Spectral Graph Theory Lecture 7

Cheeger's Inequality

Daniel A. Spielman September 23, 2009

7.1 Overview

Today, we will prove Cheeger's Inequality. I consider it to be the most important theorem in spectral graph theory. Cheeger's inequality has many variants, all of which tell us in some way that when λ_2 of a graph is small, the graph has a cut of small conductance (or ratio or sparsity).

Recall that in Lecture 5 we proved:

Theorem 7.1.1. *Let $G = (V, E)$ be a graph and let L_G be its Laplacian matrix. Let $S \subset V$ and set $\sigma = |S| / |V|$. Then,*

$$|\partial(S)| \geq \lambda_2 |S| (1 - \sigma).$$

Cheeger's inequality will provide a converse to this theorem.

7.2 Warning

I'm pretty sure that I've missed a factor of 4 in one of the bounds in this lecture.

7.3 Conductance

Different versions of Cheeger's inequality are related to different measures of the quality of a cut. In Theorem 7.1.1, we were concerned with the number of edges cut divided by the number of vertices removed. The sharpest versions of Cheeger's inequality hold when edges are treated as the most important objects, rather than vertices. That is, we weight vertices by their degrees. In this case, we will be interested in two measures of the quality of a cut, its *conductance* and its *sparsity*. These are closely related, and their names are often interchanged. For this course, I use the convention that the *conductance* of a set of vertices S is given by

$$\phi(S) \stackrel{\text{def}}{=} d(V) \frac{|\partial(S)|}{d(S) d(\bar{S}),}$$

where I define

$$d(S) = \sum_{i \in S} d(i),$$

and $d(i)$ is the degree of vertex i. The constant $d(V)$ out front just helps us normalize this quantity. Note that when all vertices have the same degree, the conductance differs from the quantity measured in Theorem 7.1.1 by precisely a factor of d.

For this class, we will define the *sparsity* of a set S to be

$$\mathrm{sp}(S) \stackrel{\text{def}}{=} \frac{|\partial(S)|}{\min\left(d(S), d(\bar{S})\right).}$$

As one of $d(S)$ or $d(\bar{S})$ is always at least half of $d(V)$, we have

$$\phi(S) \geq \mathrm{sp}(S) \geq \phi(S)/2,$$

so these two quantities will never differ by more than a factor of 2.

The conductance of a graph is defined to be the minimum of conductance of a cut, and similarly for the sparsity:

$$\phi_G \stackrel{\text{def}}{=} \min_S \phi(S)$$
$$\mathrm{sp}_G \stackrel{\text{def}}{=} \min_S \mathrm{sp}(S).$$

7.4 The Normalized Laplacian

The conductance of a graph is best approximated by the eigenvalues of the Normalized Laplacian matrix of the graph. We define the Normalized Laplacian by

$$N_G = D_G^{-1/2} L_G D_G^{-1/2} = I - M_G = I - D_G^{-1/2} A_G D^{-1/2},$$

where M_G is the normalized adjacency matrix we saw on the homework. We will always denote the eigenvalues of N_G by $\nu_1, \ldots, \nu_n$, where for a connected graph G we have

$$0 = \nu_1 < \nu_2 \leq \cdots \leq \nu_n.$$

Recall from the problem set that the eigenvector of ν_1 is $\boldsymbol{d}^{1/2}$, where this is the vector whose ith entry is the square root of $d(i)$.

We first establish an analog of Theorem 7.1.1.

Theorem 7.4.1.

$$\phi_G \geq \nu_2/2.$$

We will prove this by constructing an optimization problem whose answer is ϕ_G, and then proving that ν_2 is the solution to a relaxation of this problem. First, note that for every S,

$$d(S)d(\bar{S}) = \left(\sum_{i\in S} d(i)\right)\left(\sum_{j\notin S} d(j)\right) = \sum_{i\in S, j\notin S} d(i)d(j).$$

To create an optimization problem, we consider characteristic vectors of sets instead of sets themselves. As

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

and

$$d(S)d(\bar{S}) = \sum_{\chi_S(i) > \chi_S(j)} d(i)d(j)(\chi_S(i) - \chi_S(j))^2 = \sum_{i<j} d(i)d(j)(\chi_S(i) - \chi_S(j))^2,$$

we have

$$\phi(S) = d(V)\frac{\chi_S^T L \chi_S}{\sum_{i<j} d(i)d(j)(\chi_S(i) - \chi_S(j))^2}.$$

If we let $\boldsymbol{y}$ be a vector of $\{0,1\}$ valued variables, then we have

$$\phi_G = \min_S \phi(S) = \min_{\boldsymbol{y} \in \{0,1\}^n} d(V)\frac{\boldsymbol{y}^T L \boldsymbol{y}}{\sum_{i<j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2}.$$

Of course, we only consider the minimum over $\boldsymbol{y}$ for which the denominator is non-zero.

One of the most useful ideas in optimization is that of *relaxation.* It means that one takes a problem like the above, and removes some constraints. In this case, we remove the constraint that the values of $\boldsymbol{y}$ lie in $\{0,1\}$. As we are minimizing, and we are removing a constraint, the minimum can only become lower. Thus,

$$\phi_G \geq \min_{\boldsymbol{y} \in \mathbb{R}^n} d(V)\frac{\boldsymbol{y}^T L \boldsymbol{y}}{\sum_{i<j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2}.$$

Theorem 7.4.1 may now be seen to be a consequence of the following characterization of ν_2.

Theorem 7.4.2.

$$\nu_2 = \min_{\boldsymbol{y} \in \mathbb{R}^n} 2d(V)\frac{\boldsymbol{y}^T L \boldsymbol{y}_S}{\sum_{i<j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2}.$$

Proof. Since $\boldsymbol{d}^{1/2}$ is an eigenvector of eigenvalue 0 of N_G, the Courant-Fischer Theorem tells us that

$$\nu_2 = \min_{\boldsymbol{x} \perp \boldsymbol{d}^{1/2}} \frac{\boldsymbol{x}^T N_G \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \min_{\boldsymbol{x} \perp \boldsymbol{d}^{1/2}} \frac{\boldsymbol{x}^T D_G^{-1/2} L_G D_G^{-1/2} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Now, set $\boldsymbol{y} = D_G^{-1/2}\boldsymbol{x}$. We obtain

$$\frac{\boldsymbol{x}^T D_G^{-1/2} L_G D_G^{-1/2} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\boldsymbol{y}^T L_G \boldsymbol{y}}{\boldsymbol{y}^T D_G \boldsymbol{y}}.$$

As $\boldsymbol{x}(i) = d(i)^{1/2}\boldsymbol{y}(i)$, the condition $\boldsymbol{x} \perp \boldsymbol{d}^{1/2}$ becomes $\boldsymbol{y} \perp \boldsymbol{d}$. So,

$$\nu_2 = \min_{\boldsymbol{y} \perp \boldsymbol{d}} \frac{\boldsymbol{y}^T L_G \boldsymbol{y}}{\boldsymbol{y}^T D_G \boldsymbol{y}}.$$

The denominator of this expression is $\sum_i d(i)\boldsymbol{y}(i)^2$. I claim that, when $\boldsymbol{y} \perp \boldsymbol{d}$

$$2d(V)\sum_i d(i)\boldsymbol{y}(i)^2 = \sum_{i,j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2.$$

To prove this, compute

$$\sum_{i,j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2 = \sum_{i,j} d(i)d(j)(\boldsymbol{y}(i)^2 + \boldsymbol{y}(j)^2) - 2\sum_{i,j} d(i)d(j)\boldsymbol{y}(i)\boldsymbol{y}(j)$$
$$= 2\sum_{i,j} d(i)d(j)\boldsymbol{y}(i)^2 - 2\left(\sum_i d(i)\boldsymbol{y}(i)\right)\left(\sum_j d(j)\boldsymbol{y}(j)\right)$$
$$= 2\left(\sum_j d(j)\right)\sum_i d(i)\boldsymbol{y}(i)^2.$$

We have shown

$$\nu_2 = \min_{\boldsymbol{y}\perp\boldsymbol{d}} 2d(V)\frac{\boldsymbol{y}^T L_G \boldsymbol{y}}{\sum_{i,j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2.} \tag{7.1}$$

To finish the proof, note that for every vector $\boldsymbol{y}$, there is a c for which $\boldsymbol{y} - c\mathbf{1}$ is orthogonal to $\boldsymbol{d}$. However, the addition of a constant c to every entry of $\boldsymbol{y}$ does not change the value of (7.1). So,

$$\nu_2 = \min_{\boldsymbol{y}\in\mathbb{R}^n} 2d(V)\frac{\boldsymbol{y}^T L_G \boldsymbol{y}}{\sum_{i,j} d(i)d(j)(\boldsymbol{y}(i) - \boldsymbol{y}(j))^2.}$$

□

7.5 Cheeger's Inequality

Theorem 7.4.1 is the easy part. Cheeger's inequality provides the following converse

$$\nu_2 \geq \phi_G^2/8.$$

Moreover, an examination of most proofs of Cheeger's inequalities reveal that a cut S for which

$$\nu_2 \geq \phi_G(S)^2/8$$

may be found by an examination of $\boldsymbol{v}_2$. In fact, we can find such a set S of the form

$$S = \left\{d^{-1/2}(i)\boldsymbol{v}_2(i) \geq t\right\}$$

for some t.

We will prove a strengthening of this inequality due to Mihail, which allows us to use any vector of small Rayleigh quotient instead of $\boldsymbol{v}_2$.

Theorem 7.5.1. *Let $\boldsymbol{x}$ be a vector orthogonal to $\boldsymbol{d}^{1/2}$. There exists a t for which the set of vertices*

$$S = \left\{i : d^{-1/2}(i)\boldsymbol{x}(i) \geq t\right\}$$

satisfies

$$\frac{\boldsymbol{x}^T N_G \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \phi(S)^2/8.$$

Our proof of this theorem will use the following three inequalities.

Lemma 7.5.2 (Dan's Favorite Inequality)**.** *Let A, B, C and D be non-negative. Then*

$$\frac{A+B}{C+D} \geq \min\left(\frac{A}{C}, \frac{B}{D}\right).$$

Proof.

$$A + B = C\frac{A}{C} + D\frac{B}{D} \geq (C+D)\min\left(\frac{A}{C}, \frac{B}{D}\right).$$

□

Lemma 7.5.3 (Square of difference)**.** *For all $0 < \phi < 1$,*

$$(a-b)^2 \geq \phi a^2 - \frac{\phi}{1-\phi}b^2.$$

Proof. Using $1 + \phi/(1-\phi) = 1/(1-\phi)$, we compute

$$\begin{aligned}
(a-b)^2 - \left(\phi a^2 - \frac{\phi}{1-\phi}b^2\right) &= (1-\phi)\,a^2 - 2ab + \left(\frac{1}{1-\phi}\right)b^2 \\
&= \left(\sqrt{1-\phi}a - \frac{1}{\sqrt{1-\phi}}b\right)^2 \\
&\geq 0.
\end{aligned}$$

□

Lemma 7.5.4 (Inequality 3)**.** *Let $z_1, \ldots, z_k$ be real numbers such that*

$$z_1 \geq z_2 \geq \cdots \geq z_k \geq 0.$$

If the real numbers $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ satisfy

$$\sum_{i=1}^{k} a_i z_i \geq \sum_{i=1}^{k} b_i z_i,$$

then there exists a j for which

$$\sum_{i=1}^{j} a_i \geq \sum_{i=1}^{j} b_i.$$

Proof. Assume, by way of contradiction that

$$\left(\sum_{i=1}^{j} a_i\right) < \left(\sum_{i=1}^{j} b_i\right) \tag{7.2}$$

for all j. Define $z_{k+1} = 0$, so that we can write

$$z_i = (z_i - z_{i+1}) + (z_{i+1} - z_{i+1}) + \cdots + (z_k - z_{k+1}),$$

From (7.2) and the fact that each $(z_i - z_{i+1})$ is non-negative, we compute

$$\sum_{i=1}^{k} a_i z_i = \sum_{j=1}^{k} (z_j - z_{j+1}) \left(\sum_{i=1}^{j} a_i \right) < \sum_{j=1}^{k} (z_j - z_{j+1}) \left(\sum_{i=1}^{j} b_i \right) = \sum_{i=1}^{k} z_i b_i,$$

contradicting the hypothesis of the lemma. □

Proof of Theorem 7.5.1. Let $\boldsymbol{x}$ be any vector orthogonal to $\boldsymbol{d}^{1/2}$, and set

$$\nu = \frac{\boldsymbol{x}^T N_G \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

As before, set

$$\boldsymbol{y} = D^{-1/2} \boldsymbol{x},$$

and recall this transformation ensures

$$\boldsymbol{y}^T \boldsymbol{d} = 0.$$

Let c be a number such that

$$\sum_{i:\boldsymbol{y}(i) > c} d(i) \leq d(V)/2 \qquad \text{and}$$

$$\sum_{i:\boldsymbol{y}(i) < c} d(i) \leq d(V)/2.$$

That is, c is $\boldsymbol{d}$-weighted median of $\boldsymbol{y}$. Let $\boldsymbol{z}$ be the vector $\boldsymbol{y} - c\mathbf{1}$, so that

$$\sum_{i:\boldsymbol{z}(i) > 0} d(i) \leq d(V)/2 \qquad \text{and}$$

$$\sum_{i:\boldsymbol{z}(i) < 0} d(i) \leq d(V)/2.$$

One may easily show

$$\boldsymbol{z}^T D \boldsymbol{z} \geq \boldsymbol{y}^T D \boldsymbol{y},$$

by taking a derivative of the expression for $\boldsymbol{z}$ with respect to c. We now have

$$\nu = \frac{\boldsymbol{x}^T N \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\boldsymbol{y}^T L \boldsymbol{y}}{\boldsymbol{y}^T D \boldsymbol{y}} \geq \frac{\boldsymbol{z}^T L \boldsymbol{z}}{\boldsymbol{z}^T D \boldsymbol{z}}.$$

Define vectors $\boldsymbol{z}_+$ and $\boldsymbol{z}_-$ that capture the positive and negative parts of $\boldsymbol{z}$,

$$\boldsymbol{z}_+(i) = \begin{cases} \boldsymbol{z}(i) & \text{if } \boldsymbol{z}(i) \geq 0 \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and}$$

$$\boldsymbol{z}_-(i) = \begin{cases} \boldsymbol{z}(i) & \text{if } \boldsymbol{z}(i) \leq 0 \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\boldsymbol{z}^T D\boldsymbol{z} = \boldsymbol{z}_+^T D\boldsymbol{z}_+ + \boldsymbol{z}_-^T D\boldsymbol{z}_-.$$

On the other hand, we can show

$$\boldsymbol{z}^T L\boldsymbol{z} \geq \boldsymbol{z}_+^T L\boldsymbol{z}_+ + \boldsymbol{z}_-^T L\boldsymbol{z}_-.$$

To see this, note that if $\boldsymbol{z}(i)$ and $\boldsymbol{z}(j)$ have the same sign

$$(\boldsymbol{z}(i) - \boldsymbol{z}(j))^2 = (\boldsymbol{z}_+(i) - \boldsymbol{z}_+(j))^2 + (\boldsymbol{z}_-(i) - \boldsymbol{z}_-(j))^2,$$

whereas when they have opposite signs,

$$(\boldsymbol{z}(i) - \boldsymbol{z}(j))^2 = \boldsymbol{z}(i)^2 + \boldsymbol{z}(j)^2 - 2\boldsymbol{z}(i)\boldsymbol{z}(j) \geq \boldsymbol{z}(i)^2 + \boldsymbol{z}(j)^2 = (\boldsymbol{z}_+(i) - \boldsymbol{z}_+(j))^2 + (\boldsymbol{z}_-(i) - \boldsymbol{z}_-(j))^2.$$

So,

$$\nu \geq \frac{\boldsymbol{z}^T L\boldsymbol{z}}{\boldsymbol{z}^T D\boldsymbol{z}} \geq \frac{\boldsymbol{z}_+^T L\boldsymbol{z}_+ + \boldsymbol{z}_-^T L\boldsymbol{z}_-}{\boldsymbol{z}_+^T D\boldsymbol{z}_+ + \boldsymbol{z}_-^T D\boldsymbol{z}_-} \geq \frac{\boldsymbol{z}_s^T L\boldsymbol{z}_s}{\boldsymbol{z}_s^T D\boldsymbol{z}_s},$$

for one of $s \in \{+, -\}$, by my favorite inequality. Let's assume without loss of generality that it holds for $s = +$. Also without loss of generality, assume that

$$\boldsymbol{z}(1) \geq \boldsymbol{z}(2) \geq \cdots \geq \boldsymbol{z}(k) \geq 0$$

are exactly the positive elements of $\boldsymbol{z}$. For ease of notation, we will replace all the vertices with non-postive values in $\boldsymbol{z}$ by a vertex $k+1$, set $\boldsymbol{z}(k+1) = 0$, and let E^+ be the set of edges

$$\{(i,j) \in E : i, j \leq k\} \cup \{(i, k+1) : (i,j) \in E, i \leq k < j\}.$$

So,

$$\boldsymbol{z}_+^T L\boldsymbol{z}_+ = \sum_{(i,j)\in E^+} (\boldsymbol{z}(i) - \boldsymbol{z}(j))^2.$$

We will prove the theorem by showing that for one of the sets

$$S_j \stackrel{\text{def}}{=} \{i : i \leq j\},$$

$$\frac{\boldsymbol{z}_+^T L\boldsymbol{z}_+}{\boldsymbol{z}_+^T D\boldsymbol{z}_+} \geq \mathrm{sp}(S_j)^2/2.$$

To this end, set

$$\sigma = \min_j \mathrm{sp}(S_j).$$

By our choice of c above, $d(S_j) \leq d(V)/2$, so

$$\mathrm{sp}(S_j) = \frac{|\partial(S_j)|}{d(S_j)} \geq \sigma.$$

For each $1 \leq i \leq k$, set

$$a_i = \left|\left\{(i,j) \in E^+\right\} : j > i\right|,$$

and

$$b_i = \left|\left\{(i,j) \in E^+\right\} : j < i\right|.$$

We have $d(i) = a_i + b_i$,

$$|\partial(S_j)| = \sum_{i=1}^{j} (a_i - b_i),$$

and

$$\sum_{i=1}^{j} b_i = \frac{d(S_j) - |\partial(S_j)|}{2}.$$

By assumption,

$$\sum_{i=1}^{j} (a_i - b_i) \geq \sigma \sum_{i=1}^{j} d(i).$$

We now compute

$$\begin{aligned}
\boldsymbol{z}_+^T L \boldsymbol{z}_+ &= \sum_{i<j:(i,j)\in E} (\boldsymbol{z}(i) - \boldsymbol{z}(j))^2 \\
&\geq \sigma \boldsymbol{z}(i)^2 - \frac{\sigma}{1-\sigma} \boldsymbol{z}(j)^2 && \text{by Lemma 7.5.3} \\
&= \sum_{i=1}^{k} \left(a_i \sigma - b_i \frac{\sigma}{1-\sigma}\right) \boldsymbol{z}(i)^2.
\end{aligned}$$

So,

$$\nu \sum_{i=1}^{k} d(i) \boldsymbol{z}(i)^2 \geq \sum_{i=1}^{k} \left(a_i \sigma - b_i \frac{\sigma}{1-\sigma}\right) \boldsymbol{z}(i)^2.$$

Thus, Lemma 7.5.4 tells us that there is some j for which

$$\begin{aligned}
\nu d(S_j) = \nu \sum_{i=1}^{j} d(i) &\geq \sum_{i=1}^{j} \left(a_i \sigma - b_i \frac{\sigma}{1-\sigma}\right) \\
&= \sigma \sum_{i=1}^{j} (a_i - b_i) - \frac{\sigma^2}{1-\sigma} \sum_{i=1}^{j} b_i \\
&= \sigma \left|\partial(S_i)\right| - \frac{\sigma^2}{1-\sigma} \frac{1-\sigma}{2} d(S_i) \\
&= \frac{\sigma^2}{2} d(S_i).
\end{aligned}$$

Thus, we may conclude

$$\nu \geq \frac{\sigma^2}{2}.$$

As we chose σ to be the minimum of $\sigma(S_j)$, this tells us that there is a j for which

$$\nu \geq \frac{\sigma(S_j)^2}{2} \geq \frac{1}{8} \phi(S_j)^2.$$

□

7.6 History

Cheeger [?] originally proved his inequality for manifolds. The extensions to graphs, in various forms, were proved by [?]. I belive that these notes contain the first proof of Mihail's version of Cheeger's theorem for the normalized Laplacian, although it should have been written somewhere before.