

Lecture 7

*Lecturer: David P. Williamson**Scribe: Sam Gutekunst*

In this lecture, we introduce normalized adjacency and Laplacian matrices. We state and begin to prove Cheeger's inequality, which relates the second eigenvalue of the normalized Laplacian matrix to a graph's connectivity. Before stating the inequality, we will also define three related measures of expansion properties of a graph: conductance, (edge) expansion, and sparsity.

1 Normalized Adjacency and Laplacian Matrices

We use notation from Lap Chi Lau.

Definition 1 *The normalized adjacency matrix is*

$$\mathcal{A} \equiv D^{-1/2}AD^{-1/2},$$

where A is the adjacency matrix of G and $D = \text{diag}(d)$ for $d(i)$ the degree of node i .

For a graph G (with no isolated vertices), we can see that

$$D^{-1/2} = \begin{pmatrix} \frac{1}{\sqrt{d(1)}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{d(2)}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{d(n)}} \end{pmatrix}.$$

Definition 2 *The normalized Laplacian matrix is*

$$\mathcal{L} \equiv I - \mathcal{A}.$$

Notice that $\mathcal{L} = I - \mathcal{A} = D^{-1/2}(D - A)D^{-1/2} = D^{-1/2}L_GD^{-1/2}$, for L_G the (unnormalized) Laplacian.

Recall that for the largest eigenvalue λ of A and Δ the maximum degree of a vertex in a graph, $d_{\text{avg}} \leq \lambda \leq \Delta$. "Normalizing" the adjacency matrix makes its largest eigenvalue 1, so the analogous result for normalized matrices is the following:

Claim 1 *Let $\alpha_1 \geq \cdots \geq \alpha_n$ be the eigenvalues of \mathcal{A} and let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of \mathcal{L} . Then*

$$1 = \alpha_1 \geq \cdots \geq \alpha_n \geq -1, \quad 0 = \lambda_1 \leq \cdots \leq \lambda_n \leq 2.$$

⁰This lecture is derived from Lau's 2012 notes, Week 2, <http://appsrv.cse.cuhk.edu.hk/~chi/csc5160/notes/L02.pdf>.

Proof: First, we show that 0 is an eigenvalue of \mathcal{L} using the vector $x = D^{-1/2}e$. Then

$$\mathcal{L}(D^{1/2}e) = D^{-1/2}L_G D^{-1/2}D^{1/2}e = D^{-1/2}L_G e = 0,$$

since e is a eigenvector of L_G corresponding to eigenvalue 0. This shows that $D^{1/2}e$ is an eigenvector of \mathcal{L} of eigenvalue 0. To show that it's the smallest eigenvalue, notice that \mathcal{L} is positive semidefinite¹, as for any $x \in \mathbb{R}^n$:

$$\begin{aligned} x^T \mathcal{L} x &= x^T (I - \mathcal{A}) x \\ &= \sum_{i \in V} x(i)^2 - \sum_{(i,j) \in E} \frac{2x(i)x(j)}{\sqrt{d(i)d(j)}} \\ &= \sum_{(i,j) \in E} \left(\frac{x(i)}{\sqrt{d(i)}} - \frac{x(j)}{\sqrt{d(j)}} \right)^2 \\ &\geq 0. \end{aligned}$$

The last equality can be see “in reverse” by expanding $\left(\frac{x(i)}{\sqrt{d(i)}} - \frac{x(j)}{\sqrt{d(j)}} \right)^2$. We have now shown that \mathcal{L} has nonnegative eigenvalues, so indeed $\lambda_1 = 0$.

To show that $\alpha_1 \leq 1$, we make use of the positive semidefiniteness of $\mathcal{L} = I - \mathcal{A}$. This gives us that, for all $x \in \mathbb{R}^n$:

$$x^T (I - \mathcal{A}) x \geq 0 \implies x^T x - x^T \mathcal{A} x \geq 0 \implies 1 \geq \frac{x^T \mathcal{A} x}{x^T x}. \quad (1)$$

This Rayleigh quotient gives us the upper bound that $\alpha_1 \leq 1$. To get equality, consider again $x = D^{1/2}e$. Since, for this x ,

$$x^T \mathcal{L} x = 0 \implies x^T (I - \mathcal{A}) x = 0.$$

The exact same steps as in Equation 1 yield $\frac{x^T \mathcal{A} x}{x^T x} = 1$, as we now have equality.

To get a similar lower bound on α_n , we can show that $I + \mathcal{A}$ is positive semidefinite using a similar sum expansion². Then

$$x^T (I + \mathcal{A}) x \geq 0 \implies x^T x + x^T \mathcal{A} x \geq 0 \implies \frac{x^T \mathcal{A} x}{x^T x} \geq -1 \implies \alpha_n \geq -1.$$

¹A slick proof that does not make use of this quadratic is to use the fact that L_G is positive semidefinite. Thus $L_G = BB^T$ for some B , so that $\mathcal{L} = VV^T$ for $V = D^{-1/2}B$.

²This time, use

$$x^T (I + \mathcal{A}) x = \sum_{i \in V} x(i)^2 + \sum_{(i,j) \in E} \frac{2x(i)x(j)}{\sqrt{d(i)d(j)}} = \sum_{(i,j) \in E} \left(\frac{x(i)}{\sqrt{d(i)}} + \frac{x(j)}{\sqrt{d(j)}} \right)^2 \geq 0.$$

Finally, notice that $x^T(I + \mathcal{A})x \geq 0$ implies the following chain:

$$-x^T \mathcal{A} x \leq x^T x \implies x^T I x - x^T \mathcal{A} x \leq 2x^T x \implies \frac{x^T \mathcal{L} x}{x^T x} \leq 2 \implies \lambda_n \leq 2,$$

using the same Rayleigh quotient trick and that λ_n is the maximizer of that quotient. \square

Remark 1 Notice that, given the spectrum of \mathcal{A} , we have the following: $-\mathcal{A}$ has spectrum negatives of \mathcal{A} , and $I - \mathcal{A}$ adds one to each eigenvalue of $-\mathcal{A}$. Hence, $0 = \lambda_1 \leq \dots \leq \lambda_n \leq 2$ follows directly from $1 = \alpha_1 \geq \dots \geq \alpha_n \geq -1$.

2 Connectivity and $\lambda_2(\mathcal{L})$

Recall that $\lambda_2(L_G) = 0$ if and only if G is disconnected. The same is true for $\lambda_2(\mathcal{L})$, and we can say more!

2.1 Flavors of Connectivity

Let $S \subset V$. Recall that $\delta(S)$ denotes the set of edges with exactly one endpoint in S , and define $\text{vol}(S) \equiv \sum_{i \in S} d(i)$.

Definition 3 The conductance of $S \subset V$ is

$$\phi(S) \equiv \frac{|\delta(S)|}{\min\{\text{vol}(S), \text{vol}(V - S)\}}.$$

The edge expansion of S is

$$\alpha(S) \equiv \frac{|\delta(S)|}{|S|}, \quad \text{for } |S| \leq \frac{n}{2}.$$

The sparsity of S is

$$\rho(S) \equiv \frac{|\delta(S)|}{|S||V - S|}.$$

These measures are similar if G is d -regular (i.e., $d(i) = d$ for all $i \in V$). In this case,

$$\alpha(S) = d\phi(S), \quad \frac{n}{2}\rho(S) \leq \alpha(S) \leq n\rho(S).$$

To see the first equality, e.g., notice that the volume of S is $d|S|$.

In general, notice that $0 \leq \phi(S) \leq 1$ for all $S \subset V$.

We're usually interested in finding the sets S that minimize these quantities over the entire graph.

Definition 4 We define

$$\phi(G) \equiv \min_{S \subset V} \phi(S), \quad \alpha(G) \equiv \min_{S \subset V: |S| \leq \frac{n}{2}} \alpha(S), \quad \rho(G) \equiv \min_{S \subset V} \rho(S).$$

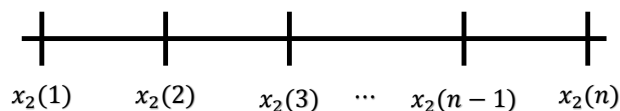
We call a graph G an *expander* if $\phi(G)$ (or $\alpha(G)$) is “large” (i.e. a constant³). Otherwise, we say that G has a sparse cut.

One algorithm for finding a sparse cut that works well in practice, but that lacks strong theoretical guarantees is called **spectral partitioning**.

Algorithm 1: Spectral Partitioning

- 1 Compute x_2 of \mathcal{L} (the eigenvector corresponding to $\lambda_2(\mathcal{L})$);
 - 2 Sort V such that $x_2(1) \leq \dots \leq x_2(n)$;
 - 3 Define the **sweep cuts** for $i = 1, \dots, n - 1$ by $S_i \equiv \{1, \dots, i\}$;
 - 4 Return $\min_{i \in \{1, \dots, n-1\}} \phi(S_i)$;
-

The following picture illustrates the idea of the algorithm; sweep cuts correspond to cuts between consecutive bars:



Cheeger’s inequality provides some insight into why this algorithm works well.

3 Cheeger’s Inequality

We now work towards proving the following:

Theorem 2 (Cheeger’s Inequality) Let λ_2 be the second smallest eigenvalue of \mathcal{L} . Then:

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}.$$

The theorem proved by Jeff Cheeger actually has to do with manifolds and hypersurfaces; the theorem above is considered to be a discrete analog of Cheeger’s original inequality. But the name has stuck.

Typically, people think of the first inequality being “easy” and the second being “hard.” We’ll prove the first inequality, and start the proof of the second inequality.

³One should then ask “A constant with respect to what?” Usually one defines families of graphs of increasing size as families of expanders, in which case we want the conductance or expansion constant with respect to the number of vertices.

Proof: Recall that

$$\lambda_2 = \min_{x: \langle x, D^{1/2}e \rangle = 0} \frac{x^T \mathcal{L}x}{x^T x} = \min_{x: \langle x, D^{1/2}e \rangle = 0} \frac{x^T D^{-1/2} L_G D^{-1/2} x}{x^T x}.$$

Consider the change of variables obtained by setting $y = D^{-1/2}x$ and $x = D^{1/2}y$:

$$\lambda_2 = \min_{y: \langle D^{1/2}y, D^{1/2}e \rangle = 0} \frac{y^T L_G y}{(D^{1/2}y)^T (D^{1/2}y)} = \min_{y: \langle D^{1/2}y, D^{1/2}e \rangle = 0} \frac{y^T L_G y}{y^T D y}.$$

The minimum is being taken over all y such that $\langle D^{1/2}y, D^{1/2}e \rangle = 0$. That is, over y such that:

$$(D^{1/2}y)^T D^{1/2}e = 0 \iff y^T D e = 0 \iff \sum_{i \in V} d(i)y(i) = 0.$$

Hence, we have that

$$\lambda_2 = \min_{y: \sum_{i \in V} d(i)y(i) = 0} \frac{\sum_{(i,j) \in E} (y(i) - y(j))^2}{\sum_{i \in V} d(i)y(i)^2}.$$

Now let S^* be such that $\phi(G) = \phi(S^*)$, and try defining

$$\hat{y}(i) = \begin{cases} 1, & i \in S^* \\ 0, & \text{else.} \end{cases}$$

It would be great if λ_2 was bounded by $\frac{|\delta(S^*)|}{\sum_{i \in S^*} d(i)} = \frac{|\delta(S^*)|}{\text{vol}(S^*)}$. However, there are two problems. We have $\sum_{i \in V} d(i)\hat{y}(i) \neq 0$; moreover $\frac{|\delta(S^*)|}{\text{vol}(S^*)}$ might not be $\phi(S^*)$, as we want the denominator to be $\min\{\text{vol}(S^*), \text{vol}(V - S^*)\}$. Hence, we redefine

$$\hat{y}(i) = \begin{cases} \frac{1}{\text{vol}(S^*)}, & i \in S^* \\ -\frac{1}{\text{vol}(V - S^*)}, & \text{else.} \end{cases}$$

Now we notice that:

$$\sum_{i \in V} d(i)\hat{y}(i) = \frac{\sum_{i \in S^*} d(i)}{\text{vol}(S^*)} - \frac{\sum_{i \notin S^*} d(i)}{\text{vol}(V - S^*)} = 1 - 1 = 0.$$

Thus, this is a feasible solution to the minimization problem defining λ_2 , and we have that the only edges contributing anything nonzero to the numerator are those with exactly one endpoint in S^* . Thus:

$$\lambda_2 \leq \frac{|\delta(S^*)| \left(\frac{1}{\text{vol}(S^*)} + \frac{1}{\text{vol}(V - S^*)} \right)^2}{\sum_{i \in S^*} d(i) \left(\frac{1}{\text{vol}(S^*)} \right)^2 + \sum_{i \notin S^*} d(i) \left(\frac{1}{\text{vol}(V - S^*)} \right)^2}$$

$$\begin{aligned}
&= \frac{|\delta(S^*)| \left(\frac{1}{\text{vol}(S^*)} + \frac{1}{\text{vol}(V-S^*)} \right)^2}{\frac{1}{\text{vol}(S^*)} + \frac{1}{\text{vol}(V-S^*)}} \\
&= |\delta(S^*)| \left(\frac{1}{\text{vol}(S^*)} + \frac{1}{\text{vol}(V-S^*)} \right) \\
&\leq 2|\delta(S^*)| \max \left\{ \frac{1}{\text{vol}(S^*)}, \frac{1}{\text{vol}(V-S^*)} \right\} \\
&= \frac{2|\delta(S^*)|}{\min\{\text{vol}(S^*), \text{vol}(V-S^*)\}} \\
&= 2\phi(G).
\end{aligned}$$

□

This completes the proof of the first inequality. To get the second, the idea is to suppose we had a y with

$$R(y) \equiv \frac{\sum_{(i,j) \in E} (y(i) - y(j))^2}{\sum_{i \in V} d(i)y(i)^2}.$$

Claim 3 *We'll be able to find a cut $S \subset \text{supp}(Y) \equiv \{i \in V : y(i) \neq 0\}$ with $\frac{\delta(S)}{\text{vol}(S)} \leq \sqrt{2R(y)}$.*

This will not suffice to prove the second part of the inequality, as $\frac{\delta(S)}{\text{vol}(S)}$ need not equal $\phi(S)$, but we'll come back to this next lecture.

Proof: Without loss of generality, we assume $-1 \leq y(i) \leq 1$, as we can scale y if not. Our trick (from Trevisan) is to pick $t \in (0, 1]$ uniformly at random, and let $S_t = \{i \in V : y(i)^2 \geq t\}$. Notice that:

$$\mathbb{E}[\text{vol}(S_t)] = \sum_{i \in V} d(i) \Pr[i \in S_t] = \sum_{i \in V} d(i)y(i)^2,$$

and assuming⁴ that $(i, j) \in E \implies y(i)^2 \leq y(j)^2$,

$$\mathbb{E}[|\delta(S_t)|] = \sum_{(i,j) \in E} \Pr[(i, j) \in \delta(S_t)] = \sum_{(i,j) \in E} \Pr[y(i)^2 < t \leq y(j)^2] = \sum_{(i,j) \in E} (y(j)^2 - y(i)^2).$$

⁴We make this assumption without loss of generality because it doesn't matter in the end and is notationally convenient.

Rewriting the above using difference of squares and using Cauchy-Schwarz,

$$\begin{aligned}
\sum_{(i,h) \in E} (y(j) - y(i))(y(j) + y(i)) &\leq \sqrt{\sum_{(i,j) \in E} (y(j) - y(i))^2} \sqrt{\sum_{(i,j) \in E} (y(j) + y(i))^2} \\
&\leq \sqrt{\sum_{(i,j) \in E} (y(j) - y(i))^2} \sqrt{2 \sum_{(i,j) \in E} (y(j)^2 + y(i)^2)} \\
&= \sqrt{\sum_{(i,j) \in E} (y(j) - y(i))^2} \sqrt{\sum_{i \in V} 2d(i)y(i)^2} \\
&= \sqrt{2R(y)} \sqrt{\sum_{i \in V} d(i)y(i)^2}.
\end{aligned}$$

This gives that

$$\frac{\mathbb{E}[|\delta(S_t)|]}{\mathbb{E}[\text{vol}(S_t)]} \leq \sqrt{R(y)} \implies \mathbb{E}[|\delta(S_t)| - \sqrt{2R(y)} \text{vol}(S_t)] \leq 0.$$

This means that there exists a t such that

$$\frac{|\delta(S_t)|}{\text{vol}(S_t)} \leq \sqrt{2R(y)}.$$

To derandomize the algorithm, look at each of the n possible cuts S_t by looking at sweep cuts for the order $y(1)^2 \leq y(2)^2 \leq \dots \leq y(n)^2$. \square