APPROXIMATING GRAPH CONDUCTANCE: FROM GLOBAL TO LOCAL

YUEHENG ZHANG

Abstract. In this expository paper, we introduce spectral algorithms and random walk based algorithms on graphs through the problem of finding sparse cuts in both global and local settings. First, we consider the global problem of finding a set with minimum conductance, and show how the Laplacian and normalized Laplacian matrices arise naturally when we relax this problem to an eigenvalue computation problem. Then we bound the relaxed quantity, completing a spectral approximation algorithm and a proof of Cheeger's inequality. Next, we discuss limitations of this global algorithm and motivate the problem of finding local conductance, for which we introduce random walks on graphs. We discuss how to gain information about conductance from the convergence of random walks, and motivate convergence bounds related to local conductance. Finally, we briefly introduce algorithms utilizing such bounds. Throughout, we place an emphasis on providing motivation and explanations from an algorithmic point of view.

Contents

1. Introduction

Graph structures are ubiquitous in computer science as well as in our daily lives, and nowadays they are getting increasingly large. An important set of problems

Written for the University of Chicago 2020 Math REU. Date updated: November 29, 2020.

when processing or performing tasks on large graphs is *graph partitioning*, where we want to partition vertices in a graph into sets of relatively even size, so that these sets are loosely connected with each other but densely connected within themselves. We informally refer to these sets by "clusters." For more precise definition of graph partitioning and examples on applications of graph partitioning, see Ch. 2, Ch.3 of [14], respectively.

In this paper, by graph partitioning we mean specifically the problem of finding sparse cuts in a graph, i.e., finding a bipartition of the vertices such that there are few connections between both sides of the cut. There are different ways to measure this objective. The conductance of a cut is one such measurement, and the conductance of a graph is defined as the minimum among conductance of all cuts.

Graph partitioning in general and finding the conductance of a graph in particular is in NP, thus believed to be hard to solve. Consequently we are interested in efficient *approximation algorithms* for it. A class of well-established approximation algorithms for this problem is based on the spectrum of matrices associated with a graph. A closely related class of approximation algorithms is based on ideas of random walks on a graph.

In monographs or graduate course lecture notes, these subjects are often discussed together as an introduction to graph partitioning, and the relationships between them are not dwelled on too much. However, readers newly introduced to algorithms on graphs, especially undergraduate students who are prone to have only studied graphs in a discrete math context, may benefit from more motivation and explanations. In this paper, we aim to elaborate on the points which graduate course notes or advanced literature often do not have enough space for, and to provide motivations from an algorithmic point of view.

Among subjects related to this topic, the Laplacian and normalized Laplacian matrices especially are often given as a predefined object because of their ties to the continuous Laplace operator, and may seem unmotivated for readers who learn about this topic from a discrete background. Although there exists an abundance of online sources that touch on motivation and insights for the Laplacian matrices, these sources often change between different versions of Laplacian matrix or different measures for sparsity of cuts, or restrict the discussion to regular graphs. Often these changes or simplifications allow insights about the problem to be better and efficiently conveyed, though for readers new to this subject, they might cause confusion for lack of coherence, motivation, or generality. We aim to provide a parallel to these resources by trying to absorb the insights from them while giving a coherent and general development.

To be more specific, we focus on the problem of finding low-conductance sets in a graph, through which we motivate spectral and random walk based methods. We state and prove all results for irregular (undirected) graphs with weighted edges. We require our graphs to be connected, because otherwise we can always look at a connected component.

First, we consider the global problem of finding the conductance of a graph. We show that the Laplacian and normalized Laplacian matrices arise naturally when we relax this computationally hard problem to an eigenvalue computation problem. More specifically, we "recover" the Laplacian matrix from considering the quadratic form corresponding to the total edge capacity across a cut. Then, we "recover" the

normalized Laplacian matrix from the process of writing our minimization problem as a generalized Rayleigh quotient, proving the easy side of Cheeger's inequality at the same time. Finally, to bound this relaxed quantity, we state and present a proof for the hard side of Cheeger's inequality, completing a spectral approximation algorithm.

Next, by discussing limitations of global algorithms and demands from applications, we motivate the problem of finding a local low-conductance set and introduce the idea of *local algorithms*, i.e., algorithms that only have access to one vertex and its neighbors at each step. We introduce random walks on graphs in this context. Using the connection between random walk transition matrix and normalized Laplacian, we prove a convergence bound which allows us to gain information about the global conductance from random walks. Then, we discuss limitations of this bound, and motivate convergence bound related to local conductance. Lastly, we give a brief introduction to such local convergence bounds as well as algorithms utilizing them, referring the reader to further sources.

We emphasize that this is only one way of introducing these rich subjects, and we will refer the reader to many sources along the way, from lecture notes we borrowed ideas or proofs from, to literature on more general or more advanced topics.

1.1. **Notation and general definitions.**

For a vector $\mathbf{x}$, we denote the i-th entry of $\mathbf{x}$ by $\mathbf{x}_i$ or $(\mathbf{x})_i$. We write $M = (m_{ij})$ to mean that the (i,j)-th entry of matrix M is m_{ij}.

We write $\mathrm{diag}(b_1, \ldots, b_m)$ for the $m \times m$ diagonal matrix with b_i's on the diagonal. For a diagonal matrix $B = \mathrm{diag}(b_1, \ldots, b_m)$ with strictly positive diagonal entries, for $a \in \mathbb{R}$, we define H^a to be the matrix $\mathrm{diag}(b_1^a, b_2^a, \ldots, b_m^a)$.

We denote the all-ones vector in $\mathbb{R}_n$ by $\mathbf{1}_n$, and omit n when it is clear from context.

Throughout this paper, by *a graph* $G = (V, E, \mathbf{w})$ we mean a connected undirected graph with vertex set V, edge set E (where self-loops are allowed), and a weight function $\mathbf{w} : E \to \mathbb{R}_{++}$ that assigns a strictly positive weight to each edge. We write $\mathbf{w}_{uv}$ or $\mathbf{w}_{vu}$ for the weight on $\{u, v\} \in E$.

We usually write n for $|V|$, and identify V with the set $[n] = \{1, 2, \ldots, n\}$. Whenver we refer to a subset of vertices, we assume the set to be nonempty and not equal to V. For a subset S of vertices, we write $\chi_S \in \mathbb{Z}^n$ for the indicator vector of S, i.e., $(\chi_S)_i = 1$ if $i \in S$ and $(\chi_S)_i = 0$ otherwise. When S consists of a single vertex $i \in V$, we use the cooresponding standard basis vector $\mathbf{e}_i$ instead of $\chi_{\{i\}}$. We write $\overline{S}$ for the complement $V - S$ of S.

Definitions 1.1. We say vertices i and j are *adjacent*, and write $i \sim j$ or $j \sim i$, if $\{i, j\} \in E$. For a vertex $i \in V$, the *degree* of i is defined as

$$\deg(i) = \sum_{j \in V,\, j \sim i} \mathbf{w}_{ij}.$$

For a set S of vertices, the *volume* of S is defined as

$$\mathrm{vol}(S) = \sum_{i \in S} \deg(i).$$

The volume of the graph, $\mathrm{vol}(G)$, is defined to be $\mathrm{vol}(V)$.

Lastly, we introduce two matrices naturally associated with a graph G.

Definitions 1.2. The *adjacency matrix* $A = (a_{ij})$ of G is the $n \times n$ matrix with

$$a_{ij} = \begin{cases} \mathbf{w}_{ij} & \text{if } i \sim j \\ 0 & \text{otherwise.} \end{cases}$$

The *degree matrix* D of G is the diagonal matrix $\mathrm{diag}(d_1, \ldots, d_n)$ where for $i \in V$,

$$d_i = \deg(i).$$

1.2. **Conductance of cuts and of a graph.**

Definition 1.3. A *cut* of a graph G defined by a subset S of vertices is a partition $V = S \sqcup \overline{S}$ of the set of vertices. We write $E(S)$ for the set of edges crossing the cut defined by S; i.e.,

$$E(S) := \{\{u, v\} : u \in S, v \in \overline{S}\}.$$

We define $\mathrm{vol}(E(S))$ by

$$\mathrm{vol}(E(S)) = \sum_{\{i,j\} \in E(S)} \mathbf{w}_{ij}.$$

We note that the definition of a cut is symmetric; that is, the cut defined by S and $\overline{S}$ are the same, moreover $E(S) = E(\overline{S})$ and $\mathrm{vol}(E(S)) = \mathrm{vol}(E(\overline{S}))$.

We can view a graph as a venue on which processes take place. The weight on an edge between two vertices can be viewed as the capacity between these vertices for a process. Similarly, $\mathrm{vol}(E(S))$ is the total capacity between two sides of the cut, and we refer to it as the *cut capacity* of a cut.

In many applications, we want to find a cut that is both sparse and balanced. That is, we want to have few connections between two sides of the cut, and we want the two sides to be of comparable size. There are different ways to define what "few", "comparable", and "size" mean, resulting in different measurements for how "good" a cut is. Below we define one such measurement.

Definition 1.4. For a subset S of vertices, the *conductance* of S is defined as

$$\phi(S) = \frac{\mathrm{vol}(E(S))}{\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}}. \tag{1.5}$$

We also refer to $\phi(S)$ as the conductance of the cut defined by S.

We recall that S is nonempty and does not equal V. The numerator in $\phi(S)$ is the cut capacity between S and $\overline{S}$, and the denominator is the total capacity of edges inside the smaller side of the cut. The smaller the conductance of a cut is, the "better" it is for our purpose. The mininum in the denominator discourages a low-conductance set from having a large volume and encourages a low-conductance cut to be balanced.

The definition of the conductance of S is again symmetric, with $\phi(S) = \phi(\overline{S})$. We also note that the range of $\phi(S)$ is $0 < \phi(S) < 1$. Indeed, this follows from that $\mathrm{vol}(E(S)) < \mathrm{vol}(S)$ and that $\phi(S) = \phi(\overline{S})$.

The conductance of a graph is defined as a minimum over the conductance of all cuts.

Definition 1.6. The conductance $\phi(G)$ of a graph G is defined as

$$\phi(G) = \min_{S \subset V} \phi(S) = \min_{S \subset V} \frac{\mathrm{vol}(E(S))}{\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}}. \tag{1.7}$$

$\phi(G)$ is also in the range $0 < \phi(G) < 1$.

To compute $\phi(G)$ or the cut that achieves it is NP-hard. Thus, we are interested in efficient algorithms that find good cuts which approximate the best cut with respect to conductance. Before we begin approximating $\phi(G)$, however, we first work out some technical details.

2. Generalized Rayleigh quotient and variational characterization of eigenvalues

In this section, we state some results we need from linear algebra. The proofs we omit can easily be found in linear algebra textbooks, or in lecture notes for spectral graph theory, e.g. Ch.1 in [13].

Theorem 2.1 (Spectral theorem for real symmetric matrices)**.** *If M is an $k \times k$ real symmetric matrix (i.e., $M = M^T$), then M has an orthonormal eigenbasis $(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ over $\mathbb{R}$. In other words, for $1 \leq i \leq k$, there is $\lambda_i \in \mathbb{R}$ such that*

$$M\mathbf{v}_i = \lambda_i \mathbf{v}_i,$$

and

$$M = \sum_{i=1}^{k} \lambda_i \mathbf{v}_i \mathbf{v}_i^T. \tag{2.2}$$

Definition 2.3. For a $k \times k$ real symmetric matrix M, and a $k \times k$ real diagonal matrix B with strictly positive diagonal entries, the *generalized Rayleigh quotient of M with respect to B* is a function

$$R(M, B, \mathbf{x}) : \mathbb{R}^k \setminus \{\mathbf{0}\} \to \mathbb{R}$$

defined by

$$R(M, B, \mathbf{x}) = \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T B \mathbf{x}}.$$

When B is the identity operator I, we write $R_M(\mathbf{x})$ for $R(M, I, \mathbf{x})$, and call it the *Rayleigh quotient* of M.

$R(M, B, \mathbf{x})$ can be interpreted as the Rayleigh quotient of M in the Hilbert space where inner product is defined by the quadratic form of $B = \operatorname{diag}(b_1, \ldots, b_k)$, i.e., in a space where the i-th coordinate is stretched by an amount defined by b_i.

Theorem 2.4 (Courant–Fischer–Weyl min-max principle)**.** *Let M be a $k \times k$ real symmetric matrix. By the spectral theorem, we know M has k real eignevalues $\nu_1 \leq \nu_2 \leq \cdots \leq \nu_k$. Then, for $1 \leq i \leq k$*

$$\nu_i = \min_{U \subset \mathbb{R}^k,\, \dim(U) = i} \quad \max_{\mathbf{x} \in U,\, \mathbf{x} \neq \mathbf{0}} \quad R_M(\mathbf{x}),$$

and

$$\nu_i = \max_{U \subset \mathbb{R}^k,\, \dim(U) = k-i+1} \quad \min_{\mathbf{x} \in U,\, \mathbf{x} \neq \mathbf{0}} R_M(\mathbf{x}).$$

Morever, optimum is attained when $\mathbf{x}$ is an eigenvector to the corresponding ν_i.

In the above context, let $\mathbf{s}$ be an eigenvector to ν_1. We have in particular that

$$\nu_1 = \min_{\mathbf{x} \neq \mathbf{0}} \quad R_M(\mathbf{x}), \tag{2.5}$$

$$\nu_2 = \min_{\mathbf{x}\not\in \mathrm{span}(\mathbf{s})} \quad R_M(\mathbf{x}), \tag{2.6}$$

and

$$\nu_n = \max_{\mathbf{x}\neq \mathbf{0}} \quad R_M(\mathbf{x}). \tag{2.7}$$

The above theorem allows us to associate the eigensystem of a real symmetric matrix with solutions to optimization problems on Rayleigh quotients. Below, we give a corollary which shows that there is a matrix whose eigensystem corresponds with solutions to optimization problems related to generalized Rayleigh quotients.

Corollary 2.8. *Let M be an $k \times k$ real symmetric matrix with eigenvalues $\nu_1 \leq \nu_2 \leq \cdots \leq \nu_k$. Let B be an $k \times k$ diagonal matrix with strictly positive diagonal entries. Then the matrix*

$$\widetilde{M} := B^{-1/2} M B^{-1/2} = (\widetilde{m}_{ij})$$

satisifes the following:

(a) $\widetilde{M}$ has k real eigenvalues $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$.

(b)

$$\mu_1 = \min_{\mathbf{x}\neq \mathbf{0}} \quad R(M, B, \mathbf{x})$$

(c)

$$\mu_2 = \min_{\mathbf{u}_1^T B^{1/2}\mathbf{x}=0,\ \mathbf{x}\neq \mathbf{0}} \quad R(M, B, \mathbf{x}). \tag{2.9}$$

Moreover, μ_2 is attained when $\mathbf{x} = B^{-1/2}\mathbf{u}_2$ where $\mathbf{u}_2$ is an eigenvector to μ_2.

(d)

$$\mu_n = \max_{\mathbf{x}\neq \mathbf{0}} \quad R(M, B, \mathbf{x}).$$

Proof. Since M is symmetric and B is diagonal, we can check by each entry that $\widetilde{m}_{ij} = \widetilde{m}_{ji}$. Then, $\widetilde{M}$ has an orthonormal real eigenbasis by the spectral theorem. Now,

$$\begin{aligned}
\mu_1 = \min_{\mathbf{y}\neq \mathbf{0}} \quad & R_{\widetilde{M}}(\mathbf{y}) = \min_{\mathbf{y}\neq \mathbf{0}} \quad \frac{\mathbf{y}^T B^{-1/2} M B^{-1/2} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \\
= \min_{\mathbf{y}\neq \mathbf{0}} \quad & \frac{(B^{-1/2}\mathbf{y})^T M (B^{-1/2}\mathbf{y})}{(B^{-1/2}\mathbf{y})^T B (B^{-1/2}\mathbf{y})} \\
= \min_{\mathbf{x}\neq \mathbf{0}} \quad & \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T B \mathbf{x}}
\end{aligned}$$

where the last equality follows from a change of variable by $B^{1/2}\mathbf{x} = \mathbf{y}$ and the fact that $B^{1/2}$ has trivial null space and thus is bijective. (d) follows from the same reasoning. Similarly,

$$\begin{aligned}
\mu_2 = \min_{\mathbf{y}\perp \mathbf{u}_1,\ \mathbf{y}\neq \mathbf{0}} \quad & \frac{\mathbf{y}^T B^{-1/2} M B^{-1/2} \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \\
= \min_{\mathbf{u}_1^T \mathbf{y}=0,\ \mathbf{y}\neq \mathbf{0}} \quad & \frac{(\mathbf{y}^T B^{-1/2}) B (B^{-1/2}\mathbf{y})}{(\mathbf{y}^T B^{-1/2}) B (B^{-1/2}\mathbf{y})} \\
= \min_{\mathbf{u}_1^T B^{1/2}\mathbf{x}=0,\ \mathbf{x}\neq \mathbf{0}} \quad & \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T B \mathbf{x}}.
\end{aligned}$$

Finally, since μ_2 is attained for the Rayleigh quotient of $\widetilde{M}$ when $\mathbf{y} = \mathbf{u}_2$, we have

$$R(M, B, B^{-1/2}\mathbf{u}_2) = R(\widetilde{M}, I, \mathbf{u}_2) = \mu_2.$$

It remains to show that $B^{-1/2}\mathbf{u}_2$ satisfies the constraint $\mathbf{u}_1^T B^{1/2}\mathbf{x} = 0$:

$$\mathbf{u}_1^T B^{1/2}(B^{-1/2}\mathbf{u}_2) = \mathbf{u}_1^T \mathbf{u}_2 = 0$$

□

We note that analogous statements hold for the other eigenvalues as well.

3. Spectral approximation of the conductance of a graph

We recall that the conductance of a graph is defined by minimizing a quotient, and that computing it is NP-hard. Eigenvalues of a symmetric matrix, on the other hand, can be computed efficiently. When dealing with a computationally hard problem, we often relax it to an efficiently-computable problem, and show that solution of the relaxed problem is not too far from the original solution. In the context of a minimization problem, this means that we expand the set on which our minimization takes place, and lower bound the minimum on this expanded set by a function related to the original minimum.

In this section, we relax the problem of finding the conductance of a graph to the problem of computing an eigenvalue, discovering the Laplacian and normalized Laplacian matrices along the way. Then we lower bound the eigenvalue by constructing a low-conductance set from an eigenvector, completing the approximation which is known as Cheeger's inequality.

3.1. **Relaxing the problem of finding the conductance of a graph.**

In this section, we relax the problem of finding the conductance of a graph by making use of the characterization of eigenvalues in relation to generalized Rayleigh quotients in Cor. 2.8. We are combining ideas in Ch.3.1 of [13], in [8], and in [15].

When finding the conductance of a graph, we are minimizing the quotient in Def. 1.6 over all subsets of V. We would like to try to relate this quotient to a generalized Rayleigh quotient. We start by converting this problem to a minimization problem over integer vectors, which we can then relax to be over real vectors.

A natural way to represent a subset S of vertices using an integer vector is to use the indicator vector χ_S. Then, 0-1 vectors in the set $\{0,1\}^n - \{\mathbf{0}, \mathbf{1}\}$ correspond bijectively with subsets of vertices. Therefore, we try to express the numerator and demonimator in $\phi(S)$ by $\phi(S)$.

For Sec. 3.1.1 and Sec. 3.1.2, we let S be a subset of vertices, and write $\mathbf{x}$ for χ_S for ease of notation.

3.1.1. *Cut capacity and the Laplacian.*

We recall that the numerator in $\phi(S)$, the cut capacity, is

$$\mathrm{vol}(E(S)) = \sum_{\{i,j\}\in E,\ i\in S, j\in\overline{S}\}} \mathbf{w}_{ij}.$$

In order to write $\phi(S)$ as a generalized Rayleigh quotient, we want to find a matrix $L = (\ell_{ij})$ such that its quadratic form satisfies

$$\mathbf{x}^T L\mathbf{x} = \mathrm{vol}(E(S)). \tag{3.1}$$

To find entries of L, we first consider the case when S consists of one single vertex i. For ease of notation, we have the following definition.

Definition 3.2. Let $g : V \to \mathbb{R}$ be defined by

$$g(i) = \begin{cases} \mathbf{w}_{ii} & \text{if } i \sim i, \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i \in V.$$

We want to have

$$\mathbf{e}_i^T L \mathbf{e}_i = \sum_{j \in V,\, j \sim i,\, j \neq i} \mathbf{w}_{ij} = \deg(i) - g(i). \tag{3.3}$$

Thus we know $\ell_{ii} = \deg(i) - g(i)$ for $i \in V$.

Meanwhile, although the demonimator in $\phi(S)$ forbids S from being empty or being equal to V, the cut capacity $\mathrm{vol}(E(S))$ is meaningful even if $S = V$. Since $\mathrm{vol}(E(V)) = 0$, we want to have

$$\mathbf{1}^T L \mathbf{1} = \mathrm{vol}(E(V)) = 0. \tag{3.4}$$

Since $L\mathbf{1} = \mathbf{0}$ implies $\mathbf{1}^T L \mathbf{1} = 0$, let us first see if assuming $L\mathbf{1} = \mathbf{0}$ works. Since

$$\begin{aligned} (L\mathbf{1})_i &= \ell_{ii} + \sum_{j \in V,\, j \sim i,\, j \neq i} \ell_{ij} + \sum_{j \in V,\, j \not\sim i} \ell_{ij} \\ &= \Big(\sum_{j \in V,\, j \sim i} \mathbf{w}_{ij} \Big) - g_i + \sum_{j \in V,\, j \sim i,\, j \neq i} \ell_{ij} + \sum_{j \in V,\, j \not\sim i} \ell_{ij}, \end{aligned}$$

a natural way to make these terms on the last line cancel out is to let $\ell_{ij} = -\mathbf{w}_{ij}$ if $j \sim i$ and $\ell_{ij} = 0$ otherwise. Now, we have decided all entries of L. We can verify that L satisfies our requirement in (3.1). Below, we give a definition of Laplacian along with some of its properties.

Definition 3.5. The (combinatorial) Laplacian L of a graph G is defined by

$$L = D - A.$$

Proposition 3.6. *The Laplacian L satisfies:*

(a)

$$L = \sum_{\{i,j\} \in E} \mathbf{w}_{ij} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T. \tag{3.7}$$

(b) L has eigevalues $0 = \xi_1 \leq \xi_2 \leq \cdots \leq \xi_n$.

(c) In fact, for a connected graph G, $\xi_1 < \xi_2$, i.e. the eigenspace of $\xi_1 = 0$ is simple, and moreover it is spanned by $\mathbf{1}$.

Proof. (a) can be checked by direct computation.

Since A and D are symmetric, L is symmetric thus has real eigenvalues $\xi_1 \leq \cdots \leq \xi_n$. Let $\mathbf{s}_1, \ldots, \mathbf{s}_n$ be a corresponding orthonormal eigenbasis. By (3.7), for any $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x}^T L \mathbf{x} = \mathbf{x}^T \left(\sum_{\{i,j\} \in E} \mathbf{w}_{ij} (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \right) \mathbf{x} = \sum_{\{i,j\} \in E} \mathbf{w}_{ij} (\mathbf{x}_i - \mathbf{x}_j)^2 \geq 0. \tag{3.8}$$

Then by Theorem 2.4, for $1 \leq i \leq n$

$$\xi_i = \frac{\mathbf{s}_i^T L \mathbf{s}_i}{\mathbf{s}_i^T \mathbf{s}_i} \geq 0.$$

Moreover, we know $L\mathbf{1} = 0$ thus 0 is an eigenvalue of L.
Finally, let $L\mathbf{x} = 0$. Then $\mathbf{x}^T L\mathbf{x} = 0$ and the inequality in (3.8) is tight, i.e. $\mathbf{w}_{ij}(\mathbf{x}_i - \mathbf{x}_j)^2 \geq 0$ for all $\{i, j\} \in E$. Since $\mathbf{w}$ is strictly positive and since G is connected, this means that all coordinates of $\mathbf{x}$ are of the same value. Therefore the null space of L is spanned by $\mathbf{1}$. □

3.1.2. *Normalized Laplacian and the easy side of Cheeger.*

Now, we try to express the denominator $\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}$ in $\phi(S)$ as a quadratic form of a positive diagonal matrix. This section is mostly based on a generalization of an argument in [8].

It is easy to express $\mathrm{vol}(S)$ as $\mathbf{x}^T D\mathbf{x}$. However, we would have trouble expressing the minimization by a quadratic form. Thus we first do a slight relaxation to make the denominator easier to deal with. Since $\mathrm{vol}(S) + \mathrm{vol}(\overline{S}) = \phi(G)$, we know $\max\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\} \geq \mathrm{vol}(G)/2$, thus

$$\phi(S) = \frac{|E(S)|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}} \geq \frac{\mathrm{vol}(G)}{2} \frac{|E(S)|}{\mathrm{vol}(S)\,\mathrm{vol}(\overline{S})}. \tag{3.9}$$

Now, we want to express $\dfrac{\mathrm{vol}(G)}{2} \dfrac{|E(S)|}{\mathrm{vol}(S)\,\mathrm{vol}(\overline{S})}$ as a quadratic form of $\mathbf{x}$.
For ease of notation, we have the following definition.

Definition 3.10. The function $\mathbb{1} : \{\text{statements}\} \to \{0, 1\}$ is defined as follows: for a statement p, $\mathbb{1}(p) = 1$ if p is true, and $\mathbb{1}(p) = 0$ otherwise.

Then,

$$\begin{aligned}
\mathrm{vol}(S)\,\mathrm{vol}(\overline{S}) &= \left(\sum_{i \in V} \deg(i) \mathbb{1}(i \in S)\right) \left(\sum_{j \in V} \deg(j) \mathbb{1}(j \in \overline{S})\right) \\
&= \sum_{i,j \in V} \deg(i) \deg(j) \mathbb{1}(i \in S,\ j \in \overline{S}) \\
&= \sum_{i<j,\ i,j \in V} \deg(i) \deg(j) \mathbb{1}(\mathbf{x}_i \neq \mathbf{x}_j) \\
&= \sum_{i<j,\ i,j \in V} \deg(i) \deg(j) (\mathbf{x}_i - \mathbf{x}_j)^2. \qquad (3.11)
\end{aligned}$$

This is still not quite a quadratic form of a positive diagonal matrix, which should look like $\sum_{i \in V} a_i \mathbf{x}_i^2$ with $a_i > 0$. Therefore, we want the terms $\mathbf{x}_i \mathbf{x}_j$ to cancel out

for $i \neq j$ when we expand (3.11). Since

$$\begin{aligned}
&\sum_{i<j,\ ,i,j\in V} \deg(i)\deg(j)(\mathbf{x}_i - \mathbf{x}_j)^2 \\
&= \sum_{i\in V} \deg(i)\mathbf{x}_i \left(\sum_{j\in V,\ j\neq i} -\deg(j)\mathbf{x}_j \right) + \sum_{i\in V} \deg(i)\mathbf{x}_i^2 \left(\sum_{j\in V,\ j\neq i} \deg(j) \right) \\
&= \sum_{i\in V} \deg(i)\mathbf{x}_i \left(-\deg(i)\mathbf{x}_i + \sum_{j\in V,\ j\neq i} -\deg(j)\mathbf{x}_j \right) + \sum_{i\in V} \deg(i)\mathbf{x}_i^2 \left(\deg(i) + \sum_{j\in V,\ j\neq i} \deg(j) \right) \\
&= \sum_{i\in V} \deg(i)\mathbf{x}_i \left(\sum_{j\in V} -\deg(j)\mathbf{x}_j \right) + \mathrm{vol}(G) \sum_{i\in V} \deg(i)\mathbf{x}_i^2,
\end{aligned}$$

we would like $\sum_{j\in V} \deg(j)\mathbf{x}_j = \mathbf{1}^T D\mathbf{x}$ to be zero. However, this is not necessarily true. Thus we need to modify $\mathbf{x}$ in some way that does not affect the numerator to allow this to hold. Since we established in Prop. 3.6 that $L\mathbf{1} = 0$, we can let

$$\widetilde{\mathbf{x}} = \mathbf{x} - c\mathbf{1}$$

for some $c \in \mathbb{R}$, and the numerator of $\phi(S)$ will not change, i.e., $\mathbf{x}^T L\mathbf{x} = \widetilde{\mathbf{x}}^T L\widetilde{\mathbf{x}}$.

Solving for

$$\mathbf{1}^T D\widetilde{\mathbf{x}} = 0 \quad \text{and} \quad \widetilde{\mathbf{x}} = \mathbf{x} - c\mathbf{1},$$

we get

$$c = \frac{\mathbf{1}^T D\mathbf{x}}{\mathrm{vol}(G)}, \tag{3.12}$$

and the denominator of our quotient becomes

$$\mathrm{vol}(G) \sum_{i\in V} \deg(i)\widetilde{\mathbf{x}}_i^2 = \mathrm{vol}(G)\widetilde{\mathbf{x}}^T D\widetilde{\mathbf{x}} \tag{3.13}$$

Continuing from (3.9), taking the minimum and relaxing to real vectors, we have

$$\begin{aligned}
\phi(G) &\geq \min_{S\subset V} \frac{\mathrm{vol}(G)}{2} \cdot \frac{\mathrm{vol}(E(S))}{\mathrm{vol}(S)\,\mathrm{vol}(\overline{S})} \\
&= \min_{\mathbf{x}\in\{0,1\}^n - \{\mathbf{0},\mathbf{1}\}} \frac{\mathrm{vol}(G)}{2} \cdot \frac{\widetilde{\mathbf{x}}^T L\widetilde{\mathbf{x}}}{\mathrm{vol}(G)\widetilde{\mathbf{x}}^T D\widetilde{\mathbf{x}}} \quad \text{where } \widetilde{\mathbf{x}} = \mathbf{x} - \frac{\mathbf{1}^T D\mathbf{x}}{\mathrm{vol}(G)}\mathbf{1} \\
&\geq \min_{\mathbf{x}\in\mathbb{R}^n,\ \mathbf{x}\neq\mathbf{0},\mathbf{1}} \frac{1}{2}\frac{\widetilde{\mathbf{x}}^T L\widetilde{\mathbf{x}}}{\widetilde{\mathbf{x}}^T D\widetilde{\mathbf{x}}} \\
&= \min_{\mathbf{1}^T D\widetilde{\mathbf{x}}=0,\ \widetilde{\mathbf{x}}\neq\mathbf{0}} \frac{1}{2}\frac{\widetilde{\mathbf{x}}^T L\widetilde{\mathbf{x}}}{\widetilde{\mathbf{x}}^T D\widetilde{\mathbf{x}}} \\
&= \min_{\mathbf{1}^T D\widetilde{\mathbf{x}}=0,\ \widetilde{\mathbf{x}}\neq\mathbf{0}} \frac{1}{2} R(L, D, \mathbf{x}). \qquad (3.14)
\end{aligned}$$

The second to last equality follows from the fact that we obtained $\widetilde{\mathbf{x}}$ from $\mathbf{x}$ by subtracting from $\mathbf{x}$ its projection onto $\mathrm{span}(D\mathbf{1})$.

Now we can recognize that the generalized Rayleigh quotient in (3.14) corresponds to the second smallest eigenvalue of the matrix $D^{-1/2}LD^{-1/2}$. We define this matrix below and verify this property among others.

Definition 3.15. The normalized Laplacian $\mathcal{L}$ of G is defined by

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}.$$

Proposition 3.16. *The normalized Laplacian $\mathcal{L}$ satisfies:*
(a)

$$\mathcal{L} = \sum_{\{i,j\}\in E} \mathbf{w}_{ij} \left(\frac{\mathbf{e}_i}{\sqrt{\deg(i)}} - \frac{\mathbf{e}_j}{\sqrt{\deg(j)}} \right) \left(\frac{\mathbf{e}_i}{\sqrt{\deg i}} - \frac{\mathbf{e}_j}{\sqrt{\deg j}} \right)^T. \tag{3.17}$$

(b) $\mathcal{L}$ is symmetric with eigevalues $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le 2$.
(c) In fact $\lambda_1 < \lambda_2$. That is, the eigenspace of $\lambda_1 = 0$ is simple, and moreover it is spanned by $D^{1/2}\mathbf{1}$.
(d)

$$\lambda_2 = \min_{\mathbf{1}^T D\mathbf{y}=0,\, \mathbf{y}\neq\mathbf{0}} \frac{\mathbf{y}^T L \mathbf{y}}{\mathbf{y}^T D \mathbf{y}}. \tag{3.18}$$

Moreover, λ_2 is obtained when $\mathbf{y} = D^{-1/2}\mathbf{v}_2$ where $\mathbf{v}_2$ is an eigenvector to λ_2.

Proof. (a) follows from the definition of $\mathcal{L}$ and (3.8). (b), except for the upper bound, and (c) can be proved by the same reasonings as in Prop. 3.6.
(d) follows from Cor. 2.8 and (c), since the constraint in (2.9) now corresponds to $(D^{1/2}\mathbf{1})^T D^{1/2}\mathbf{y} = 0$, $\mathbf{y} \neq \mathbf{0}$.
To see that $\lambda_n \le 2$, we use the fact that for $a, b \in \mathbb{R}$

$$(a-b)^2 \le 2(a^2+b^2).$$

Then, by Cor. 2.8 and (3.7),

$$\lambda_n = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T D \mathbf{x}} = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\sum_{\{i,j\}\in E} (\mathbf{x}_i - \mathbf{x}_j)^2}{\sum_{i=1}^n \deg(i)\mathbf{x}_i^2} \le 2.$$

□

We note that analogous statements as (3.18) holds for the other eigenvalues for $\mathcal{L}$ as well. We remark that $\mathcal{L}$ is *normalized* with respect to the degree matrix. More properties of the spectrum of normalized Laplacian can be found in Ch.1 of [3].

Continuing from (3.14), we have proved the following.

Lemma 3.19. *For a graph G,*

$$\phi(G) \ge \frac{1}{2}\lambda_2$$

where λ_2 is the second smallest eigenvalue of the normalized Laplacian.

3.2. **Lower bounding the relaxation and the hard side of Cheeger.**

To conclude from Lemma 3.19 a useful approximation of $\phi(G)$, we need to lower bound the relaxed quantity. In other words, we need a statement in the form of

$$\lambda_2 \overset{?}{\ge} f(\phi(G)) \tag{3.20}$$

for some function f. Equivalently, we need

$$h\left(\min_{\mathbf{1}^T D\mathbf{y}=0,\, \mathbf{y}\neq\mathbf{0}} R(L, D, \mathbf{y}) \right) \overset{?}{\ge} \min_{S\subset V} \phi(S) \tag{3.21}$$

for some functoin h. Both sides are related to optimums of minimization problems, but we usually leverage the optimality on the side which we want to be smaller.

More specifically, we want to produce a cut S from any $\mathbf{y}$ satisfying the constraint of the left hand side, and relate $R(L, D, \mathbf{y})$ and $\phi(S)$ in an inequality in the correct direction using the connection between S and $\mathbf{y}$. In particular, the optimum of the left hand side would be related with some $\phi(S)$ in an inequality in the correct direction. Since $\phi(G)$ can only be smaller than $\phi(S)$, the inequality we want would hold.

It is a common technique well motivated in practice to try to obtain such a cut S from a vector $\mathbf{y}$ by *sweep cuts* on $\mathbf{y}$. Informally, this means that for $1 \leq i \leq n$, we look at the cut defined by the vertices corresponding to the first i largest coordinates of $\mathbf{y}$, and prove that one of these cuts satisifes our requirement. In this way, we decrease the number of cuts to look at from 2^n to n.

For the problem of finding the conductance of a graph, we can indeed prove such a bound by looking at sweep cuts of vectors. There is an elegant and general proof of this result in [11]. We would like to direct the reader to the original source. However, for completeness we include the statement and proof from [11] below with some change of notation and slight elaboration.

Before we start, we note that we can look at examples to see how good a function $h(\cdot)$ we can get. For example, for paths, λ_2 is of order $1/n^2$ and the conductance is of order $1/n$, so we know we have to at least lose a square-root factor in $h(\cdot)$. This might serve as a motivation for the use of Cauchy-Schwartz inequality in the proof.

Theorem 3.22. *Let $\mathbf{y} \in \mathbb{R}^n$ be such that $\mathbf{1}^T D\mathbf{y} = 0$. Then, there is a number t for which the set $S = \{i \in V : \mathbf{y}_i < t\}$ satisfies*

$$\phi(S) \leq \sqrt{2\frac{\mathbf{y}^T L\mathbf{y}}{\mathbf{y}^T D\mathbf{y}}}. \tag{3.23}$$

Before proving the theorem, we note that the the denominator $\mathbf{y}^T D\mathbf{y}$ is minimized with regards to shifts when $\mathbf{1}^T D\mathbf{y} = 0$.

Lemma 3.24. *Let $\mathbf{y}$ be such that $\mathbf{1}^T D\mathbf{y} = 0$. Then for all $c \in \mathbb{R}$,*

$$\mathbf{y}^T D\mathbf{y} \leq (\mathbf{y} + c\mathbf{1})^T D(\mathbf{y} + c\mathbf{1}).$$

Proof. The derivative of

$$(\mathbf{y} + c\mathbf{1})^T D(\mathbf{y} + c\mathbf{1})$$

with respect to c is

$$\mathbf{1}^T 2D(\mathbf{y} + c\mathbf{1}) = \mathbf{1}^T D\mathbf{1} \cdot s,$$

(see e.g. [10]) which is zero when $s = 0$ □

Proof of Theorem 3.22. Assume without loss of generality that

$$\mathbf{y}_1 \leq \mathbf{y}_2 \leq \cdots \leq \mathbf{y}_n.$$

We begin with some normalization. Let j be the least number for which

$$\sum_{i=1}^{j} \deg(i) \geq \frac{\mathrm{vol}(G)}{2}.$$

We would prefer a vector that is centered at j. So, we set

$$\mathbf{z} = \mathbf{y} - \mathbf{y}_j \mathbf{1}.$$

$\mathbf{z}$ satisfies that $\mathbf{z}_j = 0$ and $\mathbf{z}^T L\mathbf{z} = \mathbf{y}^T L\mathbf{y}$. Then, by Lemma 3.24 and our assumption on $\mathbf{y}$,

$$R(L, D, \mathbf{z}) = \frac{\mathbf{z}^T L\mathbf{z}}{\mathbf{z}^T D\mathbf{z}} \leq \frac{\mathbf{z}^T L\mathbf{z}}{\mathbf{y}^T D\mathbf{y}} = R(L, D, \mathbf{y}). \tag{3.25}$$

Assume $\mathbf{z}$ is scaled so that

$$\mathbf{z}_1^2 + \mathbf{z}_n^2 = 1.$$

We will define a distribution on subsets S of vertices, for which we can prove that

$$\mathbb{E}[\mathrm{vol}(E(S))] \leq \sqrt{2R(L, D, \mathbf{y})}\mathbb{E}[\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}].$$

Then,

$$\mathbb{E}[\sqrt{2R(L, D, \mathbf{y})}\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\} - \mathrm{vol}(E(S))] \geq 0,$$

and it follows that there is some S for which

$$\mathrm{vol}(E(S)) \leq \sqrt{2R(L, D, \mathbf{y})}\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}.$$

Then,

$$\phi(S) \leq \sqrt{R(L, D, \mathbf{y})}$$

and we are done.

The distribution on S will be induced from a distribution on a random number t, by setting $S = \{i : \mathbf{z}_i < t\}$. We choose t between $\mathbf{z}_1$ and $\mathbf{z}_n$ with probability density $2|t|$. That is, the probability that t lies in the interval $[a, b]$ is

$$\int_{t=a}^{b} 2|t| = \mathrm{sgn}(b)b^2 - \mathrm{sgn}(a)a^2. \tag{3.26}$$

To see that the total probability is a, we observe that

$$\int_{t=\mathbf{z}_1}^{\mathbf{z}_n} 2|t| = \mathbf{z}_n^2 + \mathbf{z}_1^2 = 1,$$

as $\mathbf{z}_1 \leq \mathbf{z}_j \leq \mathbf{z}_n$ and $\mathbf{z}_j = 0$.

We now derive a formula for the denominator of $\phi(S)$. First, we observe that

$$\mathbb{E}[\mathrm{vol}(S)] = \sum_{i \in V} \Pr[i \in S]\deg(i) = \sum_{i \in V} \Pr[\mathbf{z}_i \leq t]\deg(i).$$

The result of our centering $\mathbf{z}$ at j is that

$$t < 0 \Rightarrow \mathrm{vol}(S) = \min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}, \quad \text{and}$$
$$t \geq 0 \Rightarrow \mathrm{vol}(\overline{S}) = \min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}.$$

That is, for $i < j$, i is in the smaller set if $t < 0$; and for $i \geq j$, i is in the smaller set if $t \geq 0$. So,

$$\begin{aligned}
\mathbb{E}[\min\{\mathrm{vol}(S), \mathrm{vol}(\overline{S})\}] &= \sum_{i<j} \Pr[\mathbf{z}_i < t \text{ and } t < 0]\deg(i) + \sum_{i \geq j} \Pr[\mathbf{z}_i > t \text{ and } t \geq 0]\deg(i) \\
&= \sum_{i<j} \mathbf{z}_i^2 \deg(i) + \sum_{i \geq j} \mathbf{z}_i^2 \deg(i) \\
&= \sum_{i \in V} \mathbf{z}_i^2 \deg(i) \\
&= \mathbf{z}^T D\mathbf{z}.
\end{aligned}$$

We now turn to the numerator. An edge $\{i,k\}$ with $\mathbf{z}_i \le \mathbf{z}_k$ is on the boundary of S if

$$\mathbf{z}_i \le t < \mathbf{z}_k.$$

By (3.26),

$$\operatorname{sgn}(\mathbf{z}_k)\mathbf{z}_k^2 - \operatorname{sgn}(\mathbf{z}_i)\mathbf{z}_i^2 = \begin{cases} |\mathbf{z}_i^2 - \mathbf{z}_k^2| & \text{when } \operatorname{sgn}(i) = \operatorname{sgn}(k), \\ \mathbf{z}_i^2 + \mathbf{z}_k^2 & \text{when } \operatorname{sgn}(i) \neq \operatorname{sgn}(k). \end{cases}$$

We now show that both of these terms are upper bounded by

$$|\mathbf{z}_i - \mathbf{z}_k|(|\mathbf{z}_i| + |\mathbf{z}_k|).$$

Regardless of the signs,

$$|\mathbf{z}_i^2 - \mathbf{z}_k^2| = |(\mathbf{z}_i - \mathbf{z}_k)(\mathbf{z}_i + \mathbf{z}_k)| \le |\mathbf{z}_i - \mathbf{z}_k|(|\mathbf{z}_i| + |\mathbf{z}_k|).$$

When $\operatorname{sgn}(\mathbf{z}_i) = -\operatorname{sgn}(\mathbf{z}_k)$,

$$\mathbf{z}_i^2 + \mathbf{z}_k^2 \le (\mathbf{z}_i - \mathbf{z}_k)^2 = |\mathbf{z}_i - \mathbf{z}_k|(|\mathbf{z}_i| + |\mathbf{z}_k|).$$

So,

$$\begin{aligned} \mathbb{E}[\operatorname{vol}(E(S))] &= \sum_{\{i,k\}\in E} \Pr[\{i,k\} \in E(S)]\mathbf{w}_{ik} \\ &\le \sum_{\{i,k\}\in E} |\mathbf{z}_i - \mathbf{z}_k|(|\mathbf{z}_i| + |\mathbf{z}_k|)\mathbf{w}_{ik} \\ &\le \sqrt{\sum_{\{i,k\}\in E} \mathbf{w}_{ik}(\mathbf{z}_i - \mathbf{z}_k)^2}\sqrt{\sum_{\{i,k\}\in E} \mathbf{w}_{ik}(|\mathbf{z}_i| + |\mathbf{z}_k|)^2} \quad \text{by Cauchy-Schwartz.} \end{aligned}$$

By (3.7) and (3.25), the term under the left-hand square root satisfies

$$\sum_{\{i,k\}\in E} \mathbf{w}_{ik}(\mathbf{z}_i - \mathbf{z}_k)^2 = \mathbf{z}^T L\mathbf{z} \le R(L, D, \mathbf{y}) \cdot (\mathbf{z}^T D\mathbf{z}).$$

To bound the right-hand square root, we observe that

$$\sum_{\{i,k\}\in E} \mathbf{w}_{ik}(|\mathbf{z}_i|+|\mathbf{z}_k|)^2 \le 2\sum_{\{i,k\}\in E} (\sqrt{\mathbf{w}_{ik}}\mathbf{z}_k)^2 + (\sqrt{\mathbf{w}_{ik}}\mathbf{z}_i)^2 = 2\sum_{i\in V} \mathbf{z}_i^2 \deg(i) = 2\mathbf{z}^T D\mathbf{z}.$$

Putting these inequalities together yields

$$\begin{aligned} \mathbb{E}[\operatorname{vol}(E(S))] &\le \sqrt{R(L, D, \mathbf{y}) \cdot (\mathbf{z}^T D\mathbf{z})}\sqrt{2\mathbf{z}^T D\mathbf{z}} \\ &= \sqrt{2R(L, D, \mathbf{y})}\mathbf{z}^T D\mathbf{z} \\ &= \sqrt{2R(L, D, \mathbf{y})}\mathbb{E}[\min\{\operatorname{vol}(S), \operatorname{vol}(\overline{S})\}]. \end{aligned}$$

□

We declare again that the elegant proof above is mostly verbatim from [11].

Let $\mathbf{v}_2$ be an eigenvector to the second smallest eigenvalue λ_2 of normalized Laplacian, and let $\tilde{\mathbf{v}} = D^{-1/2}\mathbf{v}_2$. By Prop. 3.16 (d), we know $\mathbf{1}^T D\tilde{\mathbf{v}} = 0$ and $R(L, D, \tilde{\mathbf{v}}) = \lambda_2$. Then, by Theorem 3.22, there is a set S such that

$$\phi(G) \le \phi(S) \le \sqrt{2R(L, D, \tilde{\mathbf{v}}} = \sqrt{2\lambda_2}.$$

We have finished the proof of the following theorem.

Theorem 3.27 (Cheeger's inequality)**.** *For a graph G,*

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2} \tag{3.28}$$

where λ_2 is the second smallest eigenvalue to the normalized Laplacian.

That is, we showed that λ_2 can approximate $\phi(G)$ within a square root factor. Moreover, Theorem 3.22 gives us an algorithm to find a set S whose conductance is within this approximation bound: we simply looking at the sweep cuts of $D^{-1/2}\mathbf{v}_2$ and use the cut with minimum conductance among them.

A natural question to ask is: how good is algorithm? We briefly discuss this point.

Firstly, Cheeger's inequality is asymptotically tight, in the sense that there are families of graphs whose conductance is of order λ_2 (e.g., complete binary trees), as well as families of graphs whose conductance is of order $\sqrt{\lambda_2}$ (e.g, paths). For examples on the values of λ_2 for various graphs, one can refer to e.g. Ch.1.2 in [3].

Secondly, the proof we presented in the previous section gives one among many algorithms and heuristics to obtain a low-conductance cut from an eigenvector. The cuts produced by these algorithms on specific graphs may have different quality with respect to sparsity, balance, etc. For a discussion on this, one can refer to e.g. [4].

4. Local low-conductance sets and random walks on a graph

In the previous sections, we have been discussing a global algorithm for a global quantity, in the sense that our algorithm requires knowledge of the whole graph, and we obtain a cut that approximately attains the minimum conductance in the whole graph. However, this may not be relevant in practice, espcially when we are dealing with large graphs.

As an example, an important class of problems in network sicence is to discover communities, i.e., sets of nodes that are densely connected internally. This can be modelled by low-conductance sets in a graph. However, the global approach we presented might be impractical or unhelpful. In massive networks, it may take too long to even process the whole graph in linear time. Moreover, often we are only interested in a local region of the graph, whose structure may not be related to the cut that attains the minimum conductance.

A particular problem people are interested in in this context is to find a local community a queried node belongs to. This corresponds to finding a low-conductance set among local sets around a vertex, i.e., sets of bounded size and consisting of vertices with short distance to this vertex. (In our discussion, we measure distance by adjacency relations.)

We note that in this case, it is not meaningful to discuss the balance of the cut as we are asking for sets much smaller than its complement. However, it still makes sense to measure the sparsity of cuts by conductance, since now the conductance is consistly the ratio between cut capacity and the volume of our set.

To efficiently find a local low-conductance set, we consider *local algorithms* , i.e., algorithms that access one vertex and its neighbors at each step. We can apply local algorithms to gain information on the local region we care about, so that the running time is related to the size of the local region instead of that of the whole graph.

A major class of local algorithms are based on ideas of random walks. In fact, random walks have rich applications in many problems related to graphs; we refer the reader to [2] for a survey on this topic. Below, we give an introduction to random walk related methods on graph partitioning.

4.1. **Preliminaries about random walks.**

First, we give definitions and state some basic results.

Definition 4.1. Given a graph G, a *probability distribution* on G is a vector $\mathbf{r} \in \mathbb{R}^n$ with $\mathbf{r}_i \geq 0$ for all i, and $\sum_{i=1}^n \mathbf{r}_i = 1$. A *transition probability matrix* on G is an $n \times n$ matrix $P = (p_{ij})$ with $p_{ij} \geq 0$ for all i, j, and $\sum_{i=1}^n p_{ij} = 1$ for all j.

Then, given an initial probability distribution $\mathbf{r}^{(0)}$, the probability distribution at step $t+1$ is

$$\mathbf{r}^{(t+1)} = P\mathbf{r}^t = P(P\mathbf{r}^{t-1}) = \cdots = P^t\mathbf{r}^{(0)}. \tag{4.2}$$

Definition 4.3. A *stationary distribution* $\overline{\mathbf{r}}$ for a transition probability matrix P is a probability distribution such that

$$P\overline{\mathbf{r}} = \overline{\mathbf{r}}. \tag{4.4}$$

We note that in literature for stochastic processes left multiplication is often used, in which case the transition probability matrix is the transpose of what we define here.

A random walk on a graph G is a process where given an initial probability distribution on G, we go randomly from one vertex to its neighbors with a transition probability determined by edge weights. More precisely, given a current probability distribution $\mathbf{q}^{(t)}$, for each $i \in V$ and each neighbor j of i, we go from i to j with probability

$$\mathbf{q}_i^{(t)}\mathbf{w}_{ij}/\deg(i). \tag{4.5}$$

The transition probability matrix corresponding to this process is

$$AD^{-1}.$$

We want to learn about the structure of our graph by observing the process of random walks on it. We usually start at one vertex, i.e., $\mathbf{q}^{(0)} = \mathbf{e}_i$ for some i, and see how quickly the probability mass dissipate among all vertices. If the graph is well-connected, this process will be fast. Meanwhile, if i is contained in a set with low conductance, the probability mass will be "trapped" within this set, and the process will be slow.

For this purpose, we would like the process to converge toward one stationary distribution. However, if we consider a bipartite graph, (e.g., a graph consists of two vertices and one edge between them,) then the probability distribution for the random walk does not necessarily converge, because the two sets that partition the set of vertices exchange their sum of probability at each step.

To circumvent this, we define lazy random walks on a graph, where given an initial probability distribution on G, at each step we perform the random walk with probability $1/2$ and stay put with probability $1/2$. A lazy random walk can also be interpreted as a random walk on a graph after preprocessing, during which we add a self-loop of weight $\deg(i)$ at each vertex i.

The lazy random walk corresponds to the following transition probability matrix.

Definition 4.6. The transition probability matrix W for lazy random walks on a graph is

$$W = \frac{1}{2}(I + AD^{-1}). \tag{4.7}$$

Theorem 4.8. *A lazy random walk on a (connected) graph G converges to a unique stationary distribution π where*

$$\pi_i = \frac{\deg(i)}{\mathrm{vol}(G)}. \tag{4.9}$$

In other words, $\pi = D\mathbf{1}/\mathrm{vol}(G)$ is the unique solution to

$$\pi = W\pi. \tag{4.10}$$

This follows from the more general fact that an irreducible, aperiodic Markov chain has a unique stationary distribution. For definitions of these terms and a sketched proof of this theorem, see e.g. Ch.1 in [7].

4.2. **A convergence bound via a spectral relation.**

Suppose we perform a lazy random walk on a graph starting from a vertex i. Let $\mathbf{p}^{(t)}$ be the probability distribution at step t. From (4.2), we know that

$$\mathbf{p}^{(t)} = W^t\mathbf{e}_i. \tag{4.11}$$

That is, for information on the process of the walk, we need to look at powers of W. For this purpose, we would like W to have an orthonormal eigenbasis in the reals. We can prove that this is true by observing its connection with normalized Laplacian:

$$D^{-1/2}WD^{1/2} = \frac{1}{2}(I + D^{-1/2}AD^{-1/2}) = \frac{1}{2}(I + (I - \mathcal{L})) = I - \frac{1}{2}\mathcal{L}. \tag{4.12}$$

That is, W is similar to the symmetric matrix $I - \frac{1}{2}\mathcal{L}$, whose eigenvalues are

$$1 \geq (1 - \frac{1}{2}\lambda_2) \geq \cdots \geq (1 - \frac{1}{2}\lambda)_n. \tag{4.13}$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be an orthonormal eigenbasis of $\mathcal{L}$. Then the spectral decompostion of $I - \frac{1}{2}\mathcal{L}$ is

$$I - \frac{1}{2}\mathcal{L} = \sum_{i=1}^{n}(1 - \frac{1}{2}\lambda_i)\mathbf{v}_i\mathbf{v}_i^T. \tag{4.14}$$

Therefore,

$$D^{-1/2}W^tD^{1/2} = (1 - \frac{1}{2}\mathcal{L})^t = \left(\sum_{i=1}^{n}(1 - \frac{1}{2}\lambda_i)\mathbf{v}_i\mathbf{v}_i^T\right)^t = \sum_{i=1}^{n}(1 - \frac{1}{2}\lambda_i)^t\mathbf{v}_i\mathbf{v}_i^T. \tag{4.15}$$

Thus

$$W^t = \sum_{i=1}^{n}(1 - \frac{1}{2}\lambda_i)^tD^{1/2}\mathbf{v}_i\mathbf{v}_i^TD^{-1/2}. \tag{4.16}$$

Now, we can obtain information on what $\mathbf{p}^{(t)}$ looks like.

Proposition 4.17. *On a graph G, after t steps of a lazy random walk starting at vertex k, for every vertex j,*

$$|\mathbf{p}_j^{(t)} - \pi_j| = |(W^t\mathbf{e}_k)_j - \pi_j| \leq \sqrt{\frac{\deg(j)}{\deg(k)}}\left(1 - \frac{\lambda_2}{2}\right)^t. \tag{4.18}$$

Proof. First, we note that by Prop. 3.16, the first term in the spectral decomposition (4.16) of W^t satisfies

$$\begin{aligned}(1-\frac{1}{2}\lambda_1)^t D^{1/2}\mathbf{v}_1\mathbf{v}_1^T D^{-1/2} &= 1 \cdot D^{1/2}\frac{(D^{1/2}\mathbf{1})(D^{1/2}\mathbf{1})^T}{||D^{1/2}\mathbf{1}||^2}D^{-1/2}\\ &= \frac{D\mathbf{1}\mathbf{1}^T}{(D^{1/2}\mathbf{1})^T(D^{1/2}\mathbf{1})}\\ &= \frac{D\mathbf{1}\mathbf{1}^T}{\mathrm{vol}(G)}.\end{aligned}$$

Since

$$\frac{D\mathbf{1}\mathbf{1}^T}{\mathrm{vol}(G)}\mathbf{e}_k = \pi, \tag{4.19}$$

We have that

$$W^t\mathbf{e}_k - \pi = \sum_{i=2}^{n}(1-\frac{1}{2}\lambda_i)^t D^{1/2}\mathbf{v}_i\mathbf{v}_i^T D^{-1/2}. \tag{4.20}$$

Thus

$$\begin{aligned}|\mathbf{p}_j^{(t)} - \pi_j| &= \left|\mathbf{e}_j^T\left(\sum_{i=2}^{n}(1-\frac{1}{2}\lambda_i)^t D^{1/2}\mathbf{v}_i\mathbf{v}_i^T D^{-1/2}\right)\mathbf{e}_k\right|\\ &= \left|\sum_{i=2}^{n}(1-\frac{1}{2}\lambda_i)^t\sqrt{\deg(j)}\mathbf{e}_j^T\mathbf{v}_i\mathbf{v}_i^T\frac{\mathbf{e}_k}{\sqrt{\deg(k)}}\right|\\ &\leq \sqrt{\frac{\deg(j)}{\deg(k)}}||\mathbf{e}_j^T||\sqrt{\left(\sum_{i=2}^{n}(1-\frac{1}{2}\lambda_i)^t\mathbf{v}_i\mathbf{v}_i^T\mathbf{e}_k\right)^T\left(\sum_{i=2}^{n}(1-\frac{1}{2}\lambda_i)^t\mathbf{v}_i\mathbf{v}_i^T\mathbf{e}_k\right)}\\ &\quad \text{by submultiplicativity of induced norms}\\ &= \sqrt{\frac{\deg(j)}{\deg(k)}}\sqrt{\sum_{i=2}^{n}(1-\frac{1}{2}\lambda_i)^{2t}(\mathbf{v}_i^T\mathbf{e}_k)^2}\\ &\leq \sqrt{\frac{\deg(j)}{\deg(k)}}\sqrt{\sum_{i=2}^{n}(1-\frac{1}{2}\lambda_2)^{2t}(\mathbf{v}_i^T\mathbf{e}_k)^2} \quad \text{since } 0<\lambda_2\leq\lambda_i\leq 2 \text{ by Prop. 3.16}\\ &= \sqrt{\frac{\deg(j)}{\deg(k)}}(1-\frac{1}{2}\lambda_2)^t\sqrt{\sum_{i=2}^{n}(\mathbf{v}_i^T\mathbf{e}_k)^2}\\ &\leq \sqrt{\frac{\deg(j)}{\deg(k)}}(1-\frac{1}{2}\lambda_2)^t.\end{aligned}$$

The last step follows from the fact that transformation by an orthogonal matrix preserves norm, or more explicitly,

(4.21)
$$1 = ||\mathbf{e}_k||^2 = \|\sum_{i=1}^{n}(\mathbf{v}_i^T\mathbf{e}_k)\mathbf{v}_i\|^2 = \left(\sum_{i=1}^{n}(\mathbf{v}_i^T\mathbf{e}_k)\mathbf{v}_i\right)^T\left(\sum_{i=1}^{n}(\mathbf{v}_i^T\mathbf{e}_k)\mathbf{v}_i\right) = \sum_{i=1}^{n}(\mathbf{v}_i^T\mathbf{e}_k)^2.$$

□

4.3. **An introduction to certifying low conductance using random walks.** We have studied some properties, in particular a convergence bound, of random walks. Now, we want to see how random walks can help us learn about the conductance of cuts and of the graph. As such, we want to relate the random walk process with conductance. Cheeger's inequality provides a way to achieve this.

Combining Cheeger's inequality (3.28) and the convergence bound (4.18), we have that

$$|\mathbf{p}_j^{(t)} - \pi_j| \leq \sqrt{\frac{\deg(j)}{\deg(k)}}\left(1 - \frac{\phi(G)^2}{4}\right)^t. \tag{4.22}$$

We can interpret this result as follows: in a graph whose conductance is unknown to us, we start a random walk at vertex k and run for a certain amount of steps. Then, if some coordinates of the current probability distribution vector are still far that of the stationary distribution, we would know that the right hand side has to be large, and as a result $\phi(G)$ has to be small. However, this is not useful in learning about the local structure around the starting vertex k, as we only get information on $\phi(G)$ which may not be related to local cuts.

Thus, we would like to relate the behavior of random walks starting at a vertex with the conductance of local regions around that vertex. It turns out that such a relation holds, as proved in [1].

To state this result, we make concrete the definition of sweep cuts.

Definition 4.23. Let $\mathbf{y}$ be a probability distribution vector on G. Let $c_1 \geq c_2 \geq \cdots \geq c_n$ be the reordered sequence of the values occuring in the coordinates of $\mathbf{y}$. We define the set $S_{\mathbf{y}}$ of vertices to be such that

$$\phi(S_{\mathbf{y}}) = \min_{1\leq i\leq n}\{\phi(T) \mid T = \{v \in V,\ \mathbf{y}_v \geq c_n\}\}.$$

Meanwhile, for a probability distribution vector $\mathbf{y}$ on G and a subset S of vertices, we define

$$\mathbf{y}(S) = \sum_{i\in S}\mathbf{y}_i. \tag{4.24}$$

Lemma 4.25 (Lovász-Simonovits)**.** *Let u be a vertex in G and let $\mathbf{p}^{(t)}$ be the probability distribution vector at step t of the lazy random walk starting at u. Define*

$$\phi_u^{(t)} = \min_{0\leq i\leq t} \quad \phi(S_{D^{-1}\mathbf{p}^{(i)}}). \tag{4.26}$$

Then, for any subset S of vertices such that $\mathrm{vol}(S) \leq \mathrm{vol}(G)/2$*,*

$$|\mathbf{p}^{(t)}(S) - \pi(S)| \leq \sqrt{\frac{\mathrm{vol}(S)}{\deg(u)}}\left(1 - \frac{(\phi_u^{(t)})^2}{8}\right)^t. \tag{4.27}$$

Intuitively, if k is inside a cluster, the probability distribution would dissipate slowly and for a certain t the left hand side of (4.27) has to be large. Then, we get an upper bound on $\phi_u^{(t)}$ that is small, and we know there is a local cut with low conductance.

The proof of this result uses a concave curve to measure the progress of the random walk, instead of using spectral methods. A clear and motivated presentation of this proof can be a subject of its own, so we do not attempt to include the proof here. We encourage the reader to look at [12] for a well-written note on this result with motivations and explanations in the context of graph partitioning. We note that the difference between the statement of this mixing result in [12] and above comes from the fact that we stated the lemma for lazy random walks, while they are stating the result for a preprocessed graph.

[5] develops an algorithm based on this mixing result to find local low-conductance sets. This algorithm is then applied in [5] to give a first nearly-linear time approximation algorithm for finding a balanced sparsest cut. Informally, this cut is found by piecing together small sets, whose conductance we know from local random walks.

Using the same approach, [6] improves results in [5] by using PageRank random walks instead of random walks. It uses an analogous mixing result as Lemma 4.25 for PageRank vectors.

We do not presume to analyze these algorithms here, but we wish to note the following: To guarantee a fast running time, we need each piece to observe a local running time. However, in a random walk the probability distribution often quickly dissipate to a large region. [5], following [1], deals with this by ignoring vertices with small probability mass and prove that it does not affect the result much. This corresponds to maintaining an approximate PageRank vector with bounded support size in [6].

Finally, we note that although we only presented how to obtain global information from the spectrum and motivated local random walk methods in the context of finding local conductance, in fact spectral methods are closely related with random walk methods and they can also obtain local information, as shown in [9].

Acknowledgments

I would like to thank Prof. J. Peter May for organizing this REU, especially during these difficult times. I thank Pallav Goyal for his help and encouragement during the REU program. I am grateful to Prof. Lorenzo Orecchia, whose guidance shaped my understanding of these subjects discussed. I thank Antares Chen and Erasmo Tani for many helpful discussions. I thank Teo F. Collin, whose nagging and support allowed this paper to take shape.

References

[1] László Lovász, Miklós Simonovits. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. *FOCS'90*, Volume I.

[2] László Lovász. Random walks on graphs. *Combinatorics: Paul Erdős is Eighty*, Volume 2, 1993.

[3] Fan. R. K. Chung. *Spectral Graph Theory.* AMS, 1997.

[4] Stephen Guattery, Gary L. Miller. On the quality of spectral separators. *SIAM J. Matrix Anal. Appl.,* 19(3), pp.701-719.

[5] Daniel A. Spielman, Shang-Hua Teng. Nearly-Linear Time Algorithms for Graph Partitioning, Graph Sparsification, and Solving Linear Systems. *STOC'04*.

[6] Reid Andersen, Fan Chuang, Kevin Lang. Local graph partitioning using PageRank vectors. *FOCS'06*.

[7] Gregory F. Lawler. *Introduction to Stochastic Processes*. Chapman & Hall/CRC, 2006.

[8] Instructor: Jonathan Kelner. Scribe: Andre Wibisono. Lec. 3 in online lecture notes for "Topics in Theoretical Computer Science: An Algorithmist's Toolkit" at MIT, Fall 2009. `https://ocw.mit.edu/courses/mathematics/18-409-topics-in-theoretical-computer-science-an-algorithmists-toolkit-fall-2009/lecture-notes`.

[9] Michael W. Mahoney, Lorenzo Orecchia, Nisheeth K. Vishnoi. A local spectral method for graphs: with applications to improving graph partitions and exploring data graphs locally. *J. Machine Learning Research*, Vol 13, No.1.

[10] Kaare Brandt Petersen, Michael Syskind Pedersen. *The Matrix Cookbook*. `https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`

[11] Daniel A. Spielman. Online lecture notes for "Spectral Graph Theory" at Yale, Fall 2012. `https://www.cs.yale.edu/homes/spielman/561/2012`

[12] Instructor: Amitabha Bagchi. Scribes: Dravyansh Sharma, Sai Praneeth. Lec. 1 in online lecture notes for "Expander graphs and their applications" at IIT Delhi, 2015 `http://www.cse.iitd.ernet.in/~bagchi/courses/CSL863_14-15/CSL863-lovasz-simonovits.pdf`

[13] Luca Trevisan. *Lecture Notes on Graph Partitioning, Expanders and Spectral Methods*. `https://people.eecs.berkeley.edu/~luca/books/expanders-2016.pdf`

[14] Aydın Buluç, Henning Meyerhenke, Ilya Safro, Peter Sanders, Christian Schulz. Recent Advances in Graph Partitioning, pp.117-158 in *Algorithm Engineering*, Springer, 2016. arXiv: `https://arxiv.org/pdf/1311.3144`

[15] Instructor: Lorenzo Orecchia. Scribe: Erasmo Tani. Third lecture in online lecture notes for "Iterative Methods for Graph Algorithms and Network Analysis" at Boston University, Fall 2018. `http://cs-people.bu.edu/orecchia/CS591fa18/course.html`