Lecture 12: Spectral clustering and Cheeger's Inequality.

Recall: $G = (V, E)$, $|V| = n$, $|E| = m$, $V = \{1, \dots, n\}$.

$$L_G = D - A$$

degree matrix → D; adjacency matrix → A

$D_{uu} = \deg(u)$

$A_{uv} = \begin{cases} 1 & (u,v) \in E \\ 0 & \text{o.w.} \end{cases}$

For any $x \in \mathbb{R}^n$,

$$(L_G x)_w = \sum_{v:(w,v)\in E} (x_w - x_v).$$ "voltage-to-current" map

$$x^T L_G x = \sum_{(u,v)\in E} (x_u - x_v)^2$$ "total power dissipation".

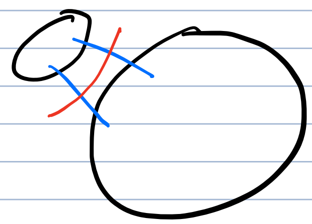
$0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ eigenvalues of L_G.

Thm: multiplicity of 0 = # connected components of L_G.

Cor: If $\lambda_2 = 0$, the graph is not connected.

This class: a robust version of this.

If λ_2 is small, then there is an isolated "cluster"

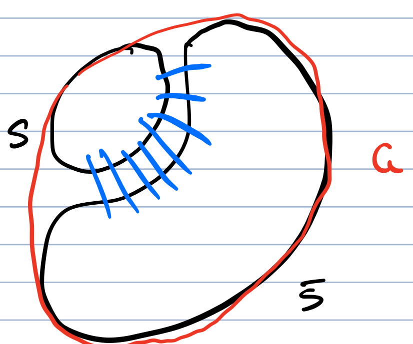


Def: For $S \subseteq V$, define $\mathrm{vol}(S) = \sum_{v \in S} \deg(v)$

Def: For $S \subseteq V$, define the <u>conductance</u> of S to be

$$\varphi_G(S) := \frac{|E(S,\bar{S})|}{\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))}$$ ← # edges between S and $\bar{S}$

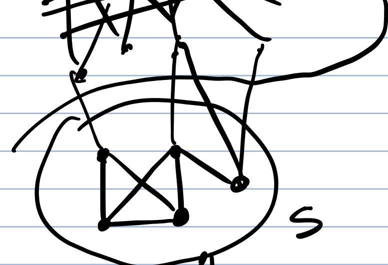
(care about $\mathrm{vol}(S) \le \mathrm{vol}(\bar{S})$).



if $\mathrm{vol}(S) \le \mathrm{vol}(\bar{S})$,

$\varphi_G(S) \approx$ fraction of edges in S that leave S.

So $\varphi_G(S)$ small $\Rightarrow$ good cluster.



Define $\rho_2(G) = \min_{S \neq \phi} \varphi_G(S)$.

$\varphi_G(S) = \frac{4}{15}$

Fact: $\rho_2(G)$ is NP-hard to compute exactly.
(see sparsest cut, graph expansion).

This class: good apx in some regimes via spectral clustering.

Recall: for a d-regular graph, the normalized Laplacian is

$$\mathcal{L}_G = \frac{1}{d} \cdot L_G$$

More generally, $\mathcal{L}_G = D^{-1/2} L_G D^{-1/2}$

Let $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq 1$ be eigenvalues of $\mathcal{L}_G$.

(discrete)
Cheeger's Inequality [Cheeger '71] For every graph G,

$$\frac{1}{2} \lambda_2(G) \leq \rho_2(G) \leq \sqrt{2 \lambda_2(G)},$$

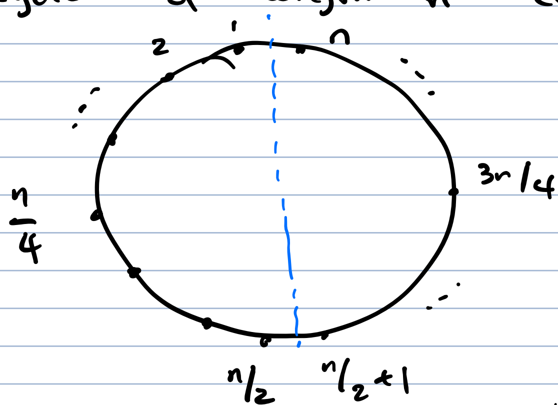
"easy side" "hard side"

and a cut achieving this upper bound can be found algorithmically.

Note: both sides are tight. i.e.

$\rho_2(G) \leq 100\, \lambda_2(G)$ not true in general.

e.g. cycle of length n C_n



$\rho_2(G) = 2/n$

but $\lambda_2(G) \leq O(1/n^2)$.

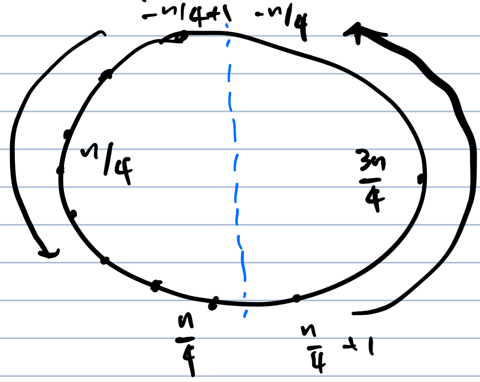
suffices to find vector u

w/ $\|u\|_2 = 1$, $u \perp \vec{1}$,

$u^T \mathcal{L}_G u = O(1/n^2)$.

$= \frac{1}{2} L_G$.

$u \perp \vec{1} \iff \langle u, \vec{1} \rangle = 0 \iff \sum u_i = 0.$

$x_i = -\frac{n}{4} + i$  $x_i = \frac{3n}{4} - i$

$\sum x_i = 0.$, and $x^T L_G x = \sum_{(u,v)} (x_u - x_v)^2 = n.$

let $u = \frac{x}{\|x\|_2}$, so $u^T L_G u = \frac{x^T L_G x}{\|x\|_2^2}$

$$\|x_i\|_2^2 = 4 \cdot \sum_{i=1}^{n/4} i^2 = 4 \cdot \frac{(n/4)\cdot(n/4+1)\cdot(n/4 \cdot 2 + 1)}{6}$$
$$= \Omega(n^3).$$

$$\Rightarrow \quad u^T L_G u = \frac{n}{\Omega(n^3)} = O\left(\frac{1}{n^2}\right).$$

Proof of "easy side": Assume for simplicity G is d-regular.

We will show that for any set $S \subseteq V$, $\lambda_2 \le 2\varphi_G(S)$.

Since G is d-regular, $\mathrm{vol}(S) = d \cdot |S|$. So wlog let's take $|S| \le n/2$, since $\varphi_G(S) = \varphi_G(\bar{S})$.

Recall variational characterization:

$$\lambda_2 = \min_{\substack{u \perp v_1 \\ \|u\|_2 = 1}} u^T \mathcal{L}_G u. \qquad v_1 = (1, \dots, 1)$$

$u \perp v_1 \iff \sum u_i = 0.$

Let $(x_S)_u = \begin{cases} 1, & \text{if } u \in S \\ 0, & \text{o.w.} \end{cases}$. To make it orthogonal to v_1, take

$y_S = x_S - \overline{x_S}$

$\overline{x_S} = \left(\frac{1}{n} \sum_u (x_S)_u\right) \cdot (1, \dots, 1) = \frac{|S|}{n} \cdot (1, \dots, 1).$

$$\frac{y_S}{\|y_S\|}^T \cdot \mathcal{L}_G \cdot \frac{y_S}{\|y_S\|}$$

$$(*) = \frac{\frac{1}{d} \cdot \sum_{(u,v) \in E} \left((y_S)_u - (y_S)_v\right)^2}{\|y_S\|^2}$$

$\left((y_S)_u - (y_S)_v\right)^2 \to 1$ if (u,v) crosses S and $\bar{S}$,

0 o.w.

$$= \frac{1}{d} \cdot |E(S, \bar{S})| \cdot \frac{1}{\|y_S\|^2}$$

$$\|y_S\|_2^2 = \sum_{v \in V} \left((x_S)_v - \frac{|S|}{n} \right)^2$$

$$= \sum_{v \in S} \left(1 - \frac{|S|}{n}\right)^2 + \sum_{v \in \bar{S}} \left(\frac{|S|}{n}\right)^2$$

$$= |S| \cdot \left(1 - \frac{|S|}{n}\right)^2 + (n - |S|) \cdot \left(\frac{|S|}{n}\right)^2$$

$$= \frac{1}{n^2} \cdot |S| \cdot (n - |S|)^2 + \frac{1}{n^2} \cdot (n - |S|) \cdot |S|^2$$

$$= \frac{1}{n^{\cancel{2}}} \cdot |S| \cdot (n - |S|) \cdot (\cancel{n - |S|} + |S|)$$

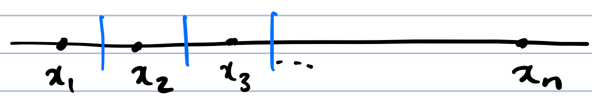
$$= \frac{1}{n} \cdot |S| \cdot (n - |S|) \geq \frac{1}{2} |S|$$

so $$(*) \leq \frac{\frac{1}{d} \cdot |E(S, \bar{S})|}{\frac{1}{2} \cdot |S|} = 2 \frac{|E(S, \bar{S})|}{\min(\mathrm{vol}(S), \mathrm{vol}(\bar{S}))}$$

$$= 2 \cdot \varphi_s(S). \quad \blacksquare$$

"Proof" of "hard side": Algorithmic in nature! Let $x \in \mathbb{R}^n$.

SWEEP(x): Sort vertices so that $x_1 \leq x_2 \leq \cdots \leq x_n$.



Consider cuts $S = \{x_1\}$

$S = \{x_1, x_2\}$

$\vdots$

Output cut of minimal conductance of these candidates.

Suppose $0 \leq x_1 \leq \cdots \leq x_n \leq 1$. For any t, let $S_t = \{x_i : x_i > t\}$. Then $\exists t$ s.t.

$$\frac{|E(S_t, \bar{S}_t)|}{d \underset{\uparrow \text{min}}{|S_t|}} \leq \frac{\sum_{\{i,j\} \in E} |x_i - x_j|^2}{d \sum_{i=1}^{n} x_i^2} \longrightarrow \frac{x^T \mathcal{L}_G x}{\|x\|_2^2}$$

If you squint, this is almost what we want!

pf: Choose $t \sim \mathrm{Unif}[0, 1]$ for now. In this case,

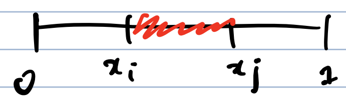
$$\mathbb{E}_t\left[\,|S_t|\,\right] = \mathbb{E}_t\left[\sum_{i=1}^{n} \mathbb{1}[x_i \in S_t]\right]$$

$$= \sum_{i=1}^{n} \Pr[x_i \in S_t]$$

$\Pr[t \le x_i]$

$$= \sum_{i=1}^{n} x_i$$

$$\mathbb{E}_t\left[\,|E(S_t, \bar{S}_t)|\,\right] = \mathbb{E}_t\left[\sum_{(i,j)\in E} \mathbb{1}[(x_i, x_j) \text{ is cut}]\right]$$



$$= \sum_{(i,j)\in E} \Pr[(x_i, x_j) \text{ is cut}]$$

$$= \sum_{(i,j)\in E} |x_i - x_j|$$

So what we've shown is that

$$(*) = \frac{\mathbb{E}_t\left[\,|E(S_t, \bar{S}_t)|\,\right]}{\mathbb{E}_t\left[\,|S_t|\,\right]} = \frac{\sum_{(i,j)\in E} |x_i - x_j|}{\sum_i x_i}$$

$\frac{\text{sum}}{\text{sum}}$.

Fact: If $a_1, \dots, a_m \ge 0$, $b_1, \dots, b_m \ge 0$, then

$$\min_i \frac{a_i}{b_i} \le \frac{a_1 + \dots + a_m}{b_1 + \dots + b_m} \le \max_i \frac{a_i}{b_i}$$

$$\Rightarrow \exists t \text{ s.t. } \frac{|E(S_t, \bar{S}_t)|}{|S_t|} \le (*).$$

the rest of the proof is a bit annoying and left to the reader!

Final claim: $\exists t$ s.t.

$$\varphi_G(S_t) \le \sqrt{2\lambda_2}$$

tl;dr: Cheeger's inequality relates spectrum of $\mathcal{L}_G$ to value of cut.

+ gives efficient algorithm for finding it!

Higher order Cheeger's Inequality.

$\lambda_2 \longrightarrow$ partitions into 2 clusters.

What about λ_k? $\longrightarrow$ partitions into k clusters.

$$\rho_k(G) := \min_{\substack{S_1, \dots, S_k \\ \text{non-empty,} \\ \text{pairwise disjoint}}} \max \{ \varphi_G(S_i), i = 1, \dots, k \}.$$

can be improved?

Thm: $\frac{\lambda_k(G)}{2} \le \rho_k(G) \le O(k^2) \sqrt{\lambda_k(G)}$

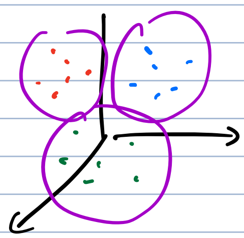
[Lee, Oveis-Gharan, Trevisan '14].

$$\rho_k(G) \le \sqrt{\lambda_{1.1 \cdot k}(G) \cdot \log k}$$

High level idea:

$F: V \longrightarrow \mathbb{R}^k$

$F(v) = ((x_1)_v, \dots, (x_k)_v)$



high-dimensional clustering

e.g. k-means.