Announcements

- start recording
- scribe: work w/other scribe to produce one set of notes
- My OH: Mon 12:30–1:30, Thu 9–10
- TF hybrid section/OH: Mon 2–3, Wed 2–3, Wed 5–6, Thu 5:30–6:30
- PS 1 due Fri : note collaboration + late day policies
- If Zoom goes down, check Piazza
- Sync whiteboard
- gather after class
- register to vote
- screenshare + jamboard ?

Agenda

- random walks on directed graphs
- Pagerank
- Markov Chain Monte Carlo

Perron-Frobenius Thm for ~~Symmetric~~ Matrices

Let M = ~~symmetric~~ nonnegative real matrix (e.g. adjacency or rw ~~normalized adj~~ mx of ~~undirected~~ directed G)

w/ corresponding graph G (i.e. $E = \{(a,b) : M(a,b) > 0\}$).

Complex Eigenvalues $\mu_1 \cancel{\geq}, \mu_2 \cancel{\geq}, \ldots \cancel{\geq}, \mu_n$ multiplicities = # occurrences as a root of $\det(xI - M)$

Then 1) $\exists$ nonnegative right e-vector v_1 w/ real e-value μ_1

2) $\mu_1 \geq |\mu_j|$ for $j > 1$

~~2) $\mu_1 \geq \mu_2$ with strict inequality if G connected + strict positivity of v_1~~

~~3) $\mu_1 \geq -\mu_n$ with strict inequality if G connected + nonbipartite~~

Moreover, if G bipartite then $\mu_{n-i} = -\mu_{i+1}$ for $i = 0, \ldots, n-1$

3) If G is strongly connected, can take v_1 to be strictly positive

4) If G is strongly connected + aperiodic (gcd of all cycle lengths = 1) then $\mu_1 > |\mu_j|$ for $j = 2, \ldots, n$

5) If G has period k, then the set of e-values is closed under mult. by $e^{2\pi i/k}$

Random Walks on ~~Un~~directed Graphs

$M(a,b) = w(b,a)$

$n = \#$ vertices

G ~~un~~directed, weighted. $W = M D_{out}^{-1}$

$p_0 =$ initial probability distribution on vertices $\in \mathbb{R}^n$

$p_t =$ prob dist. after t steps of r.w.

$= W^t p_0$

How to analyze?

Let $v_1 =$ nonneg. e-vector of W from P-F

$\pi = \frac{v_1}{|v_1|_1}$ prob dist.

$\mu_1 = 1$

$W\pi = \pi$

~~**Claim:** W has a basis of e-vectors $\phi_1, \ldots, \phi_n$~~

~~w/ evalues $\omega_1 \geq \omega_2 \geq \omega_3 \geq \cdots \geq \omega_n$~~

need not be orthogonal

$W = S L S^{-1}$ for an upper-triangular matrix whose diagonal is $\mu_1, \mu_2, \ldots, \mu_n$ (by Jordan Normal Form)

$$\begin{pmatrix} 1 & & 0 \\ & \mu_2 \ddots & \\ * & & \mu_n \end{pmatrix}$$

$\pi = S e_1$

Assume G strongly connected + aperiodic ("ergodic")

$W^t = S L^t S^{-1}$ $\quad L^t \to \begin{pmatrix} 1 & & 0 \\ & 0 \ddots & \\ 0 & & 0 \end{pmatrix}$

$$\begin{pmatrix} | & | & & | \\ c_1\pi & c_2\pi & \cdots & c_n\pi \\ | & | & & | \end{pmatrix}$$

$c_1 = c_2 = \cdots = c_n = 1$

We also don't know π

well-defnd even for directed G

Thm: $\cancel{\max_{j>1} |\omega_j|} = \max_{\text{prob. dist } p} \dfrac{\|Wp - \pi\|_{\pi^{-1}}}{\|p-\pi\|_{\pi^{-1}}} \overset{\text{def}}{=} \omega_\pi$

where $\|x\|_{\pi^{-1}} \overset{\text{def}}{=} \sqrt{\sum_{a=1}^{n} \dfrac{x(a)^2}{\pi(a)}}$

Proof: $\max_{j>1} |\omega_i| \overset{p\leq 1}{=} \max_{x \perp \psi_1} \dfrac{\|Ax\|}{\|x\|}$ $\quad (A = D^{-1/2} W D^{1/2})$

$= \max_{y : D^{-1/2} y \perp \psi_1} \dfrac{\|D^{-1/2} W y\|}{\|D^{-1/2} y\|}$ $\quad (x = D^{-1/2} y)$

$= \max_{y : \sum_a y_a = 0} \dfrac{\|Wy\|_{\vec{d}^{-1}}}{\|y\|_{\vec{d}^{-1}}}$ $\quad p = \pi + cy$

$= \max_{\text{prob dist } p} \dfrac{\|W(p-\pi)\|_{\vec{d}^{-1}}}{\|p-\pi\|_{\vec{d}^{-1}}}$

$= \max_{\text{prob dist } p} \dfrac{\|Wp - \pi\|_{\pi^{-1}}}{\|p-\pi\|_{\pi^{-1}}}$

By def: $\|W^t p_0 - \pi\|_{\pi^{-1}} \le \omega_\pi^t \cdot \|p_0 - \pi\|_{\pi^{-1}}$

- initial distance $\|p_0 - \pi\|_{\pi^{-1}} \le \|p_0\|_{\pi^{-1}} \le \sqrt{\frac{1}{\pi_{min}}}$
- final ℓ_1 distance (= 2 · total variation distance)

$$|W^t p_0 - \pi|_1 = \sum_a \frac{|W^t p_0(a) - \pi(a)|}{\pi(a)^{1/2}} \cdot \pi(a)^{1/2}$$

$$\underset{\text{Cauchy-Schwarz}}{\le} \left(\sum_a \frac{(W^t p_0(a) - \pi(a))^2}{\pi(a)}\right)^{1/2} \cdot 1 = \|W^t p_0 - \pi\|_{\pi^{-1}}$$

Thm: for every start distribution p_0

$$|W^t p_0 - \pi|_1 \le \omega_\pi^t \cdot \|p_0\|_{\pi^{-1}} \le \omega_\pi^t \cdot \frac{1}{\sqrt{\pi_{min}}}$$

~~where $\pi_{min} = \frac{d_{min}}{\sum_a d(a)} \in \left[\frac{d_{min}}{d_{max}\, n}, \frac{1}{n}\right]$~~

π_{min} can be exponentially small even in digraphs of bounded degree

Cor: "Mixing time" to get to within ℓ_1 dist. ε

is $t = O\left(\frac{\log\left(\cancel{n d_{max}/\varepsilon d_{min}}\ 1/\varepsilon\pi_{min}\right)}{1-\omega_\pi}\right)$

PS2: ω_π can be 1 even in ergodic G

Q: how to ensure $\omega_\pi < 1$?

~~A: $\omega_\pi = \max\{\omega_2, -\omega_n\}$ $\omega_2 < 1$ iff G connected, $\omega_n > -1$ iff G nonbipartite~~

Lazy Random Walk : $\tilde{W} = \frac{1}{2}W + \frac{1}{2}I$

$\tilde{\omega}_i = \frac{1+\omega_i}{2} \geq 0$, so $\tilde{\omega}_\pi = \tilde{\omega}_2 = \frac{1+\omega_2}{2}$

If G is strongly connected and $1 - \tilde{\omega}_\pi = \frac{1-\omega_2}{2} = \frac{\nu_2}{2}$

then $\tilde{\omega}_\pi < 1$ (PS2?) $\nu_2 = 2^{nd}$ smallest e-value of normalized laplacian $N = I - A$

<u>Converse</u> (maybe ps 2): $\forall$ undirected G $\exists$ start vertex s such that for $p_0 = \delta_s$:

False for ω_π

$$t = \Omega\left(\frac{\omega_\pi}{1-\omega_\pi} \cdot \log\left(\frac{1}{\varepsilon}\right)\right)$$

Not sure for $\widetilde{\omega}_\pi$

steps are needed to get to within ℓ_1 distance ε of π.

Also: for every undirected, connected, nonbipartite G,

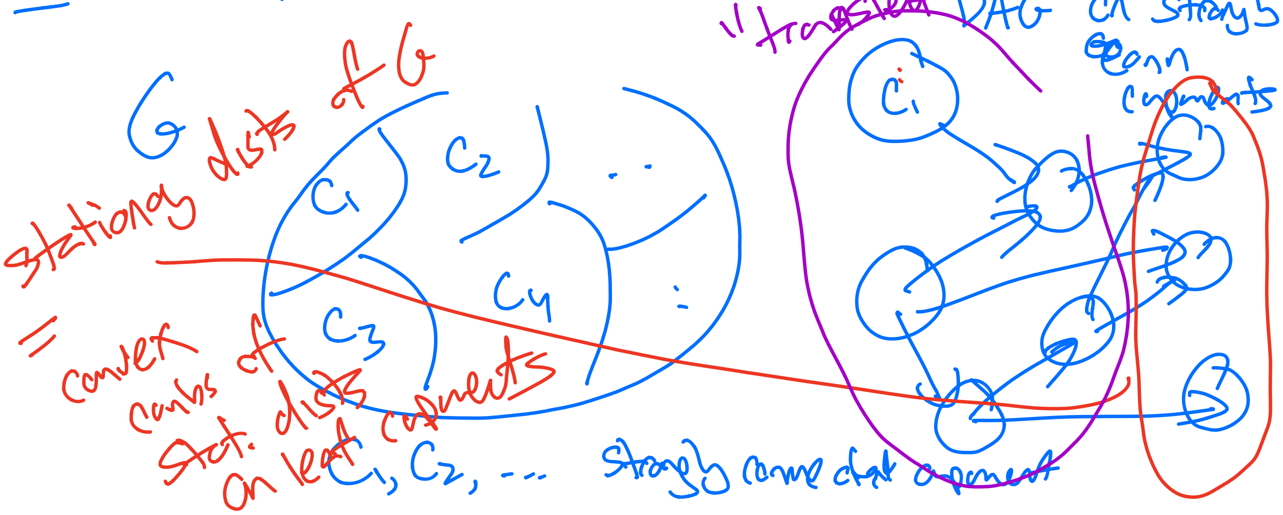
$$\omega_\pi \le 1 - 1/\mathrm{poly}(n \cdot d_{max}) \quad \text{(cf. ps 1)}$$

so mixing time $\le \mathrm{poly}(n \cdot d_{max})$.

False for ergodic directed G.

Mixing time can be $2^{\Omega(n)}$ even for bounded-degree digraphs.

Q: What happens in digraphs that are not "strongly connected"?



Pagerank

M adjacency matrix of web graph

$M(a,b) = \#$ links from page b to page a

$W = M D_{out}^{-1}$

$\tilde{W} = (1-\alpha) W + \alpha J$ ergodic

every entry $\frac{1}{n}$

(Pagerank vector = (unique) stationary distribution π of $\tilde{W}$

$\rightarrow$ Mixing time $t_{mix} = O\left(\frac{\log(n/\varepsilon)}{\alpha}\right)$

$\rightarrow$ So can calculate $\vec{u} = \frac{\vec{1}}{n}$

$W\pi = \pi$

$\Updownarrow$

$(I-W)\pi = 0$

$\pi \overset{\varepsilon}{\approx} \tilde{W}^t \vec{u}$

"Personalized" Pagerank = other restart vectors

or $\pi \overset{\varepsilon}{\approx} \sum_{j=0}^{t} \alpha \cdot (1-\alpha)^t W^t \vec{u}$

$\tilde{W}\pi = (1-\alpha) W\pi + \alpha \vec{u}$

($\tilde{W}\pi = \pi \Rightarrow \alpha \vec{u} = (I - (1-\alpha)W)\pi$

$\Rightarrow \pi = \alpha \, (I-(1-\alpha)W)^{-1} \vec{u}$)

cf. rw laplacian $I-W$

Markov Chain Monte Carlo

goal: efficiently sample from a desired probability distribution on a huge + complicated space

Why?

- TCS: approximate counting
- Stats + ML: Bayesian inference (posterior sampling)
- Statistical physics: estimate physical properties

Example: given an undirected graph G, count the # ~~spanning~~ forests of G

(acyclic subgraphs)

- NP-hard to compute exactly (even "#P-complete")
- We'll see poly time alg for # spanning trees (forests of size n-1)

Fact: MCMC can sample a near-uniform spanning forest in time poly$(n, 1/\varepsilon)$

$\Updownarrow$

can approximately count spanning forests to within factor $(1 \pm \varepsilon)$ in time poly$(n, 1/\varepsilon)$

(idea for $\Downarrow$: 1) pick an edge e

2) by repeated sampling, estimate

$$\hat{p}_e \approx \Pr\left[\begin{array}{c}\text{random spanning forest}\\ \text{contains } e\end{array}\right]$$

Rapidly Mixing Markov Chains

3a) if $\hat{p}_e < \frac{1}{2}$, recursively estimate

$\hat{Z}_{G-\{e\}} \approx$ # spanning forests in $G \setminus \{e\}$

spanning forests in G that don't contain e

output $$\frac{\hat{Z}_{G-\{e\}}}{1-\hat{p}_e} = \hat{Z}_G$$

3b) else, merge endpoints of e to get G', recursively estimate $\hat{Z}_{G'}$, output $$\frac{\hat{Z}_{G'}}{\hat{p}_e}$$

spanning forests in G that contain e

Q: how to sample a near-uniform spanning forest?

Metropolis-Hastings Markov Chain

- State space = spanning forests of $G=(V,E)$

  $N = \#$ states = as large as n^n

- Transitions: from forest $F \subseteq E$
  - pick a uniform random edge $e \in E$
  - let $F' = \underbrace{\begin{cases} F \triangle \{e\} & \text{if this is acyclic} \\ F & \text{otherwise} \end{cases}}_{\text{call this } F * e}$

- This is an $|E|$-regular graph on N vertices

- Goal: mixing time $\text{poly}(n) = \text{poly}(\log N)$

  i.e. $\omega_2 \leq 1 - 1/\text{poly}(n)$

  Proven last year! using spectral theory of high-dimensional expanders

  [Anari-Liu-Oveis Gharan, STOC 2019]

Generalizations

Statistical physics

"random cluster model"

"Gibbs distribution"

- Weighted spanning forests

$$\Pr[F] = \frac{w(F)}{Z}$$

where $w(F) = \prod_{e \in F} w_e$

$$Z = \sum_{\substack{\text{spanning} \\ \text{forests}}} w(F)$$ "partition function"

- Exercise: modify Metropolis-Hastings so that stationary distribution is above

assume all edge weights are in $[w_{min}, w_{max}]$

- (Weighted) Perfect Matchings in Bipartite graphs — "Dimer model"
- (Weighted) Matchings — "Monomer-Dimer model"
- (Weighted) Independent Sets — "Hardcore Gas Model"
- (Weighted) Colorings — "Antiferromagnetic Potts Model"