Spectral Graph Theory	Lecture 6
Conductance, the Normalized Laplacian, and Cheeger's Inequality	
Daniel A. Spielman	September 21, 2015

Disclaimer

These notes are not necessarily an accurate representation of what happened in class. The notes written before class say what I think I should say. I sometimes edit the notes after class to make them way what I wish I had said.

There may be small mistakes, so I recommend that you check any mathematically precise statement before using it in your own work.

These notes were last revised on September 21, 2015.

6.1 Overview

As the title suggests, in this lecture I will introduce conductance, a measure of the quality of a cut, and the normalized Laplacian matrix of a graph. I will then prove Cheeger's inequality, which relates the second-smallest eigenvalue of the normalized Laplacian to the conductance of a graph.

Cheeger [Che70] first proved his famous inequality for manifolds. Many discrete versions of Cheeger's inequality were proved in the late 80's [SJ89, LS88, AM85, Alo86, Dod84, Var85]. Some of these consider the walk matrix (which we will see in a week or two) instead of the normalized Laplacian, and some consider the isoperimetic ratio instead of conductance.

The proof that I present today follows an approach developed by Luca Trevisan [Tre11].

6.2 Conductance

Back in Lecture 2, we related to isoperimetric ratio of a subset of the vertices to the second eigenvalue of the Laplacian. We proved that for every $S \subset V$

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

where $s = |S| \,/\, |V|$ and

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

Re-arranging terms slightly, this can be stated as

$$|V| \frac{|\partial(S)|}{|S|\,|V-S|} \geq \lambda_2.$$

Cheeger's inequality provides a relation in the other direction. However, the relation is tighter and cleaner when we look at two slightly different quantities: the *conductance* of the set and the second eigenvalue of the normalized Laplacian.

The formula for conductance has a different denominator that depends upon the sum of the degrees of the vertices in S. I will write $d(S)$ for the sum of the degrees of the vertices in S. Thus, $d(V)$ is twice the number of edges in the graph. We define the conductance of S to be

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

Note that many similar, although sometimes slightly different, definitions appear in the literature. For example, we would instead use

$$\frac{d(V)\partial(S)}{d(S)d(V-S)},$$

which appears below in (6.5).

We define the conductance of a graph G to be

$$\phi_G \stackrel{\text{def}}{=} \min_{S \subset V} \phi(S).$$

The conductance of a graph is more useful in many applications than the isoperimetric number. I usually find that conductance is the more useful quantity when you are concerned about edges, and that isoperimetric ratio is most useful when you are concerned about vertices. Conductance is particularly useful when studying random walks in graphs.

6.3 The Normalized Laplacian

It seems natural to try to relate the conductance to the following generalized Rayleigh quotient:

$$\frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}}. \tag{6.1}$$

If we make the change of variables

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

then this ratio becomes

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

That is an ordinary Rayleigh quotient, which we understand a little better. The matrix in the middle is called the normalized Laplacian (see [Chu97]). We reserve the letter $\boldsymbol{N}$ for this matrix:

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

This matrix often proves more useful when examining graphs in which nodes have different degrees. We will let $0 = \nu_1 \leq \nu_2 \leq \cdots \leq \nu_n$ denote the eigenvalues of $\boldsymbol{N}$.

The conductance is related to ν_2 as the isoperimetric number is related to λ_2:

$$\nu_2/2 \leq \phi_G. \tag{6.2}$$

I include a proof of this in the appendix.

My goal for today's lecture is to prove Cheeger's inequality,

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

which is much more interesting. In fact, it is my favorite theorem in spectral graph theory.

The eigenvector of eigenvalue 0 of $\boldsymbol{N}$ is $\boldsymbol{d}^{1/2}$, by which I mean the vector whose entry for vertex u is the square root of the degree of u. Observe that

$$\boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\boldsymbol{d}^{1/2} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{1} = \boldsymbol{D}^{-1/2}\boldsymbol{0} = \boldsymbol{0}.$$

The eigenvector of ν_2 is given by

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

Transfering back into the variable $\boldsymbol{y}$, and observing that

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

we find

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

6.4 Cheeger's Inequality

Cheeger's inequality proves that if we have a vector $\boldsymbol{y}$, orthogonoal to $\boldsymbol{d}$, for which the generalized Rayleigh quotient (6.1) is small, then one can obtain a set of small conductance from $\boldsymbol{y}$. We obtain such a set by carefully choosing a real number t, and setting

$$S_t = \{u : \boldsymbol{y}(u) \leq t\}.$$

Theorem 6.4.1. *Let $\boldsymbol{y}$ be a vector orthogonal to $\boldsymbol{d}$. Then, there is a number t for which the set $S_t = \{u : \boldsymbol{y}(u) < t\}$ satisfies*

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

Before proving the theorem, I wish to make one small point about the denominator in the expression above. It is essentially minimized when $\boldsymbol{y}^T \boldsymbol{d} = 0$, at least with regards to shifts.

Lemma 6.4.2. *Let $\boldsymbol{v}_s = \boldsymbol{y} + z\mathbf{1}$. Then, the minimum of $\boldsymbol{v}_z^T \boldsymbol{D} \boldsymbol{v}_z^T$ is achieved at the z for which $\boldsymbol{v}_z^T \boldsymbol{d} = 0$.*

Proof. The derivative with respect to z is

$$2\boldsymbol{d}^T \boldsymbol{v}_z,$$

and the minimum is achieved when this derivative is zero. □

We begin our proof of Cheeger's inequality by defining

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

So, we need to show that there is a t for which $\phi(S_t) \leq \sqrt{2\rho}$.

By renumbering the vertices, we may assume without loss of generality that

$$\boldsymbol{y}(1) \leq \boldsymbol{y}(2) \leq \cdots \leq \boldsymbol{y}(n).$$

We begin with some normalization. Let j be the least number for which

$$\sum_{u=1}^{j} \boldsymbol{d}(u) \geq d(V)/2.$$

We would prefer a vector that is centered at j. So, set

$$\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{y}(j)\mathbf{1}.$$

This vector $\boldsymbol{z}$ satisfies $\boldsymbol{z}(j) = 0$, and, by Lemma 6.4.2,

$$\frac{\boldsymbol{z}^T \boldsymbol{L} \boldsymbol{z}}{\boldsymbol{z}^T \boldsymbol{D} \boldsymbol{z}} \leq \rho.$$

We also multiply $\boldsymbol{z}$ by a constant so that

$$\boldsymbol{z}(1)^2 + \boldsymbol{z}(n)^2 = 1.$$

Recall that

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

We will define a distribution on t for which we can prove that

$$\mathbb{E}\left[|\partial(S_t)|\right] \leq \sqrt{2\rho}\, \mathbb{E}\left[\min(d(S_t), d(V-S_t))\right].$$

This implies[1] that there is some t for which

$$|\partial(S_t)| \leq \sqrt{2\rho}\, \min(d(S_t), d(V-S_t)),$$

[1] If this is not immediately clear, note that it is equivalent to assert that $\mathbb{E}\left[\sqrt{2\rho}\min(d(S), d(V-S)) - |\partial(S)|\right] \geq 0$, which means that there must be some S for which the expression is non-negative.

which means $\phi(S) \leq \sqrt{2\rho}$.

To switch from working with $\boldsymbol{y}$ to working with $\boldsymbol{z}$, define We will set $S_t = \{u : \boldsymbol{z}(u) \leq t\}$. Trevisan had the remarkable idea of choosing t between $\boldsymbol{z}(1)$ and $\boldsymbol{z}(n)$ with probability density $2\,|t|$. That is, the probability that t lies in the interval $[a, b]$ is

$$\int_{t=a}^{b} 2\,|t|\,.$$

To see that the total probability is 1, observe that

$$\int_{t=\boldsymbol{z}(1)}^{\boldsymbol{z}(n)} 2\,|t| = \int_{t=\boldsymbol{z}(1)}^{0} 2\,|t| = + \int_{t=0}^{\boldsymbol{z}(n)} 2\,|t| = \boldsymbol{z}(n)^2 + \boldsymbol{z}(1)^2 = 1,$$

as $\boldsymbol{z}(1) \leq \boldsymbol{z}(j) \leq \boldsymbol{z}(n)$ and $\boldsymbol{z}(j) = 0$.

Similarly, the probability that t lies in the interval $[a, b]$ is

$$\int_{t=a}^{b} 2\,|t| = \mathsf{sgn}(b)b^2 - \mathsf{sgn}(a)a^2,$$

where

$$\mathsf{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0\text{, and} \\ -1 & \text{if } x < 0. \end{cases}$$

Lemma 6.4.3.

$$\mathbb{E}_t\left[|\partial(S_t)|\right] = \sum_{(u,v)\in E} Pr_t\left[(u,v) \in \partial(S_t)\right] \leq \sum_{(u,v)\in E} |\boldsymbol{z}(u) - \boldsymbol{z}(v)|\,(|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|). \tag{6.3}$$

Proof. An edge (u, v) with $\boldsymbol{z}(u) \leq \boldsymbol{z}(v)$ is on the boundary of S if

$$\boldsymbol{z}(u) \leq t < \boldsymbol{z}(v).$$

The probability that this happens is

$$\mathsf{sgn}(\boldsymbol{z}(v))\boldsymbol{z}(v)^2 - \mathsf{sgn}(\boldsymbol{z}(u))\boldsymbol{z}(u)^2 = \begin{cases} \left|\boldsymbol{z}(u)^2 - \boldsymbol{z}(v)^2\right| & \text{when } \mathsf{sgn}(u) = \mathsf{sgn}(v), \\ \boldsymbol{z}(u)^2 + \boldsymbol{z}(v)^2 & \text{when } \mathsf{sgn}(u) \neq \mathsf{sgn}(v). \end{cases}$$

We now show that both of these terms are upper bounded by

$$|\boldsymbol{z}(u) - \boldsymbol{z}(v)|\,(|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|).$$

Regardless of the signs,

$$\left|\boldsymbol{z}(u)^2 - \boldsymbol{z}(v)^2\right| = |(\boldsymbol{z}(u) - \boldsymbol{z}(v))(\boldsymbol{z}(u) + \boldsymbol{z}(v))| \leq |\boldsymbol{z}(u) - \boldsymbol{z}(v)|\,(|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|).$$

When $\mathsf{sgn}(u) = -\mathsf{sgn}(v)$,

$$\boldsymbol{z}(u)^2 + \boldsymbol{z}(v)^2 \leq (\boldsymbol{z}(u) - \boldsymbol{z}(v))^2 = |\boldsymbol{z}(u) - \boldsymbol{z}(v)|\,(|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|).$$

$\square$

We now derive a formula for the expected denominator of ϕ.

Lemma 6.4.4.

$$\mathbb{E}_t\left[\min(d(S_t), d(V - S_t))\right] = \boldsymbol{z}^T \boldsymbol{D} \boldsymbol{z}.$$

Proof. Observe that

$$\mathbb{E}_t\left[d(S_t)\right] = \sum_u \mathrm{Pr}_t\left[u \in S_t\right] d(u) = \sum_u \mathrm{Pr}_t\left[\boldsymbol{z}(u) \leq t\right] d(u).$$

The result of our centering of $\boldsymbol{z}$ at j is that

$$\begin{aligned} t < 0 &\implies d(S) = \min(d(S), d(V - S)), \quad \text{and} \\ t \geq 0 &\implies d(V - S) = \min(d(S), d(V - S)). \end{aligned}$$

That is, for $u < j$, u is in the smaller set if $t < 0$; and, for $u \geq j$, u is in the smaller set if $t \geq 0$. So,

$$\begin{aligned} \mathbb{E}_t\left[\min(d(S_t), d(V - S_t))\right] &= \sum_{u<j} \Pr\left[\boldsymbol{z}(u) < t \text{ and } t < 0\right] d(u) + \sum_{u \geq j} \Pr\left[\boldsymbol{z}(u) > t \text{ and } t \geq 0\right] d(u) \\ &= \sum_{u<j} \Pr\left[\boldsymbol{z}(u) < t < 0\right] d(u) + \sum_{u \geq j} \Pr\left[\boldsymbol{z}(u) > t \geq 0\right] d(u) \\ &= \sum_{u<j} \boldsymbol{z}(u)^2 d(u) + \sum_{u \geq j} \boldsymbol{z}(u)^2 d(u) \\ &= \sum_u \boldsymbol{z}(u)^2 d(u) \\ &= \boldsymbol{z}^T \boldsymbol{D} \boldsymbol{z}. \end{aligned}$$

□

Recall that our goal is to prove that

$$\mathbb{E}\left[|\partial(S_t)|\right] \leq \sqrt{2\rho}\, \mathbb{E}\left[\min(d(S_t), d(V - S_t))\right],$$

and we know that

$$\mathbb{E}_t\left[\min(d(S_t), d(V - S_t))\right] = \sum_u \boldsymbol{z}(u)^2 d(u)$$

and that

$$\mathbb{E}_t\left[|\partial(S_t)|\right] \leq \sum_{(u,v) \in E} |\boldsymbol{z}(u) - \boldsymbol{z}(v)| \left(|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|\right).$$

We may use the Cauchy-Schwartz inequality to upper bound the term above by

$$\sqrt{\sum_{(u,v) \in E} (\boldsymbol{z}(u) - \boldsymbol{z}(v))^2} \sqrt{\sum_{(u,v) \in E} (|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|)^2}. \tag{6.4}$$

We have defined ρ so that the term under the left-hand square root is at most

$$\rho \sum_u \boldsymbol{z}(u)^2 d(u).$$

To bound the right-hand square root, we observe

$$\sum_{(u,v)\in E} (|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|)^2 \leq 2 \sum_{(u,v)\in E} \boldsymbol{z}(u)^2 + \boldsymbol{z}(v)^2 = 2 \sum_u \boldsymbol{z}(u)^2 d(u).$$

Putting all these inequalities together yields

$$\begin{aligned}
\mathbb{E}\left[|\partial(S)|\right] &\leq \sqrt{\rho \sum_u \boldsymbol{z}(u)^2 d(u)} \sqrt{2 \sum_u \boldsymbol{z}(u)^2 d(u)} \\
&= \sqrt{2\rho} \sum_u \boldsymbol{z}(u)^2 d(u) \\
&= \sqrt{2\rho}\, \mathbb{E}\left[\min\left(d(S), d(V-S)\right)\right].
\end{aligned}$$

I wish to point out two important features of this proof:

1. This proof does not require $\boldsymbol{y}$ to be an eigenvector–it obtains a cut from any vector $\boldsymbol{y}$ that is orthogonal to $\boldsymbol{d}$.

2. This proof goes through almost without change for weighted graphs. The main difference is that for weighted graphs we measure the sum of the weights of edges on the boundary instead of their number. The main difference in the proof is that lines (6.3) and (6.4) become

$$\begin{aligned}
\mathbb{E}\left[w(\partial(S))\right] &= \sum_{(u,v)\in E} \Pr\left[(u,v) \in \partial(S)\right] w_{u,v} \\
&\leq \sum_{(u,v)\in E} |\boldsymbol{z}(u) - \boldsymbol{z}(v)| \, (|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|) w_{u,v} \\
&\leq \sqrt{\sum_{(u,v)\in E} w_{u,v} (\boldsymbol{z}(u) - \boldsymbol{z}(v))^2} \sqrt{\sum_{(u,v)\in E} w_{u,v} (|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|)^2},
\end{aligned}$$

   and we observe that

$$\sum_{(u,v)\in E} w_{u,v} (|\boldsymbol{z}(u)| + |\boldsymbol{z}(v)|)^2 \leq 2 \sum_u \boldsymbol{z}(u)^2 d(u).$$

The only drawback that I see to the approach that we took in this proof is that the application of Cauchy-Schwartz is a little mysterious. Shang-Hua Teng and I came up with a proof that avoids this by introducing one inequality for each edge. If you want to see that proof, look at my notes from 2009.

A Proof of (6.2)

Lemma A.1. *For every $S \subset V$,*

$$\phi(S) \geq \nu_2/2.$$

Proof. As in Lecture 2, we would like to again use $\boldsymbol{\chi}_S$ as a test vector. But, it is not orthogonal to $\boldsymbol{d}$. To fix this, we subtrat a constant. Set

$$\boldsymbol{y} = \boldsymbol{\chi}_S - \sigma \boldsymbol{1},$$

where

$$\sigma = d(S)/d(V).$$

You should now check that $\boldsymbol{y}^T \boldsymbol{d} = 0$:

$$\boldsymbol{y}^T \boldsymbol{d} = \boldsymbol{\chi}_S^T \boldsymbol{d} - \sigma \boldsymbol{1}^T \boldsymbol{d} = d(S) - (d(S)/d(V))d(V) = 0.$$

We already know that

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

It remains to compute $\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}$. If you remember the previous computation like this, you would guess that it is $d(S)(1-\sigma) = d(S)d(V-S)/d(V)$, and you would be right:

$$\begin{aligned}
\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} &= \sum_{u \in S} d(u)(1-\sigma)^2 + \sum_{u \notin S} d(u)\sigma^2 \\
&= d(S)(1-\sigma)^2 + d(V-S)\sigma^2 \\
&= d(S) - 2d(S)\sigma + d(V)\sigma^2 \\
&= d(S) - d(S)\sigma \\
&= d(S)d(V-S)/d(V).
\end{aligned}$$

So,

$$\nu_2 \leq \frac{\boldsymbol{y}^T \boldsymbol{L} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y}} = \frac{|\partial(S)|\, d(V)}{d(S)d(V-S)}. \tag{6.5}$$

As the larger of $d(S)$ and $d(V-S)$ is at least half of $d(V)$, we find

$$\nu_2 \leq 2 \frac{|\partial(S)|}{\min(d(S), d(V-S))}.$$

□

References

[Alo86] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[AM85] Noga Alon and V. D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Theory, Ser. B*, 38(1):73–88, 1985.

[Che70] J. Cheeger. A lower bound for smallest eigenvalue of the Laplacian. In *Problems in Analysis*, pages 195–199, Princeton University Press, 1970.

[Chu97] F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[Dod84] Jozef Dodziuk. Difference equations, isoperimetric inequality and transience of certain random walks. *Transactions of the American Mathematical Society*, 284(2):787–794, 1984.

[LS88] Gregory F. Lawler and Alan D. Sokal. Bounds on the l^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Transactions of the American Mathematical Society*, 309(2):557–580, 1988.

[SJ89] Alistair Sinclair and Mark Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation*, 82(1):93–133, July 1989.

[Tre11] Luca Trevisan. Lecture 4 from cs359g: Graph partitioning and expanders, stanford university, January 2011. available at `http://theory.stanford.edu/ trevisan/cs359g/lecture04.pdf`.

[Var85] N. Th. Varopoulos. Isoperimetric inequalities and Markov chains. *Journal of Functional Analysis*, 63(2):215 – 239, 1985.