Cheeger's Inequality

Lecturer: Sushant Sachdeva   *Scribe: Muhammed Tahsin Rahman*

1 Recall

The conductance of a graph $G = (V, E)$ is defined as

$$\Phi(G) = \min_{\substack{S \subseteq V \\ S \neq \varnothing}} \frac{\left|E(S, \overline{S})\right|}{\min\left\{\mathsf{vol}(S), \mathsf{vol}(\overline{S})\right\}}, \tag{1}$$

where $\overline{S} = S^c$, $\left|E(S, \overline{S})\right| = \sum_{(u,v)} w_{u,v} \mathbb{1}[(u, v) \in E(S, \overline{S})]$, and $\mathsf{vol}(S) = \sum_{v \in S} \deg(v)$. Cheeger's inequality states that the conductance is bounded by

$$\frac{\nu_2}{2} \leq \Phi(G) \leq \sqrt{2\nu_2}. \tag{2}$$

ν_2 is the second smallest eigenvalue of the normalized graph Laplacian, defined as

$$\nu_2 = \min_{\mathbf{y}^\top \mathbf{D}\mathbb{1} = 0} \frac{\mathbf{y}^\top \mathbf{L} \mathbf{y}}{\mathbf{y}^\top \mathbf{D} \mathbf{y}} = \lambda_2(\mathbf{N}), \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^V$, $\mathbf{D}$ is the degree matrix, $\mathbf{L}$ is the graph Laplacian, and $\mathbf{N} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ is the normalized graph Laplacian.

In the previous lecture we had proved the left side of eq. 2, and in this lecture we will prove the right side. To this end, we first state three lemmas; then, we use these lemmas to derive the bound; finally, we prove the lemmas. We will also begin the next topic on Random Walks.

2 Required Lemmas

Lemma 2.1. *Given a vector $\mathbf{y} \in \mathbb{R}^V$ s.t. $\mathbf{y}^\top \mathbf{D}\mathbb{1} = 0$, we can find a vector $\mathbf{z} \in \mathbb{R}^V$ s.t.*

$$\mathbf{z} \geq \mathbf{0}$$
$$\frac{\mathbf{z}^\top \mathbf{L} \mathbf{z}}{\mathbf{z}^\top \mathbf{D} \mathbf{z}} \leq \frac{\mathbf{y}^\top \mathbf{L} \mathbf{y}}{\mathbf{y}^\top \mathbf{D} \mathbf{y}}, \ and$$
$$\mathsf{vol}(\mathsf{supp}(\mathbf{z})) \leq \frac{1}{2}\mathsf{vol}(V),$$

where $\mathsf{supp}(\mathbf{z}) = \{v : \mathbf{z}(v) > 0\}$.

Lemma 2.2. *Given a vector* $\mathbf{z} \in \mathbb{R}^V$ *s.t.* $\mathbf{z} \geq \mathbf{0}$, *we can sample a scalar* t, *where* $S_t \subseteq V$, *s.t.* $S_t \subseteq \text{supp}(\mathbf{z})$. *Then,*

$$\frac{\mathbb{E}_t\big[\big|E(S_t, \overline{S_t})\big|\big]}{\mathbb{E}_t[\text{vol}(S_t)]} \leq \sqrt{2\frac{\mathbf{z}^\top \mathbf{L}\mathbf{z}}{\mathbf{z}^\top \mathbf{D}\mathbf{z}}}.$$

Lemma 2.3. *Given a distribution over* t *and associated random variables* X_t, Y_t *s.t.* $Y_t > 0$, $\exists\, t_0$ *s.t.*

$$\frac{X_{t_0}}{Y_{t_0}} \leq \frac{\mathbb{E}_t[X_t]}{\mathbb{E}_t[Y_t]}.$$

3   Cheeger's Upper Bound

Let $\mathbf{y} \in \mathbb{R}^V$ s.t. $\mathbf{y}^\top \mathbf{D}\mathbb{1} = 0$ and apply Lemma 2.1. We obtain $\mathbf{z} \geq \mathbf{0}$, with $\text{vol}(\mathbf{z}) \leq \frac{1}{2}\text{vol}(V)$ and $\frac{\mathbf{z}^\top \mathbf{L}\mathbf{z}}{\mathbf{z}^\top \mathbf{D}\mathbf{z}} \leq \frac{\mathbf{y}^\top \mathbf{L}\mathbf{y}}{\mathbf{y}^\top \mathbf{D}\mathbf{y}}$. Now, use this $\mathbf{z}$ in Lemma 2.2, sampling t with associated set $S_t \subseteq \text{supp}(\mathbf{z})$, to obtain $\text{vol}(S_t) \leq \frac{1}{2}\text{vol}(V)$ and $\frac{\mathbb{E}_t\big[\big|E(S_t, \overline{S_t})\big|\big]}{\mathbb{E}_t[\text{vol}(S_t)]} \leq \sqrt{2\frac{\mathbf{z}^\top \mathbf{L}\mathbf{z}}{\mathbf{z}^\top \mathbf{D}\mathbf{z}}} \leq \sqrt{2\frac{\mathbf{y}^\top \mathbf{L}\mathbf{y}}{\mathbf{y}^\top \mathbf{D}\mathbf{y}}}$. Finally, apply Lemma 2.3, in that $\exists\, t_0$ that achieves $\Phi_G(S_{t_0}) = \frac{|E(S_{t_0}, \overline{S_{t_0}})|}{\text{vol}(S_{t_0})} \leq \sqrt{2\frac{\mathbf{y}^\top \mathbf{L}\mathbf{y}}{\mathbf{y}^\top \mathbf{D}\mathbf{y}}}$. To complete the proof plug in the $\mathbf{y}$ that achieves ν_2.

4   Proof of Lemmas

4.1   Lemma 2.1

With $\mathbf{y}^\top \mathbf{D}\mathbb{1} = 0$, let

$$\tilde{\mathbf{y}} = \mathbf{y} - c\mathbb{1},$$
$$\mathcal{V}_1 = \text{vol}(\{v : \tilde{\mathbf{y}}(v) > 0\}),$$
$$\mathcal{V}_2 = \text{vol}(\{v : \tilde{\mathbf{y}}(v) < 0\}).$$

Note that $\mathcal{V}_1$ and $\mathcal{V}_2$ are strictly disjoint. As we change c over the real line, the vertices move from $\mathcal{V}_1$ to $\mathcal{V}_2$, so there will be some c s.t.

$$\mathcal{V}_1, \mathcal{V}_2 \leq \frac{1}{2}\text{vol}(V).$$

Consider what happens to the ratio $\frac{\mathbf{y}^\top \mathbf{L}\mathbf{y}}{\mathbf{y}^\top \mathbf{D}\mathbf{y}}$.

$$\tilde{\mathbf{y}}^\top \mathbf{L}\tilde{\mathbf{y}} = \sum_{(u,v)} (\tilde{\mathbf{y}}(u) - \tilde{\mathbf{y}}(v))^2 = \mathbf{y}^\top \mathbf{L}\mathbf{y},$$
$$\tilde{\mathbf{y}}^\top \mathbf{D}\tilde{\mathbf{y}} = \mathbf{y}^\top \mathbf{D}\mathbf{y} - 2c\underbrace{\mathbf{y}^\top \mathbf{D}\mathbb{1}}_{=\,0} + \underbrace{c^2 \mathbb{1}^\top \mathbf{D}\mathbb{1}}_{\geq\,0} \geq \mathbf{y}^\top \mathbf{D}\mathbf{y}, \text{ and}$$
$$\frac{\tilde{\mathbf{y}}^\top \mathbf{L}\tilde{\mathbf{y}}}{\tilde{\mathbf{y}}^\top \mathbf{D}\tilde{\mathbf{y}}} \leq \frac{\mathbf{y}^\top \mathbf{L}\mathbf{y}}{\mathbf{y}^\top \mathbf{D}\mathbf{y}}.$$

Now write $\tilde{\mathbf{y}} = \mathbf{z}^+ - \mathbf{z}^-$, where $\mathbf{z}^+, \mathbf{z}^- \geq \mathbf{0}$; then,

$$\text{vol}(\text{supp}(\mathbf{z}^+)), \ \text{vol}(\text{supp}(\mathbf{z}^-)) \leq \frac{1}{2}\text{vol}(V),$$

$$\tilde{\mathbf{y}}^\top \mathbf{D}\tilde{\mathbf{y}} = \sum_v \deg(v)\tilde{\mathbf{y}}(v)^2$$
$$= \mathbf{z}^{+\top}\mathbf{D}\mathbf{z}^+ + \mathbf{z}^{-\top}\mathbf{D}\mathbf{z}^-.$$

Note that

$$\tilde{\mathbf{y}}^\top \mathbf{L}\tilde{\mathbf{y}} = \sum_{(u,v)} (\tilde{\mathbf{y}}(u) - \tilde{\mathbf{y}}(v))^2$$
$$= \sum_{(u,v)} ((\mathbf{z}^+(u) - \mathbf{z}^+(v)) - (\mathbf{z}^-(u) - \mathbf{z}^-(v)))^2$$
$$\geq \sum_{(u,v)} (\mathbf{z}^+(u) - \mathbf{z}^+(v))^2 + (\mathbf{z}^-(u) - \mathbf{z}^-(v))^2,$$

since the cross terms of the expansion will always be ≥ 0. Now, with a proof similar to Lemma 2.3 shown in section 4.3,

$$\frac{\mathbf{z}^{+\top}\mathbf{L}\mathbf{z}^+ + \mathbf{z}^{-\top}\mathbf{L}\mathbf{z}^-}{\mathbf{z}^{+\top}\mathbf{D}\mathbf{z}^+ + \mathbf{z}^{-\top}\mathbf{D}\mathbf{z}^-} \leq \frac{\mathbf{y}^\top\mathbf{L}\mathbf{y}}{\mathbf{y}^\top\mathbf{D}\mathbf{y}}.$$

One of $\mathbf{z}^+, \mathbf{z}^-$ achieves a ratio $\leq \frac{\mathbf{y}^\top\mathbf{L}\mathbf{y}}{\mathbf{y}^\top\mathbf{D}\mathbf{y}}$.

4.2   Lemma 2.2

With $\mathbf{z} \geq \mathbf{0}$, scale $\mathbf{z}$ s.t. $\max(\mathbf{z}) = 1$. This scaling is acceptable because the claim depends on ratios with $\mathbf{z}$ and not the absolute values. Now, sample a threshold $t \in [0, 1]$ using the probability density function $f(t) = 2t$. Note that $\mathbf{Pr}[t \in [a, b]] = \int_a^b f(t)dt = b^2 - a^2$. This distribution is picked as a proof trick and does not affect the outcome of Cheeger's inequality.

Figure 1: Sampling t and the corresponding S_t

Define $S_t = \{v : \mathbf{z}(v) > t\}$. So, picking S_t depends on t discretely. If t is between a & b, then $\mathbf{Pr}[v \in S_t] = b^2 - a^2$. This selection of S_t is shown in Figure 1. Next, the denominator of Lemma 2.2

$$\mathbb{E}_t[\mathsf{vol}(S_t)] = \mathbb{E}_t\left[\sum_v \deg(v)\mathbb{1}[v \in S_t]\right],$$
$$= \sum_v \deg(v) \underbrace{\mathbb{E}_t[\mathbb{1}[v \in S_t]]}_{\substack{\mathbf{Pr}[v \in S_t] = \\ \mathbf{Pr}[t \in [0, \mathbf{z}(v)]] = \\ \mathbf{z}(v)^2}},$$
$$= \sum_v \deg(v)\mathbf{z}(v)^2$$
$$= \mathbf{z}^\top\mathbf{D}\mathbf{z}.$$

For the numerator,

$$\mathbb{E}_t\left[\left|E(S, \overline{S})\right|\right] = \mathbb{E}_t\left[\sum_{(u,v)} \mathbb{1}[(u, v) \in E(S_t, \overline{S_t})]\right]$$

3

$$= \sum_{(u,v)} \mathbf{Pr}[(u,v) \in E(S_t, \overline{S_t})]$$

$$= \sum_{(u,v)} \left| \mathbf{z}(u)^2 - \mathbf{z}(v)^2 \right|.$$

Applying Cauchy-Shwartz inequality $\left(\sum_i a_i b_i \leq \sqrt{\sum_i a_i^2} \sqrt{\sum_i b_i^2} \right)$ gives

$$\sum_{(u,v)} \left| \mathbf{z}(u)^2 - \mathbf{z}(v)^2 \right| \leq \sqrt{ (\underbrace{\sum_{(u,v)} (\mathbf{z}(u) - \mathbf{z}(v))^2}_{\text{Laplacian quadratic form}})(\sum_{(u,v)} (\mathbf{z}(u) + \mathbf{z}(v))^2) },$$

$$\sum_{(u,v)} (\mathbf{z}(u) + \mathbf{z}(v))^2 \leq 2 \sum_{(u,v)} (\mathbf{z}(u)^2 + \mathbf{z}(v)^2)$$

$$= 2 \sum_v \deg(v) \mathbf{z}(v)^2$$

$$= 2 \mathbf{z}^\top \mathbf{D} \mathbf{z}.$$

Therefore,

$$\frac{\mathbb{E}_t[|E(S, \overline{S})|]}{\mathbb{E}_t[\mathsf{vol}(S_t)]} \leq \frac{\sqrt{2(\mathbf{z}^\top \mathbf{L} \mathbf{z})(\mathbf{z}^\top \mathbf{D} \mathbf{z})}}{\mathbf{z}^\top \mathbf{D} \mathbf{z}} = \sqrt{2 \frac{\mathbf{z}^\top \mathbf{L} \mathbf{z}}{\mathbf{z}^\top \mathbf{D} \mathbf{z}}}$$

4.3   Lemma 2.3

Assuming the collection of X_t, Y_t is finite,

$$r = \min_t \frac{X_t}{Y_t} \qquad \underline{\text{s.t.}} Y_t > 0,$$

$$= \frac{X_0}{Y_0}.$$

The fact that r is a minimum means that

$$\frac{X_t}{Y_t} \geq r,$$

$$X_t \geq r Y_t,$$

$$\mathbb{E}_t[X_t] \geq r \mathbb{E}_t[Y_t]$$

5   Examples

Given a vector $\mathbf{y}$, how do we get a 'good' cut? Consider two examples that we have seen previously, a single edge between two vertices and a ring graph.

5.1 Two Vertices Connected by an Edge

We know that $\Phi(G) = 1$. Using Cheeger's inequality, we first compute the second smallest eigenvalue of the normalized graph Laplacian. Drawing from the first lecture,

$$
\begin{aligned}
\nu_2 &= \lambda_2(\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}) \\
&= \lambda_2(\mathbf{L}) \\
&= 2 \\
&= 2\Phi(G),
\end{aligned}
$$

showing that the lower bound of Cheeger's inequality is tight

5.2 Ring Graph

Lemma 5.1. *For R_n, define*

$$
\begin{aligned}
x_k(v) &= \sin(\frac{2\pi k v}{n}), \quad 0 \le k \le \frac{n}{2}, \quad \text{and} \\
y_k(v) &= \cos(\frac{2\pi k v}{n}), \quad 0 \le k \le \frac{n}{2}.
\end{aligned}
$$

$\forall k$, x_k, y_k are eigenvectors of L_{R_n}, with eigenvalues $2(1 - \cos(\frac{2\pi k}{n}))$.

Proof of the above lemma is left as an exercise.
For the simple case of n even,

$$
\Phi(R_n) = \frac{2}{2 \cdot \frac{n}{2}} = \frac{2}{n}
$$

Knowing that $\mathbf{D} = 2\mathbf{I}$,

$$
\begin{aligned}
\nu_2 &= \lambda_2(\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}) \\
&= \frac{1}{2}\lambda_2(\mathbf{L}) \\
&= \frac{1}{2}(2 - 2\cos(\frac{2\pi}{n})) \\
&= 2\sin^2(\frac{\pi}{n}) \\
&= 2(\frac{\pi}{n})^2(1 + o(1)), \quad \text{for large } n.
\end{aligned}
$$

Note that for a tight bound, we would need to take the square root of the above expression, as proven earlier.

As an exercise, try finding graph cuts using Julia. Use the previous lecture's formulation and then use Cheeger's inequality to make cuts.

6 Random Walks

We will now be looking at undirected graphs that capture reversible Markov chains. Consider the graph in Figure 2. Start with a peg on vertex a at time 0. Then, at time 1, pick uniformly between neighbours to move to; in this case, $\mathbf{Pr}[v = b \text{ at } t = 1 \mid v = a \text{ at } t = 0] = \mathbf{Pr}[v = c \text{ at } t = 1 \mid v = a \text{ at } t = 0] = 0.5$. If the graph is weighted, scale the transition probabilities using the weights. The resulting sequence of vertices that have been visited over time is called the transcript of the random walk.

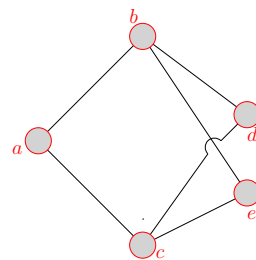Given a graph and initial vertex, we are interested in answering questions such as,

Figure 2: Example of random walk graph

- What is the distribution over vertices after a given number of steps?

- Is there a stationary distribution?

- How quickly do we converge?

The state evolution can be quantified using knowledge of the graph structure. At time t, define $\mathbf{p}_t \in \mathbb{R}^V, \mathbf{p}_t \geq \mathbf{0}$, and $\mathbb{1}^\top \mathbf{p}_t = \sum_\mathbf{v} \mathbf{p}_t(v) = 1$; let

$$\mathbf{p}_0(v) = \begin{cases} 1 & \text{if } v = a \\ 0 & \text{else} \end{cases} \quad \text{(starting at a)},$$

$$\mathbf{p}_t(v) = \mathbf{Pr}[\text{the random walk is at vertex } v \text{ at time } t],$$

$$\mathbf{p}_{t+1}(v) = \sum_{u:(u,v)} \frac{1}{\deg(u)} \mathbf{p}_t(u) \quad \text{(unweighted)}.$$

For weighted graphs, deriving the following is left as an exercise:

$$\mathbf{Pr}[v \text{ at } t+1 \mid u \text{ at } t] = \frac{w(u,v)}{\sum_{z:(u,z)} w(u,z)} = \frac{w(u,v)}{\underbrace{\deg(u)}_{\text{weighted deg}}},$$

$$\mathbf{p}_{t+1}(v) = \sum_{u:(u,v)} \frac{w(u,v)}{\deg(u)} \mathbf{p}_t(u),$$

$$\mathbf{p}_{t+1} = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}_t.$$