

Notes on Spectral Graph Theory

Mufei Li

October 2019

This notes is based on the reading group on spectral graph theory at AWS Shanghai AI Lab, Fall 2019. Our primary reference are [1] and [2], accompanied by [3]. Some sections have not been covered directly in the reading group and I mark the titles of them with an *.

Contents

1	Background	1
1.1	Linear Algebra	1
2	Introduction	2
2.1	Laplacian Matrices	2
3	Fiedler Value, Edge Expansion, and Cheeger's Inequalities	5
3.1	Motivation from Graph Cut	5
3.2	Edge Expansion	6
3.3	The Fiedler Value	6
3.4	Cheeger's Inequalities	7
3.4.1	The Case of d -regular Unweighted Graphs	7

1 Background

This section is mainly for a preparation and a refresh of the memory. You can skip it first and go back if necessary.

1.1 Linear Algebra

Theorem 1.1 (Variational Characterization of Eigenvalues). Let $M \in \mathbb{R}^{n \times n}$ be a symmetric matrix, and $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the eigenvalues of M . Then

$$\lambda_k = \min_{k\text{-dim } V} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \quad (1)$$

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

Proof. By the spectral theorem, let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be eigenvectors of M corresponding to eigenvalues $\lambda_1, \dots, \lambda_n$ that form an orthonormal basis of \mathbb{R}^n . $\forall \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \mathbf{x} = \sum_{i=1}^n a_i \mathbf{v}_i, \mathbf{x}^T \mathbf{x} = \sum_{i=1}^n a_i^2, \mathbf{x}^T M \mathbf{x} = \sum_{i=1}^n \lambda_i a_i^2$.

Consider V a space spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$. It follows that $\mathbf{x}^T M \mathbf{x} \leq \lambda_k \sum_{i=1}^n a_i^2$ and $R_M(x) \leq \lambda_k$. In particular λ_k is obtained when $\mathbf{x} = a_k \mathbf{v}_k$. Thus $\lambda_k \geq \min_{k\text{-dim } V} \max_{\mathbf{x} \in V \setminus \{\mathbf{0}\}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$.

For the other direction, let V be any k -dim space. Let S be the span of $\mathbf{v}_k, \dots, \mathbf{v}_n$. Then $\exists \mathbf{x} \in V \cap S, \mathbf{x} \neq \mathbf{0}$. Since $\mathbf{x} = \sum_{i=k}^n b_i \mathbf{v}_i, R_M(\mathbf{x}) \geq \lambda_k$. \square

Remark 1.1. Several facts can be immediately obtained from Theorem 1.1:

1. If λ_1 is the smallest eigenvalue of a real symmetric matrix M , then $\lambda_1 = \min_{\mathbf{x} \neq \mathbf{0}} R_M(\mathbf{x})$. Furthermore, every minimizer is an eigenvector of λ_1 .
2. If λ_n is the largest eigenvalue of a real symmetric matrix M , then $\lambda_n = \max_{\mathbf{x} \neq \mathbf{0}} R_M(\mathbf{x})$. Furthermore, every maximizer is an eigenvector of λ_n .
3. If λ_1 is the smallest eigenvalue of a real symmetric matrix M , and \mathbf{x}_1 is an eigenvector of λ_1 , then $\lambda_2 = \min_{\mathbf{x} \neq \mathbf{0}, \mathbf{x} \perp \mathbf{x}_1} R_M(\mathbf{x})$. Furthermore, every minimizer is an eigenvector of λ_2 .

2 Introduction

2.1 Laplacian Matrices

A graph G is specified by its vertex set V and edge set E . Given an undirected graph $G = (V, E)$, the approach of spectral graph theory is to associate a symmetric real-valued matrix to G , and to relate the eigenvalues of the matrix to combinatorial properties of G .

The most natural matrix associated to G is the weighted adjacency matrix A such that $A_{ij} = A_{ji} > 0$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise. If A_{ij} 's are constant for $(i, j) \in E$, we consider the graphs to be unweighted. Otherwise, they are considered to be weighted. Note that unless otherwise specified, the edges are unique up to symmetry in E , i.e. (i, j) and (j, i) are considered to be the same edge in E . We define the **Laplacian matrix** of G to be

$$L = D - A, \quad (2)$$

where the degree matrix D is a diagonal matrix with $D_{ii} = A_i \mathbf{1}$.

The Laplacian matrix gives rise to a useful quadratic form associated with a graph:

$$\mathbf{x}^T L \mathbf{x} = \sum_{(u,v) \in E} A_{uv} (x_u - x_v)^2. \quad (3)$$

If we consider \mathbf{x} to be a function $\mathbf{x} : V \rightarrow \mathbb{R}$, then this form measures the smoothness of the function and the evaluated result is small if \mathbf{x} does not change much over any edge (with large positive weights). From the perspective of graph cut, if $\mathbf{x} \in \{0, 1\}^V$ is a Boolean vector representing a cut in the graph and $A_{ij} = A_{ji} = 1, \forall (i, j) \in E$, then the evaluated result counts the number of edges that cross the cut. It's also quite often that we consider instead normalized Laplacian matrices $D^{-1}L$ or $D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$.

Remark 2.1 (Properties of Laplacian Matrices). Laplacian matrices have several properties:

1. They are positive semidefinite. It follows that their eigenvalues are non-negative.
2. They are additive. For a graph $G = (V, E)$, denote its Laplacian matrix by L_G . An edge (i, j) with $i, j \in V$ can be viewed as a graph $\Delta G = (V, \{(i, j)\})$. Let \tilde{G} be the graph obtained by adding the edge (i, j) to G , we have $L_{\tilde{G}} = L_G + L_{\Delta G}$.

3. Their eigenvalues are invariant under graph isomorphism. If we permute the vertices then the eigenvectors are similarly permuted. Let P be a permutation matrix, then

$$L\mathbf{v} = \lambda\mathbf{v} \text{ iff } (PLP^T)(P\mathbf{v}) = \lambda P\mathbf{v}. \quad (4)$$

Proof.

1. That Laplacian matrices are positive semi-definite comes from the quadratic form we derived above. The existence of eigenvalues is justified by the real Spectral Theorem. Since the matrices are positive semi-definite, the eigenvalues are non-negative.
2. This can be verified by applying the definition.
3. Note that $P^T P = I$.

□

Below We prove several relations between the Laplacian matrices and certain purely combinatorial properties of the corresponding graphs. We will start with unnormalized Laplacian matrices and then extend the results to the normalized cases.

Theorem 2.1. Let $G = (V, E)$ be a weighted undirected graph and denote its unnormalized Laplacian matrix by L . Let $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the real eigenvalues of L with multiplicities in nondecreasing order. Then

1. $\lambda_1 = 0$.
2. $\lambda_k = 0$ iff G has at least k connected components.
3. Let D denote the degree matrix of G and let $D_{\max} = \max_{v \in V} D_{vv}$. Then $\lambda_n \leq 2D_{\max}$. In particular, the equality holds when all nodes have the same degree (i.e. the graph is a D_{\max} -regular graph) and at least one of the connected components of G is bipartite.

Note that the second result suggests that the multiplicity of 0 as an eigenvalue is precisely the number of connected components of G .

Proof.

1. By the variational characterization of eigenvalues, we have

$$\lambda_1 = \min_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} R_L(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} \frac{\sum_{(u,v) \in E} A_{uv}(x_u - x_v)^2}{\sum_{v \in V} x_v^2} \geq 0. \quad (5)$$

If we take \mathbf{x} to be an all-one vector $\mathbf{1} = (1, \dots, 1)$, then $R_L(\mathbf{x}) = 0$. Hence $\lambda_1 = 0$.

2. By the variational characterization of eigenvalues, we have

$$\lambda_k = \min_{k\text{-dim } S} \max_{\mathbf{x} \in S \setminus \{\mathbf{0}\}} R_L(\mathbf{x}) \quad (6)$$

If $\lambda_k = 0$, then $\exists S$ a k -dim subspace such that $\forall \mathbf{x} \in S$, it is constant within each connected component. Therefore the graph has at least k connected components.

Conversely, assume G has at least k connected components. Let $\mathbf{1}^{(i)} \in \mathbb{R}^n$ be a vector such that $\mathbf{1}_v^{(i)} = 1$ if v belongs to the i -th connected component and 0 otherwise. Consider S to be the span of $\mathbf{1}^{(1)}, \dots, \mathbf{1}^{(k)}$, then $R_L(\mathbf{x}) = 0 \forall \mathbf{x} \in S$. Hence $\lambda_k = 0$.

3.

$$\mathbf{x}^T L \mathbf{x} = \sum_{(u,v) \in E} A_{uv} (x_u - x_v)^2 = \sum_{v \in V} D_{vv} x_v^2 - 2 \sum_{(u,v) \in E} A_{uv} x_u x_v \quad (7)$$

$$= 2 \sum_{v \in V} D_{vv} x_v^2 - \left(\sum_{v \in V} D_{vv} x_v^2 + 2 \sum_{(u,v) \in E} A_{uv} x_u x_v \right) \quad (8)$$

$$= 2 \sum_{v \in V} D_{vv} x_v^2 - \sum_{(u,v) \in E} A_{uv} (x_u + x_v)^2 \quad (9)$$

$$\leq 2D_{\max} \sum_{v \in V} x_v^2 - \sum_{(u,v) \in E} A_{uv} (x_u + x_v)^2 \quad (10)$$

By the variational characterization of eigenvalues,

$$\lambda_n = \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} R_L(\mathbf{x}) \leq \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} 2D_{\max} - \frac{\sum_{(u,v) \in E} A_{uv} (x_u + x_v)^2}{\sum_{v \in V} x_v^2} \leq 2D_{\max}. \quad (11)$$

The equality holds when $D_{vv} = D_{\max}, \forall v \in V$ and $\sum_{(u,v) \in E} A_{uv} (x_u + x_v)^2 = 0$. Since $\mathbf{x} \neq \mathbf{0}$, $\exists U = \{v | v \in V, x_v = c\}$ and $V = \{v | v \in V, x_v = -c\}$ for some $c \neq 0$ s.t. the nodes in U are only connected to some nodes in V and vice versa. In other words, they form a connected component that is bipartite.

□

Theorem 2.2. Let $G = (V, E)$ be a weighted undirected graph. Let $\hat{L} = D^{-1}L$ and $\tilde{L} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$ be its two normalized Laplacian matrices. Then

1.

$$\mathbf{x}^T \tilde{L} \mathbf{x} = \sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} - \frac{x_v}{\sqrt{D_{vv}}} \right)^2 \quad (12)$$

2. λ is an eigenvalue of \hat{L} with eigenvector u iff $Lu = \lambda Du$.

3. λ is an eigenvalue of \hat{L} with eigenvector u iff λ is an eigenvalue of \tilde{L} with eigenvector $w = D^{\frac{1}{2}}u$.

4. 0 is an eigenvalue of \hat{L} with eigenvector $\mathbf{1}$ and an eigenvalue of \tilde{L} with eigenvector $D^{\frac{1}{2}}\mathbf{1}$.

5. \hat{L} and \tilde{L} are positive semi-definite and have n non-negative real-valued eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_n$.

6. The multiplicity of the eigenvalue 0 of both \hat{L} and \tilde{L} equals the number of connected components in G .

7. $\lambda_n \leq 2$. In particular, the equality holds when at least one of the connected components of G is bipartite.

Proof.

1.

$$\mathbf{x}^T \tilde{L} \mathbf{x} = \mathbf{x}^T (D^{-\frac{1}{2}} L D^{-\frac{1}{2}}) \mathbf{x} = \mathbf{x}^T (D^{-\frac{1}{2}} (D - A) D^{-\frac{1}{2}}) \mathbf{x} = \mathbf{x}^T \mathbf{x} - \mathbf{x}^T D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \mathbf{x} \quad (13)$$

$$= \mathbf{x}^T \mathbf{x} - 2 \sum_{(u,v) \in E} \frac{A_{uv}}{\sqrt{D_{uu}} \sqrt{D_{vv}}} x_u x_v \quad (14)$$

$$= \sum_{(u,v) \in E} \frac{A_{uv}}{D_{uu}} x_u^2 + \frac{A_{uv}}{D_{vv}} x_v^2 - 2 \frac{A_{uv}}{\sqrt{D_{uu}} \sqrt{D_{vv}}} x_u x_v \quad (15)$$

$$= \sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} - \frac{x_v}{\sqrt{D_{vv}}} \right)^2 \quad (16)$$

2. The verification takes a direct computation.

3.

$$\hat{L}u = D^{-1}Lu = \lambda u \iff D^{-1}LD^{-\frac{1}{2}}w = \lambda D^{-\frac{1}{2}}w \iff \tilde{L}w = \lambda w \quad (17)$$

4. $\hat{L}\mathbf{1} = D^{-1}L\mathbf{1} = \mathbf{1} - D^{-1}A\mathbf{1} = 0$. Hence $\mathbf{1}$ is an eigenvector of \hat{L} with eigenvalue 0. Apply the previous result we have that $D^{\frac{1}{2}}\mathbf{1}$ is an eigenvector of \tilde{L} with eigenvalue 0.

5. That \tilde{L} is positive semi-definite is from the first result. The existence of its n eigenvalues is justified by the spectral theorem. Since it's positive semi-definite, all of its eigenvalues are non-negative. By the third result, the eigenvalues of \hat{L} are also non-negative, hence \hat{L} is positive semi-definite.

6. We only need to prove the case for \tilde{L} and the case for \hat{L} will follow from the third result.

By the variational characterization of eigenvalues, we have

$$\lambda_k = \min_{k-\dim S} \max_{\mathbf{x} \in S \setminus \{\mathbf{0}\}} R_{\tilde{L}}(\mathbf{x}) = \min_{k-\dim S} \max_{\mathbf{x} \in S \setminus \{\mathbf{0}\}} \frac{\sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} - \frac{x_v}{\sqrt{D_{vv}}} \right)^2}{\sum_{v \in V} x_v^2} \quad (18)$$

If $\lambda_k = 0$, then $\exists S$ a k -dim subspace such that $\forall \mathbf{x} \in S$, $D^{-\frac{1}{2}}\mathbf{x}$ is constant within each connected component. Therefore the graph has at least k connected components.

Conversely, assume G has at least k connected components. Let $D^{\frac{1}{2}}\mathbf{1}^{(i)} \in \mathbb{R}^n$ be a vector such that $(D^{\frac{1}{2}}\mathbf{1}^{(i)})_v = \sqrt{D_{vv}}$ if v belongs to the i -th connected component and 0 otherwise. Consider S to be the span of $D^{\frac{1}{2}}\mathbf{1}^{(1)}, \dots, D^{\frac{1}{2}}\mathbf{1}^{(k)}$, then $R_{\tilde{L}}(\mathbf{x}) = 0 \ \forall \mathbf{x} \in S$. Hence $\lambda_k = 0$.

It follows that $\lambda_k = 0$ iff G has at least k connected components. Hence the multiplicity of the eigenvalue 0 of \tilde{L} equals the number of connected components in G .

7.

$$\mathbf{x}^T \tilde{L} \mathbf{x} = \sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} - \frac{x_v}{\sqrt{D_{vv}}} \right)^2 = \sum_{v \in V} x_v^2 - 2 \sum_{(u,v) \in E} \frac{A_{uv}}{\sqrt{D_{uu}}\sqrt{D_{vv}}} x_u x_v \quad (19)$$

$$= 2 \sum_{v \in V} x_v^2 - \left(\sum_{v \in V} x_v^2 + 2 \sum_{(u,v) \in E} \frac{A_{uv}}{\sqrt{D_{uu}}\sqrt{D_{vv}}} x_u x_v \right) \quad (20)$$

$$= 2 \sum_{v \in V} x_v^2 - \sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} + \frac{x_v}{\sqrt{D_{vv}}} \right)^2 \quad (21)$$

By the variational characterization of eigenvalues,

$$\lambda_n = \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} R_{\tilde{L}}(\mathbf{x}) \leq \max_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} 2 - \frac{\sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} + \frac{x_v}{\sqrt{D_{vv}}} \right)^2}{\sum_{v \in V} x_v^2} \leq 2 \quad (22)$$

The equality holds when $\sum_{(u,v) \in E} A_{uv} \left(\frac{x_u}{\sqrt{D_{uu}}} + \frac{x_v}{\sqrt{D_{vv}}} \right)^2 = 0$. Since $\mathbf{x} \neq \mathbf{0}$, $\exists U = \{v | v \in V, \frac{x_v}{\sqrt{D_{vv}}} = c\}$ and $V = \{v | v \in V, \frac{x_v}{\sqrt{D_{vv}}} = -c\}$ for some $c \neq 0$ s.t. the nodes in U are only connected to some nodes in V , vice versa. In other words, they form a connected component that is bipartite.

□

3 Fiedler Value, Edge Expansion, and Cheeger's Inequalities

3.1 Motivation from Graph Cut

A graph cut is a partition of V into two disjoint nonempty sets (clusters) A and B . Intuitively, a good cut should minimize the connections between the sets and maximize the connections within each set. It

then seems natural to measure the quality of a cluster by

$$\sum_{u \in A} \sum_{v \in B} A_{uv} \quad (23)$$

Unfortunately this metric fails to consider the connections within a set and can be minimized by some undesired trivial cuts, e.g. a cut such that one set has only one vertex.

3.2 Edge Expansion

To address the issue, we now consider a different metric.

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

$$\phi(S) := \frac{\sum_{u \in S} \sum_{v \in V \setminus S} A_{uv}}{\sum_{u \in S} \sum_{v \in V} A_{uv}} = \frac{\sum_{u \in S} \sum_{v \in V \setminus S} A_{uv}}{\sum_{u \in S} d_u} \quad (24)$$

the **edge expansion** of S . The quantity $\phi(S)$ is the average fraction of connections pointing outside of S for a random element of S .

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

$$\phi(S, V \setminus S) := \max\{\phi(S), \phi(V \setminus S)\} = \frac{\sum_{u \in S} \sum_{v \in V \setminus S} A_{uv}}{\min\{\sum_{u \in S} d_u, \sum_{v \in V \setminus S} d_v\}}.^1 \quad (25)$$

Note that it's also quite often to call $\phi(S, V \setminus S)$ as the **conductance** of S or $V \setminus S$.

The edge expansion/conductance of the graph G is defined as

$$\phi(G) := \min_S \phi(S, V \setminus S) = \min_{S: 1 \leq |S| \leq \lfloor \frac{|V|}{2} \rfloor} \phi(S) \quad (26)$$

3.3 The Fiedler Value

We've learned that the second smallest eigenvalue λ_2 of Laplacian matrices is 0 iff the graph is disconnected. In fact, Miroslav Fiedler suggested considering λ_2 to be a measure of how well connected the graph is and called it the "algebraic connectivity" of a graph. We will follow the convention of [2] and call this value the **Fiedler value**. Fiedler proved that the further λ_2 is from 0, the better connected the graph is.

Finding cuts of small expansion is a problem with several applications. Unfortunately, computing the optimal cut with respect to edge expansion is NP-hard and it is an open question if there is a polynomial-time approximation with a constant-factor approximation ratio.

The following algorithm 1 was proposed by Fiedler, and it works well in practice when \mathbf{x} is the eigenvector corresponding to the Fiedler value.

The algorithm can be implemented in $O(|E| + |V| \log |V|)$. First we can sort the vertices in $O(|V| \log |V|)$. The cut of minimal expansion that respects the sorted order can be found in time $O(E)$. In order to find the cut, we only need to compute $e_k = E(\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_n\})$ for $k = 1, \dots, n - 1$. When $k = 1$, this is just the degree of v_1 excluding its self loops. From k to $k + 1$, we have $e_{k+1} = e_k - E(\{v_{k+1}\}, \{v_1, \dots, v_k\}) + E(\{v_{k+1}\}, \{v_{k+2}, \dots, v_n\})$. On operation, that can be done in time $O(D_{k+1, k+1})$. Thus the total running time is of order of $\sum_{v \in V} D_{vv}$, i.e. $O(|E|)$.

¹Here we are following the convention used in [1]. Note that it is common in the literature to use the notation $\phi(S)$ to refer to the quantity that we call $\phi(S, V \setminus S)$.

Algorithm 1 Approximating the Optimal Graph Cut

Require: graph $G = (V, E)$, $\mathbf{x} \in \mathbb{R}^{|V|}$

- 1: Sort the vertices according to \mathbf{x} and relabel them in order as v_1, \dots, v_n .
 - 2: $k^* = \arg \min_k \phi(\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_n\})$
 - 3: Return $(\{v_1, \dots, v_{k^*}\}, \{v_{k^*+1}, \dots, v_n\})$
-

3.4 Cheeger's Inequalities

Cheeger's inequality gives a tight connection between edge expansion and the Fiedler value.

Theorem 3.1 (Cheeger's Inequalities). Let G be an undirected regular graph and $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ be the eigenvalues of the normalized Laplacian $\tilde{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$. Then

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2} \quad (27)$$

We will start with the case of d -regular unweighted graphs and later extend the result to irregular graphs.

3.4.1 The Case of d -regular Unweighted Graphs

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

Let S be a set of vertices such that $\phi(S, V \setminus S) = \phi(G)$. Note that $\forall S \subset V$, we have

$$\phi(S) = \frac{E(S, V \setminus S)}{d|S|} = \frac{\sum_{u \in S} \sum_{v \in V \setminus S} (1 - 0)^2}{d|S|} = R_{\tilde{L}}(\mathbf{1}_S). \quad (28)$$

It follows that

$$R_{\tilde{L}}(\mathbf{1}_S) \leq \phi(G) \quad (29)$$

$$R_{\tilde{L}}(\mathbf{1}_{V \setminus S}) \leq \phi(G) \quad (30)$$

From the variational characterization of eigenvalues, we have that

$$\lambda_2 = \min_{2-\dim X} \max_{\mathbf{x} \in X \setminus \{0\}} R_{\tilde{L}}(\mathbf{x}) \quad (31)$$

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

Lemma 3.1. Let \mathbf{x} and \mathbf{y} be two orthogonal vectors, and let M be a positive semi-definite matrix. Then

$$R_M(\mathbf{x} + \mathbf{y}) \leq 2 \max\{R_M(\mathbf{x}), R_M(\mathbf{y})\} \quad (32)$$

Proof. By the spectral theorem, M has n eigenvalues corresponding to n eigenvectors that form an orthonormal basis. Since M is positive semi-definite, the eigenvalues are non-negative. For convenience, we order them as $0 \leq \lambda_1 \leq \dots \leq \lambda_n$, corresponding to eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. Let $\mathbf{x} = \sum_{i=1}^n a_i \mathbf{v}_i$, $\mathbf{y} = \sum_{i=1}^n b_i \mathbf{v}_i$.

$$R_M(\mathbf{x} + \mathbf{y}) = \frac{\sum_i \lambda_i (a_i + b_i)^2}{\|\mathbf{x} + \mathbf{y}\|^2} = \frac{\sum_i \lambda_i (a_i + b_i)^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} \leq \frac{\sum_i 2\lambda_i (a_i^2 + b_i^2)}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} \text{ (Cauchy-Schwarz inequality)} \quad (33)$$

$$= \frac{2R_M(\mathbf{x}) \cdot \|\mathbf{x}\|^2 + 2R_M(\mathbf{y}) \cdot \|\mathbf{y}\|^2}{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2} \leq 2 \max\{R_M(\mathbf{x}), R_M(\mathbf{y})\} \quad (34)$$

□

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

Below we will prove a stronger result.

Lemma 3.2. Let \mathbf{x} be a vector orthogonal to $\mathbf{1}$ and let $(S, V \setminus S)$ be the cut found by Fiedler's algorithm given \mathbf{x} . Then

$$\phi(S, V \setminus S) \leq \sqrt{2R_{\tilde{L}}(\mathbf{x})} \quad (35)$$

Assume this lemma holds, then with $\mathbf{x} = \mathbf{v}_2$ being the eigenvector of \tilde{L} corresponding to λ_2 , we have

$$\phi(G) \leq \phi(S, V \setminus S) \leq \sqrt{2R_{\tilde{L}}(\mathbf{v}_2)} = \sqrt{2\lambda_2}. \quad (36)$$

In practice, often we use the Fiedler's algorithm with an approximate eigenvector \mathbf{x} . This result is useful for giving a guarantee on the quality of the solution found by the Fiedler's algorithm.

We will prove the lemma in two steps:

1. We show that studying non-negative vectors with at most $\frac{|V|}{2}$ entries implies the cases of vectors orthogonal to $\mathbf{1}$. Recall that $\mathbf{1}$ is an eigenvector of \tilde{L} corresponding to λ_1 and $\mathbf{v}_2 \perp \mathbf{1}$.
2. We study the cases of non-negative vectors with at most $\frac{|V|}{2}$ entries.

Formally, these two steps correspond to the two lemmas below.

Lemma 3.3. Let $\mathbf{x} \in \mathbb{R}^{|V|}$ be orthogonal to $\mathbf{1}$. Then \exists a vector $\mathbf{y} \in \mathbb{R}_{\geq 0}^{|V|}$ with at most $\frac{|V|}{2}$ non-zero entries such that

$$R_{\tilde{L}}(\mathbf{y}) \leq R_{\tilde{L}}(\mathbf{x}) \quad (37)$$

Furthermore, $\forall t \in (0, \max_v y_v)$, the cut $(\{v : y_v \geq t\}, \{v : y_v < t\})$ has been considered by Fiedler's algorithm on \mathbf{x} .

Lemma 3.4. Let $\mathbf{y} \in \mathbb{R}_{\geq 0}^{|V|}$ be a vector with non-negative entries. Then \exists a $t \in (0, \max_v y_v]$ such that

$$\phi(\{v : y_v \geq t\}) \leq \sqrt{2R_{\tilde{L}}(\mathbf{y})} \quad (38)$$

Proof of Lemma 3.3.

We observe that $\forall c \in \mathbb{R}$,

$$R_{\tilde{L}}(\mathbf{x} + c\mathbf{1}) = \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{d\|\mathbf{x} + c\mathbf{1}\|^2} = \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{d(\|\mathbf{x}\|^2 + \|c\mathbf{1}\|^2)} \leq \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{d\|\mathbf{x}\|^2} = R_{\tilde{L}}(\mathbf{x}). \quad (39)$$

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

We write $\mathbf{x}' = \mathbf{x}^+ - \mathbf{x}^-$, where $\mathbf{x}_v^+ = x'_v \mathbf{1}[x'_v > 0]$ and $\mathbf{x}_v^- = -x'_v \mathbf{1}[x'_v < 0]$. Then \mathbf{x}^+ and \mathbf{x}^- are orthogonal, non-negative and each of them has at most $\frac{|V|}{2}$ nonzero entries. Note that $\forall t$, the cut defined by the set $\{v : x_v^+ \geq t\}$ is one of the cuts considered by the Fiedler's algorithm on \mathbf{x} , because it is the cut $(\{v : x_v < t + m\}, \{v : x_v \geq t + m\})$. The case for \mathbf{x}^- is similar.

It remains to show that

$$\min\{R_{\bar{L}}(\mathbf{x}^+), R_{\bar{L}}(\mathbf{x}^-)\} \leq R_{\bar{L}}(\mathbf{x}') \leq R_{\bar{L}}(\mathbf{x}) \quad (40)$$

We claim that

$$R_{\bar{L}}(\mathbf{x}') = \frac{\sum_{(u,v) \in E} (x_u - x_v)^2}{d\|\mathbf{x}'\|^2} = \frac{\sum_{(u,v) \in E} ((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2}{d(\|\mathbf{x}^+\|^2 + \|\mathbf{x}^-\|^2)} \quad (41)$$

$$\geq \frac{\sum_{(u,v) \in E} (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2}{d(\|\mathbf{x}^+\|^2 + \|\mathbf{x}^-\|^2)} \quad (42)$$

$$= \frac{R_{\bar{L}}(\mathbf{x}^+) \cdot \|\mathbf{x}^+\|^2 + R_{\bar{L}}(\mathbf{x}^-) \cdot \|\mathbf{x}^-\|^2}{d(\|\mathbf{x}^+\|^2 + \|\mathbf{x}^-\|^2)} \quad (43)$$

$$\geq \min\{R_{\bar{L}}(\mathbf{x}^+), R_{\bar{L}}(\mathbf{x}^-)\} \quad (44)$$

The only step that we need to justify is that for every edge (u, v) we have

$$((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2 \geq (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2 \quad (45)$$

This clearly holds when both vertices are positive or negative. Without loss of generality, assume u is positive and v is negative, then

$$((x_u^+ - x_v^+) - (x_u^- - x_v^-))^2 = (x_u^+ + x_v^+)^2 \geq (x_u^+)^2 + (x_v^+)^2 = (x_u^+ - x_v^+)^2 + (x_u^- - x_v^-)^2. \quad (46)$$

□

Finally, we will provide a probabilistic proof for Lemma 3.4.

Proof of Lemma 3.4.

Since multiplication by a scalar does not affect the Rayleigh quotient of a vector, without loss of generality, we may assume that $\max_v y_v = 1$.

We consider the probabilistic process in which we pick $t > 0$ in such a way that $t^2 \sim \text{Unif}((0, 1])$ and then define the non-empty set $S_t := \{v : y_v \geq t\}$.

We will prove that

$$\frac{\mathbb{E}[E(S_t, V \setminus S_t)]}{\mathbb{E}[d|S_t|]} \leq \sqrt{2R_{\bar{L}}(\mathbf{y})}. \quad (47)$$

Remark 3.1. Let X and Y be random variables such that $\mathbb{P}[Y > 0] = 1$. Then

$$\mathbb{P}\left[\frac{X}{Y} \leq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}\right] > 0 \quad (48)$$

Let $r := \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$, then by linearity $\mathbb{E}[X - rY] = 0$. It follows that $\mathbb{P}[X - rY \leq 0] > 0$. Since $Y > 0$ with probability 1, $\mathbb{P}\left[\frac{X}{Y} \leq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}\right] > 0$.

With the Remark 3.1, if equation 47 holds, then the Lemma 3.4 will be true. We will now prove equation 47.

The denominator and the numerator of the equation 47 can be transformed as follows:

$$\mathbb{E}[d|S_t|] = d \sum_{v \in V} \mathbb{P}[v \in S_t] = d \sum_{v \in V} \mathbb{P}[y_v \geq t] = d \sum_{v \in V} \mathbb{P}[y_v^2 \geq t^2] = d \sum_{v \in V} y_v^2 \quad (49)$$

$$\mathbb{E}[E(S_t, V \setminus S_t)] = \sum_{(u,v) \in E} \mathbb{P}[(u,v) \text{ is cut}] = \sum_{(u,v) \in E} |y_v^2 - y_u^2| = \sum_{(u,v) \in E} |y_v - y_u| \cdot (y_v + y_u) \quad (50)$$

Applying Cauchy-Schwarz for twice:

$$\sum_{(u,v) \in E} |y_v - y_u| \cdot (y_v + y_u) \leq \sqrt{\sum_{(u,v) \in E} (y_v - y_u)^2} \cdot \sqrt{\sum_{(u,v) \in E} (y_v + y_u)^2} \quad (51)$$

$$\leq \sqrt{\sum_{(u,v) \in E} (y_v - y_u)^2} \cdot \sqrt{\sum_{(u,v) \in E} 2y_v^2 + 2y_u^2} \quad (52)$$

$$= \sqrt{\sum_{(u,v) \in E} (y_v - y_u)^2} \cdot \sqrt{2d \sum_{v \in V} y_v^2} \quad (53)$$

Putting everything together:

$$\frac{\mathbb{E}[E(S_t, V \setminus S_t)]}{\mathbb{E}[d|S_t|]} \leq \sqrt{2 \frac{\sum_{(u,v) \in E} (y_v - y_u)^2}{d \sum_{v \in V} y_v^2}} = \sqrt{2R_L(\mathbf{y})}. \quad (54)$$

□

References

- [1] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods, 2016.
- [2] Daniel A. Spielman. Lecture notes on spectral graph theory, 2018.
- [3] Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.