CSC 2421H : Graphs, Matrices, and Optimization Lecture 2 : 17 Sep 2018

Clustering

Lecturer: Sushant Sachdeva *Scribe: Liwen Lu*

1 Clustering of a Graph and Eigenvalue

Informally, "clustering" is grouping vertices such that there are more connectivity inside each group compare to between groups.

1.1 Conductance

Definition 1.1. Volume*: let* $S \subseteq V$, $\text{Vol}(S) \stackrel{\text{def}}{=} \sum_{v \in S} d_v$ *where* $d_v = \deg(v) = \boldsymbol{D}_{v,v}$

Lemma 1.2. *Define the indicator of set* S *as*

$$\mathbf{1}_S(v) = \begin{cases} 1, & v \in S \\ 0, & \textit{otherwise} \end{cases}$$

Then

$$\mathbf{1}_S^T \boldsymbol{D} \mathbf{1}_S = \text{Vol}(S)$$

Definition 1.3. *Define* Boundary of S *as*

$$E(S, \overline{S}) = \{(u,v) | (u,v) \in E, u \in S, v \in V \setminus S\},$$

where $\overline{S}$ *is the complement set.*

Definition 1.4. *Define* conductance of S measured in graph G *to be*

$$\phi_G(S) \stackrel{\text{def}}{=} \frac{|E(S, \overline{S})|}{min\{\text{Vol}(S), \text{Vol}(V \setminus S)\}},$$

and define conductance of graph G *to be*

$$\phi(G) \stackrel{\text{def}}{=} \min_{\substack{S \subseteq V \\ S \neq \emptyset, V}} \phi_G(S).$$

Computing $\phi(G)$ is called "minimum conductance problem", and it is famously NP hard. However, we can connect conductance and eigenvalues.

1.2 Normalized Laplacian Matrix

Let $\nu_2 = \min_{y^\top \boldsymbol{D}\mathbf{1}=0} \left(\frac{y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y} \right)$.

Define $x = \boldsymbol{D}^{\frac{1}{2}} y$ (note $\boldsymbol{D}^{\frac{1}{2}}$ is replacing each of the diagonal entries in diagonal matrix $\boldsymbol{D}$ with square root), we can write $y^\top \boldsymbol{D} y = x^\top x$.

Note we are assuming the graph G is connected, then diagonal of D is strictly non-negative, so the mapping between x and y is a bijection, which gives us $x = \boldsymbol{D}^{\frac{1}{2}} y \Leftrightarrow y = \boldsymbol{D}^{-\frac{1}{2}} x \Rightarrow y^\top \boldsymbol{D} y = x^\top x$. Then we can rewrite ν_2 as

$$\nu_2 = \min_{x^\top \left(\boldsymbol{D}^{\frac{1}{2}} \mathbf{1}\right)=0} \frac{x^\top \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{L} \boldsymbol{D}^{-\frac{1}{2}} x}{x^\top x}.$$

Definition 1.5. *Define* Normalized Laplacian Matrix *as*

$$\boldsymbol{N} \stackrel{\text{def}}{=} \boldsymbol{D}^{-\frac{1}{2}} \boldsymbol{L} \boldsymbol{D}^{-\frac{1}{2}}$$

We also claim that:

$$\begin{aligned} \lambda_1(\boldsymbol{N}) &= 0 \\ \psi_1 &= \boldsymbol{D}^{\frac{1}{2}} \mathbf{1} \\ \nu_2 &= \min_{x^\top \psi_1 = 0} \frac{x^\top \boldsymbol{N} x}{x^\top x} \\ &= \lambda_2(\boldsymbol{N}) \end{aligned}$$

1.3 Cheeger's Inequality

Theorem 1.6.

$$\frac{\nu_2}{2} \le \phi(G) \le \sqrt{2\nu_2}, \nu_2 \le 2$$

For example, if $\nu_2 = 0.01 \Rightarrow 0.05 \le \phi(G) \le 0.14$
We'll first prove the left hand side, $\frac{\nu_2}{2} \le \phi(G)$.

Proof. By definition, we know $\exists S$ *s.t.* $\phi(G) = \phi_G(S) = \frac{|E(S,\overline{S})|}{\mathrm{Vol}(S)}$ and $\mathrm{Vol}(S) \le \mathrm{Vol}(\overline{S})$.
Recall $\nu_2 = \min\limits_{y^\top \boldsymbol{D}\mathbf{1}=0} \left(\frac{y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y}\right)$ and $\mathbf{1}_S^\top \boldsymbol{D} \mathbf{1}_S = \mathrm{Vol}(S)$, we have

$$\mathbf{1}_S^\top \boldsymbol{L} \mathbf{1}_S = \sum_{(u,v)\in E} \left(\mathbf{1}_S(u) - \mathbf{1}_S(v)\right)^2$$

note

$$\left(\mathbf{1}_S(u) - \mathbf{1}_S(v)\right)^2 = \begin{cases} 1, & \mathbf{1}_S(v) \ne \mathbf{1}_S(v) \Leftrightarrow (u,v) \in E(S,\overline{S}) \\ 0, & \text{otherwise} \end{cases}$$

therefore

$$\mathbf{1}_S^\top \boldsymbol{L} \mathbf{1}_S = \sum_{(u,v)\in E} \mathbf{1}[(u,v) \in E(S,\overline{S})] = |E(S,\overline{S})|$$

If we set $y = \mathbf{1}_S$, it would minimize $\frac{y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y}$, however it may not satisfy $y^\top \boldsymbol{D}\mathbf{1} = 0$. In order to make y satisfy $y^\top \boldsymbol{D}\mathbf{1} = 0$, let $y = \mathbf{1}_S + c\mathbf{1}$, note we still have $y^\top \boldsymbol{L} y = |E(S,\overline{S})|$.

$$y^\top \boldsymbol{D}\mathbf{1} = 0 \Leftrightarrow c = \frac{-\mathbf{1}_S^\top \boldsymbol{D}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{D}\mathbf{1}} = -\frac{\mathrm{Vol}(S)}{\mathrm{Vol}(V)} \qquad \text{(by previous lemma)}$$

$$
\begin{aligned}
y^\top \boldsymbol{D} y &= \mathbf{1}_S^\top \boldsymbol{D} \mathbf{1}_S + 2c\mathbf{1}_S^\top \mathbf{1} + c^2 \mathbf{1}^\top \boldsymbol{D} \mathbf{1} \\
&= \mathbf{1}_S^\top \boldsymbol{D} \mathbf{1}_S - \frac{\left(\mathbf{1}_S^\top \boldsymbol{D} \mathbf{1}\right)^2}{\mathbf{1}^\top \boldsymbol{D} \mathbf{1}} \\
&= \mathrm{Vol}(S) - \frac{\mathrm{Vol}(S)^2}{\mathrm{Vol}(V)} \\
&= \mathrm{Vol}(S)\left(1 - \frac{\mathrm{Vol}(S)}{\mathrm{Vol}(V)}\right) \\
&\geq \frac{\mathrm{Vol}(S)}{2} && \text{(since } \mathrm{Vol}(S) \leq \mathrm{Vol}(\overline{S})\text{)}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\frac{y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y} &= \frac{|E(S, \overline{S})|}{y^\top \boldsymbol{D} y} \\
&\leq \frac{2|E(S, \overline{S})|}{\mathrm{Vol}(S)} \\
&= 2\phi_G(S) \\
&= 2\phi(G) && \text{(by assumption)} \\
\Rightarrow \nu_2 &= \min_{y^\top \boldsymbol{D}\mathbf{1}=0} \frac{y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y} \leq 2\phi(G) \\
\Rightarrow \frac{\nu_2}{2} &\leq \phi(G)
\end{aligned}
$$

□

Lemma 1.7. *Given y s.t. $y^\top \boldsymbol{D}\mathbf{1} = 0$, we can find a distribution on t with $S_t \subseteq V$ s.t. $\mathrm{Vol}(S_t) \leq \frac{\mathrm{Vol}(V)}{2}$, and*

$$
\frac{\mathbb{E}_t |E(S_t, \overline{S_t})|}{\mathbb{E}_t \mathrm{Vol}(S_t)} \leq \sqrt{\frac{2y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y}}
$$

Let t be independent choices, P_t be distribution of t, X_t, Y_t are variables depend on t with $Y_t \geq 0$.
We can prove that $\exists t$ s.t.

$$
\frac{X_t}{Y_t} \leq \frac{\mathbb{E}_t X_t}{\mathbb{E}_t Y_t} = \frac{\sum_t P_t X_t}{\sum_t P_t Y_t}
$$

The proof is left as exercise. This implies that $\exists t$, s.t.

$$
\frac{|E(S_t, \overline{S_t})|}{\mathrm{Vol}(S_t)} \leq \sqrt{\frac{2y^\top \boldsymbol{L} y}{y^\top \boldsymbol{D} y}}
$$

Furthermore, if y was the minimizing vector for ν_2, then

$$
\phi_S(S_t) = \frac{|E(S_t, \overline{S_t})|}{\mathrm{Vol}(S_t)} \leq \sqrt{2\nu_2}
$$