COMS 4995-3: Advanced Algorithms — Mar 22, 2017

Lecture 16 – Cheeger's Inequality

Instructor: *Alex Andoni* — Scribes: *Sandip Sinha, Tsung-Yi Huang*

Today's lecture is about proving Cheeger's Inequality. Given a graph $G = \{V, E\}$, we can define the following:

Definition 1. *Given a set of vertices $S \subseteq V$, we define the set of edges connecting S with rest of the graph (boundary of S) to be $\partial S = \{(i,j) \in E | i \in S, j \in \bar{S}\}$ and the volume of S to be $vol(S) = \sum_{i \in S} d_i$, where d_i is the degree of vertex i.*

Definition 2. *Given a set of vertices $\phi \subsetneq S \subsetneq V$, we define the conductance of S to be $\phi(S) := \frac{|\partial S|}{\min(vol(S), vol(\bar{S}))}$, and the conductance of the graph to be $\phi(G) = \min_{\phi \subsetneq S \subsetneq V} \phi(S)$.*

Let $0 = \mu_1 \leq \mu_2 \leq \ldots \leq \mu_n$ be the eigenvalues of the normalized Laplacian matrix $\hat{L}$.

Theorem 3 (Cheeger's Inequality)**.** $\frac{\mu_2}{2} \leq \phi(G) \leq \sqrt{2\mu_2}$.

1 Lower Bound: The Proof of $\frac{\mu_2}{2} \leq \phi(G)$

Fix any set $S \subset V$ and let $s = \frac{vol(S)}{vol(V)}$. We will prove the lower bound by proving that $\phi(S) \geq (1-s)\mu_2$. We already know that

$$\mu_2 = \min_{\substack{x \neq \mathbf{0} \\ x \perp v_1}} \frac{x^T \hat{L} x}{x^T x} = \min_{\substack{x \neq \mathbf{0} \\ x \perp v_1}} \frac{x^T D^{-1/2} L D^{-1/2} x}{||x||_2^2}$$

where v_1 is the eigenvector of the smallest eigenvalue (0) of $\hat{L}$. We have found one in the last lecture which is $(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_n})$ so we can use it as v_1.

Let $y = D^{-1/2}x$ (assume that there is no isolated vertex). By changing x into y, we get

$$\mu_2 = \min_{\substack{y \neq \mathbf{0} \\ y \perp \boldsymbol{d}}} \frac{y^T L y}{y^T D y}$$

where $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$.

Since μ_2 is the smallest value of the Rayleigh quotient for all vectors satisfying the above constraints, we will construct a vector y below that satisfies these constraints and use the fact that μ_2 is no more than y's Rayleigh quotient to prove the bound.

Let $y' = \mathbb{1}_S$ ($y'_i = 1$ if vertex $i \in S$ and 0 otherwise). We have

$$y'^T L y' = \sum_{(i,j) \in E} (y'_i - y'_j)^2 = |\partial S|$$

We then set $y = y' - s\mathbb{1}$. Note that $L\mathbb{1} = 0$, so

$$y^T L y = y'^T L y' = |\partial S|$$

We also have

$$\begin{aligned} y^T D y &= \sum_{i \in V} y_i^2 d_i \\ &= \sum_{i \in S} (1-s)^2 d_i + \sum_{i \notin S} (-s)^2 d_i \\ &= (1-s)^2 vol(S) + s^2 (vol(V) - vol(S)) \\ &= (1-s) vol(S) \end{aligned}$$

Further, y is orthogonal to d because

$$y^T d = \sum_{i \in S} (1-s) d_i + \sum_{i \notin S} (-s) d_i = vol(S) - s \sum_{i \in V} d_i = 0$$

So,

$$\mu_2 \le \frac{|\partial S|}{(1-s) vol(S)} = \frac{|\partial S| \cdot vol(V)}{vol(S) \cdot vol(\bar{S})} \le 2\phi(S).$$

for any set S. The last inequality holds because $\max(vol(S), vol(\bar{S}))/vol(V) \ge 1/2$.

This implies that $\mu_2/2 \le \min_{\phi \ne S \ne V} \phi(S) = \phi(G)$.

2 Upper Bound: The Proof of $\phi(G) \le \sqrt{2\mu_2}$

Let y be the eigenvector corresponding to μ_2. By re-indexing vertices in the graph we can assume $y_1 \le y_2 \le \ldots \le y_n$.

Let $k \in [n]$ be the minimum index such that $\sum_{i=1}^k d_i \ge vol(V)/2$. We define $z := y - y_k \mathbb{1}$. Then we rescale z such that $z_1^2 + z_n^2 = 1$. Note that $z_1 \le z_k = 0 \le z_n$.

We claim that we still have $\frac{z^T L z}{z^T D z} \le \frac{y^T L y}{y^T D y}$. We prove this by calculating $\frac{z^T L z}{z^T D z}$.

$$\begin{aligned} z^T D z &= (y - y_k \mathbb{1})^T D (y - y_k \mathbb{1}) \\ &= y^T D y - 2 y_k \mathbb{1}^T D y + y_k^2 \mathbb{1}^T D \mathbb{1} \\ &\ge y^T D y \end{aligned}$$

The last inequality holds because in the second term $\mathbb{1}^T D y = \boldsymbol{d}^T y = 0$ as $y \perp \boldsymbol{d}$, and the third term is non-negative.

$$z^T L z = (y - y_k \mathbb{1})^T L (y - y_k \mathbb{1}) = y^T L y$$

This is because $L\mathbb{1} = \mathbf{0}$ so the second term and the third term are both 0. $\frac{z^T L z}{z^T D z}$ has a bigger denominator while the numerator is the same, so it is smaller.

For $t \in \mathbb{R}$, define $S_t := \{i \in V | z_i \le t\}$. We are going to prove that there exists a number $t \in \mathbb{R}$ such that $\phi(S_t) \le \sqrt{2\mu_2}$. Then we will get $\phi(G) \le \phi(S_t) \le \sqrt{2\mu_2}$. As the choice of t is not obvious, we randomly pick t from a distribution given by the pdf p, where $p(t) = 2|t|$ if $t \in [z_1, z_n]$, and $p(t) = 0$ otherwise. We verify that p is a pdf, as follows:

$$\int_{z_1}^{z_n} p(t) dt = \int_{z_1}^{0} -2t dt + \int_{0}^{z_n} 2t dt = z_1^2 + z_n^2 = 1$$

So this is a valid distribution. If we can prove that $0 \leq \mathbb{E}[\sqrt{2\mu_2}\min(vol(S), vol(\bar{S})) - |\partial S_t|]$, then we know there must be some t for which the expression inside the expectation is non-negative. This implies that for that t, $\phi(S_t) = \frac{|\partial S_t|}{min(vol(S_t), vol(\bar{S}_t))} \leq \sqrt{2\mu_2}$, and hence completes the proof of the upper bound.

Lemma 4. $\mathbb{E}_t[|\partial S_t|] \leq \sqrt{2\mu_2}\ \mathbb{E}_t[\min\{vol(S_t), vol(\bar{S}_t\}]$.

Before proving the lemma, we show that Lemma 4 implies Theorem 3.

$$\begin{aligned} 0 &\leq \sqrt{2\mu_2} \cdot \mathbb{E}_t[\min\{vol(S_t), vol(\bar{S}_t)\}] - \mathbb{E}_t[|\partial S_t|] \\ &= \mathbb{E}_t[\sqrt{2\mu_2} \cdot \min\{vol(S_t), vol(\bar{S}_t)\} - |\partial S_t|] \end{aligned}$$

Now, by the probabilistic method, there exists $t \in [z_1, z_n]$ such that

$$\begin{aligned} 0 &\leq \sqrt{2\mu_2} \cdot \min\{vol(S_t), vol(\bar{S}_t)\} - |\partial S_t| \\ \text{So, } \frac{|\partial S_t|}{\min\{vol(S_t), vol(\bar{S}_t)\}} &\leq \sqrt{2\mu_2} \end{aligned}$$

Now, we prove the lemma. To do that, we prove two claims that separately bound the left and right sides of the inequality.

Claim 5. $\mathbb{E}_t[|\partial S_t|] \leq \left(z^T L z\right)^{1/2} \cdot \sqrt{2 z^T D z} \leq \sqrt{2\mu_2} \cdot z^T D z.$

Proof. The second inequality is immediate since

$$\left(z^T L z\right)^{1/2} \cdot \sqrt{2 z^T D z} = \left(\frac{z^T L z}{z^T D z}\right)^{1/2} \cdot \sqrt{2} z^T D z \leq \sqrt{2\mu_2} \cdot z^T D z$$

Now, we prove the first inequality. Define the sign function

$$sgn(x) := \begin{cases} +1, & x > 0 \\ -1, & x < 0 \\ 0, & x = 0 \end{cases}$$

Fix an edge $(i, j) \in E$. We assume $z_i \leq z_j$ w.l.o.g, as we can interchange i and j if this does not hold. Then it can be shown, by considering various cases depending on the signs of z_i and z_j, that $Pr[z_i \leq t \leq z_j] = |sgn(z_j) z_j^2 - sgn(z_i) z_i^2|$. So, we have

$$\begin{aligned} Pr[(i,j) \in Cut] &= Pr[z_i \leq t \leq z_j] \\ &= |sgn(z_j) z_j^2 - sgn(z_i) z_i^2| \\ &= \begin{cases} |z_i^2 - z_j^2|, & sgn(z_i) = sgn(z_j) \\ z_i^2 + z_j^2, & sgn(z_i) \neq sgn(z_j) \end{cases} \\ &\leq \begin{cases} |z_i - z_j|(|z_i| + |z_j|), & sgn(z_i) = sgn(z_j) \\ (z_i - z_j)^2, & sgn(z_i) \neq sgn(z_j) \end{cases} \\ &\leq |z_i - z_j|(|z_i| + |z_j|) \end{aligned}$$

We have used triangle inequality twice. To get the last inequality when $sgn(z_i) \neq sgn(z_j)$, we upper bound one of the terms $(z_i - z_j)$ by $(|z_i| + |z_j|)$. Now, writing the random variable $|\partial S_t|$ as a sum over indicator random variables denoting whether each edge is in the cut, we observe that

$$\begin{aligned}
\mathbb{E}_t[|\partial S_t|] &= \sum_{(i,j)\in E} Pr[(i,j) \in Cut] \\
&\leq \sum_{(i,j)\in E} |z_i - z_j|(|z_i| + |z_j|) \\
&\leq \sqrt{\sum_{(i,j)\in E} (z_i - z_j)^2} \cdot \sqrt{\sum_{(i,j)\in E} (|z_i| + |z_j|)^2} && \text{(Cauchy-Schwarz)} \\
&\leq (z^T L z)^{1/2} \cdot \left(2 \sum_{(i,j)\in E} (z_i^2 + z_j^2) \right)^{1/2} && ((a+b)^2 \leq 2(a^2+b^2)) \\
&= (z^T L z)^{1/2} \cdot \left(2 \sum_{i \in V} z_i^2 d_i \right)^{1/2} \\
&= (z^T L z)^{1/2} \cdot \sqrt{2 z^T D z}
\end{aligned}$$

There are two important parts in the above derivation that deserve further explanation. First, to see the application of Cauchy-Schwarz inequality, let $m = |E|$ be the number of edges in the graph. Fix some ordering of the edges. Define $a, b \in \mathbb{R}^m$ by $a_k = |z_i - z_j|$ and $b_k = |z_i| + |z_j|$ for $k \in [m]$, where $e_k = (i,j)$ is the k^{th} edge in the order. Then Cauchy-Schwarz inequality gives

$$\sum_{(i,j)\in E} |z_i - z_j|(|z_i| + |z_j|) = |a \cdot b| \leq \|a\|_2 \cdot \|b\|_2 = \sqrt{\sum_{(i,j)\in E} (z_i - z_j)^2} \cdot \sqrt{\sum_{(i,j)\in E} (|z_i| + |z_j|)^2}.$$

The second last equality is derived by changing the sum over edges to a sum over vertices. For each edge (i,j), we assumed that $z_i \leq z_j$. This implies that we consider each edge only once in the sum. Now, fix a vertex $i \in V$. For each edge incident on i, the term z_i^2 appears exactly once. Hence, the total number of times that this term appears is the number of edges that are incident on it, or its degree. □

Claim 6. $\mathbb{E}_t[\min\{vol(S_t), vol(\bar{S}_t)\}] = z^T D z$.

Proof. This claim will be proved in the next lecture. □

Now, combining Claims 5 and 6, we can prove the lemma as follows:

$$\mathbb{E}_t[|\partial S_t|] \leq \sqrt{2\mu_2} \cdot z^T D z = \sqrt{2\mu_2} \cdot \mathbb{E}_t[\min\{vol(S_t), vol(\bar{S}_t)\}].$$

This proves Theorem 3. However, just it shows the existence of $t \in \mathbb{R}$ such that the set S_t has low conductance. We now give an efficient algorithm to compute such a set. Observe that for fixed $i \in [n]$, $S_{z_i} = S_\alpha$ for $z_i \leq \alpha < z_{i+1}$, where we define $z_{n+1} = \infty$ for convenience. So, although there are infinitely many values of t, there are effectively only n values of t that we need to check, namely $z_1, \cdots, z_n$. In other words, we know that there exists $t \in \{z_1, \cdots, z_n\}$ such that $\phi(S_t) \leq \sqrt{2\mu_2}$.

Spectral Partitioning Algorithm:

- Compute $\phi(S_t)$ for $t = z_1, \cdots, z_n$.
- Return the set S_t with minimum value of $\phi(S_t)$.

Clearly, this algorithm can be implemented in time $O(nm)$, where $n = |V|, m = |E|$. By the discussion preceding the algorithm, the set $\hat{S}$ returned by the algorithm satisfies $\phi(\hat{S}) \leq \sqrt{2\mu_2}$.