On Cheeger's Inequality and Its Improvement *

Zhenyu Sun
Department of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
z73sun@uwaterloo.ca

Haomin Li
Department of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
h439li@uwaterloo.ca

ABSTRACT

Mathematically, spectral graph theory is the study of the relationship between graphs and their linear algebra's properties. The Cheeger's inequality is a fundamental and important tool in spectral graph theory. The Cheeger's Inequality state that for any graphs G, we have $\frac{1}{2}\lambda_2 \leq \phi(G) \leq \sqrt{2\lambda_2}$, where $\phi(G)$ is the minimum conductance of an undirected graph G and $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n \leq 2$ are the eigenvalues of the normalized Laplacian matrix of G. In this report, we review most of the Cheeger's Inequality papers on its proofs to show a deeper insight of Cheeger's Inequality. Then we present the improved result from [KLLOT13] which is $\phi(G) = O(k)\frac{\lambda_2}{\sqrt{\lambda_k}}$. We show the proof of the improved result by step approximation function. We try to present the proof in a more approachable way for undergraduate readers by adding omitted proofs and details in the original paper. We introduce a simple application in finding sparsest cut in the end.

*This paper is helped by Professor Lap Chi Lau's advice and suggestions

1 Introduction

This paper surveys some different ways of proving Cheeger's Inequality and its recent improvements. Cheeger's Inequality was first proved by Jeff Chegger[J69] in manifold setting and then extended to graph setting by Alon and Milman[AM85, Alo86]. Cheeger's Inequality in graph setting is an important result in spectral graph theory with applications in spectral partition and clustering. In this paper, we assume the graph are unweighted and $d-$regular while the results shown in this paper hold for arbitrary weighted graphs. Let $G = (V, E)$ be a d-regular undirected graph. We define the $conductance$ of a subset $S \subset V$ to be

$$\phi(S) = \frac{|E(S,\overline{S})|}{d\min\{|S|,|\overline{S}|\}}$$

where $E(S, \overline{S}) = \{uv \in E(G)|u \in S, v \in \overline{S}\}$. The conductance of graph G is defined to be the smallest conductance among all subsets $S \subset V$

$$\phi(G) = \min_{S\subset V} \phi(S)$$

Note that a set of small conductance means a set with less edges crossing from the set to the rest of the graph and such set is called a sparse cut. Finding a sparse cut is an important topic in spectral partition and clustering. Cheeger's Inequality provides a quantitative result between the conductance of a graph and the second eigenvalue of its normalized Laplacian matrix. Here the normalized Laplacian matrix is a $|V| \times |V|$ matrix defined by $\mathcal{L} = I - \frac{1}{d}A$ where A is the adjacency matrix of G. It is well-known that the eigenvalues of $\mathcal{L}$ satisfy that $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|} \leq 2$.

Theorem 1.1 (Cheeger's Inequality)**.** *Let $G = (V, E)$ be a undirected graph, $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|}$ be the eigenvalues of the normalized Laplacian matrix of G, the conductance of G satisfies*

$$\frac{1}{2}\lambda_2 \leq \phi(G) \leq \sqrt{2\lambda_2} \tag{1.1}$$

There are many ways to prove Cheeger's Inequality while some of them are particularly interesting ...[F10].

However, the empirical performance of spectral partitioning is much better than the worst case performance guaranteed provided by Cheeger's Inequality. Some recent research shows that Cheeger's Inequality can be further improved[KLLOT13]. We will present the proof of the improved result.

Theorem 1.2 (Improved Cheeger's Inequality)**.** *For every undirected graph G and any $k \geq 2$, it holds that*

$$\phi(G) = O(k)\frac{\lambda_2}{\sqrt{\lambda_k}} \tag{1.2}$$

The above (1.2) gives a better approximation of $\phi(G)$ by showing λ_2 is a better approximation when there is a large gap between λ_2 and λ_k for $k \geq 3$. We will first introduce all the necessary background knowledge in real analysis and spectral graph theory in Preliminaries and then show the proof of (1.2). We give detailed proofs to all the lemma and corollaries where some proofs are omitted in the original paper[KLLOT13].

2 Preliminaries

2.1 Real Analysis

The proof of (1.2) is for arbitrary weighted graph and hence some background knowledge of weight spectral graph and real analysis is required. We give a brief introduction below.

We focus on functions on $Hilbert\ space$ where Hilbert space is a complete vector space with an associated inner product. In our paper, we mainly use the structure and properties of inner product hence it is reasonable to treat it like an Euclidean space in our setting.

Definition 2.1. *We write $\ell^2(V, w)$ for a Hilbert space of functions $f : V \to \mathbb{R}$ with associated inner product*

$$\langle f, g\rangle_w = \sum_{v\in V} w(v)f(v)g(v)$$

where $w : V \to \mathbb{R}$ is a weight function. We reserve $\langle \cdot, \cdot \rangle$ and $|| \cdot ||$ for the standard inner product and norm on $\mathbb{R}^n$ and $\ell^2(V)$.

The *support* of a function is defined to be the set of values that has non-zero images.

Definition 2.2. *Let $f : V \to \mathbb{R}$ be a function, then supp$(f) = \{v : f(v) \neq 0\}$*

We say a set of functions $\{f_1, ..., f_k\} \in \mathbb{R}^V$ are *disjointly supported* if supp$(f_i) \cap$ supp$(f_j) = \emptyset$ for all $i \neq j$

A function f is called as a *Lipschitz function* if it satisfies that $|f(x) - f(y)| \leq C|x - y|$ where C is some constant. We can extend this concept to function level and define $p - Lipschitz$.

Definition 2.3. *For $p > 0$, function $g : V \to \mathbb{R}$ is $p - Lipschitz$ with respect to function $f : V \to \mathbb{R}$ if for all $x, y \in V$,*

$$|g(u) - g(v)| \leq p|f(u) - f(v)|$$

Cauchy-Schwarz Inequality is commonly used in real analysis and will be useful in our proof.

Theorem 2.1 (Cauchy-Schwarz Inequality)**.** *Let $u_1, ..., u_n$ and $v_1, ..., v_n$ be real numbers, then*

$$(\sum_{i=1}^{n} u_i v_i)^2 \leq (\sum_{i=1}^{n} u_i^2)(\sum_{i=1}^{n} v_i^2) \tag{2.1}$$

2.2 Spectral Theory

Let $G = (V, E)$ be a graph, A be the adjacency matrix of G and L be the Laplacian matrix of G. Then $\mathcal{A} = D^{\frac{-1}{2}} A D^{\frac{-1}{2}}$ is the normalized adjacency matrix, and $\mathcal{L} = I - \mathcal{A}$ is the normalized Laplacian matrix, where D is the diagonal matrix whose i-th entry is the degree of vertex i. Note that $\mathcal{L} = I - \mathcal{A} = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$ where $L = D - A$. For any $S \subset V$, the *volume* of S is defined to be $vol(S) = \sum_{v \in S} deg(v)$. Denotes $E(S, \overline{S})$: the set of edges with one endpoint in S and one end point in $V - S$.

Definition 2.4. *The conductance of S is defined as*

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

We further define the *conductance* of graph G by $\phi(G) = \min_{S \subset V, vol(S) \leq |E|} \phi(S)$

Fact 2.1. *The conductance is a measure how well connected the graph G is.*

1. *G is disconnected if and only if there exists a set $S \neq \emptyset, S \neq V$ such that $|E(S, \overline{S})| = 0$, which means $\phi(G) = 0$*
2. *If G is a clique,*
$$\phi(G) = \min_{1 \leq k \leq n/2} \frac{k(n - k)}{(n - 1)k} = \frac{n}{2(n - 1)} \approx \frac{1}{2}$$
3. *if G is a cycle,*
$$\phi(G) = \min_{1 \leq k \leq n/2} \frac{2}{2k} = \frac{2}{n}$$

Proposition 2.2. *When the graph is d-regular, $\mathcal{A} = \frac{1}{d} A$ and $\mathcal{L} = \frac{1}{d} L$.*

Proof. Since the graph is d-regular, $\forall v_i \in V$, we have $deg(v_i) = d$. Hence the diagonal matrix D is given by

$$D = diag(deg(v_1), deg(v_2), ..., deg(v_n)) = diag(d, d, ..., d).$$

Then, $D^{-\frac{1}{2}} = diag(\frac{1}{\sqrt{d}}, ..., \frac{1}{\sqrt{d}})$ and hence $\mathcal{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = \frac{1}{\sqrt{d}} \cdot A \cdot \frac{1}{\sqrt{d}} = \frac{1}{d} A$

Similarly, we have $\mathcal{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = \frac{1}{d} L$ $\square$

2.3 Spectral of Weighted Laplacian

Let $G = (V, E)$ be a graph with associated weight function $w : E \to \mathbb{R}_+$ and hence $\ell^2(V, w)$ is a Hilbert space. We use $u \sim v$ to denote $uv \in E$ and we define the weight of a vertex v to the sum of the weights of all edges incident to v, which is $w(v) = \sum\limits_{u \sim v} w(u, v)$. For $S \subset V$, we define the $volume$ of S as $vol(S) = \sum\limits_{v \in S} w(v)$.

Definition 2.5. *Let $G = (V, E)$ be a graph with weight function w, the $Dirichlet\ conductance$ of $S \subset V$ is,*

$$\phi(S) = \frac{w(E(S,\overline{S}))}{min\{vol(S), vol(\overline{S})\}}$$

For a function $f \in \mathbb{R}^V$ and a threshold $t \in \mathbb{R}$, let $V_f(t) = \{v : f(v) \geq t\}$ be a threshold set of f and we define the conductance of the best threshold set of f by $\phi(f) = \min\limits_{t \in \mathbb{R}} \phi(V_f(t))$

We define the adjacency operator analogously as $Af(v) = \sum\limits_{u \sim v} w(u, v)f(u)$, diagonal degree operator by $Df(v) = w(v)f(v)$ and the $combinatorial\ Laplacian$ by $L = D - A$ and the $normalized\ Lapalcian$ by $\mathcal{L}_G = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$

Definition 2.6. *Let $G = (V, E)$ be a graph with weight function w, if $g : V \to \mathbb{R}$ is a non-zero function and $f = D^{-\frac{1}{2}}g$, then the Rayleigh quotient of f with respect to G is given by*

$$\mathcal{R}_G(f) = \frac{\langle g, \mathcal{L}_G g\rangle}{\langle g, g\rangle} = \frac{\langle f, Lf\rangle}{\langle D^{\frac{1}{2}}f, D^{\frac{1}{2}}f\rangle} = \frac{\sum\limits_{u \sim v} w(u, v)|f(u) - f(v)|^2}{||f||_w^2} \tag{2.2}$$

Analogously, we also know $\mathcal{L}_G$ is a positive-definite operator with eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 2$, the eigenvalues of $\mathcal{L}_G$ can be characterized by

$$\lambda_k = \min_{f_1,\ldots,f_k \in \ell^2(V,w)} \max_{f \neq 0} \{\mathcal{R}(f) : f \in \text{span}\{f_1, ..., f_k\}\} \tag{2.3}$$

where $f_1, ..., f_k$ are k non-zero orthogonal functions in $\ell^2(V, w)$

2.4 Random walk

Define function $f : V \to \mathbb{R}$: $f(u, v) = \frac{f(u)}{deg(v)}$ if $u, c \in E$. Otherwise, $f(u, v) = 0$. Define $f(S) = \sum_{(u,v) \in S} f(u, v)$ where a set of vertices $S \subseteq V$. We define a lazy random walk as following,

Definition 2.7. *If W is a lazy random walk on graph G then*

$$W = \frac{I + P}{2}$$

where $P = D^{-1}A$ is the transition probability matrix.

Definition 2.8. *For a vertex set $S \subseteq V$, we define the in set and out set as*

$$S_{in} = \{(u, v) : v \in S, \text{ u is adjacent to v}\}$$
$$S_{out} = \{(u, v) : u \in S, \text{ u is adjacent to v}\}$$

2.5 Useful Lemmas

We prove several lemmas which will be used in our paper.

Lemma 2.3. *Let $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$ be the eigenvalues of $\mathcal{A}$ and let $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$ be the eigenvalues of $\mathcal{L}$,*

$$1 = \alpha_1 \geq ... \geq \alpha_n \geq -1 \text{ , } 0 = \lambda_1 \leq ... \leq \lambda_n \leq 2,$$

where the eigenvector corresponding to λ_1 is $v_1 = \vec{1}$

Proof. Since $\mathcal{L}(D^{\frac{1}{2}}\vec{1}) = (D^{-\frac{1}{2}}LD^{-\frac{1}{2}})(D^{\frac{1}{2}}\vec{1}) = D^{-\frac{1}{2}}L\vec{1} = 0$, we have 0 as an eigenvalue and the corresponding eigenvector is $\vec{1}$. $x^T\mathcal{L}x = x^TD^{-\frac{1}{2}}LD^{-\frac{1}{2}}x = \sum_{e \in E} x^TD^{-\frac{1}{2}}L_eD^{-\frac{1}{2}}x = \sum_{i,j \in E}(\frac{x_i}{\sqrt{d_i} - \frac{x_j}{\sqrt{d_j}}})^2 \geq 0$, where $L_e =$

$b_e b_e^T$. So $\mathcal{L}$ is a positive semidefinite matrix, which shows that the smallest eigenvalue $\lambda_1 = 0$ and its corresponding eigenvector $v_1 = \vec{1}$. This also implies $I - \mathcal{A} \geq 0$. So, $\alpha \leq 1$.
Besides, $x^T(I + \mathcal{A})x = x^T \mathcal{L} x + 2x^T \mathcal{A} x = \sum_{i,j \in E}[(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}})^2 + \frac{2x_i x_j}{\sqrt{d_i d_j}}] = \sum_{i,j\in E}(\frac{x_i}{\sqrt{d_i}} + \frac{x_j}{\sqrt{d_j}})^2 \geq 0$, which implies $I + \mathcal{A} \geq 0$, and thus $\alpha_n \geq -1$ and hence$\lambda_n = 1 - \alpha_n \leq 2$.

□

Lemma 2.4. *Suppose $a_1, ..., a_n, b_1, ..., b_n > 0$, then*

$$\sum_{i=1}^{n} \frac{a_i^2}{b_i} \geq \frac{(\sum_{i=1}^{n} a_i)^2}{\sum_{i=1}^{n} b_i} \tag{2.4}$$

Proof. Since $b_i > 0$, we rewrite $\frac{a_i^2}{b_i}$ as $(\frac{a_i}{\sqrt{b_i}})^2$ and by Cauchy-Schwarz we have

$$(\sum_{i=1}^{n} \frac{a_i^2}{b_i})(\sum_{i=1}^{n} b_i) = (\sum_{i=1}^{n} (\frac{a_i}{\sqrt{b_i}})^2)(\sum_{i=1}^{n} (\sqrt{b_i})^2) \geq \sum_{i=1}^{n} ((\frac{a_i}{\sqrt{b_i}})(\sqrt{b_i}))^2 = \sum_{i=1}^{n} (a_i)^2$$

Dividing both side by $\sum_{i=1}^{n} b_i$ gives us $\sum_{i=1}^{n} \frac{a_i^2}{b_i} \geq \frac{(\sum_{i=1}^{n} a_i)^2}{\sum_{i=1}^{n} b_i}$ □

Lemma 2.5. *Suppose $a_1, ..., a_n, b_1, ..., b_n > 0$, let M be the maximum of the set $\{\frac{a_i}{b_i} | i = 1, ..., n\}$, then*

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq M \tag{2.5}$$

Proof. By definition of M, we know that $\frac{a_i}{b_i} \leq M$ for all $i \in \{1, ..., n\}$. Then $a_i \leq M(b_i)$ as $b_i > 0$. We sum over all n inequalities and get $\sum_{i=1}^{n} a_i \leq M(\sum_{i=1}^{n} b_i)$. Dividing both side by $\sum_{i=1}^{n} b_i$ gives $\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq M$ as required. □

Lemma 2.6. *For any k disjointly supported functions $f_1, ..., f_k \in \ell^2(V, w)$, we have*

$$\lambda_k \leq 2 \max_{1 \leq i \leq k} \mathcal{R}(f_i)$$

Proof. By (2.3), it suffices to prove that for any $h \in \{f_1, ..., f_k\}$, $\mathcal{R} \leq 2 \max_i \mathcal{R}(f_i)$. Notice that $\mathcal{R}(f_i) = \mathcal{R}(cf_i)$ for any $c \in \mathbb{R}$, hence we may assume $h = \sum_{i=1}^{k} f_i$. Since $f_1, .., f_k$ are disjointly supported, u, v are in the support of at most two of these functions. Hence $h(u) = f_i(u)$ and $h(v) = f_j(v)$, which gives $|h(u) - h(v)|^2 = (f_i(u) - f_j(v))^2 = f_i(u)^2 - 2f_i(u)f_j(v) + f_j(v)^2 \leq 2f_i(u)^2 + 2f_j(v)^2$. The last inequality holds since $(f_i(u) - f_j(v))^2 = f_i(u)^2 - 2f_i(u)f_j(v) + f_j(v)^2 \geq 0$. It follows that $|h(u) - h(v)|^2 \leq 2f_i(u)^2 + 2f_j(v)^2 = \sum_{i=1}^{k} 2|f_i(u) - f_i(v)|^2$ since $|f_i(u) - f_i(v)|^2 = f_i(u)^2, |f_j(u) - f_j(v)|^2 = f_j(v)^2$ and the remaining terms are all zero.
Therefore,

$$\begin{aligned}
\mathcal{R}(h) = \frac{\sum_{u \sim v} w(u, v)|h(u) - h(v)|^2}{||h||_w^2} &\leq \frac{2\sum_{u\sim v}\sum_{i=1}^{k} w(u, v)|f_i(u) - f_i(v)|^2}{||h||_w^2} \\
&= \frac{2\sum_{i=1}^{k}\sum_{u\sim v} w(u, v)|f_i(u) - f_i(v)|^2}{\sum_{i=1}^{k} ||f_i||_w^2} \leq 2 \max_{i \leq i \leq k} \mathcal{R}(f_i)
\end{aligned}$$

The last inequality can be proved using Lemma 2.3, note that the denominators can be viewed as $\{b_1 = ||f_1||_w^2, ..., b_k = ||f_k||_w^2\}$ and the numerators can be viewed as $\{a_1 = \sum_{u\sim v} w(u, v)|f_1(u) - f_1(v)|^2, ..., a_k = \sum_{u\sim v} w(u, v)|f_k(u) - f_k(v)|^2\}$. Note that $\frac{a_i}{b_i} = \mathcal{R}(f_i)$, hence by Lemma 2.3, we have $\mathcal{R}(h) \leq 2 \max_{1 \leq i \leq k} \mathcal{R}(f_i)$

□

Lemma 2.7. *For every non-negative $h \in \ell^2(V, w)$ such that $supp(h) \leq \frac{vol(V)}{2}$, the following holds,*

$$\phi(h) \leq \frac{\sum_{u\sim v} w(u,v)|h(v)-h(u)|}{\sum_v w(v)h(v)}$$

Proof. Note that in the right hand side of the equation, h is on both of the numerator and the denominator($|h(v) - h(u)| = \pm(h(v) - h(u))$), hence we can assume $\max_v h(v) \leq 1$ because we can always scale h down by multiplying a scalar without changing the result. Let $0 < t \leq 1$ be chosen uniformly random. Since $V_h(t) = \{v|h(v) \geq t\}$ and $vol(V_h(t)) = \sum\limits_{v\in V_h(t)} w(v)$. By linearity of expectation, we know that $\mathbb{E}(vol(V_h(t))) = \sum\limits_v w(v)Pr(v \in V_h(t)) = \sum\limits_v w(v)h(v)$ since we choose t randomly, the probability of $h(v) \geq t$ is precisely $h(v)$. Similarly, we get $\mathbb{E}[w(E(V_h(t), \overline{V_h(t)}))] = \sum_{u\sim v} w(u,v)|h(u) - h(v)|$. Since we know that $vol(V_h(t)) \leq \frac{vol(V)}{2}$ for all $t > 0$, we know that $\mathbb{E}[vol(V_h(t))] = \mathbb{E}[\min\{vol(V_h(t)), vol(\overline{V_h(t)})\}]$ hence $E[\phi(V_h(t))] = \frac{\sum_{u\sim v} w(u,v)|h(u)-h(v)|}{\sum\limits_v w(v)h(v)}$. This implies there exists t such that $\phi(V_h(t))$ is at most $\frac{\sum_{u\sim v} w(u,v)|h(u)-h(v)|}{\sum\limits_v w(v)h(v)}$ and the lemma follows. ∎

Lemma 2.8. *For a vertex set $S \subseteq V$,*

$$fW(S) = \frac{f(S_{in}) + f(S_{out})}{2} \leq \frac{f(vol(S)(1 + h_S)) + f(vol(S)(1 - h_S))}{2}$$

Proof.

$$\begin{aligned} fW(S) &= \frac{f(S_{out}) + f_(S_{in})}{2} \\ &= \frac{f(S_{in} \bigcup S_{out}) + f(S_{in} \bigcap S_{out})}{2} \\ &\leq \frac{f(vol(S) + |E(S, \overline{S})|) + f(vol(S) - |E(S, \overline{S})|)}{2} \\ &\leq \frac{f(vol(S)(1 + h_S)) + f(vol(S)(1 - h_S))}{2} \end{aligned}$$

∎

3 Proofs of Cheeger's Inequality

In this section, we will demonstrate two different methods to prove Cheeger's Inequality.

3.1 Proof using eigenvectors:

The idea of the following proof proves the Cheeger's Inequality when G is a d-regular graph. By fact 2.2, $\mathcal{L} = \frac{1}{d}L$. The general case is similar but slightly more involved. We separate the proof into two parts. The idea of the proof is originated by the Professor Lapchi Lau. We first show that $\frac{1}{2}\lambda_2 \leq \phi(G)$.

Proof. By Lemma 2.2, we have that the eigenvector v_1 responding to λ_1 is $\mathbf{1}$. By using Rayleigh quotient:

$$\lambda_2 = \min_{\langle x, \vec{1}\rangle = 0} \frac{x^T \mathcal{L} x}{x^T x} = \min_{\langle x, \vec{1}\rangle = 0} \frac{x^T L x}{d x^T x} = \min_{\langle x, \vec{1}\rangle = 0} \frac{\sum_{i,j\in E}(x_i - x_j)^2}{d\sum_{i\in V} x_i^2}$$

Suppose $\phi(G) = \phi(S)$ and $|S| = \frac{n}{2}$. Denote $x_i = +1$ if $i \in S$ and $x_i = -1$ otherwise.
Since $|S| = \frac{n}{2}, \sum_{i\in V} x_i = 0$, and thus $\langle x, \vec{1}\rangle = 0$. Then,

$$\lambda_2 \leq \frac{\sum_{i,j\in E}(x_i - x_j)^2}{d\sum_{i\in V} x_i^2} = \frac{4|E(S, \overline{S})|}{d|v|} = \frac{2|E(S, \overline{S})|}{d|S|} = 2\phi(S)$$

For general S, denote $x_i = \frac{+1}{|S|}$ if $i \in S$ and $x_i = \frac{-1}{|V-S|}$ otherwise.

By construction, $x \perp \vec{1}$

$$\lambda_2 \leq \frac{\sum_{i,j\in E}(x_i - x_j)^2}{d\sum_{i\in V} x_i^2} = \frac{|E(S,\overline{S})|(\frac{1}{|S|}+\frac{1}{|V-S|})^2}{d(|S|\frac{1}{|S|^2}+|V-S|\frac{1}{|V-S|^2})} = \frac{|E(S,\overline{S})||V|}{d|S||V-S|} \leq 2\phi(S)$$

Thus, we have $\frac{1}{2}\lambda_2 \leq \phi(G)$ □

We then show that $\phi(G) \leq \sqrt{2\lambda_2}$. Firstly, we preprocess the λ_2: let at most half the entries are non-zero. So the output set S has size $|S| \leq \frac{|V|}{2}$

Without loss of generality, there are fewer positive entries in x than negative entries. We would like to zero out the negative entries by constructing a new vector y corresponding to x =: $y_i = x_i$ if $x_i \geq 0$ and $y_i = 0$ otherwise.

Denote Rayleigh quotient of x by $\mathcal{R}(x) = \frac{x^T\mathcal{L}x}{x^Tx} = \frac{x^TLx}{dx^Tx}$, since $\mathcal{L} = \frac{1}{d}L$

Claim 3.1. $\mathcal{R}(y) \leq \mathcal{R}(x)$

Proof. For all i with $y_i \geq 0$, $(\mathcal{L}y)_i = y_i - \sum_{j\in Neighbour(i)} \frac{y_i}{d} \leq x_i - \sum_{j\in Neighbour(i)} \frac{x_i}{d} = (\mathcal{L}x)_i = \lambda_2 x_i$. Therefore, $y^T\mathcal{L}y = \sum_{i\in V} y_i(\mathcal{L}y)_i = \sum_{i:y_i>0} y_i(\mathcal{L}y)_i \leq \sum_{i:y_i>0}\lambda_2 x_i^2 = \sum_{i\in V}\lambda_2 y_i^2$

$\mathcal{R}(y) = \frac{y^T\mathcal{L}y}{y^Ty} \leq \lambda_2\frac{\sum_{i:y_i>0}(y_i)^2}{\sum_{i\in[n]}(y_i)^2} = \lambda_2 = \mathcal{R}(x)$ which proves the claim □

Lemma 3.2. *Given any y, there exists a subset $S \subset supply(y)$ such that*

$$\phi(S) = \frac{|E(S,\overline{S})|}{d|S|} \leq \sqrt{2\mathcal{R}(y)}, \textit{where } supply(y) = \{i|y_i \neq 0\}.$$

Proof. We can assume that $0 \leq y_i \leq 1$ by scaling if it is not. Let $t \in (0,1]$ be chosen uniformly at random. Let $S_t = \{i|y_i^2 \geq t\}$ Then, $S_t \subset Supply(y)$ by construction.

$$\begin{aligned}
E_t[|E(S,\overline{S})|] &= \sum_{i,j\in E}[\Pr(y_i^2 < t \leq y_i^2)] \text{ by linearity of expectation} \\
&= \sum_{i,j\in E}[\Pr(y_i^2 < t \leq y_i^2)] \\
&= \sum_{i,j\in E}|y_i^2 - y_j^2| \\
&= \sum_{i,j\in E}|y_i - y_j|\cdot|y_i + y_j| \\
&\leq \sqrt{\sum_{i,j\in E}(y_i - y_j)^2}\cdot\sqrt{\sum_{i,j\in E}(y_i + y_j)^2} \text{ by Cauchy-Schwarz } \langle a,b\rangle \leq ||a||\cdot||b|| \\
&\leq \sqrt{\sum_{i,j\in E}(y_i - y_j)^2}\cdot\sqrt{2\sum_{i,j\in E}(y_i + y_j)^2} \\
&= \sqrt{2\mathcal{R}(y)}(d\sum_{i\in V} y_i^2)
\end{aligned}$$

We also have that $E_t[|S_t|] = \sum_{i\in V}\Pr[y_i^2 \geq t] = \sum_{i\in V} y_i^2$

Therefore, $\frac{E_t[|E(S_t,\overline{S_t})|]}{E_t[d|s_t|]} \leq \sqrt{2\mathcal{R}(y)}$ which means $E_t[|E(S_t,\overline{S_t})| - \sqrt{2\mathcal{R}(y)}\cdot d\cdot|S_t| \leq 0$

Hence, there exists t such that $\phi(S_t) = \frac{|E(S_t,\overline{S_t})|}{d|S_t|} \leq \sqrt{2\mathcal{R}(y)}$ □

By Claim 3.1 and Lemma 3.2, we prove that $\phi((G) \leq \sqrt{2\lambda_2}$

Hence we have $\frac{1}{2}\lambda_2 \leq \phi(G) \leq \sqrt{2\lambda_2}$, which proves Theorem 1.1

3.2 Proof using random walk:

Lemma 3.3. *For any vertices u in G, a subset of vertices S with $vol(S) \leq \frac{vol(G)}{2}$ satisfies:*

$$|W^k(u,S) - \pi(S)| \leq \sqrt{\frac{vol(S)}{deg(u)}}(1 - \frac{\beta_k^2}{8})^k$$

where $\beta_k = inf\{h_f : f = \chi_u W^{k'}$ for $u \in V$ and $k' \leq k\}$. is the minimum Cheeger ratio over sets determined by the i largest values of the distribution of the lazy random walk starting at u after k steps.

Proof. For a fixed vertex u, we choose $f_k(v) = w^k(u,v) - \pi(v)$ and we prove Lemma 3.3 by induction on k

- Base case: When $k = 0$, $f_0(v) = W^0(u,v) - \pi(v) \leq \sqrt{\frac{deg(v)}{deg(u)}}$. Hence base case holds.
- Inductive hypothesis: Assume for $k \leq n$ where $n \in \mathbb{R}$, the lemma holds
- Inductive conclusion: When $k = n+1$, by lemma 2.7,

$$\begin{aligned} f_{n+1}(v) &= f_k W(v) \\ &\leq \frac{f_k(v(1+\beta_k)) + f_k(v(1-\beta_k))}{2} \\ &\leq \sqrt{\frac{deg(v)}{deg(u)}}(1 - \frac{\beta_k^2}{8})^k \frac{\sqrt{1+\beta_k} + \sqrt{1-\beta_k}}{2} \end{aligned}$$

  Since $\sqrt{1+x} + \sqrt{1-x} \leq 2 - \frac{x^2}{4}$, we get:

$$f_{k+1}(x) \leq \sqrt{\frac{deg(v)}{deg(u)}}(1 - \frac{\beta_k^2}{8})^k(1 - \frac{\beta_k^2}{8}) = \sqrt{\frac{deg(v)}{deg(u)}}(1 - \frac{\beta_k^2}{8})^{k+1}$$

  By induction, we have proved the lemma.

□

Claim 3.4. $2\sqrt{2\lambda_2} \geq \beta_G^2 \geq \phi(G) \geq \frac{\lambda_2}{2}$ *where $\beta_G = min\{\beta_t : t \leq \lceil \frac{16logn}{\lambda_2^2} \rceil\}$*

Proof. Denote the left eigenvector of $I - P$ associated with λ_2 be φ where $\varphi(I-P) = \lambda_2\varphi$. Since φ is orthogonal to $\vec{1}$ by lemma 2.2. we have:

$$||\varphi(W^t - \vec{1}^*\pi)D^{\frac{1}{2}} = ||\varphi W^t D^{1\frac{1}{2}}|| = (1 - \frac{\lambda_2}{2})^t||^{-\frac{1}{2}}||$$

The, we have:

$$\begin{aligned} ||\varphi(W^t - \vec{1}^*\pi)D^{-\frac{1}{2}}||^2 &= \sum_v (\phi(W^t - \vec{1}^*\pi)(v)\frac{1}{\sqrt{deg(v)}})^2 \\ &= \sum_v \frac{1}{deg(v)}(\sum_u \varphi(v)(W^t(u,v) - \pi(v)))^2 \\ &\leq \sum_v \frac{1}{deg(v)}(\sum_u |\varphi(v)|(1 - \frac{\beta_t^2}{8})^{2t}\sqrt{\frac{deg(v)}{deg(u)}})^2 \\ &\leq (10\frac{\beta_t^2}{8})^{2t}\sum_v n(\sum_u \frac{\varphi(u)^2}{deg(u)}) \\ &= (1 - \frac{\beta_t^2}{8})^{2t}n^2||^{-\frac{1}{2}}||^2 \end{aligned}$$

Besides, we have:

$$1 - \frac{\lambda_2}{2} \leq (1 - \frac{\beta_t^2}{8})n^{1/t}$$

We get:

$$2\phi(G) \geq \lambda_2 \geq \frac{\beta_t^2}{4} - \frac{2log(n)}{t}$$

Since $\beta_G = min\{\beta_t : t \leq \lceil \frac{16logn}{\lambda_2^2} \rceil\}$, we get:

$$2\phi(G) \geq \lambda_2 \geq \frac{\beta_G^2}{8} \geq \frac{\phi(G)^2}{8}$$

Note that $\phi(G) \leq 2\sqrt{2\lambda_2}$ follows from $\lambda_2 \geq \frac{\phi(G)^2}{8}$ ☐

By Claim 3.4, Cheeger's Inequality holds.

4 Proof of Improved Cheeger's Inequality

Throughout this section we assume $f \in \ell^2(V, w)$ to be a non-negative function of unit norm such that $\mathcal{R}(f) \leq \lambda_2$ and $vol(supp(f)) \leq \frac{vol(V)}{2}$. We first prove the existence of such f.

Proposition 4.1 (HLW06). *There are two disjointly supported functions $f_+, f_- \in \ell^2(V, w)$ such that $f_+ \geq 0$ and $f_- \leq 0$ and $\mathcal{R}(f_+) \leq \lambda_2$ and $\mathcal{R}(f_-) \leq \lambda_2$*

Proof. Let $g \in \ell^2(V)$ be the second eigenfunction of $\mathcal{L}$ such that $\mathcal{L}g = \lambda_2 g$, we define $g_+ \in \ell^2(V)$ by $g_+(u) = \max\{g(u), 0\}$ and $g_- \in \ell^2(V)$ by $g_-(u) = \min\{g(u), 0\}$. Let u be any vertex in $supp(g_+)$, then $g_+(u) = g(u) > 0$. Consider $(\mathcal{L}g_+)(u) = g_+(u) - \sum_{v:v\sim u} \frac{w(u,v)g_+(v)}{\sqrt{w(u)w(v)}} = g(u) - \sum_{v:v\sim u} \frac{w(u,v)g_+(v)}{\sqrt{w(u)w(v)}}$. Note that if $g(v) < 0$ then $g(v) < g_+(v) = 0$, hence $g(v) \leq g_+(v)$ for all v. Hence we have

$$(\mathcal{L}g_+)(u) = g(u) - \sum_{v:v\sim u} \frac{w(u,v)g_+(v)}{\sqrt{w(u)w(v)}} \leq g(u) - \sum_{v:v\sim u} \frac{w(u,v)g(v)}{\sqrt{w(u)w(v)}} = (\mathcal{L}g)(u) = \lambda_2 g(u)$$

Therefore,

$$\begin{aligned}\langle g_+, \mathcal{L}g_+ \rangle = \sum_{u\in V} g_+(u) \cdot (\mathcal{L}g_+)(u) &= \sum_{u\in supp(g_+)} g_+(u) \cdot (\mathcal{L}g_+)(u) \\ &\leq \sum_{u\in supp(g_+)} g_+(u) \cdot \lambda_2 g(u) \\ &\leq \sum_{u\in supp(g_+)} \lambda_2 \cdot g_+(u)^2 \\ &= \lambda_2 ||g_+||^2\end{aligned}$$

Let $f_+ = D^{-\frac{1}{2}} g_+$, we get

$$\lambda_2 \geq \frac{\langle g_+, \mathcal{L}g_+ \rangle}{||g_+||^2} = \frac{\langle f_+, \mathcal{L}f_+ \rangle}{||f_+||_w^2} = \mathcal{R}(f_+)$$

The above follows exactly from the definition of (2.2). Similarly, we can let $f_- = D^{-\frac{1}{2}} g_-$ and show $\mathcal{R}(f_-) \leq \lambda_2$ ☐

We get a immediate corollary that shows the existence of our required f.

Corollary 4.1.1. *There exists function $f \in \ell^2(V, w)$ such that $f \geq 0, \mathcal{R}(f) \leq \lambda_2, supp(f) \leq \frac{vol(V)}{2}$ and $||f||_w = 1$*

Proof. By our construction of f_+ and f_-, we know that $|supp(f_+) \cup supp(f_-)| \leq |V|$ and hence $vol(supp(f_+) \cup supp(f_-)) \leq vol(V)$ Without loss of generality, we assume $vol(supp(f_+)) \leq vol(supp(f_{)})$. Hence $vol(supp(f_+)) \leq \frac{vol(V)}{2}$. We let $f' = \frac{f_+}{||f_+||_w}$, then we have $||f'||_w = 1$ and by Proposition 4.1, we know $f' \geq 0$ and $\mathcal{R}(f') = \mathcal{R}(f_+) \leq \lambda_2$ as required. □

Hence by corollary 4.1.1 we know our required f exists. To prove Theorem 1.2, we can instead prove a stronger result, that is $\phi(f)$ is upper-bounded by $O(k)\frac{\lambda_2}{\sqrt{\lambda_k}}$

Theorem 4.2. *For any undirected graph G, and $k \geq 2$,*

$$\phi(f) = O(k)\frac{\lambda_2}{\sqrt{\lambda_k}}$$

We prove Theorem 4.2 by approximating f by a $2k+1$ step function g. We say a function $g \in \ell^2(V,w)$ is a l-step approximation of f, if there exists l thresholds $0 = t_0 \leq t_1 \leq \cdots \leq t_{l-1}$ such that for all vertex v,

$$g(v) = \psi_{t_0,t_1,\ldots,t_{l-1}}(f(v))$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is defined by

$$\psi_{t_1,\ldots,t_l}(x) = \text{argmin }_{t_i}|x - t_i|$$

In words, $\psi_{t_1,\ldots,t_l}(x)$ is the value of t_i where t_i is closest to x. g is defined to be the threshold t_i such that t_i is closest to $f(v)$. To be specifically, g is similar to a step function such that only takes value from $\{t_0,\ldots,t_{l-1}\}$ and for each v, $g(v)$ is within the closest distance from $f(v)$.

We show that f is well approximated by step function g with at most $2k+1$ steps, provided $|\lambda_2 - \lambda_k|$ is large. In the next lemma we show if there is a large gap between $\mathcal{R}(f)$ and λ_k, then there is a $2k+1$ step function g such that $||f-g||_w^2 = O(\frac{\mathcal{R}(f)}{\lambda_k})$

Lemma 4.3. *There exists a $2k+1$ step approximation of f, call g, such that*

$$||f-g||_w^2 \leq \frac{4\mathcal{R}(f)}{\lambda_k} \tag{4.1}$$

Proof. Let $M = \max_v f(v)$. We claim that we can find $2k+1$ thresholds $0 = t_0 \leq t_1 \leq \cdots \leq t_{2k} = M$ such that the $2k+1$ step approximation g satisfies $||f-g||_w^2 \leq \frac{4\mathcal{R}(f)}{\lambda_k}$. We choose thresholds by the following algorithm.

Algorithm 1: Find required $2k+1$ approximation thresholds

Result: $0 = t_0 \leq t_1 \leq \cdots \leq t_{2k} = M$
$t_0 = 0$;
$i = 1$;
$C = \frac{2\mathcal{R}(f)}{k\lambda_k}$;
while $i < 2k+1$ **do**
  Find smallest $m \geq t_{i-1}$ such that $\sum\limits_{v:t_{i-1} \leq f(v) \leq m} w(v)|f(v) - \psi_{t_{i-1},m}(f(v))|^2 = C$;
  if *such m exists* **then**
    $t_i = m$;
  else
    $t_i = M$;
  end
  $t_i + = 1$;
end

We first show that if Algorithm 1 terminates successfully then these thresholds form our required approximation and then we show Algorithm 1 always terminates successfully.

Suppose Algorithm 1 terminates successfully, then we can define g to be the $2k+1$ step approximation with respect to $t_0,\ldots,t_{2k}$, then $||f-g||_w^2 = \sum\limits_{i=1}^{2k} \sum\limits_{v:t_{i-1} \leq f(v) \leq t_i} w(v)|f(v) - \psi_{t_{i-1},t_i}(f(v))|^2$. Note that if $t_{i-1} = t_i = M$ for any i, then there is either no v satisfies $f(v) = t_i$ or we have $\sum\limits_{v:t_{i-1} \leq f(v) \leq t_i} w(v)|f(v) - \psi_{t_{i-1},t_i}(f(v))|^2 = w(v)|t_i - t_i| = 0$

Hence we have $\sum_{v:t_{i-1}\le f(v)\le t_i} w(v)|f(v)-\psi_{t_{i-1},t_i}(f(v))|^2 \le C$ for all i. It follows that

$$||f-g||_w^2 = \sum_{i=1}^{2k} \sum_{v:t_{i-1}\le f(v)\le t_i} w(v)|f(v)-\psi_{t_{i-1},t_i}(f(v))|^2 \le 2kC = \tfrac{4\mathcal{R}(f)}{\lambda_k}$$

Next we prove Algorithm 1 always terminates successfully.

Suppose by contradiction that Algorithm 1 does not terminate successfully, then we know that $t_{2k} < M$ and in this case $\sum_{v:t_{i-1}\le f(v)\le t_i} w(v)|f(v)-\psi_{t_{i-1},t_i}(f(v))|^2 = C$ for all i. For $1\le i\le 2k$, we define $2k$ disjointly supported functions $f_1, ..., f_{2k}$ by

$$f_i(v) = \begin{cases} |f(v)-\psi_{t_{i-1},t_i}(f(v))| & \text{if } t_{i-1}\le f(v)\le t_i \\ 0 & \text{otherwise} \end{cases}$$

Notice that $||f_i||_w^2 = \sum_{v:t_{i-1}\le f(v)\le t_i} w(v)|f(v)-\psi_{t_{i-1},t_i}(f(v))|^2 = C$ and $||f_i||_w^2$ is exactly the denominator of $\mathcal{R}(f_i)$ by (2.2). It remains to bound the numerator of $\mathcal{R}(f_i)$. We show that for all pairs of u, v

$$\sum_{i=1}^{2k} |f_i(u)-f_i(v)|^2 \le |f(u)-f(v)|^2 \tag{4.2}$$

Since $f_1, ..., f_{2k}$ are disjointly supported, u, v can be in the support of distinct f_i, f_j or in the support of same function f_i. If u, v are in the support of the same function, say f_i, then $\sum_{i=1}^{2k} |f_i(u)-f_i(v)|^2 = |f_i(u)-f_i(v)|^2$ By Reversed Triangle Inequality, we have $|f_i(u)-f_i(v)| \le |f_i(u)-\psi_{t_{i-1},t_i}(f(u)) - (f_i(v)-\psi_{t_{i-1},t_i}(f(v)))| \le |f(u)-f(v)|$ The above inequality holds since the distance between $f(u)$ and $f(v)$ is always at least as large as the difference of the distance between $f_i(v)$ and $\psi_{t_{i-1},t_i}(v)$. An illustration of this is provided below.

If u, v are in different supports, then suppose $u \in supp(f_i), v \in supp(f_j)$ where $i < j$, then (3.3) becomes $|f_i(u)-f_i(v)|^2 + |f_j(u)-f_j(v)|^2 = |f_i(u)|^2 - |f_j(v)|^2 \le |f(u)-f(v)|^2$. Hence (4.3) holds for all pairs of u, v. Summing (4.3) we have

$$\sum_{i=1}^{2k} \mathcal{R}(f_i) = \sum_{i=1}^{2k} \frac{\sum_{u\sim v} w(u,v)|f_i(u)-f_i(v)|^2}{||f_i||_w^2} \le \frac{\sum_{u\sim v} w(u,v)|f(u)-f(v)|^2}{C} = \frac{\mathcal{R}(f)}{C} = \frac{k\lambda_k}{2} \tag{4.3}$$

Hence there are at least k disjointly supported functions $f'_1, ..., f'_k$ of Rayleight quotients less than $\frac{\lambda_k}{2}$. Otherwise if there are strictly more than k functions with Rayleight quotient at least $\frac{\lambda_k}{2}$ then the sum of Rayleight quotients of these functions is strictly more than $k(\frac{\lambda_k}{2}) = \frac{k\lambda_k}{2}$,which contradicts (4.3)

But these k disjointly supported functions should satisfy $\lambda_k \le 2 \max_{1\le i\le k} \mathcal{R}(f_i)$ by Lemma 2.4. Since we just showed that $\mathcal{R}(f_i) \le \frac{\lambda_k}{2}$ for all i, we get a contradiction and hence Algorithm 1 always terminates with required thresholds, hence Lemma 4.3 follows. □

Next we show that we can use any $2k+1$ approximation of f to upper-bound $\phi(f)$

Proposition 4.4. *For any $2k+1$ step approximation of f, called g,*

$$\phi(f) \le 4k\mathcal{R}(f) + 4\sqrt{2k}||f-g||_w\sqrt{\mathcal{R}(f)}$$

Let g be the $2k+1$ approximation of f defined as before($g(v) = \psi_{t_0,...,t_{2k}}(f(v))$) with thresholds $0 = t_0 \le \cdots \le t_{2k}$. We will find a function $h \in \ell^2(V, w)$ such that each threshold set of h is also a threshold set of $f(supp(h) = supp(f))$ and

$$\frac{\sum_{u\sim v} w(u,v)|h(u)-h(v)|}{\sum_v w(v)h(v)} \le 4k\mathcal{R}(f) + 4\sqrt{2}k||f-g||_w\sqrt{\mathcal{R}(f)} \tag{4.4}$$

Then we can apply Lemma 2.6 to (4.4) and Proposition 4.2 follows.

Let $\mu : \mathbb{R} \to \mathbb{R}, \mu(x) = |x - \psi_{t_0,\dots,t_{2k}}(x)|$, note that $\mu(f(v)) = |f(v) - \psi_{t_0,\dots,t_{2k}}(f(v))| = |f(v) - g(v)|$.

We define $h \in \ell^2(V, w)$ as

$$h(v) = \int_0^{f(v)} \mu(x)dx$$

Notice that $h(u) \geq h(v)$ if and only if $f(u) \geq f(v)$ since $\mu(x)$ is non-negative. This implies that the threshold sets of h and the threshold sets of f are the same. We prove (4.4) by making two claims.

Claim 4.5. *For every vertex v,*

$$h(v) \geq \tfrac{1}{8k} f^2(v)$$

Proof. When $f(v) = 0$, we have $h(v) = 0$ and hence our inequality holds. Suppose $f(v) \in [t_i, t_{i+1}]$. Note that t_i can be rewritten as $\sum\limits_{j=0}^{i-1}(t_{j+1} - t_j)$ and hence by Cauchy-Schwarz inequality

$$f^2(v) = (t_i + f(v) - t_i)^2 = (\sum_{j=0}^{i-1}(t_{j+1} - t_j) + (f(v) - t_i))^2 \leq 2k \cdot (\sum_{j=0}^{i-1}(t_{j+1} - t_j)^2 + (f(v) - t_i)^2)$$

On the other hand, by the definition of h, $h(v) = \sum\limits_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} \mu(x)dx + \int_{t_i}^{f(v)} \mu(x)dx$. Note that when $t_j < x \leq \frac{t_j + t_{j+1}}{2}$, we have $\mu(x) = x - t_j$ and when $\frac{t_j + t_{j+1}}{2} < x \leq x_{j+1}$, we have $\mu(x) = t_{j+1} - x$. Hence

$$\begin{aligned}
\int_{t_j}^{t_{j+1}} \mu(x)dx &= \int_{t_j}^{\frac{t_j+t_{j+1}}{2}} (x - t_j)dx + \int_{\frac{t_j+t_{j+1}}{2}}^{t_{j+1}} (t_{j+1} - x)dx \\
&= \frac{1}{2}(\frac{t_j + t_{j+1}}{2})^2 - t_j(\frac{t_j + t_{j+1}}{2}) - \frac{1}{2}t_j^2 + t_j^2 + t_{j+1}^2 - \frac{1}{2}(t_{j+1})^2 - t_{j+1}(\frac{t_j + t_{j+1}}{2}) + \frac{1}{2}(\frac{t_j + t_{j+1}}{2})^2 \\
&= \frac{1}{4}(t_{j+1} - t_j)^2
\end{aligned}$$

We can expand $\int_{t_i}^{f(v)} \mu(x)dx$ similarly and get $\int_{t_i}^{f(v)} \mu(x)dx \geq \frac{1}{4}(f(v) - t_i)^2$

Hence we get $h(v) \geq \sum\limits_{j=0}^{i-1} \frac{1}{4}(t_{j+1} - t_j)^2 + \frac{1}{4}(f(v) - t_i)^2 = \frac{1}{8k} f^2(v)$

$\square$

Claim 4.6. *For any pair of vertices $u, v \in V$,*

$$|h(v) - h(u)| \leq \tfrac{1}{2}|f(v) - f(u)| \cdot (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - g(u)|)$$

Proof. Without loss of generality, we assume $f(u) \leq f(v)$. By definition of μ, for $x \in [f(u), f(v)]$,

$$\begin{aligned}
\mu(x) \leq \min\{|x - g(u)|, |x - g(v)|\} &\leq \frac{|x - g(u)| + |x - g(v)|}{2} \\
&= \frac{|(x - f(u) + (f(u) - g(u)))| + |(x - f(v)) + (f(v) - g(v))|}{2} \\
&\leq \frac{1}{2}(|x - f(u)| + |f(u) - g(u)| + |x - f(v)| + |f(v) - g(v)|) \\
&= \frac{1}{2}(x - f(u) + f(v) - x + |f(u) - g(u)| + |f(v) - g(v)|) \\
&= \frac{1}{2}(|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - g(u)|)
\end{aligned}$$

where the third inequality follows from Triangle inequality and the last two equality follows by $x \in [f(u), f(v)]$. Therefore,

$$h(v) - h(u) = \int_{f(u)}^{f(v)} \mu(x)dx \leq |f(v) - f(u)| \cdot \max_{x\in[f(u),f(v)]} \mu(x)$$
$$\leq \frac{1}{2}|f(v) - f(u)| \cdot (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - g(u)|)$$

Since we assumed $f(v) \geq f(u)$, we know $h(v) \geq h(u)$, hence the above result also holds for $|h(v) - h(u)|$ $\square$

Now we are ready to prove Proposition 4.2

Proof. By Claim 4.4,

$$\begin{aligned}
\sum_{u\sim v} w(u,v)|h(u) - h(v)| &\leq \sum_{u\sim v} \frac{1}{2} w(u,v)|f(v) - f(u)| \cdot (|f(u) - g(u)| + |f(v) - g(v)| + |f(v) - f(u)|) \\
&= \frac{1}{2}\sum_{u\sim v} w(u,v)|f(v) - f(u)|^2 + \frac{1}{2}\sum_{u\sim v} w(u,v)|f(v) - f(u)| \cdot (|f(u) - g(u) + f(v) - g(v)|) \\
&\leq \frac{1}{2}\mathcal{R}(f) + \frac{1}{2}\sqrt{\sum_{u\sim v} w(u,v)|f(v) - f(u)|^2}\sqrt{\sum_{u\sim v} w(u,v)(|f(u) - g(u)| + |f(v) - g(v)|)^2} \\
&\leq \frac{1}{2}\mathcal{R}(f) + \frac{1}{2}\sqrt{\mathcal{R}(f)} \cdot \sqrt{2\sum_{u\sim v} w(u,v)(|f(u) - g(u)|^2 + |f(v) - g(v)|^2)} \\
&= \frac{1}{2}\mathcal{R}(f) + \frac{1}{2}\sqrt{\mathcal{R}(f)} \cdot \sqrt{2||f - g||_w^2}
\end{aligned}$$

The second inequality follows from Cauchy-Schwarz inequality and the third inequality holds since $(x+y)^2 \leq 2x^2 + y^2$ for all x, y. On the other hand, by Claim 4.3,

$$\sum_v w(v)h(v) \geq \tfrac{1}{8k}\sum_v w(v)f^2(v) = \tfrac{1}{8k}||f||_w^2 = \tfrac{1}{8k}$$

Hence $\frac{\sum_{u\sim v} w(u,v)|h(v)-h(u)|}{\sum_v w(v)h(v)} \leq \frac{\frac{1}{2}\mathcal{R}(f) + \frac{1}{2}\sqrt{\mathcal{R}(f)}\cdot\sqrt{2||f-g||_w^2}}{\frac{1}{8k}} = 4k\mathcal{R}(f) + 4\sqrt{2}k||f - g||_w\sqrt{\mathcal{R}(f)}$

Since h and f have the same threshold sets, we know that $\phi(f) = \phi(h)$ and hence by Lemma 2.6,

$$\phi(f) = \phi(h) \leq \tfrac{\sum_{u\sim v} w(u,v)|h(v)-h(u)|}{\sum_v w(v)h(v)} \leq 4k\mathcal{R}(f) + 4\sqrt{2}k||f - g||_w\sqrt{\mathcal{R}(f)}$$

$\square$

Now we can prove Theorem 1.2,

Proof. Let g be as defined in Lemma 4.3. By Proposition 4.2, we get

$$\phi(f) \leq 4k\mathcal{R}(f) + 4\sqrt{2}k||f - g||_w\sqrt{\mathcal{R}(f)} \leq 4k\mathcal{R}(f) + \tfrac{8\sqrt{2}k\mathcal{R}(f)}{\sqrt{\lambda_k}} \leq \tfrac{12\sqrt{2}k\mathcal{R}(f)}{\sqrt{\lambda_k}} \leq O(k)\tfrac{\lambda_2}{\sqrt{\lambda_k}}$$

The last inequality since $\mathcal{R}(f) \leq \lambda_2$ by our assumption $\square$

5 Spectral partition problem

Given data points $X_1, X_2, ..., X_n$ and their similarities $f(X_i, X_j)$. Partition the data points into different groups so that similar points are in the same group and dissimilar points are in the different groups. This is a very common problem in machine learning area. We can convert this problem into a graph problem by constructing a graph $G = (V, E, W)$ such that each point is assigned as a vertex and the weight of edges is given by the similarity of the two vertices. Suppose we would like to partition the graph into two groups. We can find the cut of the graph such that we can partition the graph into two sets S and $\overline{S}$ where the weight of edges connecting vertices in A to vertices in B are minimum, where

$$cut(A, B) = \sum_{i \in S, j \in \overline{S}} w_{ij}$$

And hence we can further define Cheeger's Cut as follows,

Definition 5.1 (Cheeger's Cut)**.** *Given a graph G and a vertex partition $(S, \overline{S})$, the cheeger's cut is given by*

$$h(S) = \frac{cut(S, \overline{S})}{\min\{vol(S), vol(\overline{S})\}}$$

Notice that Cheeger's Cut is actually equivalent to the conducatance of a graph. However, $A = \emptyset$ would be an optimal solution but it is not practical. Hence we would like to find a balanced solution S such that $|S| = |\overline{S}|$. Hence we can rewrite finding the cut as

$$\begin{aligned} \min_{|S|=|\overline{S}|} cut(S, \overline{S}) &= \min_{x \in \{+1,-1\}^n, \vec{1} \cdot x = 0} \frac{1}{4} w_{ij}(y_i - y_j)^2 \\ &= \frac{1}{4} \min_{x \in \{+1,-1\}^n, \vec{1} \cdot x = 0} x^T L_G x \end{aligned}$$

where $L_G = D - W$ is the Laplacian matrix and W is the weight matrix. Note for $i \in S$, x_i = 1. For $i \in \overline{S}$, $x_i = -1$. Requiring the partition to be completely balanced is too somewhat restrictive. The conductance of the graph gives a $O(\sqrt{\phi(G)})$ approximation to ensure size of set A and size of set B not to be too small and the proof of Cheeger's Inequality leads to the flowing algorithm.

Algorithm 2: Cheeger's sparsest cut algorithm

For a graph G, calculate the second eigenvector v_2 for the Lapalcian matrix $\mathcal{L}$ of G
Sort the vertices of V according to v_2: $v_2(u_1) \leq v_2(u_2) \leq ... \leq v_2(u_n)$
Return the threshold cut $(v_1, ..., v_i, v_{i+1}, ..., v_n)$ with the smallest conductance

The algorithm returns a vertex set S with $vol(S) \leq \frac{vol(V)}{2}$. By the right side of the Cheeger's Inequality, we have $\phi(S) \leq \sqrt{2\lambda_2}$. The above result can be improved by our improved Cheeger's Inequality. We consider the general case that we want to parition the graph into k parts. We define $\phi_k(G) = \min_{S_1,...,S_k} \max_{1 \leq i \leq k} \phi(S_i)$ where $S_1, ..., S_k$ are non-empty subsets of V and we can get following results[KLLOT13]:

Corollary 5.0.1. *For every undirected graph G and $l > k \geq 2$, it holds that*

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

This shows that λ_k is a better approximation of $\phi_k(G)$ when there is a large gap between λ_k and λ_l. Hence this gives a better performance guarantee of the spectral partition algorithm.

The benefits of the idea of Cheeger's Cut does not limited to partition a graph. For example, combining k-means and Laplacian eigenvectors can complete the clustering with non-convex boundaries. This gives a better performance than normal k-means partitioning. The below pictures are an example of partitioning a set of points into 2 clusters.

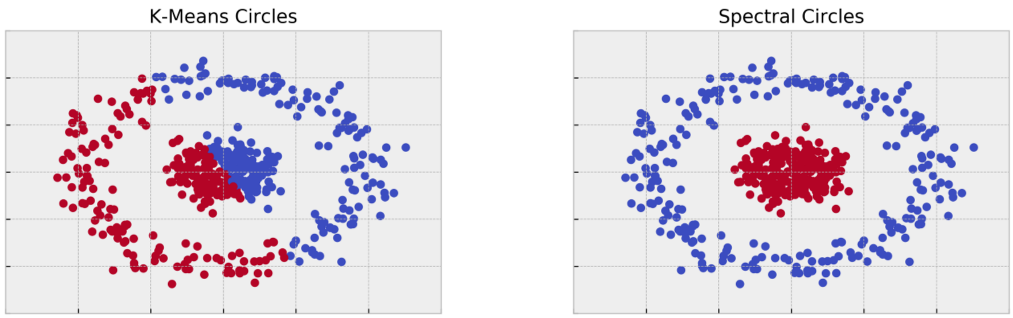

Figure 1: (a)k-mean clustering (Fleshman, 2019) (b) Spectral clustering (Fleshman, 2019)

References

[Alo86] N.Alon. Eogenvalues and expanders. Combinatorica, 6, 83-96, 1986

[AM85] N.Alon, V. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. Journal of Combinatorial Theory, Series B, 38(1), 73-88, 1985

[KLLOT13] T.C. Kwok, L.C. Lau, Y.T. Lee, S. Oveis Gharan, L. Trevisan. Improved Cheeger's inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. In Proceedings of the 45th Annual Symposium on Theory of Computing (STOC), 11–20, 2013.

[HLW06] S.Horry, N.Linial, A. Wigderson. Expander graphs and their applications. Bulletin of the American Mathematical Society 43(4), 439-561, 2006

[KLL16] Tsz Chiu Kwok, Lap Chi Lau, Yin Tat Lee. Improved Cheeger's inequality and analysis of local graph partitioning using vertex expansion and expansion profile SIAM Journal on Computing, 46(3), 890-910, 2017. (Conference version in Proceedings of the 27th ACM-SIAM Symposium on Discrete Algorithms (SODA), 1848-1861, 2016.

[CF05] Chung, F. (2005). Laplacians and the Cheeger Inequality for Directed Graphs. Annals of Combinatorics, 9(1), 1-19.

[LT11] Louis, A., Raghavendra, P., Tetali, P., & Vempala, S. (2011). Algorithmic Extensions of Cheeger's Inequality to Higher Eigenvalues and Partitions. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques Lecture Notes in Computer Science, 315-326.

[F10] F. Chung, Four proofs for the Cheeger inequality and graph partition algorithms, Fourth International Congress of Chinese Mathematicians, 2010, pp. 331–349.

[J69] J. Cheeger, A lower bound for the smallest eigenvalue of the Laplacian, Problems in Analysis, Papers dedicated to Salomon Bochner, 1969, Princeton University Press, Princeton, 195-199.

[F19] Fleshman, W. (2019, February 21). Spectral Clustering. Retrieved August 14, 2020, from https://towardsdatascience.com/spectral-clustering-aba2640c0d5b