

# Spectral Graph Theory

Kelvin Li

November 9, 2019

## Abstract

These are the study notes of the course *Spectral Graph Theory* by Daniel Spielman at the Yale University, Fall 2015. There are 26 lectures. I try to summarise the important results and topics in each of the lectures and my own interpretations of these results.

A reference book that I use is *Spectra of Graphs* by Andries E. Brouwer and Willem H. Haemers, 2010.

## 1 Lecture 1 Introduction

Spectral graph theory has applications in graph drawing, graph isomorphism graph approximation and etc. This area uses a lot of results from linear algebra (in particular matrix theory).<sup>1</sup>

There are two natural matrices corresponding to a graph  $G(V, E)$ , the adjacency matrix  $A$  and the degree (or diagonal) matrix  $D$ . These two matrices themselves are boring, but some interesting matrices are defined based on them. For example, we can define the **diffusion matrix** as

$$W = D^{-1}A$$

that describes the diffusion of stuff (e.g., gas) among the vertices of a graph. We can also define an important matrix called the **Laplacian matrix**

$$L = D - A.$$

Given a function (or vector)  $x \in \mathbb{R}^V$  on the vertices, the **Laplacian quadratic form**<sup>2</sup> is

$$x^T L x = \sum_{(u,v) \in E} (x(u) - x(v))^2.$$

---

<sup>1</sup>The time complexity of finding eigenvalues of a  $n \times n$  symmetric matrix is  $O(n^3)$ .

<sup>2</sup>A function  $Q : \mathbb{R}^n \rightarrow \mathbb{R}$  is a **quadratic form** if  $Q(\vec{x}) = \vec{x}^T A \vec{x}$ , where  $\vec{x}$  is a column vector of length  $n$ .  $A$  is said to be the matrix of this quadratic form. Quadratic form defines a quadratic surface in a  $n$ -dimensional space.

It measures the smoothness of a vector  $\vec{x}$ .<sup>3</sup> The Laplacian quadratic form is sort of similar to the notion of **Laplacian operator** of a function, both of which measure the smoothness of a function. Hence, the Laplacian quadratic form measures the smoothness of the eigenvectors of the Laplacian. The quadratic form is important when maximizing (respectively minimizing) the Rayleigh quotient of the Laplacian matrix.<sup>4</sup>

A function in the context of graphs can be thought as a real value function that maps the vertices of a graph to real values. Imagine we a graph in a graph in a 2-dimensional space. A graph function maps each vertex of the graph to a vertical line, the height of which is the value of the vertex. Now we have a discrete “surface”, whose domain are the graph vertices. Using the Laplacian quadratic form we can measure how smooth or bumpy the “surface” is. Later, we will see that the eigenvectors of the Laplacian are indeed this type of **graph functions**.

**Definition 1.1.** A matrix  $A$  is **similar** to a matrix  $B$  if there is a non-singular<sup>5</sup> matrix  $M$  such that  $M^{-1}AM = B$ .

If  $A$  and  $B$  are similar, then they have the same set of eigenvalues. Almost all the matrices consider in this course are *symmetric* or similar to symmetric matrices.

**Definition 1.2.** Let  $M$  be an  $n \times n$  matrix. The **characteristic polynomial** of  $M$  is  $p_M(x) = \det(xI - M)$ .

The **eigenvalues** of  $M$  are the roots of the equation  $p_M(x) = 0$ .

**Definition 1.3.** A vector  $\psi$  is an **eigenvector** of a matrix  $M$  with eigenvalue  $\lambda$  if  $M\psi = \lambda\psi$ .

A matrix represents a linear transformation. An eigenvector of a matrix  $M$  is a non-zero vector that is invariant under the linear transformation by  $M$ , up to a scalar. This scalar is the eigenvalue, to which the eigenvector correspond. In other words, under the linear transformation by  $M$ , the eigenvectors do not change directions, but their scale may or may not change, depending on their corresponding eigenvalues. If an eigenvector is for the eigenvalue 1, then its magnitude does not change under  $M$ . If its eigenvalue is 3, its magnitude is scaled 3 times.

The eigenvectors of a given eigenvalue are only determinant up to an orthogonal transformation.

**Theorem 1.1. (Spectral theorem)** An  $n \times n$  real symmetric matrix has  $n$  eigenvalues and  $n$  eigenvectors that are unit vectors and mutually orthogonal.<sup>6,7</sup>

<sup>3</sup>Smoothness here corresponds to the values assigned to each pair of adjacent vertices. The smoothest vector is the constant vector, in which every element has the same value.

<sup>4</sup>Normalized Laplacian is introduced in Lecture 6.

<sup>5</sup>A matrix is **singular** if and only its determinant is zero.

<sup>6</sup>The **spectral theorem** states that any symmetric matrix is diagonalizable.

<sup>7</sup>A matrix  $A$  is **diagonalizable** if it is similar to a diagonal matrix  $B$ . That is, there exists an invertible matrix  $P$  such that  $B = P^{-1}AP$ .

If a matrix is not symmetric, it may not have  $n$  eigenvalues. Even if it has  $n$  eigenvalues, the corresponding  $n$  eigenvectors will not be mutually orthogonal.

**Definition 1.4.** The **Rayleigh quotient** of a vector  $x$  w.r.t. a matrix  $M$  is the ratio

$$\frac{x^T M x}{x^T x}.$$

By definition, if  $\psi$  is an eigenvector of  $M$  for the eigenvalue  $\lambda$ , then the Rayleigh quotient of  $\psi$  is  $\lambda$ . The notion of Rayleigh quotient is important, because it allows one to prove more insights about the eigenvalues (in particular the largest and smallest) of a matrix by connecting with the Laplacian quadratic form.

**Theorem 1.2. (Rayleigh quotient maximization)** Let  $M$  be a symmetric matrix and  $x$  be a non-zero vector that maximizes the Rayleigh quotient w.r.t.  $M$ . Then  $x$  is an eigenvector of  $M$  w.r.t. the eigenvalue that is equal to the Rayleigh quotient, and this eigenvalue is the largest eigenvalue of  $M$ . That is,

$$\psi = \arg \max_x \frac{x^T M x}{x^T x}$$

is an eigenvector of  $M$  for the eigenvalue  $\lambda$ , where

$$\lambda = \max_x \frac{x^T M x}{x^T x}.$$

The theorem can also be proved for **minimization**, in which the Rayleigh quotient is the smallest eigenvalue of  $M$ .

**Lemma 1.1. (Smallest eigenvalue of Laplacian)** The smallest eigenvalue of the Laplacian is zero. The corresponding eigenvector is a constant vector  $x = (c, \dots, c)$ .

The lemma makes sense intuitively. The sum of each column (or row) of the Laplacian is 0, so for the constant vector  $\vec{1} = (1, \dots, 1)$ , we have  $0 = L\vec{1} = \lambda\vec{1}$ . Hence,  $\lambda = 0$ . Since the determinant of a matrix is the product of all its eigenvalues, this also implies that the **Laplacian matrix is singular**.

*Proof.* Using Rayleigh quotient, in which  $M$  is the graph Laplacian, so the numerator is just the Laplacian quadratic form, in which each term in the summation is non-negative. So the smallest numerator is 0, which is obtained when  $x$  is a constant vector (or function). By the above theorem, this minimized Rayleigh quotient is the smallest eigenvalue of  $M$ .  $\square$

Assume without loss of generality that the eigenvalues of  $M$  are sorted  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ .

**Lemma 1.2. (Second smallest eigenvalue of Laplacian)**  $\lambda_2 > 0$  if and only if the graph  $G$  is connected.

*Proof.* If  $G$  contains two components, construct a vector for each component with constant values for vertices in the component and zero value for vertices out of the component. So the dot product of these two vectors is zero, which implies they are orthogonal. Both have corresponding zero eigenvalue, so  $\lambda_2 > 0$ .

If  $G$  is connected, then there must exist an eigenvector  $\vec{x}$  that is orthogonal to the constant eigenvector. This implies that  $x(u) \neq x(v)$  for some vertices  $u, v$ , because we want zero dot product. Since  $G$  is connected, there is a path between  $u$  and  $v$ . This implies that there must be an edge on the path such that the two end nodes have different values, hence the Laplacian quadratic form is positive, so is  $\lambda_2$ .  $\square$

The second smallest eigenvalue  $\lambda_2$  of a graph Laplacian is called the **algebraic connectivity of a graph**. Later, we will see that the smallest eigenvalue of the diffusion matrix is zero if and only if the graph is bipartite.

The  $i^{th}$  eigenvector of the graph Laplacian is (approximately?) a polynomial of order  $i - 1$  of node index. That is, when plotting the second eigenvector (on the Y-axis) against node indices (on the X-axis), the plot is roughly a straight line. When plotting the third, the plot is a quadratic curve. See plots near the end of Lecture 1.

## 1.1 Notes

Some additional notes worth mentioning:

1. By Spectral theorem, the adjacency and Laplacian matrices' eigenvectors form  $n$  orthonormal basis. Hence, their eigenvalues' **algebraic** and **geometric multiplicities**<sup>8</sup> are identical. Hence, we use the term multiplicity without referencing to algebraic or geometric. Note geometric multiplicity is **bounded above** by algebraic multiplicity.
2. The **(ordinary) spectrum** of a graph is the spectrum of the graph's adjacency matrix, that is, its sets of eigenvalues and multiplicities. The **Laplace spectrum** is the spectrum of the Laplacian matrix.
3. The eigenvalues and eigenvectors are of a adjacency or Laplacian matrix are **invariant under vertex numbering**.

---

<sup>8</sup>**Algebraic multiplicity** of an eigenvalue  $\lambda$  refers to the degree of the root  $\lambda$  in the characteristic polynomial. For example, if the characteristic polynomial of a matrix is  $(1 - \lambda)^2 = 0$ , then the eigenvalue  $\lambda = 1$  has algebraic multiplicity 2. **Geometric multiplicity** of  $\lambda$  is the dimension of the null space of the matrix. For example, if a matrix  $A = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$ , its eigenvalue  $\lambda = 5$  has geometric multiplicity 2. If  $A = \begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$ , its eigenvalue  $\lambda = 5$  has geometric multiplicity 1. Both cases have algebraic multiplicity 2.

## 2 Lecture 2 The Laplacian

**Theorem 2.1. (Eigenvalues and eigenvectors optimization)** Let  $M$  be an  $n \times n$  symmetric matrix with eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  with corresponding eigenvectors  $\psi_1, \dots, \psi_n$ . Then

$$\lambda_i = \min_{x \perp \psi_1, \dots, \psi_{i-1}} \frac{x^T M x}{x^T x},$$

$$\psi_i = \arg \min_{x \perp \psi_1, \dots, \psi_{i-1}} \frac{x^T M x}{x^T x}.$$

The **Laplacian quadratic form** for a weighted graph  $G = (V, E)$  is

$$x^T L x = \sum_{(uv) \in E} w_{uv} (x(u) - x(v))^2. \quad (1)$$

Let  $L_{G_{uv}}$  be the Laplacian of the graph  $G_{uv}$ , in which the graph has only one edge  $E(G_{uv}) = \{uv\}$  and possibly more than two vertices. Hence,  $L_{G_{uv}}$  has the form that the intersections between the rows and columns indexed by  $u$  and  $v$  are  $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$  and everywhere else is 0. Hence, the Laplacian of a graph  $G$  can be obtained by summing  $L_{G_{uv}}$  over all edges  $uv$  in  $G$ . That is,

$$L_G = \sum_{uv \in E} w_{uv} L_{G_{uv}}.$$

Intuitively this makes sense too, because if the vertex  $u$  has degree  $d(u) = 3$ , the corresponding row and column for the vertex  $u$  will appear 3 times in the summation.

Another important interpretation of the Laplacian matrix is that it is an operator

$$(L_G x)(u) = d(u)x(u) - \sum_{uv \in E} w_{uv} x(v) = \sum_{uv \in E} w_{uv} (x(u) - x(v)).$$

### 2.1 Graph drawing

When drawing a graph, take a vector  $x \in \mathbb{R}^V$  that maps each vertex of the graph  $G$  to a real value. We would like two adjacent vertices to have close values.<sup>9</sup> So this becomes a constraint optimization problem. That is, find a vector  $x$  such that Equation (1) is minimized, where the constraints are  $\|x\|^2 = 1$  to avoid  $x$  be a zero vector and  $\sum_{i=1}^V x(i) = 0$  to avoid  $x$  being a constant vector with  $x(i) = 1/\sqrt{n}$ .

When drawing the graph in one dimension, the optimal  $x$  is the eigenvector of norm 1 for the second smallest eigenvalue  $\lambda_2$ . When drawing in two dimensions, we need to find two optimal vectors  $x, y$  with an additional constraint that they are orthogonal. So the optimal  $x, y$  are the unit eigenvectors for  $\lambda_2, \lambda_3$  respectively.

---

<sup>9</sup>Probably because we do not want an edge to cross half of the space to connect two vertices.

## 2.2 Isoperimetry, expander graph and $\lambda_2$

**Definition 2.1.** Let  $S \subseteq V$  be a subset of vertices of the graph  $G = (V, E)$ . The **boundary** of  $S$  is

$$\partial(S) = \{uv \in E \mid u \in S, v \notin S\}$$

the set of edges connecting  $S$  and the rest of the graph.

**Definition 2.2.** Let  $S \subseteq V$  be a non-empty subset of vertices of the graph  $G = (V, E)$ . The **isoperimetric ratio** of  $S$  is

$$\theta(S) = \frac{|\partial(S)|}{|S|}.$$

**Definition 2.3.** The **isoperimetric number (or edge expansion or Cheeger constant)** of a graph  $G$  with  $n$  vertices is

$$\theta_G = \min_{S: |S| \leq n/2} \theta(S).$$

the minimum isoperimetric ratio over all non-empty subsets of at most half the vertices in  $G$ .

The condition  $|S| \leq n/2$  says we partition the graph vertices into two subsets and take the smaller one.

Based on these definitions, we can prove the next theorem, which gives a **lower bound** on the isoperimetric number of a graph  $G$  in terms of the second smallest eigenvalue  $\lambda_2$  of the graph Laplacian.

**Theorem 2.2. (Isoperimetric number lower bound)** Let  $G = (V, E)$  be a graph. For every  $S \subsetneq V$ , the isoperimetric ratio of  $S$  satisfies

$$\theta(S) \geq \lambda_2(1 - s),$$

where  $s = |S|/|V|$  and  $\lambda_2$  is the second smallest eigenvalue of the graph Laplacian. In particular, the isoperimetric number of  $G$  satisfies

$$\theta_G \geq \frac{\lambda_2}{2}.$$

*Proof.* The proof relies on the **characteristic vector (or indicator vector)** of the set  $S$

$$\chi_S(u) = \begin{cases} 1 & \text{if } u \in S \\ 0 & \text{otherwise.} \end{cases}$$

It shows that the characteristic vector is not orthogonal to the  $\psi_1$  that is a constant vector. But it can be used to create another vector  $x$  that is orthogonal to  $\psi_1$ .  $\square$

The theorems says that if  $\lambda_2$  is large, then the graph  $G$  is very well connected, because the boundary of any small subset of vertices is large. This relates to the concept of **expander graph**, which is vaguely understood as a finite undirected graph in which every small subset of vertices has large boundary. Every connected graph is an expander graph. A disconnect graph is not an expander graph, because there exists a component which has empty boundary.

**Remark 2.1.** Let  $S \subseteq V$  and  $s = |S|/|V|$ . Then

$$||\chi_S - s\vec{1}||^2 = s(1-s)|V|.$$

## 2.3 Eigenvalues and eigenvectors of the Laplacian of some graphs

For complete graphs  $K_n$ .

**Theorem 2.3.** The Laplacian of the complete graph  $K_n$  has eigenvalue 0 with multiplicity 1 and  $n$  with multiplicity  $n-1$ .

For star graph  $S_n$ .

**Lemma 2.1.** Let  $G = (V, E)$  be a graph, and let  $v, w$  be vertices of degree one that are both connected to another vertex  $z$ . Then, the vector  $\psi = \delta_v - \delta_w$  is an eigenvector of  $L_G$  of eigenvalue 1.

**Theorem 2.4.** The graph  $S_n$  has eigenvalue 0 with multiplicity 1, eigenvalue 1 with multiplicity  $n-2$ , and eigenvalue  $n$  with multiplicity 1.

Hypercube graphs, skipped.

## 3 Lecture 3 The adjacency matrix and graph coloring

### 3.1 Adjacency matrix

The adjacency matrix  $A$  of a weighted graph can be viewed as an operator that acts on a vector  $x \in \mathbb{R}^V$  by

$$(Ax)(u) = \sum_{uv \in E} w_{uv}x(v).$$

Denote the eigenvalues of  $A$  by  $\mu_1, \dots, \mu_n$  and order them in the opposite direction  $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ .

If  $G$  is a  $d$ -regular graph, then  $L = Id - A$ . The eigenvalues satisfy  $\lambda_i = d - \mu_i$ . The largest  $\mu_i = d$ , which corresponds to the smallest  $\lambda_i = 0$ . Hence, the eigenvector for the largest eigenvalue of the adjacency matrix is a constant vector.

**Lemma 3.1. (Bounds on the largest eigenvalue)** Let  $d_{max}$  and  $d_{ave}$  be the maximum and average degree of a vertex in a graph  $G$ . Then

$$d_{ave} \leq \mu_1 \leq d_{max}.$$

**Lemma 3.2. (Better lower bound)** For every  $S \subseteq V$ , let  $d_{ave}(S)$  be the average degree of vertices in the induced subgraph  $G[S]$  by the vertices in  $S$ . Then

$$d_{ave}(S) \leq \mu_1.$$

This lemma is proved using the following result from linear algebra.

**Lemma 3.3.** Let  $A$  be a symmetric matrix and let  $S$  be a subset of its row and columns indices. Then the eigenvalues satisfy

$$\lambda_{max}(A) \geq \lambda_{max}(A(S)) \geq \lambda_{min}(A(S)) \geq \lambda_{min}(A).$$

**Lemma 3.4. (Largest eigenvalue)** If  $G$  is connected and  $\mu_1 = d_{max}$ , then  $G$  is  $d_{max}$ -regular.

Although the largest eigenvalue of the adjacency matrix of a regular graph is 0 and hence its corresponding eigenvector is a constant vector, the eigenvector is usually not constant for general graphs.

**Lemma 3.5. (Non-negative implies positive)** Let  $G$  be a connected (non-negative) weighted graph. If  $\phi$  is a non-negative eigenvector of its adjacency matrix  $A$ , then  $\phi$  is strictly positive.

**Theorem 3.1. (Perron-Frobenius for symmetric graphs)** Let  $G$  be a connected weighted symmetric graph. Let  $A$  be its adjacency matrix and  $\mu_1 \geq \dots \geq \mu_n$  be its eigenvalues. Then

1.  $\mu_1 \geq -\mu_n$ .
2.  $\mu_1 > \mu_2$ .
3. The eigenvalue  $\mu_1$  has a strictly positive eigenvector.

*Proof.* To prove part 3, we can construct a vector  $x = |\phi_1|$  that is the absolute value of the eigenvector for the largest eigenvalue  $\mu_1$ . Since  $x$  is non-negative, once we can prove that  $x$  is also an eigenvector of  $\mu_1$ , we can apply the above lemma to conclude that  $\mu_1$  has a positive eigenvector.  $\square$

**Proposition 3.1. (Bipartite graph)** If  $G$  is connected, then  $\mu_n = -\mu_1$  if and only if  $G$  is bipartite.

The  $n^{th}$  eigenvalue  $\mu_n$  (smallest of  $A$ ) and  $\lambda_n$  (largest of  $L$ ) correspond to the highest **frequency vibration**<sup>10</sup> in a graph. The corresponding eigenvector tries to assign as different as possible values to neighbouring vertices. This is consistent with the plot for the largest eigenvector in Lecture 1, in which at each node index, the plot has a local optimal.<sup>11</sup>

<sup>10</sup>What is it?

<sup>11</sup>Because eigenvectors are invariant under vertex numbering.



### 3.2 Graph coloring

**Definition 3.1.** A *k-coloring* of a graph  $G = (V, E)$  is a function  $c : V \rightarrow \{1, \dots, k\}$  so that for all edges  $uv \in E$ , we have  $c(u) \neq c(v)$ .

**Definition 3.2.** A graph is *k-colorable* if it has a *k-coloring*.

**Definition 3.3.** The *chromatic number* of a graph  $G$ , written  $\chi_G$ , is the least  $k$  for which  $G$  is *k-colorable*.

**Lemma 3.6. (2-colorable)** A graph  $G$  is 2-colorable if and only if it is bipartite.

**Lemma 3.7. (3-colorable)** Determining whether or not a graph is 3-colorable is an NP-complete problem.

**Theorem 3.2. (4-colorable)** Every planar graph is 4-colorable.

**Remark 3.1.** Every graph is  $d_{\max} + 1$ -colorable.

**Remark 3.2.** Order the vertices of a graph  $1, \dots, n$ . Let  $k$  be an upper bound on the number of neighbours of every vertex  $u \in V$  that has smaller labels. That is,

$$|\{v \mid uv \in E \text{ and } v < u\}| \leq k, \quad \forall u \in V.$$

Then  $G$  can be colored with at most  $k + 1$  colors.

**Theorem 3.3. (Wilf's theorem)** Let  $\mu_1$  be the largest eigenvalue of the adjacency matrix of a graph  $G$ . The chromatic number of  $G$  satisfies

$$\chi_G \leq \lfloor \mu_1 \rfloor + 1.$$

*Proof.* The proof relies on the above remark. It first finds a vertex with degree at most  $\mu_1$ . Such a vertex exists due to the bound that  $d_{\text{ave}}(G) \leq \mu_1$ . Label this vertex  $n$ . Then use the fact that the largest eigenvalue of any subgraph is bounded above by  $\mu_1$ .  $\square$

**Lemma 3.8. (Block matrix)** Let

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,k} \\ A_{1,2}^T & A_{2,2} & \cdots & A_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,k}^T & A_{2,k}^T & \cdots & A_{k,k} \end{bmatrix}$$

be a block-partitioned symmetric matrix with  $k \geq 2$ . Then

$$(k-1)\lambda_{\min}(A) + \lambda_{\max}(A) \leq \sum_i \lambda_{\max}(A_{i,i}).$$

**Theorem 3.4. (Hoffman's bound)** Let  $\mu_1, \mu_n$  be the largest and smallest eigenvalues of the adjacency matrix of a graph  $G$ . The chromatic number of  $G$  satisfies

$$\chi_G \geq \frac{\mu_1 - \mu_n}{-\mu_n} = 1 + \frac{\mu_1}{-\mu_n}.$$

## 4 Lecture 4 Bounding eigenvalues

The Courant-Fischer theorem turns an eigenvalue of a symmetric matrix into the solution of optimization problems, hence are useful when proving bounds on the largest eigenvalue.

**Theorem 4.1. (Courant-Fischer theorem)** Let  $L$  be a symmetric matrix with eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ . Then,

$$\lambda_k = \min_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \max_{x \in S} \frac{x^T L x}{x^T x} = \max_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T)=n-k+1}} \min_{x \in T} \frac{x^T L x}{x^T x}.$$

(Alternatively) The way to interpret the first equality is that for all choices of a vector  $x$  that maximize the Rayleigh quotient, find the one that is orthogonal to those that have been found and minimizes the total value. The max is used to ensure  $x$  is an eigenvector. For example, when  $k = 2$ , each of the  $n - 1$  eigenvectors is orthogonal to the constant vector  $v_1$ . But the one that minimizes the total value is  $v_2$ .

The way to interpret the theorem is that  $S$  is a  $k$ -dimensional subspace. Within  $S$ , we want to find a vector  $x$  that is orthogonal to a  $k - 1$ -dimensional subspace and maximize the Rayleigh quotient. For every choice of  $S$  we can find such a vector  $x$ , so we want to find the  $S$  that minimizes the maximized Rayleigh quotient. The second max min can be interpreted similarly, but starting from  $n$  down to 1-dimensional subspace. For example, when  $k = 1$ ,  $S$  is a 1-dimensional subspace. That is,  $S$  can be the span of any eigenvector  $\psi_i$  of  $L$ . Within each of the  $n$  choices of  $S$ , the max Rayleigh quotient is achieved when  $x$  is an eigenvector. By Rayleigh quotient minimization, we know the smallest Rayleigh quotient is the smallest eigenvalue of  $L$ .

*Proof.* The theorem is proved in two steps. Firstly, show that for any  $k$ -dimensional subspace  $S$ , it follows that  $\max_{x \in S} \frac{x^T L x}{x^T x} \geq \lambda_k$ . Then, prove that the minimum equals  $\lambda_k$ , which happens when  $S$  is the span of eigenvectors  $\{\psi_1, \dots, \psi_k\}$ .  $\square$

Some additional useful notations. Use the **partial order** notation  $\succeq$  to compare matrices and graphs.

- $A \succeq 0$  denotes the matrix  $A$  is positive semi-definite.<sup>12</sup> That is,  $v^T A v \geq 0$  for all  $v$ .
- If  $A, B$  are  $n \times n$  square matrices, write  $A \succeq B$  if  $A - B \succeq 0$ , which is equivalent to  $v^T A v \geq v^T B v$  for all  $v$ .
- For (weighted) graphs  $G$  and  $H$ , write  $G \succeq H$  if  $L_G \succeq L_H$ .

<sup>12</sup>That is, for all vector  $v$ , we have  $v^T A v \geq 0$ . By Rayleigh quotient, this implies that all eigenvalues of  $A$  are non-negative, because the denominator of the Rayleigh quotient is the length of  $v$ .

- For a weighted graph  $H$ , write  $c \cdot H$  to denote the same graph but the weight of every edge in  $H$  is multiplied by a constant  $c > 0$ .

If  $H \subseteq G$  is a subgraph of  $G$ , then the Laplacian quadratic form implies that  $L_G \succeq L_H$ , because dropping an edge from  $G$  will not make the summation any larger due to the square difference.

**Lemma 4.1.** *If  $G$  and  $H$  are graphs such that  $G \succeq c \cdot H$ , then  $\lambda_k(G) \geq c\lambda_k(H)$ .*

*Proof.* Prove using Courant-Fischer theorem and the fact that  $L_G \succeq cL_H$ .  $\square$

**Corollary 4.1.** *Let  $G$  be a graph and let  $H$  be obtained by either adding an edge to  $G$  or increasing the weight of an edge in  $G$ . Then,  $\lambda_i(G) \leq \lambda_i(H)$  for all  $i$ .*

## 4.1 Approximation of graphs

**Definition 4.1.** *A graph  $H$  is a  $c$ -approximation of  $G$  if  $cH \succeq G \succeq H/c$ .*

**Theorem 4.2.** (*Regular graph approximates complete graph*) *For every  $\epsilon > 0$ , there exists a  $d > 0$  such that for all sufficiently large  $n$  there is a  $d$ -regular graph  $G_n$  that is a  $(1 + \epsilon)$ -approximation of  $K_n$ .*

Skipped the rest of this lecture.

## 5 Lecture 5 Rings, paths and Cayley graphs

This lecture is devoted to an examination of some special graphs and their eigenvalues. So I skipped it for now!

## 6 Lecture 6 Conductance, the normalized Laplacian and Cheeger's inequality

**Definition 6.1.** *Let  $G = (V, E)$  be a graph. The **conductance** of a non-empty proper subset  $S \subsetneq V$  is*

$$\phi(S) = \frac{|\partial(S)|}{\min\{d(S), d(V - S)\}},$$

where  $d(S) = \sum_{u \in S} d(u)$  is the sum of degrees of vertices in  $S$ .

Conductance is a measure of how “connected” a vertex subset  $S$  is to the rest of the graph. It ranges in  $[0, 1]$ , where 0 is when  $S$  is disconnected from the rest of the graph and 1 is when  $S$  contains a single vertex only. In the context of clustering, when  $\phi(S) \approx 0$ , the vertices in  $S$  form its own cluster and are less connected to the rest of the graph. When  $\phi(S) = 1$ , most of the edges in  $S$  also leave  $S$ , so the whole graph forms a single cluster.

**Definition 6.2.** Let  $G = (V, E)$  be a graph. The **conductance** of the graph  $G$  is

$$\phi_G = \min_{S \subsetneq V} \phi(S)$$

the minimum conductance over all non-empty proper subset  $S$  of  $V$ .

The conductance of a graph is more useful when concerning about the edges, whilst the isoperimetric number is more useful when concerning about vertices.

**Definition 6.3.** The **normalized Laplacian** is

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

The diagonal entries of  $N$  equal to 1, because the degrees are normalized. The entries correspond to adjacent vertices  $v_i, v_j$  equal  $1/\sqrt{d(v_i)d(v_j)}$ . The rest of  $N$  equals to 0. The normalized Laplacian is more useful when examining graphs in which nodes have different degrees. The corresponding eigenvalues of a normalized Laplacian are denoted by  $0 = \nu_1 \leq \nu_2 \leq \dots \leq \nu_n$ .<sup>13</sup> The conductance is related to the second smallest eigenvalue  $\nu_2$  as the isoperimetric number is related to  $\lambda_2$  of  $L$ .

**Lemma 6.1.** (**Conductance lower bound**) The conductance of a graph  $G$  is

$$\phi_G \geq \frac{\nu_2}{2}.$$

**Remark 6.1.** (**Smallest eigenvector of  $N$** ) The eigenvector for the smallest eigenvalue  $\nu_1 = 0$  is

$$d^{1/2} = (\sqrt{d(u_1)}, \dots, \sqrt{d(u_n)}).$$

*Proof.*

$$D^{-1/2} L D^{-1/2} d^{1/2} = D^{-1/2} L 1 = D^{-1/2} 0 = 0 = \lambda d^{1/2},$$

where 1 is the constant eigenvector corresponds to  $\lambda_1$ . So  $\lambda = 0$ .  $\square$

By Rayleigh quotient, the eigenvector for the second smallest eigenvalue of  $N$  is given by

$$\arg \min_{x \perp d^{1/2}} \frac{x^T N x}{x^T x}.$$

The eigenvalue  $\nu_2$  can also be expressed as

$$\nu_2 = \min_{y \perp d} \frac{y^T L y}{y^T D y}$$

---

<sup>13</sup>In fact, the range of the eigenvalues of the normalized Laplacian is between  $[0, 2]$ . This is another reason to work with normalized Laplacian.

in terms of  $L$  and  $D$ , where  $x = D^{1/2}y$  and the term

$$\frac{y^T Ly}{y^T Dy}$$

is defined as the **generalized Rayleigh quotient**.

**Theorem 6.1. (Cheeger's inequality)** *Let  $y$  be a vector orthogonal to  $d$ . Then, there is a number  $t$  for which the set  $S_t = \{u \mid y(u) < t\}$  satisfies*

$$\phi(S_t) \leq \sqrt{2 \frac{y^T Ly}{y^T y}}.$$

By definition of the conductance of a graph  $G$ , we get

$$\phi_G \leq \sqrt{2\nu_2}.$$

*Proof.* The proof is quite involved, refer to the lecture slides.  $\square$

**Corollary 6.1. (Conductance boundary)**

$$\frac{\nu_2}{2} \leq \phi_G \leq \sqrt{2\nu_2}.$$

**Remark 6.2. (Disconnected graphs)** *A graph  $G$  is disconnected if and only if  $\nu_2 = 0$ .*

**Corollary 6.2.**  *$G$  graph  $G$  is disconnected if and only if its conductance  $\phi_G = 0$ .*

*Proof.* By the remark and the range of  $\nu_i \in [0, 2]$ .  $\square$

In principle, the second eigenvector  $v_2$  of the normalized Laplacian  $N$  can be used to find the partition  $S$  by finding the most appropriate cut value  $t$ .

## 7 Lecture 7 Spring and resistor networks

Imagine that every edge of a graph  $G = (V, E)$  is a spring or rubber band. Pick a subset  $F \subseteq V$  of vertices. Fix the location of each vertex  $u \in F$  at  $\vec{x}(u)$ . The **spring constant** of each edge is the weight of that edge. By **Hook's law**, the force the rubber band between vertices  $u$  and  $v$  exerts on  $u$  is

$$w_{uv}\vec{x}(u) - \vec{x}(v)$$

that is proportional to the distance between  $u$  and  $v$ .

In a stable condition, a non-fixed vertex  $u$  must experience a zero net force (or potential energy?). That is,

$$\begin{aligned} \sum_{v \in N(u)} [w_{uv}\vec{x}(v) - \vec{x}(u)] &= 0 \\ \implies \sum_{v \in N(u)} w_{uv}\vec{x}(v) &= d(u)\vec{x}(u) \\ \implies \vec{x}(u) &= \frac{1}{d(u)} \sum_{v \in N(u)} w_{uv}\vec{x}(v) \end{aligned} \tag{2}$$

where  $d(u) = \sum_{v \in N(u)} w_{uv}$ . A function  $\vec{x}$  is **harmonic** on  $S = V - F$  if  $\vec{x}(u)$  satisfies the above equation for every  $u \in S$ .

Equation (2) can be written as

$$d(u)\vec{x}(u) - \sum_{v \in S \cap N(u)} w_{uv} \vec{x}(v)$$

in the form of matrix multiplication

$$L(u)\vec{x} = 0.$$

## 8 Lecture 10 Random walks on graphs

Define **random walk** on weighted undirected graphs  $G = (V, E, w)$  to be a process that starts from a vertex  $u$  then moves to an adjacent vertex with probability proportional to the weight of each edge incident to  $u$ . The behaviour of a random walk is controlled by the **probability distribution over vertices** at a certain time. Define this probability distribution at a discrete time step  $t$  to be a function

$$p_t : V \rightarrow \mathbb{R}$$

such that

$$p_t(u) \geq 0 \quad \text{and} \quad \sum_{u \in V} p_t(u) = 1,$$

where  $p_t(u)$  is the probability of being at vertex  $u$  at time  $t$ . The initial distribution  $p_0$  is concentrated at one vertex  $u$  i.e.,  $p_0(u) = 1$ , which is the starting point of the random walk.

The probability at vertex  $u$  at time  $t + 1$  can be derived from the previous time step probability by

$$p_{t+1}(u) = \sum_{v \in N(u)} \frac{w_{uv}}{d(v)} p_t(v),$$

where  $d(v) = \sum_{s \in N(v)} w_{sv}$ . The above is interpreted as the weight fraction of an edge  $uv$  over the degree of a vertex  $v$  times the probability the walk is at vertex  $v$  at time  $t$ .

**Lazy random walk** is a random walk that has  $1/2$  probability of staying at the current vertex and the other  $1/2$  probability of being a random walk. Hence,

$$p_{t+1}(u) = \frac{1}{2} p_t(u) + \frac{1}{2} \sum_{v \in N(u)} \frac{w_{uv}}{d(v)} p_t(v). \quad (3)$$

**Diffusion** can be defined the same as lazy random walk. That is, half of the substance stays at the current vertex and the other half distributed randomly (according to some probability) to neighbouring vertices.

Equation (3) can be written in the following matrix form

$$p_{t+1} = \frac{1}{2}(I + AD^{-1})p_t.$$

Hence, define the **lazy walk matrix** as

$$\begin{aligned} W_G &= \frac{1}{2}(I + AD^{-1}) \\ &= I - \frac{1}{2}D^{1/2}ND^{-1/2}, \end{aligned}$$

where  $N = I - D^{-1/2}ND^{-1/2}$  is the normalized Laplacian. Note  $W_G$  is an **asymmetric** matrix, so the eigenvectors are not necessarily orthogonal. One reason of using the lazy walk is because the matrix  $W$ 's **spectrum are bounded between 0 and 1**. Denote the eigenvalues of  $W$  as

$$1 = \omega_1 \geq \omega_2 \geq \dots \geq \omega_n \geq 0,$$

where  $\omega_i = 1 - \nu_i/2$ . Another reason is that lazy random walk will always converge to the stable distribution on a connected graph.

**Lemma 8.1.** (*Right eigenvector of lazy walk matrix*) *For every eigenvector  $\psi_i$  of  $N$  with eigenvalue  $\nu_i$ , the vector  $D^{1/2}\psi_i$  is a right-eigenvector of  $W$  for the eigenvalue  $\omega_i = 1 - \nu_i/2$ .*

*Proof.*

$$\begin{aligned} W(D^{1/2}\psi_i) &= \left(I - \frac{1}{2}D^{1/2}ND^{-1/2}\right)(D^{1/2}\psi_i) \\ &= D^{1/2}\psi_i - \frac{1}{2}D^{1/2}N\psi_i \\ &= D^{1/2}\psi_i - \frac{1}{2}D^{1/2}\nu_i\psi_i \\ &= \left(1 - \frac{\nu_i}{2}\right)D^{1/2}\psi_i \end{aligned}$$

□

As mentioned above, the lazy random walk will always converge to the **stable distribution**

$$\vec{\pi} = \frac{\vec{d}}{\vec{1}^T \vec{d}}$$

regardless of the original distribution over vertices. The stable distribution is a vector of each vertex's degree proportional to the total degree.

**Lemma 8.2.**  *$\vec{\pi}$  is a right-eigenvector of  $W$  for the eigenvalue 1.*

*Proof.* WTS  $W\pi = \pi$ .

□

**Lemma 8.3.** (*Lazy random walk convergence*) *The lazy random walk converges to the stable distribution  $\vec{\pi}$ .*

*Proof.* WTS  $p_t = W^t p_0 = \vec{\pi}$ . That is, apply the lazy walk matrix  $t$  time will get to the distribution  $\vec{\pi}$ .

Here  $W$  can be seen as an operator that is applied to a vector  $p_0$  repeatedly.  $\square$