[CS1961: Lecture 16] Cheeger's Inequality, Kirchhoff's Theorem

Instructor: Chihao Zhang;
Scribed by Yuchen He

1 Cheeger's Inequality

In the last lecture, we introduced the notion of expansion, which measures how well the graph is connected or how fast the Markov chain converges.

Let P be the transition matrix of a reversible Markov chain whose eigenvalues are $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Then P equates to an undirected weighted graph $G = (V, E)$. Consider the Laplacian matrix $L = I - P$ with eigenvalues $\gamma_1 \leqslant \gamma_2 \leqslant \cdots \leqslant \gamma_n$. We have shown that $\gamma_1 = 0$ and $\gamma_i = 1 - \lambda_i$ for all $i \in [n]$. Then the Cheeger's inequality can be written in terms of γ_2.

Theorem 1 (Cheeger's Inequality) $\frac{\gamma_2}{2} \leqslant \Phi(P) \leqslant \sqrt{2\gamma_2}$.

We prove $\frac{\gamma_2}{2} \leqslant \Phi(P)$ (①) and $\Phi(P) \leqslant \sqrt{2\gamma_2}$ (②) respectively.

Proof of ①. We relate $\Phi(P)$ with γ_2 using the variational characterization. Note that

$$\gamma_2 = \min_{\substack{U \subseteq \mathbb{R}^n \\ \dim(U)=2}} \max_{\mathbf{x} \in U \setminus \{0\}} \frac{\langle \mathbf{x}, L\mathbf{x} \rangle_\Pi}{\langle \mathbf{x}, \mathbf{x} \rangle_\Pi}.$$

Let S be the subset of V such that $\Phi(P) = \max\left\{\Phi(S), \Phi(\overline{S})\right\}$. Let $U = \mathrm{span}\left(\mathbf{1}_S, \mathbf{1}_{\overline{S}}\right)$. For any $\mathbf{x} \in U$, we can write $\mathbf{x}$ as $a\mathbf{1}_S + b\mathbf{1}_{\overline{S}}$ for some constants a and b. Then

$$\begin{aligned}
\frac{\langle \mathbf{x}, L\mathbf{x} \rangle_\Pi}{\langle \mathbf{x}, \mathbf{x} \rangle_\Pi} &= \frac{\sum_{\{i,j\} \in E} \pi(i)P(i,j)(x_i - x_j)^2}{\sum_i \pi(i)x_i^2} = \frac{\sum_{i \in S, j \in \overline{S}} \pi(i)P(i,j)(a-b)^2}{\pi(S)a^2 + \pi(\overline{S})b^2} \\
&\leqslant \frac{2\sum_{i \in S, j \in \overline{S}} \pi(i)P(i,j)(a^2+b^2)}{\pi(S)a^2 + \pi(\overline{S})b^2} \\
&\leqslant 2\max\left\{\frac{\sum_{i \in S, j \in \overline{S}} \pi(i)P(i,j)}{\pi(S)}, \frac{\sum_{i \in S, j \in \overline{S}} \pi(i)P(i,j)}{\pi(\overline{S})}\right\} = 2\Phi(P)
\end{aligned}$$

where the second inequality follows from the fact that for positive real numbers z_1, z_2, y_1, y_2, $\frac{z_1+z_2}{y_1+y_2} \leqslant \max\left\{\frac{z_1}{y_1}, \frac{z_2}{y_2}\right\}$. □

By definition, $\Phi(P) = \min_{\substack{S \subseteq V \\ \pi(S) \leqslant \frac{1}{2}}} \Phi(S)$. To prove ②, we only need to find a $S \subseteq V$ such that $\Phi(S) \leqslant \sqrt{2\gamma_2}$. Such S can be generated using the Fiedler's algorithm. With input $\mathbf{x} \in \mathbb{R}^V$:

- sort V according to $\mathbf{x}$, get $V = \{v_1, \ldots, v_n\}$ where $\mathbf{x}(v_1) \leqslant \mathbf{x}(v_2) \leqslant \mathbf{x}(v_n)$;
- for each $i \in [n]$, let $S_i = \{v_1, \ldots, v_i\}$;

- return the S_i with the minimum $\Phi(S_i) \vee \Phi(\overline{S}_i)$.[1]

[1] $a \vee b$ means $\max\{a, b\}$.

Theorem 2 *For any* $\mathbf{x} \perp \mathbf{1}$, *assume the Fiedler's algorithm returns* S *with input* $\mathbf{x}$. *Then* $\Phi(S) \leqslant \sqrt{2R_L(\mathbf{x})}$.

With Theorem 2, the proof of ② is straightforward. Note that $\mathbf{v}_2$, the eigenvector of L corresponding to eigenvalue γ_2, is the minimizer of $R_L(\mathbf{x})$ on the constraint that $\mathbf{x} \perp \mathbf{1}$. We can divide the graph into different blocks where each block is well connected inside. Intuitively, to get smaller $\sum_{\{i,j\}\in E} \pi(i)P(i,j)(x_i - x_j)^2$, we tend to assign the same value to the x_i's in the same block. The Fiedler's algorithm will return a partition that divides the blocks into two groups. This indicates that $\mathbf{v}_2$ contains the information to find the bottleneck of the graph.

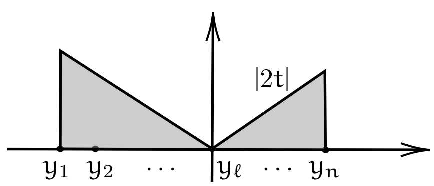
Proof of ②. Run the Fiedler's algorithm with input $\mathbf{x} = \mathbf{v}_2$ and get output S. By Theorem 2, $\Phi(S) \leqslant \sqrt{2R_L(\mathbf{v}_2)} = \sqrt{2\gamma_2}$.

□

It remains to prove Theorem 2.

Proof of Theorem 2 Input $\mathbf{x}$ and run the Fiedler's algorithm. W.l.o.g., assume $x_1 \leqslant \cdots \leqslant x_n$.

Define ℓ be the minimum k such that $\sum_{i=1}^{k} \pi_i \geqslant \frac{1}{2}$. Let $\mathbf{y} = (y_1, \ldots, y_n) = \mathbf{x} - x_\ell \cdot \mathbf{1}$. That is, $y_i = x_i - x_\ell$ for all $i \in [n]$. Rescale $\mathbf{y}$ such that $y_1^2 + y_n^2 = 1$. We randomly pick $t \in [y_1, y_n]$ with density $2|t|$ and set



$S_t = \{i \in [n] \mid y_i \leqslant t\}$. Then

$$\max\left\{\Phi(S_t), \Phi(\overline{S}_t)\right\} = \frac{\sum_{i\in S_t, j\in \overline{S}_t} \pi(i)P(i,j)}{\min\left\{\pi(S_t), \pi(\overline{S}_t)\right\}}.$$

Let $A \triangleq \sum_{i\in S_t, j\in\overline{S}_t} \pi(i)P(i,j)$ and $B \triangleq \min\left\{\pi(S_t), \pi(\overline{S}_t)\right\}$. We claim that $\frac{\mathbf{E}[A]}{\mathbf{E}[B]} \leqslant \sqrt{2R_L(\mathbf{x})}$.

By definition,

$$\mathbf{E}[A] = \sum_{\substack{\{i,j\}\in E \\ i<j}} \pi(i)P(i,j)\mathbf{Pr}\left[i \in S_t, j \in \overline{S}_t\right]. \tag{1}$$

Note that $\mathbf{Pr}\left[i \in S_t, j \in \overline{S}_t\right]$ is the probability that $t \in [y_i, y_j]$, which can be calculated directly by integration:

$$\begin{aligned}
\text{Equation (1)} &= \sum_{\substack{\{i,j\}\in E\\ i<j}} \pi(i)P(i,j)\int_{y_i}^{y_j} 2|t|\,dt \\
&= \sum_{\substack{\{i,j\}\in E\\ i<j}} \pi(i)P(i,j)\left(\operatorname{sgn}(y_j)y_j^2 - \operatorname{sgn}(y_i)y_i^2\right) \\
&\leqslant \sum_{\substack{\{i,j\}\in E\\ i<j}} \pi(i)P(i,j)\left(|y_i|+|y_j|\right)(y_j-y_i) \\
&= \sum_{\substack{\{i,j\}\in E\\ i<j}} \left(\pi(i)P(i,j)\right)^{\frac{1}{2}}\left(|y_i|+|y_j|\right)\cdot\left(\pi(i)P(i,j)\right)^{\frac{1}{2}}(y_j-y_i)\,.
\end{aligned} \tag{2}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\text{Equation (2)} &\leqslant \sqrt{\sum_{\substack{\{i,j\}\in E\\ i<j}} \pi(i)P(i,j)\left(|y_i|+|y_j|\right)^2}\cdot\sqrt{\sum_{\substack{\{i,j\}\in E\\ i<j}} \pi(i)P(i,j)\left(y_j-y_i\right)^2} \\
&\leqslant \sqrt{2\sum_{\substack{\{i,j\}\in E\\ i<j}} \pi(i)P(i,j)\left(y_i^2+y_j^2\right)}\cdot\sqrt{\langle \mathbf{y}, L\mathbf{y}\rangle_\Pi} \\
&= \sqrt{2\langle \mathbf{y},\mathbf{y}\rangle_\Pi}\cdot\sqrt{\langle \mathbf{y}, L\mathbf{y}\rangle_\Pi}
\end{aligned}$$

Recall that ℓ is a middle line of π. Therefore, if $t<0$, $\pi(S_t)\leqslant\pi(\overline{S}_t)$ and otherwise $\pi(S_t)>\pi(\overline{S}_t)$. Then we have

$$\mathbf{E}[B] = \underbrace{\mathbf{Pr}[t<0]\,\mathbf{E}[\pi(S_t)\mid t<0]}_{(3)} + \underbrace{\mathbf{Pr}[t\geqslant 0]\,\mathbf{E}\left[\pi(\overline{S}_t)\;\middle|\; t\geqslant 0\right]}_{(4)}.$$

Note that

$$\begin{aligned}
(3) &= \mathbf{Pr}[t<0]\sum_{i=1}^{\ell-1}\pi(i)\mathbf{Pr}[i\in S_t\mid t<0] \\
&= \sum_{i=1}^{\ell-1}\pi(i)\int_{y_i}^{0}2|t|\,dt = \sum_{i=1}^{\ell-1}\pi(i)y_i^2.
\end{aligned}$$

Similarly, $(4)=\sum_{i=\ell}^{n}\pi(i)y_i^2$. Summing up the two terms, we have $\mathbf{E}[B] = \sum_{i=1}^{n}\pi(i)y_i^2 = \langle \mathbf{y},\mathbf{y}\rangle_\Pi$. Therefore,

$$\frac{\mathbf{E}[A]}{\mathbf{E}[B]} \leqslant \sqrt{\frac{2\langle \mathbf{y}, L\mathbf{y}\rangle_\Pi}{\langle \mathbf{y},\mathbf{y}\rangle_\Pi}} = \sqrt{2R_L(\mathbf{y})}.$$

Since $\mathbf{y}$ is obtained by adding a constant offset to $\mathbf{x}$ and $\mathbf{x}\perp\mathbf{1}$, we have $\langle \mathbf{y},\mathbf{y}\rangle_\Pi \geqslant \langle \mathbf{x},\mathbf{x}\rangle_\Pi$ and $\langle \mathbf{y}, L\mathbf{y}\rangle_\Pi = \langle \mathbf{x}, L\mathbf{x}\rangle_\Pi$. Thus

$$\frac{\mathbf{E}[A]}{\mathbf{E}[B]} \leqslant \sqrt{2R_L(\mathbf{y})} \leqslant \sqrt{2R_L(\mathbf{x})},$$

or equivalently

$$\mathbf{E}\left[A - B\sqrt{2R_L(\mathbf{x})}\right] \leqslant 0.$$

Therefore, the probability of choosing $t \in [y_1, y_n]$ such that $A - B\sqrt{2R_L(\mathbf{x})} \leqslant 0$ is nonzero. This proves the existence of S_t that $\max\left\{\Phi(S_t), \Phi(\overline{S}_t)\right\} \leqslant \sqrt{2R_L(\mathbf{x})}$ and thus indicates the correctness of Theorem 2. $\square$

2 *Kirchhoff's Theorem*

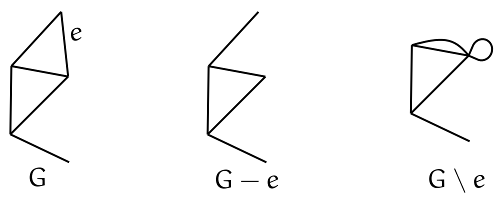
Denote the number of spanning trees in graph $G = (V, E)$ as $\tau(G)$. W.l.o.g., assume $V = [n]$. Let $L(u)$ be the matrix obtained by removing the u-th row and column in the Laplacian matrix of G. Recall that in the first lecture, we introduced the Kirchhoff's theorem.

Theorem 3 (Kirchhoff's Theorem) *The number of spanning trees in* G *equals to* $\det(L(u))$ *for any* $u \in [n]$.

For example, when G is K_3, we have $A_G = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ and $L_G = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$. Choosing $u = 1$, $L_G(1) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. By the Kirchhoff's theorem, $\tau(G) = \mathrm{det}(L_G(1)) = 3$.

Proof. We prove this by induction on the number of edges in G. When there are only one edge in G, it is easy to verify that $\tau(G) = \det(L(u))$ for any u.

We define the deletion operation $G - e$ and contraction operation $G \setminus e$ as the following figure shows. Note that for any $e \in E$, $\tau(G - e)$ is the number



of spanning trees in G without choosing edge e and $\tau(G \setminus e)$ is the spanning trees containing e. Therefore, we have

$$\tau(G) = \tau(G - e) + \tau(G \setminus e)$$

for any $e \in E$.

When G contains more than one edges, w.l.o.g., assume there exists an edge $e = (2, 1)$. Consider $L_G(2)$ and $L_{G-(2,1)}(2)$. There is a little difference between these two matrices since the degree of vertex 1 is different in G and $G - (2, 1)$. To be specific,

$$L_G(2) = L_{G-(2,1)}(2) + \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & & & 0 \end{bmatrix}.$$

Note that

$$\det(L_G(2)) = \sum_{j=1}^{n-1} (-1)^{j+1} \cdot L_G(2)(1, j) \cdot \det\left(L_G(1, 2, j)\right). \tag{5}$$

The right hand side of Equation (5) is exactly the sum of $\det\left(L_{G-(2,1)}(2)\right)$ and $\det\left(L_G(1,2)\right)$. Furthermore, $L_G(1,2)$ is equivalent with $L_{G\setminus(2,1)}(1)$. By the induction hypothesis,

Here $L_G(2)(1,j)$ is the element on the first row and j-th column in $L_G(2)$ and $L_G(1,2,j)$ is the matrix obtained by deleting the first, second and j-th rows and columns in L_G.

$$\det\left(L_{G-(2,1)}(2)\right) = \tau\left(G-(2,1)\right) \quad \text{and} \quad \det\left(L_{G\setminus(2,1)}(1)\right) = \tau\left(G\setminus(2,1)\right).$$

Therefore,

$$\det(L_G(2)) = \tau\left(G-(2,1)\right) + \tau\left(G\setminus(2,1)\right) = \tau(G).$$

□

With the Kirchhoff's theorem, we can derive the Cayley's formula efficiently.

Theorem 4 (Cayley's formula) *The number of labeled trees with n vertices is n^{n-2}.*

Proof. Note that the number of labeled trees with n vertices is exactly the number of spanning trees in K_n. Applying the Kirchhoff's theorem on K_n, we have

$$\begin{aligned}\tau(K_n) &= \det\left(L_{K_n}(1)\right) = \det\left(nI_{n-1} - J_{n-1}\right)\\ &= \det\left(\begin{bmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & & \\ \vdots & & \ddots & \\ -1 & & & n-1 \end{bmatrix}\right).\end{aligned} \tag{6}$$

The above determinant equals to n^{n-2} by direct calculation. □

Here I_{n-1} is the $(n-1)$-dimensional identity matrix and J_{n-1} is a $(n-1)$-dimensional matrix whose entries are all ones.