Mathematics for Data Sciences **October 20, 2011**

Lecture 08. Random Walk on Graphs: Fiedler Vector, Cheeger inequality

Instructor: Yuan Yao, Peking University *Scribe: Wang, Zhiming; Lin, Feng*

Introduction

In this class, we introduced the random walk on graphs. The last lecture shows Perron-Frobenius theory to the analysis of primary eigenvectors which is the stationary distribution. In this lecture we will study the second eigenvector. To analyze the properties of the graph, we construct two matrices: one is (unnormalized) graph Laplacian and the other is normalized graph Laplacian. In the first part, we introduce Fiedler Theory for the unnormalized graph Laplacian, which shows the second eigenvector can be used to bipartite the graph into two connected components. In the second part, we study the eigenvalues and eigenvectors of normalized Laplacian matrix to show its relations with random walks or Markov chains on graphs. In the third part, we will introduce the Cheeger Inequality for second eigenvector of normalized Laplacian, which leads to an approximate algorithm for Normalized graph cut (NCut) problem, an NP-hard problem itself.

1 Fiedler Theory

Let $G = (V, E)$ be an undirected, unweighted simple[1] graph. Although the edges here are unweighted, the theory below still holds when weight is added. We can get a similar conclusion with the weighted adjacency matrix. However the extension to directed graphs will lead to different pictures.

We use i∼j to denote that node $i \in V$ is a neighbor of node $j \in V$.

Definition (Adjacency Matrix)**.**

$$A_{ij} = \begin{cases} 1 & i \sim j \\ 0 & otherwise \end{cases} .$$

Remark. We can use the weight of edge $i \sim j$ to define A_{ij} if the graph is weighted. That indicates $A_{ij} \in \mathbb{R}^+$. We can also extend A_{ij} to $\mathbb{R}$ which involves both positive and negative weights, like correlation graphs. But the theory below can not be applied to such weights being positive and negative.

The degree of node i is defined as follows.

$$d_i = \sum_{j=1}^{n} A_{ij}.$$

Define a diagonal matrix $D = \mathrm{diag}(d_i)$. Now let's come to the definition of Laplacian Matrix L.

Definition (Graph Laplacian)**.**

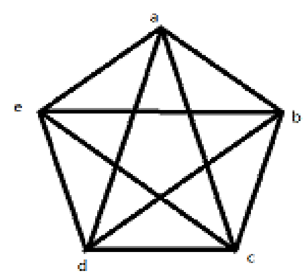
$$L_{ij} = \begin{cases} d_i & i = j, \\ -1 & i \sim j \\ 0 & otherwise \end{cases}$$

[1] Simple graph means for every pair of nodes there are at most one edge associated with it; and there is no self loop on each node.

This matrix is often called *unnormalized graph Laplacian* in literature, to distinguish it from the normalized graph Laplacian below. In fact, $L = D - A$.

Example 1. $V = \{1, 2, 3, 4\}$, $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}$. This is a linear chain with four nodes.

$$L = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$



Example 2. A complete graph of n nodes, K_n. $V = \{1, 2, 3...n\}$, every two points are connected, as the figure above with $n = 5$.

$$L = \begin{pmatrix} n-1 & -1 & -1 & ... & -1 \\ -1 & n-1 & -1 & ... & -1 \\ -1 & ... & -1 & n-1 & -1 \\ -1 & ... & -1 & -1 & n-1 \end{pmatrix}.$$

From the definition, we can see that L is symmetric, so all its eigenvalues will be real and there is an orthonormal eigenvector system. Moreover L is positive semi-definite (p.s.d.). This is due to the fact that

$$\begin{aligned} v^T L v &= \sum_i \sum_{j:j\sim i} v_i(v_i - v_j) = \sum_i \left(d_i v_i^2 - \sum_{j:j\sim i} v_i v_j \right) \\ &= \sum_{i\sim j} (v_i - v_j)^2 \geq 0, \qquad \forall v \in \mathbb{R}^n. \end{aligned}$$

In fact, L admits the decomposition $L = BB^T$ where $B \in \mathbb{R}^{|V|\times|E|}$ is called *incidence matrix* (or *boundary map* in algebraic topology) here, for any $1 \leq j < k \leq n$,

$$B(i, \{j, k\}) = \begin{cases} 1, & i = j, \\ -1, & i = k, \\ 0, & \text{otherwise} \end{cases}$$

These two statements imply the eigenvalues of L can't be negative. That is to say $\lambda(L) \geq 0$.

Theorem 1.1 (Fiedler theory)**.** Let L has n eigenvectors

$$Lv_i = \lambda_i v_i, \quad v_i \neq 0, \quad i = 0, \ldots, n-1$$

where $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$. For the second smallest eigenvector v_1, define

$$N_- = \{i : v_1(i) < 0\},$$

$$N_+ = \{i : v_1(i) > 0\},$$

$$N_0 = V - N_- - N_+.$$

We have the following results.

1. $\#\{i, \lambda_i = 0\} = \#\{connected\ components\ of\ G\}$;
2. If G is connected, then both N_- and N_+ are connected. $N_- \cup N_0$ and $N_+ \cup N_0$ might be disconnected if $N_0 \neq \emptyset$.

This theorem tells us that the second smallest eigenvalue can be used to tell us if the graph is connected, *i.e.* G is connected iff $\lambda_1 \neq 0$, *i.e.*

$$\lambda_1 = 0 \Leftrightarrow \textit{there are at least two connected components.}$$
$$\lambda_1 > 0 \Leftrightarrow \textit{the graph is connected.}$$

Moreover, the second smallest eigenvector can be used to bipartite the graph into two connected components by taking N_- and N_+ when N_0 is empty. For this reason, we often call the second smallest eigenvalue λ_1 as the *algebraic connectivity*.

We can calculate eigenvalues by using Rayleigh Quotient. This gives a sketch proof of the first part of the theory.

Proof of Part I. Let (λ, v) be a pair of eigenvalue-eigenvector, *i.e.* $Lv = \lambda v$. Since $L1 = 0$, so the constant vector $1 \in \mathbb{R}^n$ is always the eigenvector associated with $\lambda_0 = 0$. In general,

$$\lambda = \frac{v^T L v}{v^T v} = \frac{\sum\limits_{i \sim j} (v_i - v_j)^2}{\sum\limits_{i} {v_i}^2}.$$

Note that

$$0 = \lambda_1 \Leftrightarrow v_i = v_j \ (j \textit{ is path connected with } i).$$

Therefore v is a piecewise constant function on connected components of G. If G has k components, then there are k independent piecewise constant vectors in the span of characteristic functions on those components, which can be used as eigenvectors of L. In this way, we proved the first part of the theory. □

2 Normalized graph Laplacian

Definition (Normalized Graph Laplacian)**.**

$$\mathcal{L}_{ij} = \begin{cases} 1 & i = j, \\ -\frac{1}{\sqrt{d_i d_j}} & i \sim j, \\ 0 & \textit{otherwise}. \end{cases}$$

In fact $\mathcal{L} = D^{-1/2}(D-A)D^{-1/2} = D^{-1/2}LD^{-1/2} = I - D^{-1/2}(D-A)D^{-1/2}$. From this one can see the relations between eigenvectors of normalized $\mathcal{L}$ and unnormalized L. For eigenvectors $\mathcal{L}v = \lambda v$, we have

$$(I - D^{-1/2}LD^{-1/2})v = \lambda v \Leftrightarrow Lu = \lambda Du, \quad u = D^{-1/2}v,$$

whence eigenvectors of $\mathcal{L}$, v after rescaling by $D^{-1/2}v$, become generalized eigenvectors of L.

We can also use the Rayleigh Quotient to calculate the eigenvalues of $\mathcal{L}$.

$$\begin{aligned}
\frac{v^T\mathcal{L}v}{v^Tv} &= \frac{v^TD^{-\frac{1}{2}}(D-A)D^{-\frac{1}{2}}v}{v^v} \\
&= \frac{u^TLu}{u^TDu} \\
&= \frac{\sum\limits_{i\sim j}(u_i-u_j)^2}{\sum\limits_j {u_j}^2 d_j}.
\end{aligned}$$

Similarly we get the relations between eigenvalue and the connected components of the graph.

$$\#\{\lambda_i(\mathcal{L}) = 0\} = \#\{connected\ components\ of\ G\}.$$

Next we show that eigenvectors of $\mathcal{L}$ are related to random walks on graphs. This will show you why we choose this matrix to analysis the graph.

We can construct a random walk on G whose transition matrix is defined by

$$P_{ij} \sim \frac{A_{ij}}{\sum\limits_j A_{ij}} = \frac{1}{d_i}.$$

By easy calculation, we see the result below.

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

Hence P is similar to $I - \mathcal{L}$. So their eigenvalues satisfy $\lambda_i(P) = 1 - \lambda_i(\mathcal{L})$. Consider the right eigenvector ϕ and left eigenvector ψ of P.

$$\begin{aligned}
u^TP &= \lambda u, \\
Pv &= \lambda v.
\end{aligned}$$

Due to the similarity between P and $\mathcal{L}$,

$$u^TP = \lambda u^T \Leftrightarrow u^TD^{-1/2}(I-\mathcal{L})D^{1/2} = \lambda u^T.$$

Let $\bar{u} = D^{-1/2}u$, we will get:

$$\begin{aligned}
&\bar{u}^T(I-\mathcal{L}) = \lambda\bar{u}^T \\
&\Leftrightarrow \mathcal{L}\bar{u} = (1-\lambda)\bar{u}.
\end{aligned}$$

You can see $\bar{u}$ is the eigenvector of $\mathcal{L}$, and we can get left eigenvectors of P from $\bar{u}$ by multiply it with $D^{1/2}$ on the left side. Similarly for the right eigenvectors $v = D^{-1/2}\bar{u}$.

If we choose $u_0 = \pi_i \sim \frac{d_i}{\sum d_i}$, then:

$$\begin{aligned}
\bar{u}_0(i) &\sim \sqrt{d_i}, \\
\bar{u}_k^T\bar{u}_l &= \delta_{kl},
\end{aligned}$$

$$u_k^T D v_l = \delta_{kl},$$
$$\pi_i P_{ij} = \pi_j P_{ji} \sim A_{ij} = A_{ji},$$

where the last identity says the Markov chain is time-reversible.

All the conclusions above show that the normalized graph Laplacian $\mathcal{L}$ keeps some connectivity measure of unnormalized graph Laplacian L. Furthermore, $\mathcal{L}$ is more related with random walks on graph, through which eigenvectors of P are easy to check and calculate. That's why we choose this matrix to analysis the graph.

3 Cheeger Inequality

Let G be a graph, $G = (V, E)$ and S is a subset of V whose complement is $\bar{S} = V - S$. We define $Vol(S)$, $CUT(S)$ and $NCUT(S)$ as below.

$$Vol(S) = \sum_{i \in S} d_i.$$
$$CUT(S) = \sum_{i \in S, j \in \bar{S}} A_{ij}.$$
$$NCUT(S) = \frac{CUT(S)}{\min(Vol(S), Vol(\bar{S}))}.$$

$NCUT(S)$ is called normalized-cut. We define the Cheeger constant $h_G = \min_S NCUT(S)$. Finding minimal normalized graph cut is NP-hard.

Cheeger Inequality says the second smallest eigenvalue provides both upper and lower bounds on the minimal normalized graph cut. Its proof gives us a constructive polynomial algorithm to achieve such bounds.

Theorem 3.1 (Cheeger Inequality)**.** If G is connected, then

$$\frac{h_G^2}{2} \le \lambda_1(\mathcal{L}) \le 2h_G.$$

Proof. (1) Upper bound:

Assume the following function f realizes the optimal normalized graph cut,

$$f(i) = \begin{cases} \frac{1}{Vol(S)} & i \in S, \\ \frac{-1}{Vol(\bar{S})} & i \in \bar{S}, \end{cases}$$

By using the Rayleigh Quotient, we get

$$\begin{aligned} \lambda_1 = & \inf_{g \perp D^{1/2} e} \frac{g^T \mathcal{L} g}{g^T g} \\ \le & \frac{\sum_{i \sim j} (f_i - f_j)^2}{\sum f_i^2 d_i} \\ = & \frac{(\frac{1}{Vol(S)} + \frac{1}{Vol(\bar{S})})^2 CUT(S)}{Vol(S) \frac{1}{Vol(S)^2} + Vol(\bar{S}) \frac{1}{Vol(\bar{S})^2}} \\ = & (\frac{1}{Vol(S)} + \frac{1}{Vol(\bar{S})}) CUT(S) \\ \le & \frac{2CUT(S)}{\min(Vol(S), Vol(\bar{S}))} =: 2h_G. \end{aligned}$$

span

which gives the upper bound.

(2) Lower bound: the proof of lower bound actually gives a constructive algorithm to compute an approximate optimal cut as follows.

Let $\lambda_1 v = \mathcal{L}v$. Then we reorder node set V such that $v_1 \leq v_2 \leq ... \leq v_n$. Now consider a series of particular subsets of V: $S_i = \{v_1, v_2, ...v_i\}$, and define

$$\alpha_G = \min_i NCUT(S_i).$$

Clearly finding the optimal value α just requires comparison over $n-1$ NCUT values.

Below we shall show that

$$\frac{h_G^2}{2} \leq \frac{\alpha_G^2}{2} \leq \lambda_1.$$

Denote $V_- = \{i; v_i \leq 0\}, V_+ = \{i; v_i > 0\}$. Without lose of generality, we assume $Vol(V_-) \leq Vol(V_+)$. Define $\widetilde{Vol}(S) = \min(Vol(S), Vol(\bar{S}))$.

We write $R(v)$ for

$$R(v) = \frac{\sum_{i\sim j}(v_i - v_j)^2}{\sum v_i^2 d_i}.$$

We will get the following results.

$$\begin{aligned}
\lambda_1 = & \quad R(v) \\
\geq & \quad \frac{\sum_{i,j\in V_+, i\sim j}(v_i - v_j)^2 + \sum_{i,j\in V_-, i\sim j}(v_i - v_j)^2}{\sum_{i\in V_+} v_i^2 d_i + \sum_{i\in V_-} v_i^2 d_i} \\
\geq & \quad \frac{\sum_{i,j\in V_+, i\sim j}(v_i - v_j)^2}{\sum_{i\in V_+} v_i^2 d_i} \qquad \frac{a+b}{c+d} \geq \min\left(\frac{a}{c}, \frac{b}{d}\right) \\
= & \quad \frac{(\sum_{i,j\in V_+, i\sim j}(v_i - v_j)^2)(\sum_{i,j\in V_+, i\sim j}(v_i + v_j)^2)}{(\sum_{i\in V_+} v_i^2 d_i)(\sum_{i,j\in V_+, i\sim j}(v_i + v_j)^2)} \\
\geq & \quad \frac{(\sum_{i,j\in V_+, i\sim j}(v_i - v_j)^2)^2}{2(\sum_{i\in V_+} v_i^2 d_i)^2},
\end{aligned}$$

where the numerator is due to the Cauchy-Schwartz inequality $|\langle x, y\rangle|^2 \leq \langle x, x\rangle \cdot \langle y, y\rangle$, and the denominator is due to $\sum_{i\sim j\in V_+}(v_i + v_j)^2 = \sum_{i\sim j\in V_+}(v_i^2 + v_j^2 + 2v_i v_j) \leq 2\sum_{i\sim j\in V_+}(v_i^2 + v_j^2) \leq 2\sum_{i\in V_+} v_i^2 d_i$. Continued from the last inequality,

$$\begin{aligned}
\lambda_1 \geq & \quad \frac{(\sum_{i,j\in V_+, i\sim j}(v_i - v_j)^2)^2}{2(\sum_{i\in V_+} v_i^2 d_i)^2}, \\
\geq & \quad \frac{(\sum_{i\in V_+} |v_i^2 - v_{i+1}^2| CUT(S_i))^2}{2(\sum_{i\in V_+} v_i^2 d_i)^2}, \qquad \text{since } v_1 \leq v_2 \leq \ldots \leq v_n \\
\geq & \quad \frac{(\sum_{i\in V_+} |v_i^2 - v_{i+1}^2| \alpha_G |\widetilde{Vol}(S_i)|)^2}{2(\sum_{i\in V_+} v_i^2 d_i)^2} \\
\geq & \quad \frac{\alpha_G^2}{2} \frac{(\sum_{i\in V_+} v_i^2 (\widetilde{Vol}(S_i) - \widetilde{Vol}(S_{i-1})))^2}{(\sum_{i\in V_+} v_i^2 d_i)^2} \\
= & \quad \frac{\alpha_G^2}{2} \frac{(\sum_{i\in V_+} v_i^2 d_i)^2}{(\sum_{i\in V_+} v_i^2 d_i)^2} = \frac{\alpha_G^2}{2}.
\end{aligned}$$

This completes the proof. □

Reference

[1] Chung, Fan R.K. Spectral Graph Theory. 1997, AMS-CBMS.

[2] Chung, Fan R.K. (2007) "Four proofs for the Cheeger inequality and graph partition algorithms". ICCM 2007.

[3] Jim Demmel's Lecture notes on Fiedler Theory at UC Berkeley: why we use unnormalized Laplacian eigenvectors for spectral partition.

[4] Chung and Lu. Complex Graphs and Networks. 2006, AMS-CBMS.