ORIE 6334 Bridging Continuous and Discrete Optimization Oct. 2, 2019

Lecture 9

Lecturer: David P. Williamson *Scribe: Yilun Chen*

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*

$$\mathscr{A} \triangleq D^{-1/2} A D^{-1/2},$$

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

For a graph G (with no isolated vertices)

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

Definition 2 *The* **normalized Laplacian matrix** *is*

$$\mathscr{L} \triangleq I - \mathscr{A}.$$

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

We now give simple bounds on the eigenvalues of $\mathscr{A}$ and $\mathscr{L}$. Recall that for the largest eigenvalue λ of A, $d_{avg} \leq \lambda \leq \Delta$(the maximum degree). "Normalizing" the adjacency matrix makes its largest eigenvalue 1, so the analogous result for normalized matrices is the following:

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

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

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

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

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

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

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

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

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

$$x^T(I - \mathscr{A})x \geq 0 \implies x^Tx - x^T\mathscr{A}x \geq 0 \implies 1 \geq \frac{x^T\mathscr{A}x}{x^Tx}. \tag{1}$$

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,

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

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

[1] A 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 $\mathscr{L} = VV^T$ for $V = D^{-1/2}B$.

To get a similar lower bound on α_n, we can show that $I+\mathscr{A}$ is positive semidefinite using a similar sum expansion[2]. Then

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

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

$$-x^T\mathscr{A}x \leq x^Tx \implies x^TIx - x^T\mathscr{A}x \leq 2x^Tx \implies \frac{x^T\mathscr{L}x}{x^Tx} \leq 2 \implies \lambda_n \leq 2,$$

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

□

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

Recall that $\lambda_2(L_G)=0$ if and only if G is disconnected. The same is true for $\lambda_2(\mathscr{L})$. In today (and next time)'s lecture, we demonstrate that $\lambda_2(\mathscr{L})$ actually shows how well G is connected.

2 Measures of Connectivity

We introduce three related measures of connectivity of a graph G. 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*

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

The **sparsity** *of S is*

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

[2]This time, use

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

Some direct observations that relate the above three measures of connectivity:

- $\frac{n}{2}\rho(S) \leq \alpha(S) \leq n\rho(S)$.
- If G is d-regular (i.e. $d(i) = d$ for all $i \in V$), then $\alpha(S) = d\phi(S)$.
- $0 \leq \phi(S) \leq 1$ 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*

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

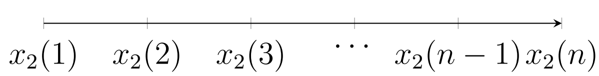
We call a graph G an *expander* if $\phi(G)$ (or $\alpha(G)$) is "large" (i.e. a constant[3]). Otherwise, we say that G has a sparse cut.

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

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

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 Connectivity and $\lambda_2(\mathscr{L})$: Cheeger's Inequality

We now work towards proving the following Cheeger's inequality, which links the second eigenvalue of $\mathscr{L}$ with the connectivity of the graph. For simplicity, we here only consider **d-regular graphs**. For a proof in the general case one may check the scribe notes from three years ago[4].

[3] One 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 to be constant with respect to the number of vertices.

[4] https://people.orie.cornell.edu/dpw/orie6334/Fall2016/

Theorem 2 (Cheeger's Inequality) [5] *Let λ_2 be the second smallest eigenvalue of $\mathscr{L}$. Then:*

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

Remark 2 *Typically, people think of the lower bound being "easy" and the upper bound being "hard." We'll prove the lower bound, and start the proof of the upper bound (continued next time).*

Let's first prove the easier one of the inequalities, namely $\frac{\lambda_2}{2} \le \phi(G)$.

Proof: Since the graph is assumed to be d-regular, $\mathscr{L} = \frac{1}{d}L_G$. Therefore, $\mathscr{L}e = \frac{1}{d}L_G e = 0$, where e, the all ones vector, is an eigenvector of L_G associated with the eigenvalue 0. (recall from the previous lectures.) We thus conclude that e is also an eigenvector of $\mathscr{L}$ associated with the eigenvalue 0. Now with the Raleigh quotient representation of eigenvalues:

$$\lambda_2 = \min_{z:<z,e>=0} \frac{z^\intercal \mathscr{L} z}{z^\intercal z} \le \frac{x^\intercal \mathscr{L} x}{x^\intercal x}, \qquad \text{for any } x:\ \langle x, e\rangle = 0.$$

The high level idea behind the proof is to find a proper x, so that $\frac{x^\intercal \mathscr{L} x}{x^\intercal x}$ is related in some way to the conductance $\phi(G)$ and that the chains of inequalities will then lead to the desired result. For simplicity, let $\bar{S}$ be such that $\phi(G) = \phi(\bar{S})$. We set $x(i) = \frac{1}{|\bar{S}|}$ for $i \in \bar{S}$ and $x(i) = -\frac{1}{|V-\bar{S}|}$ otherwise. Then it's not hard to verify that $\langle x, e\rangle = \frac{|\bar{S}|}{|\bar{S}|} - \frac{|V-\bar{S}|}{|V-\bar{S}|} = 0$. We thus have

$$\lambda_2 \le \frac{x^\intercal \mathscr{L} x}{x^\intercal x} = \frac{\frac{1}{d}x^\intercal L_G x}{x^\intercal x} = \frac{\frac{1}{d}\sum_{(i,j)\in E}(x(i)-x(j))^2}{\sum_{i\in V} x^2(i)}.$$

Recall the definition of x, we have that the above right hand side equals

$$\frac{\frac{1}{d}|\delta(\bar{S})|\big(\frac{1}{|\bar{S}|} + \frac{1}{|V-\bar{S}|}\big)^2}{\sum_{i\in\bar{S}} \frac{1}{|\bar{S}|^2} + \sum_{i\in V-\bar{S}} \frac{1}{|V-\bar{S}|^2}} = \frac{1}{d}|\delta(\bar{S})|\left(\frac{1}{|\bar{S}|} + \frac{1}{|V-\bar{S}|}\right) = \frac{n|\delta(\bar{S})|}{d|\bar{S}||V-\bar{S}|}$$

We notice that the above right hand side can be upper bounded by

$$\frac{n|\delta(\bar{S})|}{d|\bar{S}||V-\bar{S}|} \le \frac{2|\delta(\bar{S})|}{\min(\mathrm{vol}(\bar{S}), \mathrm{vol}(V-\bar{S}))} = 2\phi(\bar{S}),$$

where the bound follows from the facts:

[5] 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.

- $\min(\mathrm{vol}(\bar{S}), \mathrm{vol}(V - \bar{S})) = d \min(|\bar{S}|, |V - \bar{S}|)$ (since G is d-regular.)
- $\max(|\bar{S}|, |V - \bar{S}|) \geq n/2$.
- $|\bar{S}||V - \bar{S}| = \max(|\bar{S}|, |V - \bar{S}|) \times \min(|\bar{S}|, |V - \bar{S}|)$.

Combining the chains of inequalities, we finally conclude that

$$\frac{\lambda_2}{2} \leq \phi(\bar{S}).$$

□

This completes the proof of the lower bound. To get the upper bound, the idea is to find a subset S, such that the following bounds hold:

- $|S| \leq \frac{n}{2}$.
- $\frac{|\delta(S)|}{d|S|} \leq \sqrt{2\lambda_2}$.

By definition of conductance $\phi(G)$, the existence of such subset S immediately implies the upper bound. Let's formalize the above high level idea. To begin with, we introduce some additional notation.

Definition 5 *For any vector $x \in R^n$, we define*

$$supp(x) \triangleq \{i : x(i) \neq 0\}, \;\; supp^+(x) \triangleq \{i : x(i) > 0\}, \;\; supp^-(x) \triangleq \{i : x(i) < 0\}.$$

Definition 6 *For any vector $x \in R^n$, we define*

$$R(x) \triangleq \frac{x^\intercal \mathscr{L} x}{x^\intercal x}.$$

Let x_2 be the eigenvector associated with λ_2, then $R(x_2) = \lambda_2$. Without loss of generality, we may assume $|\mathrm{supp}^+(x_2)| \leq |\mathrm{supp}^-(x_2)|$ (for otherwise just setting $x_2 \leftarrow -x_2$), then $|\mathrm{supp}^+(x_2)| \leq n/2$. We introduce an auxiliary vector $y \triangleq x_2^+$, or equivalently $y(i) = \max(0, x_2(i))$ for all $i \in [n]$. The vector y shall serve as a bridge as we march towards the bound $\frac{|\delta(S)|}{d|S|} \leq \sqrt{2\lambda_2}$. More specifically, we seek to prove $\frac{|\delta(S)|}{d|S|} \leq \sqrt{2R(y)}$ for some $|S| \leq n/2$, and $R(y) \leq R(x_2) = \lambda_2$.

Claim 3 $R(y) \leq R(x_2)$.

Proof: For all i such that $y(i) > 0$, we have

$$(\mathscr{L}y)(i) = (Iy - \frac{1}{d}Ay)(i) = y(i) - \frac{1}{d}\sum_{j:(i,j)\in E} y(j) \leq x_2(i) - \frac{1}{d}\sum_{j:(i,j)\in E} x_2(j),$$

where the last inequality follows from the definition of y. The above right-hand side is equal to

$$(Ix_2 - \frac{1}{d}Ax_2)(i) = (\mathscr{L}x_2)(i) = \lambda_2 x_2(i)$$

since x_2 is the eigenvector. We thus have

$$y^\intercal \mathscr{L} y = \sum_{i\in[n]} y(i)(\mathscr{L}y)(i) \leq \sum_{i:y(i)>0} \lambda_2 (x_2(i))^2 = \lambda_2 \sum_{i:y(i)>0} (y(i))^2.$$

As a result

$$R(y) = \frac{y^\intercal \mathscr{L} y}{y^\intercal y} \leq \lambda_2 \frac{\sum_{i:y(i)>0}(y(i))^2}{\sum_{i\in[n]}(y(i))^2} = \lambda_2 = R(x_2).$$

□

Next time we are going to prove the following

Lemma 4 *For any $x \in R^n$, there exists $\bar{S} \subseteq supp(x)$ such that*

$$\frac{|\delta(\bar{S})|}{d|\bar{S}|} \leq \sqrt{2R(x)}.$$

Notice that if we apply the above Lemma with y defined previously, the set $\bar{S}$ found must satisfy $|\bar{S}| \leq |\text{supp}(y)| \leq n/2$. Combining with the claim $R(y) \leq \lambda_2$ and the definition of $\phi(G)$, we will have proved the desired upper bound.