Lecture 24: Spectral Graph Theory I

[Questions motivating us now...]: Given undirected graph G...

- how to "cluster" vertices into 2+ groups?
- how to "visualize" in 2 or 3 dims?
- how long does random walk take to get "close" to stationary dist π?

Henceforth for simplicity: G is d-regular. [Everything has appropriate analogues for non-regular graphs, but math & formulas get a little annoying; let's keep things simple!]

$$K = \begin{bmatrix} & & \\ & \text{adj. matrix} & \\ & & \end{bmatrix} = \frac{1}{d}A$$
$\div d_1 \ \div d_2 \ \cdots \ \div d_n$

symmetric, $K = K^T$

[Remember I told you linear algebra is 100X easier with symmetric matrices? This is one example...]

Stationary distribution: $\pi[u]$ proportional to $\deg[u]$ ... $\therefore$ $\pi = \begin{bmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{bmatrix}$ = uniform dist. ☺

$$K\pi = \pi \Rightarrow \pi \text{ is eigenvec of } K \text{, eigenvalue } 1.$$

[we'll talk about eigvecs/vals more soon]

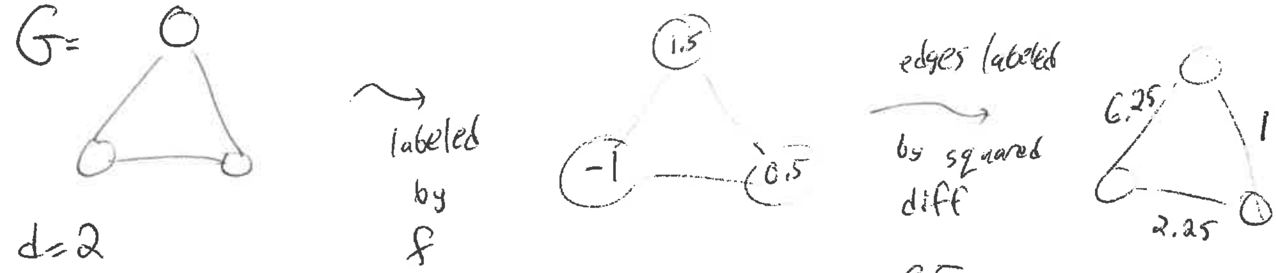
Say for each vertex $v \in V$ we have a real # data value $f(v)$.
[Could be anything — temperature, color id, height, x-coord. in a plot...] $f: V \to \mathbb{R}$

define: $$\mathcal{E}[f] = \frac{1}{2} \mathop{\mathbb{E}}_{(u,v)\text{ edge}} \left[(f(u) - f(v))^2 \right]$$

average

"energy"
["quadratic form"]

"average variance on edges"

[don't worry about too much; just convenient to have this factor] e.g.: 

$$\text{avg} = \frac{9.5}{3} \approx 3.166$$

$$\Rightarrow \mathcal{E}[f] \approx 1.584$$

[whatever, just an example]

** e.g. ** Let $S \subseteq V$. Let $f(v) = \begin{cases} 1 & \text{if } v \in S \\ 0 & \text{if } v \notin S \end{cases}$ What's $\mathcal{E}[f]$?

[S is a "cut"; f is 1 on one side, 0 on other]

$$\mathcal{E}[f] = \frac{1}{2} \mathop{\mathbb{E}}_{\substack{(u,v) \\ \text{edge}}} \Big[\underbrace{(f(u)-f(v))^2}_{} \Big]$$

0 if $u,v \in S$ or $u,v \notin S$

$(\pm 1)^2 = 1$ if u,v on "opp. sides": (u,v) crosses "cut" $(S, \bar{S})$

$$= \frac{1}{2} \mathop{\Pr}_{\substack{(u,v) \\ \text{edge}}} \left[u \in S, v \notin S \text{ or } u \notin S, v \in S \right]$$

$$= \mathop{\Pr}_{\substack{(u,v) \\ \text{edge} \\ \text{(directed)}}} \left[u \in S, v \notin S \right]$$

"$u \to v$ is stepping out of S"

[= "fractional size of boundary of S"]

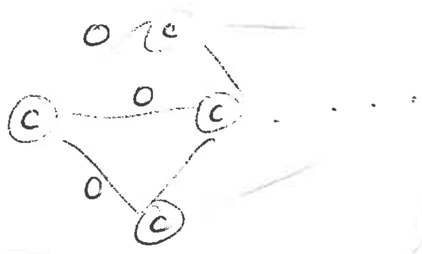
[Has something to do with how long it takes random walk to "escape out of S".]

Fact: $\mathcal{E}[f] \geq 0$ always. [Why? Well, it's avg. of squares, ≥ 0.]

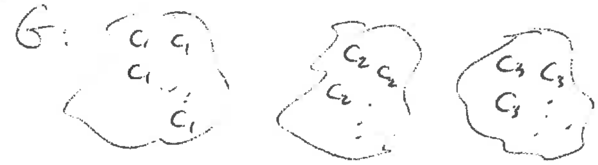
Q: For what f is $\mathcal{E}[f] = 0$?

A: $f(v) = 0 \ \forall v$, $f(v) = 1 \ \forall v$, $f(v) = c \ \forall v$ [Anything else?]

That's it if G is connected.



Say now G might be disconnected.



We'll think of $f: V \to \mathbb{R}$ as a vector: $\begin{bmatrix} f(v_1) \\ f(v_2) \\ \vdots \\ f(v_n) \end{bmatrix} \in \mathbb{R}^n$.

Say G has 3 connected comps, S_1, S_2, S_3.

Then $\exists$ 3 linearly independent f's with $\mathcal{E}[f] = 0$:

$$\begin{array}{l} S_1\{ \\ \\ \\ \\ \end{array}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \begin{array}{l} \\ S_2\{ \\ \\ \end{array}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \begin{array}{l} \\ \\ \\ S_3\{ \end{array}\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

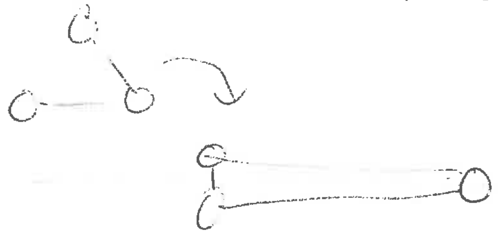
[and every f with $\mathcal{E}[f] = 0$ is a lin. combo of them, coeffs c_1, c_2, c_3.

Connection between linear algebra & graph theory!]

What's largest $\mathcal{E}[f]$ can be? [Well... ∞. We need to fix some "scaling".]

$\sim$ Multiplying f by k multiplies $\mathcal{E}[f]$ by k^2. [So the natural way to scale is also quadratic.]

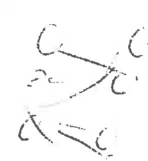
Say f "normalized" if $\underset{v \text{ vertex}}{\overset{(\text{avg})}{\mathbb{E}}}[f(v)^2] = 1$. [Can always mult by a const. to achieve this.]



[Trying to map vertices $v \in V$ into real line — avg. squared dist to $0 = 1$ — so that edges "stretched" as much as possible.]

[What graph might allow for a very large $\mathcal{E}$?]

Say G is bipartite: vertices U, V, all edges between them



Let $f(w) = \begin{cases} +1 & \text{if } w \in U \\ -1 & \text{if } w \in V. \end{cases}$ Normalized? ✓ $\mathbb{E}[f(w)^2] = \mathbb{E}[(\pm 1)^2] = 1$.

$$\mathcal{E}[f] = \frac{1}{2} \underset{\substack{(u,v) \\ \text{edge}}}{\mathbb{E}}\Big[\underbrace{(f(u) - f(v))^2}_{(\pm 2)^2}\Big] = 2.$$

Prop: For normalized f, $\mathcal{E}[f] \le 2$ always; equality possible iff G is bipartite.

Proof:

$$\mathcal{E}[f] = \frac{1}{2}\Big(\underbrace{\mathop{\mathbb{E}}_{(u,v)\text{ edge}}[f(u)^2]}_{\hookrightarrow u \text{ is a unif. rand. vertex [regularity]}} - 2\mathop{\mathbb{E}}_{(u,v)\text{ edge}}[f(u)f(v)] + \mathop{\mathbb{E}}_{(u,v)\text{ edge}}[f(v)^2]\Big)$$

normalized

$$= \frac{1}{2}(1) - \frac{1}{2}\cdot 2 \mathop{\mathbb{E}}_{(u,v)\text{ edge}}[f(u)f(v)] + \frac{1}{2}(1) = 1 - \mathbb{E}[f(u)f(v)] \le 2$$

provided...

Claim: $\left|\mathop{\mathbb{E}}_{(u,v)\text{ edge}}[f(u)f(v)]\right| \le 1$. **Proof:** ~~Cauchy-Schwarz~~ "Cauchy-Schwarz": $|\mathbb{E}[XY]| \le \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$

$$\le \sqrt{\mathbb{E}[f(u)^2]}\sqrt{\mathbb{E}[f(v)^2]} = \sqrt{1}\sqrt{1} = 1$$

$\blacksquare$

Spectral Graph Theory:

- G has k "almost disconnected" clusters $\iff \mathcal{E}[f_1], \dots, \mathcal{E}[f_k] \approx 0$ for k lin. independent $f_1, \dots, f_k$.

- G is almost bipartite (= huge max cut) $\iff \mathcal{E}[f] \approx 2$ for some f

[Where do eigenvalues come in?] [Almost there...] [E.g., if G not bip., how to find f maximizing $\mathcal{E}[f]$?]

What is

$$\begin{array}{cc} & \cdots v \cdots \\ u \rightarrow & \begin{bmatrix} & & \\ & K & \\ \frac{1}{d} & \frac{1}{d} & \frac{1}{d} \end{bmatrix} \end{array} \begin{bmatrix} \vdots \\ f \\ \vdots \end{bmatrix} \begin{array}{c} \leftarrow v_1 \\ \leftarrow v_2 \\ \vdots \\ \leftarrow v_d \end{array} \overset{?}{=} \begin{bmatrix} \\ h \\ \\ \end{bmatrix} \leftarrow u$$

at any $(u,v_1), (u,v_2), \dots, (u,v_d)$

$$h(u) = \frac{1}{d} f(v_1) + \cdots + \frac{1}{d} f(v_d), \quad v_i\text{'s are neighbors of } u.$$

$$= \mathop{\mathbb{E}}_{v:\ (u,v)\text{ edge}}[f(v)].$$

[Applying K to f: replace each data value with average over neighbors.]

def: Inner product: $\langle f, g\rangle = \frac{1}{n}(f \cdot g) = \frac{1}{n}(f(v_1)g(v_1) + \dots + f(v_n)g(v_n))$

$= \underset{v \text{ vertex}}{\mathbb{E}}[f(v)g(v)]$

("nonstandard")

["correlation between f & g"]

$\therefore \langle f, f\rangle = \underset{v}{\mathbb{E}}[f(v)^2] = 1$ iff f "normalized"

[sorta like "unit vec"]

Q: What is $\langle f, Kf\rangle$?

A: $\underset{u}{\mathbb{E}}[f(u) \cdot (Kf)(u)] = \underset{u}{\mathbb{E}}\left[f(u) \cdot \underset{\substack{v: \\ (u,v) \text{ edge}}}{\mathbb{E}}[f(v)]\right] = \underset{(u,v) \text{ edge}}{\mathbb{E}}[f(u)f(v)]$

[Pretty natural].

Q: What is $\mathcal{E}[f]$?

A: $= \frac{1}{2} \underset{(u,v) \text{ edge}}{\mathbb{E}}[(f(u) - f(v))^2] = \frac{1}{2}\left(\underset{u}{\mathbb{E}}[f(u)^2] - 2\underset{(u,v) \text{ edge}}{\mathbb{E}}[f(u)f(v)] + \underset{v}{\mathbb{E}}[f(v)^2]\right)$

$= \frac{1}{2}\langle f, f\rangle - \langle f, Kf\rangle + \frac{1}{2}\langle f, f\rangle$

$= \langle f, f\rangle - \langle f, Kf\rangle$

$= \langle f, f - Kf\rangle = \langle f, \overset{\text{ident mtx}}{I}f - Kf\rangle$

$= \langle f, (I-K)f\rangle$

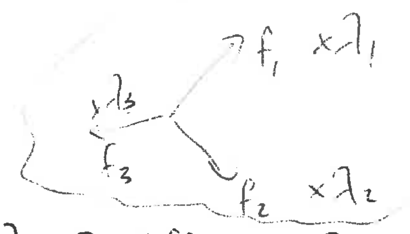
def: (normalized) Laplacian matrix of G: $L := I - K = I - \frac{1}{d}A = \frac{1}{d}(dI - A)$

[symmetric]

[The reason to care about it is not that $I-K$ is inherently meaningful; just that...]

$\mathcal{E}[f] = \langle f, Lf\rangle = \frac{1}{n} [\; f \;] [\; L \;] [f]$.

__Recall__: A symmetric matrix like L stretches $\mathbb{R}^n$ by stretch factors (eigenvalues) $\lambda_1, \dots, \lambda_n$ along axes (orthonormal eigenvectors) $f_1, \dots, f_n$.



Say f_i normalized. $Lf_i = \lambda_i \cdot f_i$

$\Rightarrow \langle f_i, Lf_i\rangle = \langle f_i, \lambda f_i\rangle$

$= \underset{v}{\mathbb{E}}[\lambda_i f_i(v)^2] = \lambda_i \mathbb{E}[f_i(v)^2] = \lambda_i$

$\therefore \mathcal{E}[f_i] = \lambda_i$. $\therefore$ all λ_i's are between 0 & 2.

$\therefore$ G has k connected components

[we'll review] $\iff$ k of the λ_i's are 0.

G is bipartite $\iff$ one of the λ_i's is 2. !