CS 229r Spectral Graph Theory in Computer Science, Lecture 8	Sep 29, 2020

Lecture 8

Instructor: Salil Vadhan *Scribe: Luyi Qiu & Juspreet Singh Sandhu*

1 Agenda

1. Review the notions of *Conductance*($\phi(G)$) & *Mixing-Time*($t_{mix}(\varepsilon)$) and contrast the latter's relationship with ν_2 with the former's.
2. Introduce **Cheeger's Inequality**.
3. Briefly review the proof for the easy-direction and flesh out the details for the hard-direction.

2 Conductance & Mixing-Time

For the purposes of this lecture, every graph G is implicitly assumed to be undirected. Cheeger's inequality bounds the conductance $\phi(G)$ of a graph from above and below using the second-eigenvalue ν_2 of the normalized laplacian N. We shall see examples of two graphs (section 4) that exhibit that the inequality is tight (up-to constant factors) on both sides. The conductance of a graph measures the minimum weight that escapes a partitioning of the vertices of the graph by a corresponding cut.

Definition 1 (Conductance)**.** The conductance of a graph $G = (V, E, w)$ is:

$$\phi(G) = \min_{S \subset V} \frac{w(\partial S)}{\min(d(S), d(V - S))} . \tag{1}$$

where $\partial S = \{(a, b) | a \in S, b \notin S\}$ and $d(S) = \sum_{s \in S} d(s)$.

Another way to interpret this quantity is that the conductance of a graph G is the minimum "surface area" (escaping edges) to "volume" (incident edges on S) ratio under a vertex partitioning.

Let us define a slightly different version of conductance.

Definition 2 (Normalized Conductance)**.** The normalized conductance of a cut induced by a partition $(S, V - S)$ is:

$$\hat{\phi}(S) = \frac{w(\partial S) d(V)}{d(S) d(V - S)} . \tag{2}$$

This conductance is convenient to work with for the purposes of evaluating the Rayleigh quotient when trying to prove Cheeger's inqequality.
The mixing time of a random walk on a graph $G = (V, E, w)$ is the minimum number of steps ($t_{mix}(\varepsilon)$) it takes for the probability distribution on vertices to get ε-close in TVD to the stationary distribution across all possible starting distributions.

Definition 3 (ε Mixing-Time)**.** The minimum number of steps $t_{mix}(\varepsilon)$ needed by a random walk across any initial distribution p_0 on the vertices V of a graph with stationary distribution π, such that:

$$\left| W^t p_0 - \pi \right|_1 \leq \varepsilon .$$

The mixing-time is also related to the second eigenvalue ν_2 of the normalized laplacian as we saw before. Intuitively, the better connected a graph is (larger ν_2) the faster it mixes. Combining the result of *Problem-5* on *Pset-2* with the upper-bound from *Lecture-5*, we know that:

$$\Omega\left(\frac{\log\left(\frac{1}{\varepsilon}\right)}{\nu_2}\right) \leq \tilde{t}_{mix}(\varepsilon) \leq \mathcal{O}\left(\frac{\log(\frac{nd_{max}}{\varepsilon d_{min}})}{\nu_2}\right). \tag{3}$$

where we utilize the fact that $\tilde{\omega}_\pi = \tilde{\omega_2} = \frac{1}{2}(1+\omega_2) = 1 - \frac{\nu_2}{2}$. Cheeger's inequality allows one to similarly sandwich the conductance of a graph by functions of ν_2.

3 Cheeger's Inequality vs. Mixing-Time

We state Cheeger's inequality, with the underlying assumption that the graphs being considered are undirected.

Theorem 4 (Cheeger's Inequality). *For all graphs $G = (V, E, w)$, the following holds:*

$$\frac{\nu_2}{2} \leq \phi(G) \leq \sqrt{2\nu_2}\,. \tag{4}$$

The left-hand side of Cheeger's inequality says that if there is a cut which is small, then there is a vector orthogonal to $\vec{\mathbf{d}}$ that has a comparably small eigenvalue (for N). The right-hand side of Cheeger's inequality says that if there is ν_2 is small, then there is a cut that is sparse. Contrast this with Equation 3, where $t_{mix}(\varepsilon)$ is bounded by ν_2^{-1} on both sides. Notice that if $\nu_2 > 1$, then Cheeger's inequality is trivial. This can be seen by interpreting $\phi(G)$ as the probability of escaping the most "poorly" connected set S during a random walk (and that can never be more than 1). Nevertheless, we shall now see that there are graphs that make Cheeger's inequality tight on both sides (upto constant factors).

4 Two Examples

Consider two examples of graphs: C_n (the undirected n-cycle) and an imbalanced dumbell D_{n_1,n_2}. The latter is formed by taking two complete graphs K_{n_1} and K_{n_2} with $n_1 < n_2$, and removing 2 edges from each at random and connecting the vertices with reduced degree by a single edge.

Example 5 (C_n). Notice that the undirected n-cycle has $\nu_2 = 1 - \frac{1}{2}(\mu_1)$ (since it is 2-regular). Since $C_n = \mathtt{Cay}(\mathbb{Z}_n, \{-1, 1\})$, this immediately implies $\mu_1 = 2\cos(\frac{2\pi}{n}) \approx 2 - \Theta\left(\frac{1}{n^2}\right)$ and $\nu_2 = \Theta\left(\frac{1}{n^2}\right)$. Note that the cut that minimizes expansion is the balanced cut with 2 escaping edges. This yields that $\phi(C_n) = \Theta(\frac{1}{n})$, and notice that this combination of ν_2 and $\phi(C_n)$ are tight instantiations of the right-hand side of Cheeger's inequality (up to constant factors).

Example 6 (D_{n_1,n_2}). It's easy to see that the conductance cut is the one which slices the dumbell into two parts, one with n_1 vertices and the other with n_2 vertices. Since the boundary has just 1 edge, the conductance is $\phi(D_{n_1,n_2}) = \Theta\left(\frac{1}{n_1^2}\right)$. It is slightly non-trivial to see, but a random walk will have mixing time $t_{mix}(\frac{1}{4}) \approx \Theta\left(\frac{1}{n_1^2}\right)$ since one needs to reach the boundary vertex (which happens with probability $\frac{1}{n_1-1}$) and then take the escaping edge (which also happens with probability $\frac{1}{n_1-1}$). Notice that since since the graph is well-connected in both local clusters, Equation 3 tells us that $\nu_2 \approx \Theta\left(\frac{1}{n_1^2}\right)$ and this shows that the left-hand side of Cheeger's inequality is tight (upto constant factors).

5 Easy-Direction

In this section (section 5) we will briefly review Spielman's argument for proving the left-hand side (easy) of Cheeger's inequality.

Proof. The argument involves constructing a characteristic vector for the set S, making it orthogonal to $\vec{\mathbf{d}}$ and comparing its normalized Rayleigh quotient with the second eigenvalue (ν_2), and then bounding that quotient from above by $2 \cdot \phi(G)$ via the use of the Courant-Fischer Theorem.

5.1 Construct the orthogonalized indicator vector

Let the cut of minimal conductance be given by partitioning G into S and $V - S$. Choose the test vector as $y_S = \vec{1}_S - \frac{d(S)}{d(V)}\vec{1}$, where $\vec{1}_S$ is the indicator vector of a vertex i being in the set S.

$$(\vec{1}_S)(i) = \begin{cases} 1 & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases} . \tag{5}$$

It is easy to verify that $y_S \perp \vec{d}$, where $\vec{d}$ is the degree vector of the graph G (and the eigenvector of L corresponding to eigenvalue 0).

5.2 Evaluate the normalized Rayleigh Quotient of $\vec{1}_s$

Claim 7. *For the orthogonalized indicator vector y_S:*

$$y_S^T L y_S = w(\partial S) .$$

Proof. Notice that y_S has (by construction) the same value $1 - \frac{d(S)}{d(V)}$ at every coordinate (equivalently vertex) $i \in S$ and value $-\frac{d((S)}{d(V)}$ at every $j \in V - S$. This implies that $y_S^T L y_s$, which measures the quadratic flow along every edge in G, now only accumulates weights for edges (i, j) where $i \in S$ and $j \in (V - S)$. Consequently:

$$\begin{aligned} y_S^T L y_S &= \sum_{i,j \in E} w_{ij} (y_S(i) - y_S(j))^2 \\ &= \sum_{i,j \in E, i \in S, j \in (V-S)} w_{ij} \left(1 - \frac{d(S)}{d(V)} - \left(-\frac{d(S)}{d(V)}\right)\right)^2 \\ &= \sum_{i,j \in E, i \in S, j \in (V-S)} w_{ij} \\ &= w(\partial S) . \end{aligned} \tag{6}$$

□

The Rayleigh quotient of a vector of the laplacian L normalized by the vector weighted sum of the weights of the degrees corresponds to the Rayleigh quotient of the normalized laplacian N on a related vector. This is clearly seen by scaling the vector y_S (on which the laplacian L acts) in the standard basis by $D^{\frac{1}{2}}$ (setting $x = D^{\frac{1}{2}} y_S$):

$$\begin{aligned} \frac{y_S^T L y_S}{y_S^T D y_S} &= \frac{(D^{-\frac{1}{2}}x)^T L (D^{-\frac{1}{2}}x)}{(D^{-\frac{1}{2}}x)^T (D^{-\frac{1}{2}}x)} \\ &= \frac{x^T D^{-\frac{1}{2}} L D^{-\frac{1}{2}} x}{x^T D^{-\frac{1}{2}} D D^{-\frac{1}{2}} x} \\ &= \frac{x^T N x}{x^T x} . \end{aligned} \tag{7}$$

Using the intuition from the argument above, once can compute that the normalization of the laplacian via the degrees to be:

$$y_S^T D y_S = \left(1 - \frac{d(S)}{d(V)}\right)^2 \sum_{i \in S} d(i) + \left(\frac{d(S)}{d(V)}\right)^2 \sum_{j \in V-S} d(j)\,. \tag{8}$$

Under some algebra, Equation 8 simplifies to be equal to the normalized ratio of the product of the volume of the respected vertex components:

$$y_S^T D y_S = \frac{d(S)d(V-S)}{d(V)}\,. \tag{9}$$

5.3 Upper-Bound via Courant-Fischer Theorem

Theorem 7 and Equation 9 together yield:

Lemma 8 (Equivalence of $\hat{\phi}(S)$ and RQ_N)**.** *Given some partition of the vertices S and $V-S$, the Rayleigh Quotient of N on the orthognalized indicator vector and the normalized conductance of an undirected graph are equal:*

$$\hat{\phi}(S) = \frac{w(\partial S)d(V)}{d(S)d(V-S)} = \frac{y_S^T L y_S}{y_S^T D y_S} = \frac{x^T N x}{x^T x}\,. \tag{10}$$

Here, $x = D^{\frac{1}{2}} y_S$.

The Courant-Fischer Theorem implies that the second eigenvector ν_2 of the normalized laplacian N will be the minimal Rayleigh quotient of N across all $x' \perp \vec{d}^{\frac{1}{2}}$. Note that $x \perp \vec{d}^{\frac{1}{2}}$ by construction, and:

$$\frac{x^T N x}{x^T x} \geq \min_{x' \perp \vec{d}^{\frac{1}{2}}} \frac{x'^T N x'}{x'^T x'} = \nu_2\,. \tag{11}$$

where the inequality follows because x is a test vector in the appropriate subspace. One can assume WLOG that $d(S) \geq d(V-S)$ which immediately implies that $\frac{d(V-S)}{d(V)} \leq \frac{1}{2}$ which yields that $\hat{\phi}(S) \leq 2\phi(S)$. This proves the result. □

6 Hard-Direction

In this section we wish to prove the hard-direction of Cheeger's inequality. Our goal is: Given $\mathbf{y} \perp \mathbf{d}$ s.t. $\frac{\mathbf{y}^T \mathbf{L} \mathbf{y}}{\mathbf{y}^T \mathbf{D} \mathbf{y}} \leq \rho$, the goal is to "round" $\mathbf{y}$ to obtain a set S of small conductance:

$$\phi(G) = \frac{w(\partial S)}{\min\{d(S), d(V-S)\}} \leq \sqrt{2\rho}. \tag{12}$$

Note if $\mathbf{y} = \mathbf{D}^{1/2}\boldsymbol{\phi}_{\mathbf{2}}$, then $\rho = \nu_2$, where $\boldsymbol{\phi}_{\mathbf{2}}$ is the second eigenvector of N.

Proof. The idea behind this "rounding" is that we wish to group things that are close to each other as one set. Recall that the second eigenvector is supposed to put things connected by an edge close to each other (in terms of their weights). Therefore, we first sort the second eigenvector $\mathbf{y}$ so that $\mathbf{y}(1) \leq \mathbf{y}(1) \leq ... \leq \mathbf{y}(n)$, and try a cut τ somewhere between $\mathbf{y}(1)$ and $\mathbf{y}(n)$. The two groups are defined by $L = \{a : \mathbf{y}(a) \leq \tau\}$ and $R = \{a : \mathbf{y}(a) > \tau\}$. We will use probabilistic method for this proof. We will construct a probability distribution that satisfies the inequality.

6.1 Center the orthogonalized vector y

Let $\mathbf{z} = \mathbf{y} - s\mathbf{1}$ for appropriate s, so that the volume contribution by the vertices of the positive and negative entries of z are "centered":

$$\begin{aligned} \sum_{a:z(a)<0} d(a) \leq \frac{d(V)}{2} \\ \sum_{a:z(a)>0} d(a) \leq \frac{d(V)}{2}, \end{aligned} \tag{13}$$

Here, we just shift the vector so that half of its entries are negative while the other half is positive (weighted by the number of edges connected to each entry). It suffices to choose $j \in [n]$ to be the smallest value that satisfies Equation 13 (which also asserts that $\mathbf{z}(j) = 0$). At the same time, we can bound the normalized Rayleigh quotient of $\mathbf{z}$ from above by ρ:

$$\frac{\mathbf{z}^T\mathbf{L}\mathbf{z}}{\mathbf{z}^T\mathbf{D}\mathbf{z}} = \frac{(\mathbf{y} - s\mathbf{1})^T\mathbf{L}(\mathbf{y} - s\mathbf{1})}{(\mathbf{y} - s\mathbf{1})^T\mathbf{D}(\mathbf{y} - s\mathbf{1})} = \frac{\mathbf{y}^T\mathbf{L}\mathbf{y}}{\mathbf{y}^T\mathbf{D}\mathbf{y} + s^2\mathbf{1}^T\mathbf{D}\mathbf{1}} \leq \frac{\mathbf{y}^T\mathbf{L}\mathbf{y}}{\mathbf{y}^T\mathbf{D}\mathbf{y}} \leq \rho. \tag{14}$$

We assume without lost of generality that $\mathbf{z}(1)^2 + \mathbf{z}(n)^2 = 1$.

6.2 Construct distribution with desired threshold

We now construct a desired probability distribution of threshold τ so that

$$\mathbb{E}_\tau[\omega(\partial S_\tau)] \leq \sqrt{2\rho} \cdot \mathbb{E}_\tau[\min\{d(S_\tau), d(V - S_\tau)\}], \tag{15}$$

Since the expectation is over τ, if a predicate holds in expectation it would imply that there exists some τ for which Equation 15 would hold with vertex set S_τ. This is a simple use of the probabilistic method. We will show that the distribution $\mathbf{P}(\mathbf{t}) = \mathbf{2}|\mathbf{t}|$, with $-1 \leq t \leq 1$ meets the requirement. As a sanity check, the probability that τ lies in the interval $[\mathbf{z}(1), \mathbf{z}(n)]$ is

$$\int_{\mathbf{z}(1)}^{\mathbf{z}(n)} 2|t|dt = \int_{\mathbf{z}(1)}^{0} 2|t|dt + \int_{0}^{\mathbf{z}(n)} 2|t|dt = \mathbf{z}(1)^2 + \mathbf{z}(n)^2 = 1. \tag{16}$$

The goal now is to relate the average quantities in the left-hand side and right-hand side of Equation 15 to the normalized Rayleigh quotient of the "centered" orthogonal vector $\mathbf{z}$. Since Equation 14 bounds the normalized Rayleigh quotient on $\mathbf{z}$ from above, it would suffice to bound the expected cut $\mathbb{E}_\tau[\phi(S_\tau]$ from above by $\sqrt{2} \cdot \mathrm{RQ}_N(\mathbf{z})$.
The next two claims formalize the bounds on the aforementioned average quantities.

Claim 9.

$$\mathbb{E}_\tau[\min\{d(S_\tau), d(V - S_\tau)\}] = \mathbf{z}^T\mathbf{D}\mathbf{z}. \tag{17}$$

Proof.

$$\begin{aligned} \text{LHS} &= \sum_a d(a) \cdot \Pr[\text{a is in the "smaller" part of } S_\tau \text{ and } V - S_\tau] \\ &= \sum_a d(a) \cdot \Pr[\tau \text{ is between } \mathbf{z}(a) \text{ and } 0] \\ &= \sum_a d(a) \cdot \mathbf{z}(a)^2 \\ &= \mathbf{z}^T\mathbf{D}\mathbf{z}. \end{aligned} \tag{18}$$

The first line in Equation 18 is by definition. The second line follows because $\mathbf{z}$ has been centered: when $\tau < 0$, $\{a : \mathbf{z}(a) < \tau\}$ is the smaller part; when $\tau > 0$, $\{a : \mathbf{z}(a) > \tau\}$ is the larger part. Note that for the probability distribution we constructed, τ can't be zero. We can calculate $\Pr[\tau$ is between $\mathbf{z}(a)$ and $0]$ directly since we know the probability distribution $\mathbf{P(t)} = \mathbf{2|t|}$. Lastly, the degree of a averaged by $\mathbf{z}(a)^2$ is just the quadratic form under the degree matrix $\mathbf{D}$. □

Claim 10.

$$\mathbb{E}_\tau[w(\partial S_\tau)] \leq \sqrt{2} \cdot \sqrt{\mathbf{z}^T\mathbf{L}\mathbf{z}}\sqrt{\mathbf{z}^T\mathbf{D}\mathbf{z}}. \tag{19}$$

Proof.

$$\begin{aligned} \text{LHS} &= \sum_{(a,b)\in E} w_{ab} \cdot \Pr[\tau \text{ is between } \mathbf{z}(a) \text{ and } \mathbf{z}(b)] \\ &\leq \sum_{(a,b)\in E} w_{ab} \cdot |\mathbf{z}(b) - \mathbf{z}(a)| \cdot (|\mathbf{z}(b)| + |\mathbf{z}(a)|), \end{aligned} \tag{20}$$

The edge is counted in $w(\partial S_\tau)$ when the two end points are separated by the threshold. The inequality can be derived from a visualization of $\mathbf{P(t)}$=$\mathbf{2}|t|$: When $z(a)$ and $z(b)$ are of the same sign, the probability that τ is between z(a) and z(b) is presented by a right trapezoid with an area of $|z(b) - z(a)| \cdot (|z(b)| + |z(a)|)$. When $z(a)$ and $z(b)$ are of opposite sign, this probability is presented by two triangles with a sum of areas less than $|z(b) - z(a)| \cdot (|z(b)| + |z(a)|)$.
Using the Cauchy-Schwartz inequality, we can upper bound the term above by:

$$\sqrt{\sum_{(a,b)\in E} w_{ab}(|\mathbf{z}(a) - \mathbf{z}(b)|)^2}\sqrt{\sum_{(a,b)\in E} w_{ab}(|\mathbf{z}(a)| + |\mathbf{z}(b)|)^2}, \tag{21}$$

For the left-hand square root, we have

$$\sum_{(a,b)\in E} w_{ab}(|\mathbf{z}(a) - \mathbf{z}(b)|)^2 = \mathbf{z}^T\mathbf{L}\mathbf{z} \leq \rho\mathbf{z}^T\mathbf{D}\mathbf{z}, \tag{22}$$

For the right-hand square root, we have

$$\sum_{(a,b)\in E} w_{ab}(|\mathbf{z}(a)| + |\mathbf{z}(b)|)^2 \leq 2 \sum_{(a,b)\in E} w_{ab}(\mathbf{z}(a)^2 + \mathbf{z}(b)^2) = 2\sum_{a} \mathbf{z}(a)^2 d(a) = 2\mathbf{z}^T\mathbf{D}\mathbf{z}, \tag{23}$$

Putting them together, we have

$$\mathbb{E}_\tau[w(\partial S_\tau)] \leq \sqrt{2} \cdot \sqrt{\mathbf{z}^T\mathbf{L}\mathbf{z}}\sqrt{\mathbf{z}^T\mathbf{D}\mathbf{z}}. \tag{24}$$

So Theorem 10 is proved. □

Putting the two claims Theorem 9 and Theorem 10 together, and using Equation 14:

$$\mathbf{z}^T\mathbf{L}\mathbf{z} \leq \rho\mathbf{z}^T\mathbf{D}\mathbf{z} \tag{25}$$

we have

$$\mathbb{E}_\tau[w(\partial S_\tau)] \leq \sqrt{2\rho}\mathbf{z}^T\mathbf{D}\mathbf{z} = \sqrt{2\rho}\ \mathbb{E}_\tau[\min\{d(S_\tau), d(V - S_\tau)\}] \tag{26}$$

6.3 Conclusion

This equation of expectation Equation 26 implies that there is some threshold τ for which

$$w(\partial S_\tau) \leq \sqrt{2\rho}\min\{d(S_\tau), d(V - S_\tau)\} \tag{27}$$

which means

$$\phi(S) = \frac{w(\partial S_\tau)}{\min\{d(S_\tau), d(V - S_\tau)\}} \leq \sqrt{2\rho} \tag{28}$$

since $\rho = \nu_2$ and $\mathbf{y} = D^{\frac{1}{2}}\phi_{\mathbf{2}}$, where $\phi_{\mathbf{2}}$ is the second eigenvector of $\mathbf{N}$. □