.

We assume that G is a d-regular graph, as in [58] where the small set expansion conjecture was formulated. Suppose there is a subset S^* with $|S^*| = k$ and $\phi(S^*) < \epsilon$. First we use the algorithm to obtain a set S with $\phi(S) \leq f(\epsilon)$ and assume $|S| = Mk$ (instead of $|S| \leq Mk$). Next we show that a random subset $S' \subseteq S$ of size exactly k will have $\phi(S') < 1 - \epsilon$ with a constant probability for a small enough ϵ. Let $E(S)$ be the set of edges with both endpoints in S. Each edge in $E(S)$ has probability $2(1/M)(1-1/M)$ to be in $\delta(S')$. So, the expected value of

$$w(\delta(S')) \leq w(\delta(S)) + 2(\frac{1}{M})(1 - \frac{1}{M})|E(S)|.$$

By construction $\mathrm{vol}(S') = kd$, and so the expected value of

$$\phi(S') \leq \frac{w(\delta(S))}{kd} + \frac{2}{kd}(\frac{1}{M})(1 - \frac{1}{M})|E(S)|.$$

Note that $|E(S)| \leq Mkd/2$ and $w(\delta(S))/(kd) = M\phi(S) \leq Mf(\epsilon)$, so the expected value of

$$\phi(S') \leq Mf(\epsilon) + 1 - \frac{1}{M}.$$

For a small enough ϵ depending only on M, the expected value of $\phi(S') \leq 1 - 10\epsilon$. Therefore, with a constant probability, we have $\phi(S') < 1 - \epsilon$.

We show that random walks can be used to obtain nontrivial bicriteria approximation algorithms for the small sparsest cut problem. We do not know of an example showing that our analysis is tight. It would be interesting to find examples showing the limitations of random walk algorithms (e.g. showing that they fail to disprove the small set expansion conjecture).

Simpler proof of a result by Arora, Barak and Steurer

Arora et al. [7] proves a structural showing that we can obtain small sparse cuts efficiently if there are many small eigenvalues (Theorem 2.7). The result is equivalent as the statement

$$\phi_{O(k^{-1/100})} = O(\sqrt{\lambda_k \log_k n}).$$

As observed by Oveis Gharan and Trevisan, our results can actually give a simpler yet stronger proof of Theorem 2.7.

Theorem 4.5 ([55]). *For any undirected graph G and $\epsilon > 0$, we can efficiently find a set S with size $O(n/k^{1-\epsilon})$ and $\phi(S) = O(\sqrt{\lambda_k \log_k n/\epsilon})$.*

Proof. We will use Theorem 4.3 to upper bound $\chi_u^T W^t \chi_u$ and use the eigenvalues to lower bound the trace of W^t, which is equal to $\sum_u \chi_u^T W^t \chi_u$. Hence we obtain the bound relating the small set conductance and the eigenvalues.

Applying Theorem 4.3 with $l = 2\,\mathrm{vol}(V)/k^{1-\epsilon}$ and $t = 16\log n/\phi_1^2$, we have

$$\chi_u^T W^t \chi_u \le C(W^t\chi_u, \deg(u)) \le \frac{\deg(u)}{l} + \sqrt{\deg(u)}(1-\frac{\phi_1^2}{8})^t \le \frac{\deg(u)}{l} + n^{-1.5}.$$

Hence $\mathrm{tr}(W^t) = \sum_u \chi_u^T W^t \chi_u \le \mathrm{vol}(V)/l + 1/\sqrt{n} \le k^{1-\epsilon}$.

On the other hand, $\mathrm{tr}(W^t) = \sum_i (1-\lambda_i/2)^t \ge k(1-\lambda_k/2)^t$. Therefore

$$k^{1-\epsilon} \ge k(1-\frac{\lambda_k}{2})^t \ge k\exp(-\frac{16\log n\lambda_k}{\phi_1^2}).$$

Hence $-\epsilon\log k \ge -16\log n\lambda_k/\phi_1^2$, and hence $\phi_1 = O(\sqrt{\lambda_k \log_k n/\epsilon})$. This completes the proof. □

4.2 Expansions of graph powers

4.2.1 Introduction

A well-known operation to improve the graph expansion for regular graphs is by taking the t-th power of G, which has a natural correspondence to simulating the random walk on G for t steps. In this section, we assume that G is 1-regular, that is, $\deg(u) = \sum_{v\in V} w(u,v) = 1$ for every $u \in V$. We also assume that G is lazy, that is, $w(u,u) \ge 1/2$ for every $u \in V$. Since G is 1-regular and lazy, we use column vectors for probability distributions in this section and hence $A^t p$ (instead of pW^t) denotes the probability distribution after t steps of lazy random walks. The t-th power of G, denoted by G^t, is defined as the undirected graph with adjacency matrix $A(G)^t$, which corresponds to the transition matrix of the t-step random walk of G. Note that G^t is also 1-regular if G is.

The question we study is to prove lower bounds on $\phi(G^t)$ in terms of $\phi(G)$. Besides being a basic graph theoretical question, proving lower bounds on $\phi(G^t)$ has applications in hardness of approximation [25, 57]. Our main result is a tight lower bound on the expansion of the graph powers of a lazy 1-regular graph.

Previous work

There is a spectral argument to show that $\phi(G^t)$ is larger than $\phi(G)$ for large enough t. Let $1 = \alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n \ge 0$ be the eigenvalues of the adjacency matrix A_G of

G, where $\alpha_1 = 1$ because G is 1-regular and $\alpha_n \geq 0$ because G is lazy. Note that the eigenvalues of A^t is $1 = \alpha_1^t \geq \alpha_2^t \geq \ldots \geq \alpha_n^t \geq 0$, and thus the i-th eigenvalue of the Laplacian matrix of G^t is $1 - \alpha_i^t = 1 - (1 - \lambda_i)^t$. Therefore, by the Cheeger's inequality, we have

$$\phi(G^t) \geq \frac{1}{2}(1 - (1 - \lambda_2)^t) \geq \frac{1}{2}(1 - (1 - \frac{1}{2}t\lambda_2)) = \frac{1}{4}t\lambda_2 \geq \frac{1}{8}t \cdot \phi(G)^2 = \Omega(t \cdot \phi(G)^2),$$

where the second inequality follows from Fact 4.1 when $t\lambda_2 < 1/2$.

Recently, the spectral method was extended to prove lower bounds on the small set expansion of a graph. Raghavendra and Schramm [57] proved an analog of the above bound for small set expansion:

$$\phi_{\Omega(\delta)}(G^t) = \Omega(t \cdot \phi_\delta(G)^2),$$

when G is a lazy 1-regular graph and $t = O(1/\phi_\delta(G)^2)$. The proof is based on the techniques developed in [7] relating higher eigenvalues to small set expansion. They used this lower bound to amplify the hardness of the small set expansion problem; see Section 4.2.3 for more discussions.

Our results

Our main result is a tight lower bound on the expansion of the graph powers of a lazy 1-regular graphs.

Theorem 4.6. *Let G be an undirected 1-regular lazy graph. For any non-negative integer t, we have*

$$\phi(G^t) \geq \frac{1}{20}(1 - (1 - \phi(G))^{\sqrt{t}}) = \Omega(\min(\sqrt{t} \cdot \phi(G), 1)).$$

This is a quadratic improvement of the previous bound. This bound is tight up to a constant factor for all t as we will show examples (e.g. cycles) in Section 4.2.2.

Observe that the spectral argument only shows that $\phi(G^t) > \phi(G)$ when $t = \Omega(1/\phi(G))$ but does not show that $\phi(G^t) > \phi(G)$ for small t. Theorem 4.6 implies that $\phi(G^t) > \phi(G)$ for some small constant t. Actually, we can show that $\phi(G^3) > \phi(G)$ when $\phi(G) < 1/2$ by a more explicit calculation.

Theorem 4.7. *Let G be an undirected 1-regular lazy graph with even n. We have*

$$\phi(G^3) \geq \frac{3}{2}\phi(G) - 2\phi(G)^3.$$

One advantage of our approach is that Theorem 4.6 can be extended easily to small set expansion.

Theorem 4.8. *Let G be an undirected 1-regular lazy graph. For any non-negative integer t, we have*

$$\phi_{\delta/2}(G^t) \geq \frac{1}{20}(1-(1-2\phi_\delta(G))^{\sqrt{t}}) = \Omega(\min(\sqrt{t}\cdot\phi_\delta(G), 1)).$$

We show some applications of our results in Subsection 4.2.3, including the gap amplification result in [57] for small set expansion and some reductions for proving Cheeger-type inequalities [4, 43].

Techniques

We use the Lovász-Simonovits curve (see Subsection 2.3.2) for analyzing random walks. As it turns out, this more combinatorial approach has the advantage of directly reason about graph expansions without having the quadratic loss in the spectral method.

For an initial probability distribution p on the vertex set, $C^{(t)}(x)$ is the sum of the probability of the x largest vertices after t steps of random walk on G. First, we observe in Lemma 4.6 that when the initial distribution is $\chi_S/|S|$,

$$\phi_{G^t}(S) \geq 1 - C^{(t)}(|S|).$$

Hence, to lower bound $\phi_{G^t}(S)$, we can instead upper bound $C^{(t)}(|S|)$. Imprecisely, by Lemma 2.4, we can essentially argue that for all S with $|S| \leq n/2$,

$$\begin{aligned} C^{(t)}(|S|) &\lesssim \frac{1}{2^t}\sum_{i=0}^{t}\binom{t}{i} C^{(0)}\left((1-\phi(G))^i(1+\phi(G))^{t-i}|S|\right) \\ &= \frac{1}{2^t}\sum_{i=0}^{t}\binom{t}{i}\min\{(1-\phi(G))^i(1+\phi(G))^{t-i}, 1\}, \end{aligned}$$

where the equality holds because $C^{(0)}(x) = \min\{x/|S|, 1\}$ as the initial distribution is $\chi_S/|S|$. Since there is at least a $1/10$ fraction of terms in the summation with $i \geq t/2 + \sqrt{t}$, we have

$$C^{(t)}(|S|) \lesssim \frac{1}{10}(1-\phi(G))^{\sqrt{t}} + \frac{9}{10} \leq \frac{1}{10}(1-\frac{1}{2}\sqrt{t}\cdot\phi(G)) + \frac{9}{10},$$

where the last inequality is by Fact 4.1 when $\sqrt{t}\cdot\phi(G) \leq 1/2$. Therefore, for all S with

$|S| \le n/2$, we have

$$\phi_{G^t}(S) \ge \frac{1}{20}\sqrt{t} \cdot \phi(G), \text{ and therefore } \phi(G^t) = \Omega(\sqrt{t} \cdot \phi(G)).$$

We need to be careful to make the arguments in $\lesssim$ precise and this is some technicality of the proof, but the main ideas are pretty accurately summarized in this section.

4.2.2 Expansion of graph power

The following fact is used frequently in the proof.

Fact 4.1. For any $z \in [0,1]$, we have

$$(1-z)^t \ge 1 - zt, \quad \text{or} \quad 1-(1-z)^t \le zt.$$

For any $zt \in [0, 1/2]$, we have

$$(1-z)^t \le \exp(-zt) \le 1 - \frac{1}{2}zt, \quad \text{or} \quad 1-(1-z)^t \ge \frac{1}{2}zt.$$

Lovász-Simonovits curve

We do our analysis with the function introduced by Lovász and Simonovits as presented in Subsection 2.3.2, except for a subtle difference. In this result, we assume the graph to be lazy and 1-regular. Hence it is more natural to define $C^{(t)}(x)$ to be $C(A^t p, x)$ instead of $C(pW^t, x)$, since the graph is already lazy.

We use $\overline{x}$ to denote $\min(x, n-x)$ for $x \in [0, n]$. This notation is frequently used and should be interpreted as the distance to the boundary. With the new definition of $C^{(t)}(x)$, we change Lemma 2.4 accordingly and obtain the following:

Lemma 4.4. *If G is a lazy 1-regular graph, then for any integer $t \ge 0$ and any integer $x \in [0, n]$, we have*

$$C^{(t+1)}(x) \le \frac{1}{2}\left(C^{(t)}(x - 2\phi\overline{x}) + C^{(t)}(x + 2\phi\overline{x})\right).$$

Comparing to the original lemma Lemma 2.4, the difference is here we have 2ϕ instead of ϕ. This is due to the conductance of a lazy graph is exactly half the conductance of the original graph. We remark that both lemmas only give bounds on integral values[1]. In our proof, however, we require bounds for all $x \in [0, n]$. The following lemma

[1]It was claimed in [51] that the lemma holds for any $x \in [0, n]$, but later it was pointed out in [67] that the lemma only holds for extreme points x, which is integral when the graph is lazy 1-regular.

provides a slightly weaker bound that also holds for fractional x when the graph is lazy 1-regular.

Lemma 4.5. *If G is a lazy 1-regular graph, then for any integer $t \geq 0$ and $x \in [0, n]$, we have*

$$C^{(t+1)}(x) \leq \frac{1}{2}\left(C^{(t)}(x - \phi\overline{x}) + C^{(t)}(x + \phi\overline{x})\right).$$

Proof. Since $C^{(t)}$ is concave, for $\beta > \gamma$ we have

$$C^{(t)}(x - \beta\overline{x}) + C^{(t)}(x + \beta\overline{x}) \leq C^{(t)}(x - \gamma\overline{x}) + C^{(t)}(x + \gamma\overline{x}), \tag{4.1}$$

We will prove that

$$C^{(t+1)}(x) \leq \frac{1}{2}\left(C^{(t)}(x - 2\phi'\overline{x}) + C^{(t)}(x + 2\phi'\overline{x})\right) \tag{4.2}$$

where

$$\phi' = \frac{n-1}{n}\phi,$$

and this would imply the lemma by Equation 4.1 since $\phi' \geq \phi/2$.

Note that for any integral $x \in [0, n-1]$ and any $\alpha \in [0, 1]$,

$$\begin{aligned}
C^{(t+1)}(x+\alpha) &= (1-\alpha)C^{(t+1)}(x) + \alpha C^{(t+1)}(x+1) \\
&\leq (1-\alpha)\left(C^{(t)}(x - 2\phi\overline{x}) + C^{(t)}(x + 2\phi\overline{x})\right) \\
&\qquad + \alpha\left(C^{(t)}(x+1-2\phi\overline{(x+1)}) + C^{(t)}(x+1+2\phi\overline{(x+1)})\right) \\
&= \left((1-\alpha)C^{(t)}(x - 2\phi\overline{x}) + \alpha C^{(t)}(x+1-2\phi\overline{(x+1)})\right) \\
&\qquad + \left((1-\alpha)C^{(t)}(x + 2\phi\overline{x}) + \alpha C^{(t)}(x+1+2\phi\overline{(x+1)})\right) \\
&\leq C^{(t)}(x+\alpha-2\phi((1-\alpha)\overline{x} + \alpha\overline{(x+1)})) \\
&\qquad + C^{(t)}(x+\alpha+2\phi((1-\alpha)\overline{x} + \alpha\overline{(x+1)})),
\end{aligned}$$

where the first inequality follows from Lemma 4.4, and last inequality holds because $C^{(t)}$ is concave. If $(1-\alpha)\overline{x} + \alpha\overline{(x+1)} = \overline{(x+\alpha)}$, then Lemma 4.4 holds and the lemma follows by Equation 4.1.

Note that the only case where $(1-\alpha)\overline{x} + \alpha\overline{(x+1)} \neq \overline{(x+\alpha)}$ is when n is odd and $x = (n-1)/2$. At that time, $\overline{x} = \overline{(x+1)} = x$ and thus $(1-\alpha)\overline{x} + \alpha\overline{(x+1)} = x$.

Therefore, when n is odd and $x = (n-1)/2$, we have

$$\begin{aligned}
& C^{(t+1)}(x+\alpha) \\
\leq\ & \frac{1}{2}\left(C^{(t)}(x+\alpha-2\phi x) + C^{(t)}(x+\alpha+2\phi x)\right) \\
\leq\ & \frac{1}{2}\left(C^{(t)}\left(x+\alpha-2(\frac{n-1}{n})\cdot\phi\cdot\overline{(x+\alpha)}\right) + C^{(t)}\left(x+\alpha+2(\frac{n-1}{n})\cdot\phi\cdot\overline{(x+\alpha)}\right)\right) \\
=\ & \frac{1}{2}\left(C^{(t)}\left(x+\alpha-2\phi'\cdot\overline{(x+\alpha)}\right) + C^{(t)}\left(x+\alpha+2\phi'\cdot\overline{(x+\alpha)}\right)\right),
\end{aligned}$$

where the later inequality holds because $C^{(t)}$ is concave and $\overline{x+\alpha} \leq x+1/2 = n/2$. □

Proof of the main theorem

As mentioned in the proof outline in Section 4.2.1, we first show that we can prove a lower bound on $\phi(G^t)$ by proving an upper bound on $C^{(t)}(|S|)$ for the initial distribution $\chi_S/|S|$.

Lemma 4.6. *Suppose that for any set $S \subseteq V$ with $|S| \leq n/2$, we have $C^{(t)}(|S|) \leq 1-\alpha$ for the initial distribution $p = \chi_S/|S|$, then we can conclude that $\phi(G^t) \geq \alpha$.*

Proof. Let S be the set attaining minimum expansion in G^t, that is, $|S| \leq n/2$ and $\phi_{G^t}(S) = \phi(G^t)$. For the initial distribution $p = \chi_S/|S|$,

$$C^{(t)}(|S|) = C(A^t p, |S|) \geq \chi_S^T A^t p = \frac{\chi_S^T A^t \chi_S}{|S|} = 1 - \frac{\chi_S^T (I_{n\times n} - A^t)\chi_S}{|S|} = 1 - \phi_{G^t}(S).$$

Therefore, we have $\phi(G^t) = \phi_{G^t}(S) \geq 1 - C^{(t)}(|S|) \geq \alpha$. □

With Lemma 4.6, it remains to upper bound $C^{(t)}(|S|)$ for the initial distribution $\chi_S/|S|$ for any S with $|S| \leq n/2$. It turns out that there is a good upper bound independent of $|S|$.

Lemma 4.7. *For any S with $|S| \leq n/2$, for the initial distribution $p = \chi_S/|S|$, for any non-negative integer t, we have*

$$C^{(t)}(|S|) \leq 1 - \frac{1}{20}(1-(1-\phi)^{\sqrt{t}}).$$

Proof. For technical reasons, we consider $D^{(t)}(x) = C^{(t)}(x) - x/n$ instead to make the argument more symmetric. See Figure 4.1 for the definition of $D^{(0)}$. Note that

Lemma 4.5 still holds for $D^{(t)}$ since x/n is linear. So, we have

$$D^{(t+1)}(x) \le \frac{1}{2}\left(D^{(t)}(x-\phi\overline{x}) + D^{(t)}(x+\phi\overline{x})\right).$$

By applying this equation repeatedly, we have

$$D^{(t)}(x) \le \frac{1}{2^t}\sum_{T\in\{-1,1\}^t} D^{(0)}(f_T(x)), \tag{4.3}$$

where T is a sequence of t ±1-bits and f_T is defined recursively as follows. In the base case, when the sequence is empty, we define $f_{()}(x) = x$ for any $x \in [0,n]$. For any partial sequence T', we define

$$f_{(T',+1)}(x) = \begin{cases} f_{T'}(x) - \phi\cdot f_{T'}(x) & \text{if } f_{T'}(x) \le n/2 \\ f_{T'}(x) + \phi\cdot \overline{f_{T'}(x)} & \text{if } f_{T'}(x) > n/2, \end{cases}$$

and

$$f_{(T',-1)}(x) = \begin{cases} f_{T'}(x) + \phi\cdot f_{T'}(x) & \text{if } f_{T'}(x) \le n/2 \\ f_{T'}(x) - \phi\cdot \overline{f_{T'}(x)} & \text{if } f_{T'}(x) > n/2, \end{cases}$$

We can view $+1$ as moving in the direction towards boundary and -1 as moving in the direction towards center. Recall that $\overline{x} = \min\{x, n-x\}$ can be viewed as the distance to the boundary. In the following, we focus on the distance to the boundary of a point rather than its actual location. It follows from the definition that for any $x \in [0,n]$, we have

$$\overline{f_{+1}(x)} = \overline{x} - \phi\overline{x} = (1-\phi)\cdot\overline{x},$$

and

$$\overline{f_{-1}(x)} \le \overline{x} + \phi\overline{x} = (1+\phi)\overline{x} \le (1-\phi)^{-1}\cdot\overline{x}.$$

Therefore, $\overline{f_{T_i}(x)} \le (1-\phi)^{T_i}\cdot\overline{x}$ where T_i is the i-th bit in the sequence T, and hence

$$\overline{f_T(x)} = \overline{f_{T_t}\circ f_{T_{t-1}}\circ\cdots\circ f_{T_1}(x)} \le (1-\phi)^{T_t}\cdot\overline{f_{T_{t-1}}\circ\cdots\circ f_{T_1}(x)} \le \cdots \le (1-\phi)^{\sum_{i=1}^t T_i}\cdot\overline{x}.$$

We call a sequence T good if $\sum_{i=1}^t T_i \ge \sqrt{t}$, otherwise we call it bad. For a good T, we have $\overline{f_T(x)} \le (1-\phi)^{\sqrt{t}}\cdot\overline{x}$, and thus

$$\overline{f_T(|S|)} \le (1-\phi)^{\sqrt{t}}\cdot|S| \quad \text{for } |S| \le n/2 \text{ and } T \text{ good.} \tag{4.4}$$

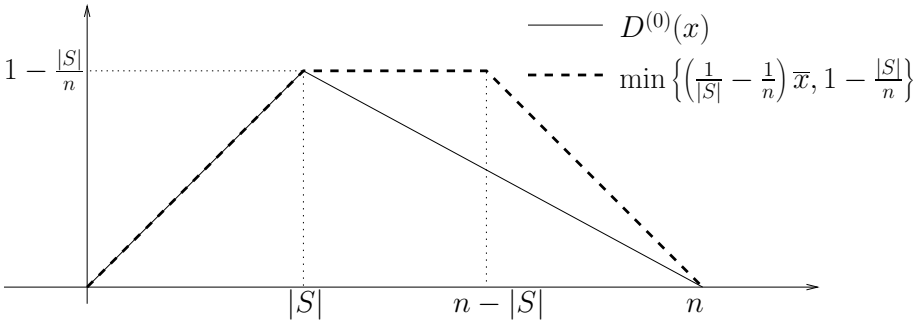


Figure 4.1: The solid line is the curve $D^0(x)$ and the dotted line is the upper bound on $D^{(0)}(x)$ that is stated in Equation 4.5.

As the initial distribution is $\chi_S/|S|$, for $t = 0$, we have

$$D^{(0)}(x) \leq \min\left\{\left(\frac{1}{|S|}-\frac{1}{n}\right)\overline{x}, 1-\frac{|S|}{n}\right\}. \tag{4.5}$$

See Figure 4.1 for an illustration of the inequality. The advantage of using $D^{(t)}$ instead of $C^{(t)}$ is that we could bound $D^{(0)}(x)$ using $\overline{x}$ as shown in the above inequality.

Finally, we know that at least a $1/10$ fraction of T are good. So, for S with $|S| \leq n/2$,

$$\begin{aligned}
D^{(t)}(|S|) &\leq \frac{1}{2^t}\sum_{T\in\{-1,1\}^t} D^{(0)}(f_T(|S|)) && \text{(by } \textit{Equation 4.3}\text{)}\\
&= \frac{1}{2^t}\sum_{T:\text{good}} D^{(0)}(f_T(|S|)) + \frac{1}{2^t}\sum_{T:\text{bad}} D^{(0)}(f_T(|S|))\\
&\leq \frac{1}{2^t}\sum_{T:\text{good}}\left(\frac{1}{|S|}-\frac{1}{n}\right)\overline{f_T(|S|)} + \frac{1}{2^t}\sum_{T:\text{bad}}\left(1-\frac{|S|}{n}\right) && \text{(by } \textit{Equation 4.5}\text{)}\\
&\leq \frac{1}{2^t}\sum_{T:\text{good}}\left(\frac{1}{|S|}-\frac{1}{n}\right)(1-\phi)^{\sqrt{t}}|S| + \frac{1}{2^t}\sum_{T:\text{bad}}\left(1-\frac{|S|}{n}\right) && \text{(by } \textit{Equation 4.4}\text{)}\\
&\leq \frac{1}{10}\left(\left(\frac{1}{|S|}-\frac{1}{n}\right)(1-\phi)^{\sqrt{t}}|S|\right) + \frac{9}{10}\left(1-\frac{|S|}{n}\right)\\
&= \left(1-\frac{|S|}{n}\right) - \frac{1}{10}\left(1-\frac{|S|}{n}-\left(\frac{1}{|S|}-\frac{1}{n}\right)(1-\phi)^{\sqrt{t}}|S|\right)\\
&= \left(1-\frac{|S|}{n}\right) - \frac{1}{10}\left(1-\frac{|S|}{n}\right)\left(1-(1-\phi)^{\sqrt{t}}\right)\\
&\leq \left(1-\frac{|S|}{n}\right) - \frac{1}{20}(1-(1-\phi)^{\sqrt{t}}).
\end{aligned}$$

Therefore,

$$C^{(t)}(|S|) = D^{(t)}(|S|) + |S|/n \leq 1 - \frac{1}{20}(1 - (1-\phi)^{\sqrt{t}}).$$

□

Combining Lemma 4.6 and Lemma 4.7, we have

$$\phi(G^t) \geq \frac{1}{20}(1 - (1-\phi)^{\sqrt{t}}) \geq \frac{1}{40}\sqrt{t} \cdot \phi,$$

where the last inequality is by Fact 4.1 for $\sqrt{t} \cdot \phi \leq 1/2$. This completes the proof of Theorem 4.6.

Proof of graph cube

Theorem 4.6 showed that $\phi(G^t) > \phi(G)$ for a small constant t. To prove that this is true even for $t = 3$, we need to do a more explicit calculation. We use the bound

$$C^{(t+1)}(x) \leq \frac{1}{2}\left(C^{(t)}(x - 2\phi'\overline{x}) + C^{(t)}(x + 2\phi'\overline{x})\right)$$

for $\phi' = \frac{n-1}{n}\phi$ as was shown in Equation 4.2 in the proof of Lemma 4.5. When $t = 3$, we have

$$\begin{aligned} C^{(3)}(|S|) &\leq \frac{1}{8}C^{(0)}((1-2\phi')^3|S|) + \frac{3}{8}C^{(0)}((1-2\phi')^2(1+2\phi')|S|) + \frac{4}{8} \\ &= \frac{1}{8}(1-2\phi')^3 + \frac{3}{8}(1-2\phi')^2(1+2\phi') + \frac{4}{8} \\ &= 1 - \frac{3}{2}\phi' + 2\phi'^3. \end{aligned}$$

Thus we conclude $\phi(G^3) \geq \frac{3}{2}\phi' - 2\phi'^3$. Therefore, for a large graph with small conductance, taking cube increases the conductance by a factor of almost $\frac{3}{2}$. When n is even, we can replace ϕ' by ϕ as was shown in the proof of Lemma 4.5, and this proves Theorem 4.7.

Proof of small set conductance

Our result can be easily extended to the case of small set expansion with a little loss in size. More precisely, suppose G is an undirected 1-regular lazy graph such that all sets of size at most δn have conductance ϕ_δ, where $\delta \leq 1/2$. In this setting, the following lemma holds in place of Lemma 4.4.

Lemma 4.8. *If G is a lazy 1-regular graph, then for any integer $t \geq 0$ and any $x \in [0, \delta n]$,*

$$C^{(t+1)}(x) \leq \frac{1}{2}\left(C^{(t)}(x - 2\phi_\delta \cdot \overline{x}) + C^{(t)}(x + 2\phi_\delta \cdot \overline{x})\right),$$

where $\overline{x} = \min(x, \delta n - x)$ here.

We remark that we do not need to fix the non-integral problem as in Lemma 4.5 because we only consider $x \leq \delta n \leq n/2$ (see the proof of Lemma 4.5).

Lemma 4.6 can be restated as follows with the same proof.

Lemma 4.9. *Suppose that for any set $S \subseteq V$ with $|S| \leq \delta n/2$ with the initial distribution $p = \chi_S/|S|$, we have $C^{(t)}(|S|) \leq 1 - \alpha$, then we can conclude that $\phi_{\delta/2}(G^t) \geq \alpha$.*

Finally, in Lemma 4.7, we consider $D^{(t)}(x) = C^{(t)}(x) - \frac{x}{\delta n}$ instead, and we use the new $\overline{x}$ in the analysis. Observe that $f_T(x)$ can never leave the range $[0, \delta n]$ when x starts in the range. Therefore the same analysis applies and we have the following lemma.

Lemma 4.10. *For any S with $|S| \leq \delta n/2$, for the initial distribution $p = \chi_S/|S|$, for any non-negative integer t, we have*

$$C^{(t)}(|S|) \leq 1 - \frac{1}{20}(1 - (1 - 2\phi_\delta)^{\sqrt{t}}).$$

Theorem 4.8 follows by combining Lemma 4.9 and Lemma 4.10.

Tight examples

We show that the dependence on t in Theorem 4.6 is tight up to a constant factor. The tight example we use is a lazy cycle. Intuitively, after t steps of random walk on a lazy cycle, the final position with high probability only differs from the initial position by $O(\sqrt{t})$, and therefore the expansion should be bounded by $O(\sqrt{t})$ times the original expansion. It turns out that we can easily justify this intuition through Cheeger's inequality.

Proposition 4.1. *Let C_n be the lazy cycle. Then we have $\phi(C_n^t) = O(\sqrt{t} \cdot \phi(C_n))$.*

Proof. As in Section 4.2.1, we have

$$\lambda_2(C_n^t) = 1 - (1 - \lambda_2(C_n))^t \leq t\lambda_2(C_n) = O(t \cdot \phi(C_n)^2),$$

where the inequality is by Fact 4.1 and the last equality is by the spectrum of the cycle. By Cheeger's inequality (Theorem 2.1), $\phi(C_n^t) = O(\sqrt{\lambda_2(C_n^t)})$, and thus $\phi(C_n^t) = O(\sqrt{t} \cdot \phi(C_n))$. □

We remark that tight examples of Theorem 4.6 must have many small eigenvalues. By the improved Cheeger's inequality (Theorem 3.1), we have $\phi(G) = O(k\lambda_2/\sqrt{\lambda_k})$ for any k. Therefore, by the same calculation as in Section 4.2.1, we have that for any k,

$$\phi(G^t) \geq \frac{1}{4}t\lambda_2 = \Omega(\frac{t \cdot \phi(G) \cdot \sqrt{\lambda_k}}{k}),$$

and therefore a graph G with $\lambda_k(G)$ large for a small k could not be a tight example for Theorem 4.6.

Irregular graphs

Theorem 4.6 showed that $\phi(G^t) = \Omega(\sqrt{t} \cdot \phi(G))$ for a regular graph. There are different ways to generalize the statement to irregular graphs. In the following, we show that the generalization is not true if we replace expansion by conductance, and we show that the generalization is true if we replace expansion by the escape probability of a t-step random walk.

Consider the graph G consisting of a regular complete graph with self loops $(2I_{n\times n} + \frac{1}{n}K_n)$ and an extra vertex u. The extra vertex only connects to a single vertex v in the complete graph with edge weight 1 and it has a self loop of weight m. We assume the complete graph is so large that $n > 2m^4$. Then $\phi(G) = \phi(\{u\}) = 1/m + o(1/m)$. Consider G^3. Since $\deg_{G^3}(u) = m^3 + o(m^3) < n/2$, the set achieving minimum conductance is still $\{u\}$. In G^3, the total weight of edges between u and the complete graph is $m^2 + o(m^2)$. Therefore $\phi(G^3) = 1/m + o(1/m)$. Note that the same argument applies for any G^t if we set n to be large enough. Therefore, no matter how small $\phi(G)$ is or how large t is, we cannot argue that $\phi(G^t) > (1+\epsilon)\phi(G)$ for a positive constant ϵ when we replace expansion by conductance in irregular graphs.

On the other hand, our results can be extend to another natural generalization of expansion. Consider the definition

$$\varphi(G^t) = \min_{S\subseteq V, |S|\leq n/2} \varphi_{G^t}(S) = \min_{S\subseteq V, |S|\leq n/2} (1 - \frac{\chi_S^T (D^{-1}A_G)^t \chi_S}{|S|}),$$

where $\varphi_{G^t}(S)$ is the probability that a t-step random walk starting from a random vertex in S escapes S. With this definition and assuming that the graph does not contain a vertex of degree more than half of the total degrees, we can show that Lemma 4.5 still holds, with a extended definition for $C^{(t)}$. Therefore, $\varphi(G^t) = \Omega(\min\{\sqrt{t} \cdot \varphi(G), 1\})$ follows.

4.2.3 Applications

In this section, we discuss some consequences of our main theorem. We show that proving the general cases of Cheeger's inequalities can be reduced to proving the special cases where the eigenvalues are constants. Similar arguments can be used to deduce the recent result on gap amplification for small set expansion in [57].

Cheeger's inequalities

Let G be an undirected 1-regular lazy graph. The following result shows that if one could prove Cheeger's inequality when λ_2 is a constant, then one could prove Cheeger's inequality for all λ_2. One consequence is that if one could prove that say $\phi(G) = O((\lambda_2)^{1/100})$ (so that Cheeger's inequality is true when λ_2 is a constant), then it actually implies that $\phi(G) = O(\sqrt{\lambda_2})$.

Corollary 4.1. *Suppose one could prove that $\lambda_2(H) \geq C$ for some constant $C \leq 1/2$ whenever $\phi(H) \geq 1/40$, then it implies that $\phi(G) \leq \sqrt{\lambda_2(G)/C}$ for any G and any $\lambda_2(G)$.*

Proof. Given G, we assume that $\lambda_2(G) \leq \phi(G)^2/2$, as otherwise the statement is trivial. Consider $H = G^{1/\phi(G)^2}$. By Theorem 4.6, we have

$$\phi(H) \geq \frac{1}{20}(1 - (1 - \phi(G))^{\sqrt{1/\phi(G)^2}}) \geq \frac{1}{40}.$$

Therefore, if we could prove that $\lambda_2(H) \geq C$, then we could conclude that

$$C \leq \lambda_2(H) = 1 - (1 - \lambda_2(G))^{1/\phi(G)^2} \leq \frac{\lambda_2(G)}{\phi(G)^2},$$

where the last inequality is by Fact 4.1. Hence the corollary follows. □

Improved Cheeger's inequality

The improved Cheeger's inequality (Theorem 3.1) states that $\phi(G) = O(k\lambda_2/\sqrt{\lambda_k})$ for any k. Using similar arguments as above, the following result shows that if one could prove this improved Cheeger's inequality when λ_3 is a constant, then one could prove it for all λ_3. For instance, if one could prove that say $\phi(G) = O(\lambda_2/\lambda_3^{100})$, then it actually implies that $\phi(G) = O(\lambda_2/\sqrt{\lambda_3})$.

Corollary 4.2. *Suppose one could prove that $\phi(H) \leq C\lambda_2(H)$ for some $C \geq 1/10$ whenever $\lambda_3(H) \geq 1/2$, then it implies that $\phi(G) \leq 40C\lambda_2(G)/\sqrt{\lambda_3(G)}$ for any G and*

any $\lambda_3(G)$.

Proof. We assume that $\phi \leq \sqrt{\lambda_3}/2$, as otherwise, by Cheeger's inequality, $2\lambda_2(G) \geq \phi(G)^2 \geq \frac{1}{2}\phi(G)\sqrt{\lambda_3}$ and the statement is true. Consider $H = G^{1/\lambda_3(G)}$. Then

$$\lambda_3(H) = 1 - (1 - \lambda_3(G))^{1/\lambda_3} \geq 1 - e^{-1} \geq 1/2.$$

Therefore, if one could prove that $\phi(H) \leq C\lambda_2(H)$, then

$$C\lambda_2(H) \geq \phi(H) \geq \frac{1}{20}(1 - (1 - \phi(G))^{\sqrt{1/\lambda_3(G)}}) \geq \frac{\phi(G)}{40\sqrt{\lambda_3(G)}},$$

where the second inequality is by Theorem 4.6 and the last inequality is by Fact 4.1. On the other hand,

$$\lambda_2(H) = 1 - (1 - \lambda_2(G))^{1/\lambda_3(G)} \leq \frac{\lambda_2(G)}{\lambda_3(G)},$$

and the corollary follows by combining the two inequalities. □

Gap amplification for small set expansion

Consider the small set expansion problem $\mathrm{SSE}_{\delta,\delta'}(c, s)$: Given a graph G, distinguish whether $\phi_\delta(G) \leq c$ or $\phi_{\delta'}(G) \geq s$. The small set expansion conjecture [58] states that for any $\epsilon > 0$, there exists $\delta > 0$ such that $\mathrm{SSE}_{\delta,\delta}(\epsilon, 1 - \epsilon)$ is NP-hard.

Let f be a function such that $f(x) = \omega(\sqrt{x})$. Raghavendra and Schramm [57] showed that if for all $\epsilon > 0$ there exists $\delta > 0$ such that $\mathrm{SSE}_{\delta,\delta}(\epsilon, f(\epsilon))$ is NP-hard, then for all $\epsilon > 0$ there exists $\delta > 0$ such that $\mathrm{SSE}_{\delta,\delta/8}(\epsilon, 1/2)$ is NP-hard.

We would show that our techniques can be easily applied to get a similar result.

Theorem 4.9. *If for all $\epsilon > 0$ there exists $\delta > 0$ such that $\mathrm{SSE}_{\delta,\delta}(\epsilon, f(\epsilon))$ is NP-hard, then for all $\epsilon > 0$ there exists $\delta > 0$ such that $\mathrm{SSE}_{\delta,\delta/2}(\epsilon, \Omega(1))$ is NP-hard.*

Proof. Given an instance G that we would like to distinguish whether $\phi_\delta(G) \leq \epsilon$ or $\phi_\delta(G) \geq f(\epsilon)$, we consider the graph $H = G^{O(1/f(\epsilon)^2)}$. In the case when $\phi_\delta(G) \geq f(\epsilon)$, by Theorem 4.8, we have

$$\phi_{\delta/2}(H) = \Omega(\sqrt{1/f(\epsilon)^2} \cdot f(\epsilon)) = \Omega(1).$$

In the case when $\phi_\delta(G) \leq \epsilon$, we have

$$\phi_\delta(H) \leq (1/f(\epsilon)^2) \cdot \epsilon = o_\epsilon(1) \leq \epsilon',$$

where the equality holds because $f(\epsilon) = \omega(\sqrt{\epsilon})$ and the first inequality holds because

$$\phi_{G^t}(S) = 1 - \frac{\chi_S^T A^t \chi_S}{|S|} \leq t \cdot \phi_G(S),$$

where the inequality is proven (see Proposition 2.5 in [67]) by a simple induction. Therefore, if $\mathrm{SSE}_{\delta,\delta}(\epsilon, f(\epsilon))$ is NP-hard, then $\mathrm{SSE}_{\delta,\delta/2}(\epsilon', \Omega(1))$ is NP-hard. □

Finally, we remark that it is easier to bound $\phi_\delta(G^t)$ for large t using Lovász-Simonovits curve. Using the techniques in Section 4.1, we have the following bound for $C^{(t)}$ when the initial probability vector is $\chi_S/|S|$:

$$C^{(t)}(x) \leq \frac{x}{\delta n} + \sqrt{\frac{x}{|S|}}(1 - \frac{\phi^2}{2})^t.$$

Therefore,

$$\phi_{G^t}(S) \geq 1 - C^{(t)}(|S|) \geq 1 - \frac{|S|}{\delta n} - (1 - \frac{\phi^2}{2})^t,$$

where the first inequality follows from Lemma 4.6. Set $t = 100/\phi^2$, then for $|S| \leq \delta n/4$, we have $\phi_{G^t}(S) \geq \frac{3}{4} - \exp(-50)$. Therefore, if $\mathrm{SSE}_{\delta,\delta}(\epsilon, f(\epsilon))$ is NP-hard, then $\mathrm{SSE}_{\delta,\delta/4}(\epsilon', 1/2)$ is NP-hard. This recovers the result of Raghavendra and Schramm with better constant.

Chapter 5

Matrix rank

In the chapter, we present a fast algorithm to compute the matrix rank using probabilistic vertex expanders. The results presented in this chapter are mainly based on joint work with Ho Yee Cheung and Lap Chi Lau [17].

5.1 Introduction

Given an $m \times n$ matrix A over a field F, the rank of A, denoted by $\text{rank}(A)$, is the maximum number of linearly independent columns of A. We consider the problem of computing $\text{rank}(A)$ and finding a set of $\text{rank}(A)$ linearly independent columns efficiently. It is a basic computational problem in numerical linear algebra that is used as a subroutine for other problems [72, 76]. It also has a number of applications in graph algorithms and combinatorial optimization: Some of the fastest algorithms for graph matching [53, 32], graph connectivity [16, 61, 18], matroid optimization problems [32, 19] are based on fast algorithms for computing matrix rank and finding linearly independent columns.

5.1.1 Previous works

The traditional approach to compute $\text{rank}(A)$ is by Gaussian elimination. For an $m \times n$ matrix with $m \leq n$, it is known that this approach can be implemented in $O(nm^{\omega-1})$ field operations [11, 37], where $\omega < 2.373$ is the matrix multiplication exponent [23, 78, 44]. More generally, given an $m \times n$ matrix and a parameter $k \leq m \leq n$, one can compute $\min(\text{rank}(A), k)$ in $O(nmk^{\omega-2})$ field operations [70]. The time complexity can be improved somewhat for sparse matrices [79]. The Gaussian elimination approach has the advantage that it can also find a set of $\min(\text{rank}(A), k)$ linearly independent columns in the same time. These algorithms are deterministic.

There are also randomized algorithms to compute the value of $\text{rank}(A)$ more efficiently. There are at least three approaches.

1. The first approach is to do an efficient preconditioning [39, 15]. Let $B = T_1AT_2$ where T_1 and T_2 are Toeplitz matrices with entries chosen uniformly and randomly from a large enough subset of the field. Then B can be computed in $\tilde{O}(mn)$ time because of the structure of T_1 and T_2. Let $r = \text{rank}(A)$. It is proven that [39] the leading $r \times r$ minor of B is of full rank with high probability. Thus $\text{rank}(A)$ can be computed in $\tilde{O}(mn+r^\omega)$ field operations. There is another efficient preconditioner based on butterfly network [15] with similar property and running time. This approach works for any field.

2. There is a black-box approach that computes $\text{rank}(A)$ in $O(m \cdot \text{nnz}(A))$ field operations [77, 76, 63] where $\text{nnz}(A)$ is the number of non-zero entries of A. The method is based on computing the minimal polynomial of A for Krylov subspaces. It does not require to store A explicitly, as long as there is an oracle to compute Ab for any vector b. This approach is fast when the matrix is sparse, and it works for any field.

3. Another approach is based on random projection for matrices over real numbers. Given an $m \times n$ matrix A over $\mathbb{R}$, one can reduce A into an $m \times (m \log m)$ matrix A' so that $\text{rank}(A) = \text{rank}(A')$ with high probability [62] by the Johnson-Lindenstrauss lemma. The matrix A' can be computed efficiently using fast Johnson-Lindenstrauss transform [2, 3], and this implies an $\tilde{O}(nm + m^\omega)$ randomized algorithm to compute $\text{rank}(A)$. This approach is only known to work for matrices over real numbers.

We note that only the Gaussian elimination approach can also find a set of $\text{rank}(A)$ linearly independent columns, while other approaches can only compute the value of $\text{rank}(A)$.

5.1.2 Our results

Our main result is a faster randomized algorithm to compute matrix rank. We assume that there is at least one non-zero entry in each row and each column, and thus $\text{nnz}(A) \geq \max\{m, n\}$.

Theorem 5.1. *Given an $m \times n$ matrix A over a field F and a parameter k where $k \leq \min(m, n)$, there is a randomized algorithm to compute $\min(\text{rank}(A), k)$ in $O(\text{nnz}(A) +$*

$\min(k^\omega, k\,\mathrm{nnz}(A)))$ *field operations where* $\mathrm{nnz}(A)$ *denotes the number of non-zeros in* A*. Furthermore, there is a randomized algorithm to find a set of* $\min(\mathrm{rank}(A), k)$ *linearly independent columns in* $\tilde{O}(\mathrm{nnz}(A) + k^\omega)$ *field operations.*

For computing $\min(\mathrm{rank}(A), k)$, previous algorithms require $\tilde{O}(mn + k^\omega)$ field operations, while we replace the mn term by $\mathrm{nnz}(A)$ and remove the (small) polylog factor. Moreover, we can also find a set of $\min(\mathrm{rank}(A), k)$ linearly independent columns in about the same time, which is considerably faster than the $O(mnk^{\omega-2})$ algorithm by Gaussian elimination when k is small. For instances, we can find a set of $k = n^{1/\omega} \approx n^{0.42}$ linearly independent columns in $\tilde{O}(\mathrm{nnz}(A))$ field operations, and a set of $k = n^{1/(\omega-1)} \approx n^{0.72}$ linearly independent columns in $\tilde{O}(mn)$ field operations, while previously it was possible only for $k = O(\mathrm{polylog}(n))$. The algorithm for finding linearly independent columns is needed in applications on various problems in exact linear algebra and combinatorial optimization [17].

5.2 Fast matrix rank algorithms

In this section, we will prove Theorem 5.1. First, we state the setting in Subsection 5.2.1 and present an outline of our approach in Subsection 5.2.2. Then, we define magical graphs in Subsection 5.2.3, and use them to obtain the compression algorithm in Subsection 5.2.4. Finally, we present the algorithms to computing the matrix rank and finding a maximum set of independent columns in Subsection 5.2.5 and Subsection 5.2.6 respectively.

5.2.1 Setting

Let A be an $m \times n$ matrix over a field F. We will assume that A is given by a list of the value and the position of its non-zero entries, and each row and column of A contains at least one non-zero entry, so $\mathrm{nnz}(A) \geq \max(n, m)$.

When F is a finite field, we will assume that $|F| = \Omega(n^4)$ by the following lemma using an extension field.

Lemma 5.1. *Let* A *be an* $m \times n$ *matrix over a field* F *with* p^c *elements. We can construct a finite field* F' *with* $p^{ck} = \Omega(n^4)$ *elements and an injective mapping* $f : F \to F'$ *so that the image of* F *is a subfield of* F'*. Then the* $m \times n$ *matrix* A' *where* $A'(i, j) = f(A(i, j))$ *satisfies the property that* $\mathrm{rank}(A') = \mathrm{rank}(A)$*. This preprocessing step can be done in* $O(\mathrm{nnz}(A))$ *field operations. Each field operation in* F' *can be done in* $\tilde{O}(\log |F| + \log n)$ *steps.*

The proof is omitted here since it is not exacted related to this thesis. When F is an infinite field, we will assume the exact arithmetic model where each field operation can be done at unit cost. In the algorithms, we will need to choose a random element from F. When F is an infinite field, we just choose an arbitrary subset $S \subset F$ with $|S| = \Omega(n^4)$, and pick a uniformly random element from S. This will be enough for our applications of the Schwartz-Zippel lemma.

Lemma 5.2 (Schwartz-Zippel lemma [64])**.** *Let $P \in F[x_1, \ldots, x_n]$ be a non-zero polynomial of total degree d over a field F. Let S be a finite subset of F and let $r_1, \ldots, r_n$ be selected randomly from S. Then the probability that $P(r_1, \ldots, r_n) = 0$ is at most $d/|S|$.*

5.2.2 Outline

Suppose a parameter k is given and the task is to compute $\min(\text{rank}(A), k)$. Our approach is to compress the matrix into a $O(k) \times O(k)$ matrix whose rank is at least $\min(\text{rank}(A), k)$ with high probability. Our method is inspired by the random linear coding algorithm [33, 18] in network coding [1]. We can construct an $m \times k$ matrix B where each column of B is a random linear combination of the columns of A, i.e. $B_i = \sum_{j=1}^{n} r_{j,i} A_j$ where A_j and B_i denote the j-th column of A and the i-th column of B respectively, and $r_{j,i}$ is a random element in F. In other words, $B = AR$ where R is an $n \times k$ matrix where each entry is an independent random element in F. It can be shown that $\text{rank}(B) = \min(\text{rank}(A), k)$ with high probability using the Schwartz-Zippel lemma (see Lemma 5.4), but the problem is that it requires a rectangular matrix multiplication algorithm [36] to compute B and it is not efficient enough.

We observe that this way of constructing B is the same as doing the random linear coding algorithm in a single vertex with n incoming edges and k outgoing edges. And so the idea of using a superconcentrator to do the random linear coding efficiently [18] can be applied to construct an $m \times k$ matrix B in $O(mn)$ field operations, while $\text{rank}(B) = \min(\text{rank}(A), k)$ with high probability. We can apply the same procedure to reduce the matrix B into a $k \times k$ matrix C in $O(mk)$ field operations while $\text{rank}(C) = \text{rank}(B)$ with high probability, and then $\text{rank}(C)$ can be computed directly. The technical point here is that a superconcentrator is a sparse graph that has a strong connectivity property. The sparsity allows for fast computation. And the strong connectivity property ensures that any set of k linearly independent columns in A can be mapped to the k columns in B bijectively by some linear combinations, and random linear combinations ensure that $\text{rank}(B) = \min(\text{rank}(A), k)$ with high probability by the Schwartz-Zippel lemma. This implies that $\min(\text{rank}(A), k)$ can be computed in $O(mn + k^\omega)$ field operations

with high probability, improving the existing algorithms by removing the polylog factor. There are, however, two disadvantages of this method. One is that the compression algorithm requires $\Theta(mn)$ field operations even when A is a sparse matrix. Another is that we do not know how to find a set of $\min(\text{rank}(A), k)$ linearly independent columns of A using this method. See appendix of [17] for the full proof of computing rank by superconcentrators.

To improve the compression algorithm, we choose R to be a sparse $n \times l$ matrix (indeed $l = O(k)$ would be enough), with at most two non-zeros per row and about $2n/l$ non-zeros per column. Their locations are chosen at random, so that with high probability they form a "magical graph" (a sparse vertex expander used in the construction of a superconcentrator) when the matrix R is viewed as a bipartite graph with n vertices on one side and l vertices on the other side. The property of the magical graph ensures that with high probability any set of k linearly independent columns in A can be mapped to some set of k columns in B bijectively by some linear combinations. Again, the non-zero values are chosen randomly from the field, so that $\min(\text{rank}(B), k) = \min(\text{rank}(A), k)$ with high probability by the Schwartz-Zippel lemma. Since there are only two non-zeros per row of R, we can compute $B = AR$ easily in $O(\text{nnz}(A))$ time.

Furthermore, since there are about $2n/l$ non-zeros per column of R, from any set of at most k linearly independent columns in B, we can identify a subset of at most $2nk/l$ columns in A with the same number of linearly independent columns. By choosing $l \approx 11k$, we can (1) guarantee with high probability that R is a magical graph, (2) compute the rank of the compressed matrix in $O(k^\omega)$ field operations, and (3) remove a constant fraction of the columns of A while keeping $\min(\text{rank}(A), k)$ unchanged. Therefore, we can repeat this procedure for $O(\log n)$ times to reduce the number of columns in A to be $O(k)$, and the total running time is $O((\text{nnz}(A) + k^\omega) \log n)$ field operations.

5.2.3 Magical graphs

Our construction requires a probability distribution of bipartite graphs with the following properties.

Definition 5.1 (Magical Graphs)**.** A probability distribution of bipartite graphs with vertex set $X \cup Y$ is (k, ϵ)-magical if for any given subset $S \subseteq X$ with $|S| = k$, the probability that there is a matching in which every vertex in S is matched is at least $1 - \epsilon$.

We note that this definition only requires any particular subset S of size k can be matched to the other side with high probability, while the definition in [34] requires that

all subsets up to certain size can be matched to the other side. This is enough for us to show that for any particular set of k linearly independently columns in the original matrix, with high probability there exist some linear combinations that will map it to some set of k columns bijectively in the compressed matrix.

We show that a graph from a magical distribution with good parameters can be generated efficiently.

Lemma 5.3. *For any values of $|X| \geq |Y| \geq ck$ where $c \geq 11$, there is a $(k, O(1/k))$-magical distribution with the additional properties that each vertex of X is of degree 2 and each vertex of Y is of degree at most $2\lceil |X|/|Y| \rceil$. Moreover, there is a randomized $O(|X|)$ time algorithm to generate a graph from this distribution.*

We note that the magical graphs in [34] cannot be used directly because of the following reasons: (1) the failure probability in [34] is a constant while we need a much smaller failure probability in order to find a set of linearly independent columns, (2) we need the additional property that the graph is almost regular to find a set of linearly independent columns. The proof is by a standard probabilistic argument, which can be skipped in the first reading.

Proof. The generation algorithm is simple. We assume that $|X|$ is a multiple of $|Y|$; otherwise we construct a slightly larger graph and delete the extra vertices. We first construct a 2-regular graph G' with $|X|$ vertices on both sides, by taking the union of two random perfect matchings independently from $|X|$ vertices to $|X|$ vertices. Then we divide the $|X|$ vertices on one side into $|Y|$ groups where each group has $|X|/|Y|$ vertices. We obtain G by merging each group into a single vertex, and so each vertex in Y is of degree $2|X|/|Y|$.

For any $S \subseteq X$ with $|S| = k$, we analyze the probability that there is a matching in G in which every vertex in S is matched. By Hall's theorem, we need to show that for any $S' \subseteq S$, the neighbor set of S' in G is of size at least $|S'|$. To analyze the probability that the neighbor set of S' is at least $|S'|$ for a fixed $S' \subseteq S$, we consider the equivalent random process where the $2|S'|$ edges incident on S' are added one by one. Consider the i-th edge added. We say that it is a bad edge if the other endpoint falls in the same group with some previously added edges. If the neighbor set size of S' is less than $|S'|$, then there must be at least $|S'| + 1$ bad edges out of the $2|S'|$ edges, and the probability that an edge is bad is less than $|S'|/|Y|$. So the probability that the neighbor set size of

S' is less than $|S'|$ is less than

$$\binom{2|S'|}{|S'|+1} \times \left(\frac{|S'|}{|Y|}\right)^{|S'|+1}$$

by a union bound on the possible $|S'|+1$ bad edges. Summing over the choices of the size of S' and the choices of S' with that size, we have that the probability that there is a subset $S' \subseteq S$ with less than $|S'|$ neighbors is at most

$$\begin{aligned}\sum_{z=0}^{k} \binom{2z}{z+1} \left(\frac{z}{|Y|}\right)^{z+1} \binom{k}{z} &\leq \sum_{z=0}^{k} 2^{2z} \left(\frac{z}{|Y|}\right)^{z+1} \left(\frac{ke}{z}\right)^{z} \\ &\leq \sum_{z=0}^{k} \left(\frac{4e}{c}\right)^{z} \frac{z}{ck} \\ &= O(1/k),\end{aligned}$$

using $|Y| \geq ck$ and the identity $\sum_{z=0}^{\infty} r^z \cdot z = r/(1-r)^2$ for $r < 1$, and setting $r = 4e/c$ as $c \geq 11 > 4e$ by our assumption. Therefore, by Hall's theorem, the probability that there is a matching in which every vertex in S is matched is at least $1 - O(1/k)$. □

5.2.4 Compression algorithm by magical graph

In the following we use a graph from a magical distribution to do an efficient rank-preserving compression. The algorithm is shown in Algorithm 2 and illustrated in Figure 5.1.

Algorithm 2: Compression algorithm by magical graphs

Input: An $m \times n$ matrix A over a field F, and a bipartite graph $G = (X \cup Y, E)$ with $|X| = n$ and $|Y| = l$ sampled from a (k, ϵ)-magical distribution.

Output: An $m \times l$ matrix B over the field F with $\min(\text{rank}(B), k) = \min(\text{rank}(A), k)$.

1 Let $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_l\}$;

2 Each column of A corresponds to a vertex in X and each column of B corresponds to a vertex in Y. Let A_j be the j-th column of A for $1 \leq j \leq n$ and B_i be the i-th column of B for $1 \leq i \leq l$;

3 Construct B by writing B_i as a random linear combination of those columns of A whose corresponding vertices have an edge to y_i. More precisely, we write $B_i = \sum_{e=x_j y_i \in E} c_e A_j$ for $1 \leq i \leq l$ where c_e is an independent random element in F for each edge $e \in G$;

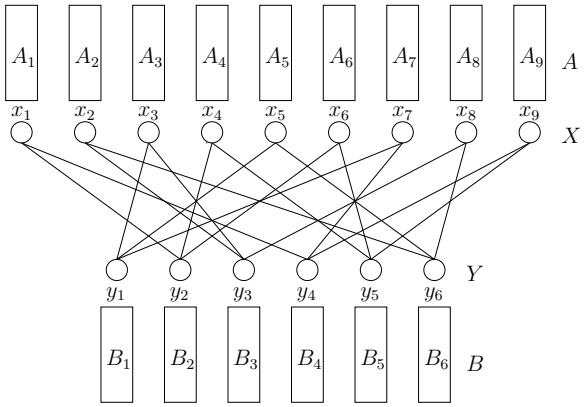


Figure 5.1: The notations used are the same as in Algorithm 2. The bipartite graph $G = (X \cup Y, E)$ is used to compress the matrix A into matrix B. Each column of B is a random linear combination of the columns of its neighbors, e.g. B_3 is a random linear combination of A_2, A_3 and A_8.

The following lemma uses the Schwartz-Zippel lemma to prove that the compression algorithm is rank-preserving with high probability.

Lemma 5.4. *The probability that the algorithm in Algorithm 2 returns a matrix B such that* $\min(\text{rank}(B), k) = \min(\text{rank}(A), k)$ *is at least* $1 - \epsilon - k/|F|$.

Proof. Clearly $\text{rank}(B) \leq \text{rank}(A)$ since the column space of B is a subspace of the column space of A. So $\min(\text{rank}(B), k) \leq \min(\text{rank}(A), k)$, and it remains to show that $\text{rank}(B) \geq \min(\text{rank}(A), k)$ with high probability.

Let $k' = \min(\text{rank}(A), k)$. Let S be a set of linearly independent columns of A with $|S| = k'$, and let $A_{U,S}$ be a $k' \times k'$ submatrix of A with $\text{rank}(A_{U,S}) = k'$. We overload notation to also use S to denote the subset of vertices in G corresponding to those columns. Since G is sampled from a (k, ϵ)-magical distribution, the probability that there is a matching M in which every vertex in S is matched is at least $1 - \epsilon$. Suppose such a matching M exists and let T be the neighbors of S in M with $|T| = |S| = k'$. (In the example in Figure 5.1, suppose $S = \{A_1, A_2, A_3\}$, then M could be $\{x_1y_2, x_2y_3, x_3y_1\}$ and $T = \{B_1, B_2, B_3\}$.) If we view each c_e as a variable, then $\det(B_{U,T})$ is a multivariate polynomial with total degree k'. By setting $c_e = 1$ for each $e \in M$ and $c_e = 0$ for each $e \in E - M$, we get that $B_{U,T} = A_{U,S}$ and thus $\det(B_{U,T})$ is a non-zero multivariate polynomial as $A_{U,S}$ is of full rank. By the Schwartz-Zippel lemma, if we substitute each variable c_e by a random element in a field F, then the probability that $\det(B_{U,T}) = 0$ is at most $k'/|F| \leq k/|F|$. So, if G has a matching that matches every vertex in S, then $\text{rank}(B) \geq \text{rank}(B_{U,T}) = k'$ with probability at least $1 - k/|F|$. Therefore the algorithm

succeeds with probability at least $1 - \epsilon - k/|F|$. □

We can combine Lemma 5.3 and Lemma 5.4 to obtain an efficient compression algorithm.

Theorem 5.2. *Suppose an $m \times n$ matrix A over a field F is given. Given k, there is an algorithm that constructs an $m \times O(k)$ matrix B over F with the following properties.*

1. $\min(\text{rank}(A), k) = \min(\text{rank}(B), k)$ *with probability at least* $1 - O(1/k) - O(k/|F|)$.
2. $\text{nnz}(B) = O(\text{nnz}(A))$ *and B can be constructed in $O(\text{nnz}(A))$ field operations.*

Proof. We can assume $n \geq 11k$; otherwise we can just let $B = A$. We sample a bipartite graph $G = (X \cup Y, E)$ with $|X| = n$ and $|Y| = 11k$ from a $(k, O(1/k))$-magical distribution in $O(n)$ time by Lemma 5.3, with the additional property that each vertex in X is of degree two. We use G in the algorithm in Algorithm 2 to obtain an $m \times 11k$ matrix B over F. Since each vertex of X is of degree two, each entry of A is related to two entries in B. We can represent B by listing the value and position of its non-zero entries without handling duplicate positions, i.e. each non-zero entry in A introduces exactly two entries in B. Therefore, $\text{nnz}(B) = 2\,\text{nnz}(A)$ and B can be constructed in $O(\text{nnz}(A))$ field operations. The probability that $\min(\text{rank}(A), k) = \min(\text{rank}(B), k)$ is at least $1 - O(1/k) - O(k/|F|)$ by Lemma 5.4. □

5.2.5 Computing matrix rank

With the compression algorithm, the first part of Theorem 5.1 follows easily.

Theorem 5.3. *Suppose an $m \times n$ matrix A over a field F is given with $m \leq n$. There is an algorithm to compute* $\min(\text{rank}(A), k)$ *for a given $k \leq m$ in* $O(\text{nnz}(A) + \min(k^\omega, k\,\text{nnz}(A)))$ *field operations with success probability at least* $1 - O(1/n^{1/3})$.

Proof. We can assume that $|F| = \Omega(n^4)$ by Lemma 5.1. We also assume that $k \geq n^{1/3}$; otherwise if $k < n^{1/3}$ we just reset k to be $n^{1/3}$. We apply Theorem 5.2 to compress the $m \times n$ matrix A into an $m \times O(k)$ matrix B. Then $\min(\text{rank}(B), k) = \min(\text{rank}(A), k)$ with probability at least $1 - O(1/k) - O(k/|F|) = 1 - O(1/n^{1/3})$ since $n^{1/3} \leq k \leq n$ and $|F| = \Omega(n^4)$. And B can be constructed in $O(\text{nnz}(A))$ field operations with $\text{nnz}(B) = O(\text{nnz}(A))$. We then apply Theorem 5.2 again on B^T to compress the $m \times O(k)$ matrix B into an $O(k) \times O(k)$ matrix C. Then $\min(\text{rank}(C), k) = \min(\text{rank}(B), k)$ with probability at least $1 - O(1/n^{1/3})$ and C can be constructed in $O(\text{nnz}(A))$ field operations with $\text{nnz}(C) = O(\text{nnz}(A))$. Now we can compute $\text{rank}(C)$ in $O(k^\omega)$ field operations

by using fast matrix multiplication [11]. Alternatively, we can compute $\text{rank}(C)$ in $O(k \,\text{nnz}(C)) = O(k \,\text{nnz}(A))$ field operations using the black box approach [63, 76]. Thus $\min(\text{rank}(A), k)$ can be computed in $O(\text{nnz}(A)+\min(k^\omega, k \,\text{nnz}(A)))$ field operations with success probability $1 - O(1/n^{1/3})$. □

We do not know the rank of the matrix A directly from Theorem 5.3. However we can do so efficiently by searching for a good value of k.

Corollary 5.1. *Given the same setting as in Theorem 5.3, there is an algorithm to compute* $r = \text{rank}(A)$ *in* $O(\text{nnz}(A) \log r + \min(r^\omega, r \,\text{nnz}(A)))$ *field operations with success probability* $1 - O(1/n^{1/3})$.

Proof. To compute $\text{rank}(A)$, we can simply apply Theorem 5.3 with $k = n^{1/3}, 2n^{1/3}, 4n^{1/3}, \ldots, 2^{\log n^{2/3}} n^{1/3}$ until the algorithm returns an answer smaller than k or A is of full rank. Let $r = \text{rank}(A)$. The failure probability is bounded by $O(1/n^{1/3})$ since sum of $1/k$ is less than $2/n^{1/3}$. The number of field operations needed is $O(\text{nnz}(A) \log r + \min(r^\omega, r \,\text{nnz}(A)))$, since the sum of k^ω is $O(r^\omega)$ and the sum of $k \,\text{nnz}(A)$ is $O(r \,\text{nnz}(A))$. □

We can improve Corollary 5.1 slightly and reduce the field operations needed to be $O(\min(\text{nnz}(A) \log r, nm) + \min(r^\omega, r \,\text{nnz}(A)))$. This is done by computing the compressed matrices aggregately and we omit the details here.

5.2.6 Finding independent set

In this subsection, we will find a set of $\min(\text{rank}(A), k)$ linearly independent columns of A, by applying the compression algorithm iteratively to reduce the number of columns of A progressively. In the following, we let $c = 11$, and assume without loss of generality that $k \geq n^{1/3}$ (as in Theorem 5.3). First, we compress the rows while preserving the position of a set of at most k independent columns.

Lemma 5.5. *Suppose an* $m \times n$ *matrix* A *over a field* F *is given. There is an algorithm to return a* $ck \times n$ *matrix* A' *in* $O(\text{nnz}(A))$ *field operations with* $\text{nnz}(A') = O(\text{nnz}(A))$, *such that if* S *is a set of at most* k *linearly independent columns in* A, *then* S *is also a set of linearly independent columns in* A' *with probability at least* $1 - O(1/n^{1/3})$.

Proof. If $m > ck$, we apply the algorithm in Theorem 5.2 to A^T to compress A into a $ck \times n$ matrix A' in $O(\text{nnz}(A))$ field operations, such that $\text{nnz}(A') = O(\text{nnz}(A))$. Let S be a set of at most k linearly independent columns in A, i.e. $|S| \leq k$. By Theorem 5.2,

we have $\text{rank}(A'_{[ck],S}) = \text{rank}(A_{[m],S}) = |S|$ with probability at least $1 - O(1/n^{1/3})$, and thus S is a set of linearly independent columns in A'. □

Next, given a $ck \times n$ matrix A, we show how to find a submatrix A' of A with at most $n/5$ columns in $O(\text{nnz}(A) + k^\omega)$ field operations, such that $\min(\text{rank}(A), k) = \min(\text{rank}(A'), k)$ with high probability. The bounded degree condition of magical graphs is important in the following lemma.

Lemma 5.6. *Given a $ck \times n$ matrix A over a field F where $ck \leq n$, there is an algorithm to find a $(ck) \times (n/5)$ submatrix A' of A in $O(\text{nnz}(A) + k^\omega)$ field operations, such that $\min(\text{rank}(A), k) = \min(\text{rank}(A'), k)$ with probability at least $1 - O(1/n^{1/3})$.*

Proof. We use the algorithm in Theorem 5.2 to compress A into a $ck \times ck$ matrix B in $O(\text{nnz}(A))$ field operations, while $\min(\text{rank}(A), k) = \min(\text{rank}(B), k)$ with probability at least $1 - O(1/n^{1/3})$. Since B is a $ck \times ck$ matrix, we can directly find a set S of $\min(\text{rank}(B), k)$ linearly independent columns in B in $O(k^\omega)$ field operations using fast matrix multiplication [11]. Let $G = (X \cup Y, E)$ be the bipartite graph used in the compression algorithm with $|X| = n$ and $|Y| = ck$. Let T be the set of columns in A that correspond to the neighbors of the vertices corresponding to S in G. By the bounded degree condition of G, each vertex corresponding to a column in S is of degree at most $2|X|/|Y| = 2n/(ck)$ and hence $|T| \leq 2n|S|/(ck) \leq 2n/c < n/5$. We have that the $ck \times |T|$ submatrix $A' := A_{[ck],T}$ is of rank at least $\min(\text{rank}(A), k)$, since the column space of S in B is spanned by the column space of $A_{[ck],T}$. □

Applying Lemma 5.6 repeatedly gives us the second part of Theorem 5.1.

Theorem 5.4. *Suppose an $m \times n$ matrix A over a field F is given. There is an algorithm to find a set of $\min(\text{rank}(A), k)$ linearly independent columns of A for a given k in $O((\text{nnz}(A) + k^\omega) \log n)$ field operations with success probability at least $1 - O(\log n/n^{1/3})$. When F is a finite field, each field operation can be done in $\tilde{O}(\log n + \log |F|)$ steps.*

Proof. First, we apply Lemma 5.5 to reduce the number of rows to ck. Then, we apply Lemma 5.6 repeatedly until the number of columns is reduced to $O(k)$. Since each time we can reduce the number of columns by a constant factor, we need to repeat the algorithm in Lemma 5.6 at most $O(\log n)$ times. Finally, we find a set of $\min(\text{rank}(A), k)$ linearly independent columns by Gaussian elimination in the $ck \times O(k)$ submatrix in $O(k^\omega)$ time. So, the whole algorithm can be done in at most $O((\text{nnz}(A) + k^\omega) \log n)$ field operations, and the failure probability is at most $O(\log n/n^{1/3})$. □

We remark that c can be arbitrarily chosen as long as $c \geq 11 > 4e$. Hence when $\mathrm{nnz}(A) \gg k$, we can choose c to be larger so that the number of columns reduces to $O(k)$ faster. For example, if we choose $c = 2n^\epsilon$, then Lemma 5.6 find a $(ck) \times (n/n^\epsilon)$ submatrix in $O(\mathrm{nnz}(A)+(ck)^\omega)$ field operations, and so we only need to apply Lemma 5.6 $1/\epsilon$ times. Therefore if $k = O(\mathrm{nnz}(A)^{1/\omega-\epsilon})$, then we can find a set of $\min(\mathrm{rank}(A), k)$ linearly independent columns in $O(\mathrm{nnz}(A)/\epsilon)$ field operations, saving the additional log factor.

5.3 Applications

The matrix rank algorithms can be readily applied to various problems in numerical linear algebra, combinatorial optimization, and dynamic data structure. In this subsection, we will state some applications of our results without proofs. Interested readers may find more details in our paper [17].

First we show that the algorithms can be applied to computing a rank-one decomposition, finding a basis of the null space, and performing matrix multiplication for a low rank matrix.

Theorem 5.5. *Let A be an $m \times n$ matrix over a field F. Let $r = \mathrm{rank}(A)$. Let $m' = \min\{m, n\}$. Let $\omega(a, b, c)$ be the exponent for multiplying an $n^a \times n^b$ matrix with an $n^b \times n^c$ matrix.*

1. *There is a randomized algorithm to compute an $m \times r$ matrix X and an $r \times n$ matrix Y such that $A = XY$ in $\tilde{O}(\mathrm{nnz}(A) + r^{\omega(1,1,\log_r m')}) = \tilde{O}(\mathrm{nnz}(A) + m'r^{\omega-1})$ steps.*

2. *There is a randomized algorithm to find a basis of the null space of A in $\tilde{O}(\mathrm{nnz}(A)+ r^{\omega(1,1,\log_r n)}) = \tilde{O}(\mathrm{nnz}(A) + nr^{\omega-1})$ steps.*

3. *Let A and B be $n \times n$ matrices. There is a randomized algorithm to compute AB in $\tilde{O}(n^{\omega(\log_n r,1,1)}) = \tilde{O}(n^2 r^{\omega-2})$ steps.*

The success probability for all three tasks is at least $1 - O(\log(nm)/\mathrm{nnz}(A)^{1/3})$.

Previously the best known algorithms require $\tilde{\Theta}(mnr^{\omega-2})$ for the first two tasks, and $\tilde{\Theta}(n^2 r^{\omega-2})$ for the third task. Our algorithms are faster than the existing algorithms, especially when r is small. The statement about matrix multiplication essentially says that the problem of multiplying two $n \times n$ matrices while one matrix is of rank r can be reduced to the problem of multiplying an $r \times n$ matrix and an $n \times n$ matrix.

	graph matching	linear matroid intersection	linear matroid union
combinatorial	$O(\sqrt{\mathsf{opt}}\lvert E\rvert)$ [52, 28]	$\tilde{O}(nr\mathsf{opt}^{1/(4-\omega)})$ [27]	$\tilde{O}(nr\mathsf{b}\mathsf{opt} + \mathsf{nb}^2\mathsf{opt}^2)$ [24]
algebraic	$O(\lvert V\rvert^\omega)$ [53]	$O(nr^{\omega-1})$ [32]	–
this paper	$\tilde{O}(\lvert E\rvert + \mathsf{opt}^\omega)$	$\tilde{O}(nr + n\mathsf{opt}^{\omega-1})$	$\tilde{O}(nr\mathsf{opt} + \mathsf{b}^3\mathsf{opt}^3)$

Table 5.1: Time complexity of algorithms for some problems in combinatorial optimization

In combinatorial optimization, there are algebraic formulations of the problems that relate the optimal value to the rank of an associated matrix. Using this connection, we can apply the algorithm in Theorem 5.1 to obtain fast algorithms for graph matching and matroid optimization problems.

Theorem 5.6. *Let* opt *be the optimal value of an optimization problem.*

1. *Given an undirected graph* $G = (V, E)$*, there is a randomized algorithm to find a matching of size* $\min\{\mathsf{opt}, \mathsf{k}\}$ *in* $\tilde{O}(|E| + k^\omega)$ *time.*

2. *Given a linear matroid intersection problem or a linear matroid parity problem with an* $r \times 2n$ *matrix* A*, there is a randomized algorithm to find a solution of size* $\min\{\mathsf{opt}, \mathsf{k}\}$ *in* $\tilde{O}(\mathrm{nnz}(A) + nk^{\omega-1})$ *time.*

3. *Given a linear matroid union problem with an* $r \times n$ *matrix* $|A|$*, there is a randomized algorithm to find* $\min\{\mathsf{opt}, \mathsf{k}\}$ *disjoint bases in* $\tilde{O}(k\,\mathrm{nnz}(A) + \min\{k^{\omega+1}b^\omega, k^3b^3\})$ *time, where* b *denotes the size of a basis.*

Table 5.1 lists the time complexity of the best known combinatorial algorithms and algebraic algorithms for these problems. Notice that previous algebraic algorithms have the same time complexity even when the optimal value is small. On the other hand, combinatorial algorithms for these problems are based on finding augmenting structures iteratively, and thus the number of iterations and the overall complexity depend on the optimal value. While the previous algebraic algorithms are faster than combinatorial algorithms only when the optimal value is large, the results in Theorem 5.6 show that the algebraic approach can be faster for any optimal value. For the matroid optimization problems, the algorithms in Theorem 5.6 are faster than previous algorithms in any setting.

Bibliography

[1] Rudolf Ahlswede, Ning Cai, S-YR Li, and Raymond W Yeung. Network information flow. *Information Theory, IEEE Transactions on*, 46(4):1204–1216, 2000.

[2] Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast johnson-lindenstrauss transform. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 557–563. ACM, 2006.

[3] Nir Ailon and Edo Liberty. An almost optimal unrestricted fast johnson-lindenstrauss transform. *ACM Transactions on Algorithms (TALG)*, 9(3):21, 2013.

[4] Noga Alon and Vitali D Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B*, 38(1):73–88, 1985.

[5] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 475–486. IEEE, 2006.

[6] Reid Andersen and Yuval Peres. Finding sparse cuts locally using evolving sets. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 235–244. ACM, 2009.

[7] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 563–572. IEEE, 2010.

[8] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *Journal of the ACM (JACM)*, 56(2):5, 2009.

[9] Nikhil Bansal, Uriel Feige, Robert Krauthgamer, Konstantin Makarychev, Viswanath Nagarajan, Joseph Seffi, and Roy Schwartz. Min-max graph partitioning and small set expansion. *SIAM Journal on Computing*, 43(2):872–904, 2014.

[10] Paul Bonsma, Hajo Broersma, Viresh Patel, and Artem Pyatkin. The complexity status of problems related to sparsest cuts. In *Combinatorial Algorithms*, pages 125–135. Springer, 2011.

[11] James R Bunch and John E Hopcroft. Triangular factorization and inversion by fast matrix multiplication. *Mathematics of Computation*, 28(125):231–236, 1974.

[12] D Burago. Periodic metrics. In *Seminar on Dynamical Systems*, pages 90–95. Springer, 1994.

[13] Shuchi Chawla, Robert Krauthgamer, Ravi Kumar, Yuval Rabani, and D Sivakumar. On the hardness of approximating multicut and sparsest-cut. *computational complexity*, 15(2):94–114, 2006.

[14] Jeff Cheeger. A lower bound for the smallest eigenvalue of the laplacian. *Problems in analysis*, 625:195–199, 1970.

[15] Li Chen, Wayne Eberly, Erich Kaltofen, B David Saunders, William J Turner, and Gilles Villard. Efficient matrix preconditioners for black box linear algebra. *Linear Algebra and its Applications*, 343:119–146, 2002.

[16] Joseph Cheriyan. Randomized $\tilde{O}(M(|V|))$ algorithms for problems in matching theory. *SIAM Journal on Computing*, 26(6):1635–1655, 1997.

[17] Ho Yee Cheung, Tsz Chiu Kwok, and Lap Chi Lau. Fast matrix rank algorithms and applications. *Journal of the ACM (JACM)*, 60(5):31, 2013.

[18] 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.

[19] Ho Yee Cheung, Lap Chi Lau, and Kai Man Leung. Algebraic algorithms for linear matroid parity problems. *ACM Transactions on Algorithms (TALG)*, 10(3):10, 2014.

[20] Fan Chung. Random walks and local cuts in graphs. *Linear Algebra and its applications*, 423(1):22–32, 2007.

[21] Fan Chung. A local graph partitioning algorithm using heat kernel pagerank. *Internet Mathematics*, 6(3):315–330, 2009.

[22] Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

[23] Don Coppersmith and Shmuel Winograd. Matrix multiplication via arithmetic progressions. *Journal of symbolic computation*, 9(3):251–280, 1990.

[24] William H Cunningham. Improved bounds for matroid partition and intersection algorithms. *SIAM Journal on Computing*, 15(4):948–957, 1986.

[25] Irit Dinur. The pcp theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12, 2007.

[26] Martin Dyer, Alan Frieze, and Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM (JACM)*, 38(1):1–17, 1991.

[27] Harold N Gabow and Ying Xu. Efficient theoretic and practical algorithms for linear matroid intersection problems. *Journal of Computer and System Sciences*, 53(1):129–147, 1996.

[28] Andrew V Goldberg and Alexander V Karzanov. Maximum skew-symmetric flows and matchings. *Mathematical Programming*, 100(3):537–568, 2004.

[29] Venkatesan Guruswami and Ali Kemal Sinop. Approximating non-uniform sparsest cut via generalized spectra. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 295–305. SIAM, 2013.

[30] Lars Hagen and Andrew B Kahng. New spectral methods for ratio cut partitioning and clustering. *Computer-aided design of integrated circuits and systems, ieee transactions on*, 11(9):1074–1085, 1992.

[31] Olle Häggström. *Finite Markov chains and algorithmic applications*, volume 52. Cambridge University Press, 2002.

[32] Nicholas JA Harvey. Algebraic algorithms for matching and matroid problems. *SIAM Journal on Computing*, 39(2):679–702, 2009.

[33] Tracey Ho, Muriel Médard, Ralf Koetter, David R Karger, Michelle Effros, Jun Shi, and Ben Leong. A random linear network coding approach to multicast. *Information Theory, IEEE Transactions on*, 52(10):4413–4430, 2006.

[34] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[35] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[36] Xiaohan Huang and Victor Y Pan. Fast rectangular matrix multiplication and applications. *Journal of complexity*, 14(2):257–299, 1998.

[37] Oscar H Ibarra, Shlomo Moran, and Roger Hui. A generalization of the fast lup matrix decomposition algorithm and applications. *Journal of Algorithms*, 3(1):45–56, 1982.

[38] Mark Jerrum and Alistair Sinclair. The markov chain monte carlo method: an approach to approximate counting and integration. *Approximation algorithms for NP-hard problems*, pages 482–520, 1996.

[39] Erich Kaltofen and B David Saunders. On wiedemann's method of solving sparse linear systems. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, pages 29–38. Springer, 1991.

[40] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. *Journal of the ACM (JACM)*, 51(3):497–515, 2004.

[41] Tsz Chiu Kwok and Lap Chi Lau. Finding small sparse cuts by random walk. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 615–626. Springer, 2012.

[42] Tsz Chiu Kwok and Lap Chi Lau. Lower bounds on expansions of graph powers. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 395–410. Springer, 2014.

[43] 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 *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 11–20. ACM, 2013.

[44] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, pages 296–303. ACM, 2014.

[45] James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multi-way spectral partitioning and higher-order cheeger inequalities. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1117–1130. ACM, 2012.

[46] Tom Leighton and Satish Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999.

[47] Jure Leskovec, Kevin J Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, pages 631–640. ACM, 2010.

[48] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.

[49] Anand Louis, Prasad Raghavendra, Prasad Tetali, and Santosh Vempala. Many sparse cuts via higher eigenvalues. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1131–1140. ACM, 2012.

[50] László Lovász and Ravi Kannan. Faster mixing via average conductance. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 282–287. ACM, 1999.

[51] László Lovász and Miklós Simonovits. The mixing rate of markov chains, an isoperimetric inequality, and computing the volume. In *Foundations of Computer Science, 1990. Proceedings., 31st Annual Symposium on*, pages 346–354. IEEE, 1990.

[52] Silvio Micali and Vijay V Vazirani. An $O(\sqrt{V}E)$ algorithm for finding maximum matching in general graphs. In *Foundations of Computer Science, 1980., 21st Annual Symposium on*, pages 17–27. IEEE, 1980.

[53] Marcin Mucha and Piotr Sankowski. Maximum matchings via gaussian elimination. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 248–255. IEEE, 2004.

[54] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[55] Shayan Oveis Gharan and Luca Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 187–196. IEEE, 2012.

[56] Ryan O'Donnell and David Witmer. Improved small-set expansion from higher eigenvalues. *CoRR, abs/1204.4688*, 152, 2012.

[57] Prasad Raghavendra and Tselil Schramm. Gap amplification for small-set expansion via random walks. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 381–391. Springer, 2014.

[58] Prasad Raghavendra and David Steurer. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 755–764. ACM, 2010.

[59] Prasad Raghavendra, David Steurer, and Prasad Tetali. Approximations for the isoperimetric and spectral profile of graphs and related parameters. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 631–640. ACM, 2010.

[60] Prasad Raghavendra, David Steurer, and Madhur Tulsiani. Reductions between expansion problems. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 64–73. IEEE, 2012.

[61] Piotr Sankowski. Faster dynamic matchings and vertex connectivity. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 118–126. Society for Industrial and Applied Mathematics, 2007.

[62] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 143–152. IEEE, 2006.

[63] B David Saunders, Arne Storjohann, and Gilles Villard. Matrix rank certification. *Electronic Journal of Linear Algebra*, 11:16–23, 2004.

[64] Jacob T Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM (JACM)*, 27(4):701–717, 1980.

[65] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.

[66] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82(1):93–133, 1989.

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

[68] Daniel A Spielman and Shang-Hua Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM Journal on Matrix Analysis and Applications*, 35(3):835–885, 2014.

[69] David Steurer. *On the complexity of unique games and graph expansion.* PhD thesis, Citeseer, 2010.

[70] Arne Storjohann. Integer matrix rank certification. In *Proceedings of the 2009 international symposium on Symbolic and algebraic computation*, pages 333–340. ACM, 2009.

[71] David A Tolliver and Gary L Miller. Graph partitioning by spectral rounding: Applications in image segmentation and clustering. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 1053–1060. IEEE, 2006.

[72] Lloyd N Trefethen and David Bau III. *Numerical linear algebra*, volume 50. Siam, 1997.

[73] Luca Trevisan. Max cut and the smallest eigenvalue. *SIAM Journal on Computing*, 41(6):1769–1786, 2012.

[74] Luva Trevisan. The spectral partitioning algorithm. http://lucatrevisan.wordpress.com/2008/05/11/the-spectral-partitioning-algorithm/, 2008.

[75] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[76] Joachim Von Zur Gathen and Jürgen Gerhard. *Modern computer algebra*. Cambridge university press, 2013.

[77] Douglas Wiedemann. Solving sparse linear equations over finite fields. *Information Theory, IEEE Transactions on*, 32(1):54–62, 1986.

[78] Virginia Vassilevska Williams. Multiplying matrices faster than coppersmith-winograd. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 887–898. ACM, 2012.

[79] Raphael Yuster. Generating a d-dimensional linear subspace efficiently. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 467–470. Society for Industrial and Applied Mathematics, 2010.