Graphs and Networks Lecture 7

Spectral Partitioning

Lecturer: Daniel A. Spielman September 27, 2007

7.1 Appology

I want to begin this lecture by admiting that the last lecture was not well-organized. My plan is to spend this lecture fixing the damage.

7.2 The Laplacian

Let me begin by recalling the definition of the Laplacian matrix of a graph $G = (V, E)$. Formally, it is given by

$$\boldsymbol{L}(u, v) = \begin{cases} -1 & \text{if } (u, v) \in E \\ d(u) & \text{if } u = v \\ 0 & \text{otherwise.} \end{cases}$$

One can also define the Laplacian by

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}.$$

The key property of the Laplacian that we exploit is that for $\boldsymbol{x} \in \mathbb{R}^V$

$$\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = \sum_{(u,v) \in E} (\boldsymbol{x}(u) - \boldsymbol{x}(v))^2 . \tag{7.1}$$

Let me give you a simple proof of this. First, consider a graph with one edge. It's Laplacian matrix is

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

So, for any vector $\boldsymbol{x} \in \mathbb{R}^2$,

$$\begin{pmatrix} \boldsymbol{x}(1) & \boldsymbol{x}(2) \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{x}(1) \\ \boldsymbol{x}(2) \end{pmatrix} = \boldsymbol{x}(1)^2 - 2\boldsymbol{x}(1)\boldsymbol{x}(2) + \boldsymbol{x}(2)^2 = (\boldsymbol{x}(1) - \boldsymbol{x}(2))^2 .$$

Now, to derive (7.1), for any two vertices u and v define $\boldsymbol{L}_{u,v}$ to be the Laplacian matrix of the graph containing just the edge from u to v. It is easy to show that

$$\boldsymbol{L} = \sum_{(u,v) \in E} \boldsymbol{L}_{u,v},$$

so that

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

Note that if $\boldsymbol{x}$ is the characteristic vector of a set S, that is

$$\boldsymbol{x}(u) = \begin{cases} 1 & \text{if } u \in S \\ 0 & \text{otherwise,} \end{cases}$$

then

$$\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = |\partial(S)| .$$

Let me state a few elementary facts about the Laplacian matrix. Their proofs are similar to proofs we did for the walk matrix.

Proposition 7.2.1. *Let $\boldsymbol{L}$ be the Laplacian of connected graph $G = (V, E)$.*

1. *The smallest eigenvalue of $\boldsymbol{L}$ is 0.*
2. $\boldsymbol{L}\boldsymbol{1} = \boldsymbol{0}$.
3. *The second-smallest eigenvalue of $\boldsymbol{L}$ is greater than* 0.

7.3 The Courant-Fischer Theorem

Theorem 7.3.1 (Courant-Fischer). *Let A be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then, for all $1 \leq k \leq n$,*

$$\lambda_k = \min_{S \text{ of dimension } k} \max_{x \in S} \frac{x^T A x}{x^T x}.$$

Note that the term

$$\frac{x^T A x}{x^T x}$$

is usually called the Rayleigh quotient of x. For an eigenvector v, the Rayleigh quotient of v is its eigenvalue.

We often exploit this theorem through the following corollary.

Corollary 7.3.2. *Let A be a symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. and corresponding eigenvectors $v_1, \ldots, v_n$. Then, for all $1 \leq k \leq n$,*

$$\lambda_k = \min_{x \perp v_1, \ldots, v_{k-1}} \frac{x^T A x}{x^T x}, \text{ and}$$

$$v_k = \arg \min_{x \perp v_1, \ldots, v_{k-1}} \frac{x^T A x}{x^T x}.$$

I'll now show you how the Courant-Fischer theorem is usually applied, and then I will sketch a proof of it for you.

The typical use of the Courant-Fischer theorem is to prove upper bounds on small eigenvalues (or lower bounds on large eigenvalues). For example, for the Laplacian we have $\boldsymbol{v}_1 = \mathbf{1}$. So, this tells us that the second-smallest eigenvalue of the Laplacian satisfies

$$\gamma_2 = \min_{\boldsymbol{x} \perp \mathbf{1}} \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

This means that for every vector $\boldsymbol{x}$ that is orthogonal to $\mathbf{1}$, we have

$$\gamma_2 \leq \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

So, to prove an upper bound on γ_2, it suffices to evidence an $\boldsymbol{x}$ orthogonal to $\mathbf{1}$ of small Rayleigh quotient.

Recall that last class we defined the *sparsity* of a cut S to be

$$\mathrm{sp}(S) \stackrel{\mathrm{def}}{=} \frac{|\partial(S)|}{\min(|S|, |V - S|)}.$$

We will now use Corollary 7.3.2 to show that

$$\gamma_2 \leq 2 \min_{S \subset V} \mathrm{sp}(S).$$

Let S be a set for of size at most $|V|/2$, and let χ_S be the characteristic vector of S. We then have

$$\chi_S^T \boldsymbol{L} \chi_S = |\partial(S)|,$$

and

$$\chi_S^T \chi_S = |S|.$$

This is almost what we want, but we don't have $\chi_S^T \mathbf{1} = 0$. So, lets modify χ_S to make it orthogonal to $\mathbf{1}$. Set

$$\boldsymbol{x} = \chi_S - c\mathbf{1},$$

where

$$c = |S| / |V|.$$

We now have $\boldsymbol{x}^T \mathbf{1} = 0$,

$$\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x} = \chi_S^T \boldsymbol{L} \chi_S = |\partial(S)|,$$

and

$$\boldsymbol{x}^T \boldsymbol{x} = |S| (1 - |S| / |V|)^2 + (|V - S|)(-|S| / |V|)^2 = |S| (1 - |S| / |V|) \geq |S|/2.$$

So,

$$\gamma_2 \leq \frac{\boldsymbol{x}^T \boldsymbol{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \frac{|\partial(S)|}{|S|/2} = 2\mathrm{sp}(S).$$

7.4 A proof of Courant-Fischer

I'll just prove the Corollary, which is equivalent to the Theorem. Recall that we can choose eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ so that

$$\boldsymbol{v}_i^T \boldsymbol{v}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Also recall that every vector $\boldsymbol{x}$ may be expressed as

$$\boldsymbol{x} = \sum_i \boldsymbol{v}_i \alpha_i, \quad \text{where} \quad \alpha_i = \boldsymbol{v}_i^T \boldsymbol{x}.$$

So, we have

$$\boldsymbol{A}\boldsymbol{x} = \sum_i \lambda_i \boldsymbol{v}_i \alpha_i,$$

and

$$\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} = \left(\sum_i \boldsymbol{v}_i \alpha_i \right)^T \left(\sum_i \lambda_i \boldsymbol{v}_i \alpha_i \right) = \sum_i \lambda_i \alpha_i^2.$$

Similarly, we have

$$\boldsymbol{x}^T \boldsymbol{x} = \sum_i \alpha_i^2.$$

So,

$$\frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_i \lambda_i \alpha_i^2.}{\sum_i \alpha_i^2.} \geq \frac{\sum_i \lambda_1 \alpha_i^2.}{\sum_i \alpha_i^2.} = \lambda_1.$$

Moreover, if $\boldsymbol{x} = \boldsymbol{v}_1$, then we achieve equality. This shows that

$$\min_{\boldsymbol{x}} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \lambda_1.$$

To extend to the case in which $\boldsymbol{x}$ is orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{k-1}$, note that this just implies that $\alpha_1 = \cdots = \alpha_{k-1} = 0$, and so we get

$$\frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \lambda_i,$$

and we get equality when $\boldsymbol{x} = \boldsymbol{v}_i$.

7.5 The Normalized Laplacian

We are now going to look at a matrix called the *normalized Laplacian* which turns out to usually be more useful than the ordinary Laplacian. It is defined by

$$\mathcal{L} \stackrel{\text{def}}{=} \boldsymbol{D}^{-1/2} \boldsymbol{L} \boldsymbol{D}^{-1/2}.$$

This is a symmetric matrix. Note that it has the same eigenvalues as (because it is similar to)

$$\boldsymbol{L}\boldsymbol{D}^{-1} = (\boldsymbol{D} - \boldsymbol{A})\boldsymbol{D}^{-1} = \boldsymbol{I} - \boldsymbol{A}\boldsymbol{D}^{-1} = \boldsymbol{I} - \boldsymbol{M},$$

where $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{D}^{-1}$ was the transition matrix of the random walk on G. So, the eigenvalues of $\mathcal{L}$ are just one minus the eigenvalues of $\boldsymbol{M}$, and there is also a translation between their eigenvectors.

Where the Laplacian is related to the sparsity and ratio of cuts, the normalized Laplacian is related to the conductance, where I recall that we defined the conductance of a set to be

$$\phi(S) \stackrel{\text{def}}{=} \frac{w(\partial(S))}{d(S)d(V-S)}.$$

To see why these should be related, consider the Rayleigh quotient with respect to $\mathcal{L}$:

$$\frac{\boldsymbol{x}^T \mathcal{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\boldsymbol{x}^T \boldsymbol{D}^{-1/2} \boldsymbol{L} \boldsymbol{D}^{-1/2} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

This is bit of a mess, and is much improved if we set $\boldsymbol{y} = \boldsymbol{D}^{-1/2}\boldsymbol{x}$, to get

$$\frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}} = \frac{\sum_{(u,v)\in E} (\boldsymbol{y}(u) - \boldsymbol{y}(v))^2}{\sum_u d(u)(\boldsymbol{y}(u))^2}.$$

So, where the Rayleigh quotient of the Laplacian had the sum of the terms squared, the Rayleigh quotient of the normalized Laplacian weights these by the degrees.

Let's see how the Courant-Fischer Theorem applies in this later form. The eigenvector of eigenvalue 0 of $\mathcal{L}$ is $\boldsymbol{d}^{1/2}$. So,

$$\boldsymbol{v}_2 = \arg \min_{\boldsymbol{x} \perp \boldsymbol{d}^{1/2}} \frac{\boldsymbol{x}^T \mathcal{L} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Setting $\boldsymbol{y}_1 = \boldsymbol{D}^{-1/2}\boldsymbol{v}_1 = \boldsymbol{D}^{-1/2}\boldsymbol{d}^{1/2} = \boldsymbol{1}$, and $\boldsymbol{y}_2 = \boldsymbol{D}^{-1/2}\boldsymbol{v}_2$, the condition that $\boldsymbol{v}_2$ be orthogonal to $\boldsymbol{v}_1$ becomes

$$0 = \boldsymbol{v}_2^T \boldsymbol{v}_1 = \left(\boldsymbol{y}_2^T \boldsymbol{D}^{1/2}\right)\left(\boldsymbol{D}^{1/2}\boldsymbol{y}_1\right) = \left(\boldsymbol{y}_2^T \boldsymbol{D}^{1/2}\right)\left(\boldsymbol{D}^{1/2}\boldsymbol{1}\right) = \boldsymbol{y}_2^T \boldsymbol{D}\boldsymbol{1} = \boldsymbol{y}_2^T \boldsymbol{d}.$$

So,

$$\boldsymbol{y}_2 = \arg \min_{\boldsymbol{y} \perp \boldsymbol{d}} \frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}}.$$

7.6 The Normalized Laplacian and Conductance

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the eigenvalues of the normalized Laplacian. As in the previous section, one can show that

$$\lambda_2 \leq d(V) \min_S \phi(S) = d(V)\Phi(S).$$

One of the greatest theorems in spectral graph theory is Cheeger's inequality (as proved for graphs by Jerrum and Sinclair), which says that

$$2\sqrt{2}\sqrt{\lambda_2} \geq d(V)\Phi(G).$$

So, whenever λ_2 is small, we know that $\Phi(G)$ is small as well.

Even more is true. If $\boldsymbol{v}_2$ is the eigenvector of eigenvalue λ_2 of $\mathcal{L}$, then one can use $\boldsymbol{v}_2$ to find a cut in G of conductance at most

$$\frac{2\sqrt{2}}{d(V)}\sqrt{\lambda_2}. \tag{7.2}$$

The way we do it is to instead consider the vector $\boldsymbol{y}_2 = \boldsymbol{D}^{-1/2}\boldsymbol{v}_2$, sort the vertices according to $\boldsymbol{y}_2$, and take some prefix. That is, the set has the form

$$S = \{u : \boldsymbol{y}_2(u) \leq t\}$$

for some threshold t. Note that $\boldsymbol{y}_2$ is one of the vectors that we used when we drew spectral pictures of graphs. So, this means that a cut of small conductance may be found by considering a vertical cut in the spectral embedding.

By slightly extending a theorem of Mihail (who proved it for the ordinary Laplacian) we can show that one only needs an approximation of the eigenvector. For any vector $\boldsymbol{y}$ that is orthogonal to $\boldsymbol{d}$, we can can use $\boldsymbol{y}$ to find a cut whose conductance will be bounded by (7.2), but with λ_2 replaced by

$$\frac{\boldsymbol{y}^T\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}}.$$