CPSC 531F   March 4, 2021

Today: Concrete examples of Markov chains and mixing times, esp.

- reversible chains
- "Metropolis algorithm" - uses reversibility

We said in P is an irreducible Markov matrices (chains), then P has a unique stationary distribution:

π s.t. (1) π is stochastic

(2) $\pi^T P = \pi^T$

also $\pi P = \pi$ (drop T)

All stuff today in Chapters 1–4 of [Levin–Peres]. In [L-P],

$$\pi P = \pi \quad \text{rather than} \quad \pi^T P = \pi^T$$

P is <u>reversible</u> if for all $i, j \in [n]$

($P \in M_n(\mathbb{R})$, $n = \#$ states, row stochastic, irreducible)

$$\pi_i P_{ij} = \pi_j P_{ji}$$

If P is reversible:

(i) P is the same as "running P backwards"

(2) P is symmetric wrt the "appropriately" weighted inner product

[notion of self-adjoint matrices, Rayleigh quotients, ...]

(3) You can get new Markov chains from P that have any stationary distribution you want on the same underlying digraph, in a "purely local" fashion ... { Metropolis (et al.) algorithm, simmulated annealing, ⋮

P irreducible Markov matrix, stationary distribution $\vec{\pi}$:

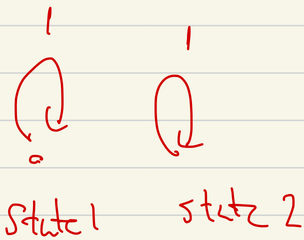
$$d(t) = \max_{\text{stoch } \vec{\mu}} \|\vec{\mu} P^t - \vec{\pi}\|_{TV}$$

$$= \max_{\text{stoch } \vec{\mu}} \left(\|\vec{\mu} P^t - \vec{\pi}\|_1 / 2 \right)$$

$$= \max_{i \in [n]} \|\vec{e}_i P^t - \vec{\pi}\|_{TV}$$

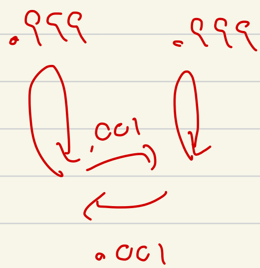
2 state examples:

extreme 1: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$



not irreducible

near extreme: $\begin{bmatrix} .999 & .001 \\ .001 & .999 \end{bmatrix}$



true $\pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

mixing occurs slowly...

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} : \quad a+b, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad a-b, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \qquad \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$= (a+b) \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} + (a-b) \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}^t = (a+b)^t \begin{bmatrix} \; \end{bmatrix} + (a-b)^t \begin{bmatrix} \; \end{bmatrix}$$

$$\left\| \vec{e}_1 P^t - \begin{bmatrix} \tfrac{1}{2} & \tfrac{1}{2} \end{bmatrix} \right\|_{TV}$$

$$P = \begin{bmatrix} .999 & .001 \\ .001 & .999 \end{bmatrix}$$

$a+b=1$

$a-b=.998$

$$= \| \qquad - \cancel{\begin{bmatrix} \tfrac{1}{2} & \tfrac{1}{2} \end{bmatrix}} \|, \quad \left(\tfrac{1}{2}\right)$$

$$\cancel{(a+b)^t \begin{bmatrix} 1/2 & 1/2 \end{bmatrix}} + (a-b)^t \begin{bmatrix} 1/2 & -1/2 \end{bmatrix}$$

1^t

$$= \left\| (a-b)^t \begin{bmatrix} \tfrac{1}{2} & -\tfrac{1}{2} \end{bmatrix} \right\|, \quad \left(\tfrac{1}{2}\right)$$

$$= |a-b|^t \underbrace{\left\| \begin{bmatrix} \tfrac{1}{2} & -\tfrac{1}{2} \end{bmatrix} \right\|}_{1}, \quad \underbrace{\left(\tfrac{1}{2}\right)}_{1}$$

$|a-b|^t$

$$d(t) = |a-b|^t \frac{1}{2} ,$$ Markov mat $\begin{pmatrix} a & b \\ b & a \end{pmatrix}$

here $a = .999$, $b = .001$

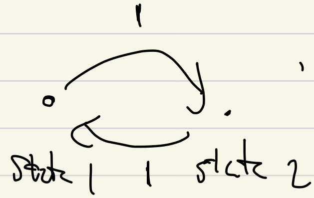
$$d(t) = (.998)^t \frac{1}{2}$$

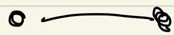
$$d(500) = \left(1 - \frac{1}{500}\right)^{500} \frac{1}{2}$$

$$\approx \frac{1}{e} \cdot \frac{1}{2} \approx \frac{1}{5.4\ldots}$$

$$d(500\,l) \approx \frac{1}{e^l} \cdot \frac{1}{2}$$

Extreme 2! $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$



$\left(= A_{B'} \right.$  $\left. \right)$

$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\lambda = a+b,\ a-b$

$= 1,\ -1$

$$\left\| \vec{e}_1 P^t - \pi \right\|_{TV} =$$

$$\left\| \begin{cases} \vec{e}_1 & t \text{ even} \\ \vec{e}_2 & t \text{ odd} \end{cases} - \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \right\|_{TV} \xrightarrow{\overbrace{t \to \infty}^{\text{any } t}} \frac{1}{2}$$

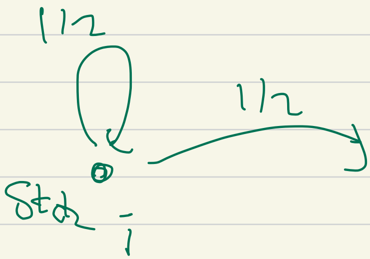
P aperiodic $\iff$ $P^t = \begin{bmatrix} - & \pi & - \\ - & \pi & - \\ & \vdots & \end{bmatrix}$

$\iff$

To make $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ not periodic,

make P "<u>lazy</u>"!

$$P \longrightarrow I\left(\tfrac{1}{2}\right) + P\left(\tfrac{1}{2}\right)$$



do what P would do

"$\frac{1}{10}$ lazy(P)" $= \frac{1}{10} I + \frac{9}{10} P$

Def If P is Markov matrix and $0 < \alpha < 1$, the **α-lazy version of P** is

$$I(\alpha) + P(1-\alpha).$$

Any lazy version is aperiodic

E.g. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ make it $\frac{1}{100}$ lazy!

$$P_{\text{lazy}, \frac{1}{100}} = \begin{bmatrix} .01 & .99 \\ .99 & .01 \end{bmatrix}$$

So for

$$P_{(c_2)} = \begin{bmatrix} .01 & .99 \\ .99 & .01 \end{bmatrix}, \quad \lambda = a+b, a-b$$

$$= 1, -.98$$

$$d(t) = |a-b|^t \frac{1}{2} = |-.98|^t \frac{1}{2}$$

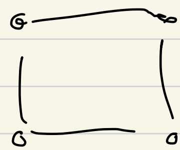
$$= (.98)^t \frac{1}{2}$$

$$= \left(1 - \frac{1}{50}\right)^t \frac{1}{2}$$

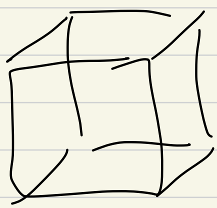
$$\approx \frac{1}{e^{t/50}} \frac{1}{2}$$

Hypercube!

$\mathbb{B}^1 \rightarrow \frac{A_{\mathbb{B}^1}}{1}$



$\mathbb{B}^2 \qquad \frac{A_{\mathbb{B}^2}}{2}$



$\mathbb{B}^3 \qquad \frac{A_{\mathbb{B}^3}}{3}$

etc.

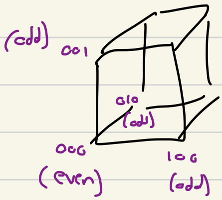
=

ε-mixing time, $t_{mix}(\varepsilon)$ is

$$\min\{t \in \mathbb{N} \mid d(t) \le \varepsilon\}$$

(e.g. $d(t) \approx \frac{1}{e^{t/50}} \frac{1}{2}$, $t_{mix} \approx \frac{\log_e 1/\varepsilon}{50}$)

Hypercube:



$\lambda = 3, \quad 1, \quad -1, \quad -3$

mult 1   mult 3   mult 1

d-regular

$d = 3$

$-3 = -d$

$\frac{1}{10}$ Lazy-hypercube

$$\frac{1}{10} \cdot I + \frac{9}{10} \left(\frac{A_{\mathbb{B}^3}}{3} \right)$$

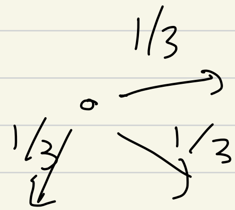
$$\lambda = \frac{1}{10} + \frac{9}{10} \left(\qquad \right)$$

$$= \frac{1}{10} + \frac{9}{10} \left\{ 1, \frac{1}{3}, -\frac{1}{3}, -1 \right\}$$

$$= \frac{1}{10} + \left\{ \frac{9}{10}, \frac{3}{10}, \frac{-3}{10}, \frac{-9}{10} \right\}$$

$$= \left\{ 1, \frac{4}{10}, \frac{-2}{10}, \frac{-8}{10} \right\}$$

$\frac{\mathbb{A}_{B_3}}{3}$:

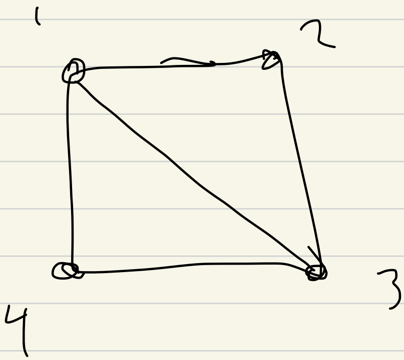


Markov

-1 moved into

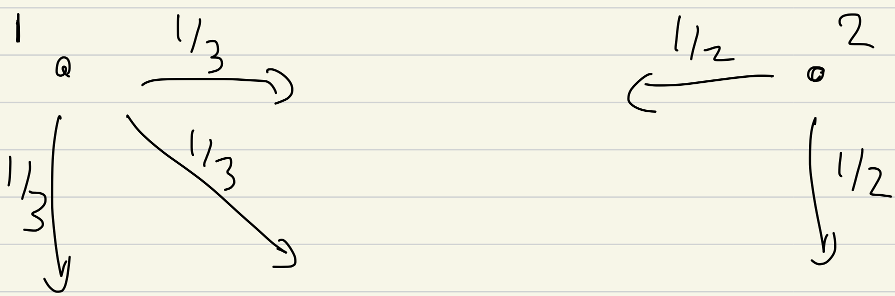
$(-1, 1)$

Take any graph!



G

Markov chain



Markov(G): $p_{ij} = \begin{cases} 1/d_i & \text{if } i \sim j \\ 0 & \text{if } i \not\sim j \end{cases}$

$i \sim j$ i shares an edge with j

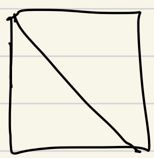
$i \not\sim j$ i ~~does not~~ doesn't.

If G is d-regular,

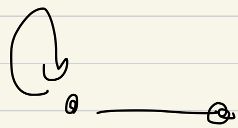
$$\mathrm{Markov}(G) = \frac{A_G}{d}$$

<u>Claim</u>: If G is any graph, then stationary distribution, π, of $P = \mathrm{Markov}(G)$ is given by

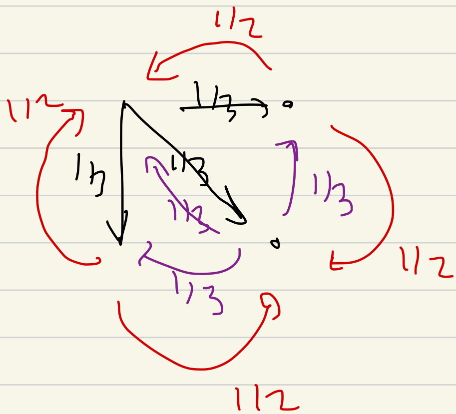
$$\pi_i = ???$$

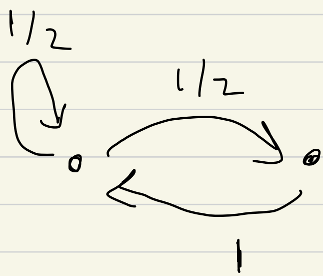


non-regular



Fibonacci graph





Is there a nice formula for π?

π_i = something simple ??

4-minute break

How about:

$$\pi_i = \frac{d_i}{d_1 + \cdots + d_n}$$

$$\vec{\pi} = \underbrace{\frac{\vec{d}}{\quad}}_{\text{normalize to make stochastic}} = \frac{\vec{d}}{\vec{1} \cdot \vec{d}} \;??$$

Prop: Say that P is an irreducible Markov matrix, say that $\vec{v} \in \mathbb{R}^n$ $(P \in M_n(\mathbb{R}))$ s.t. $\vec{v}$ has all positive components, and $\forall\, i,j \in [n]$

$$v_i P_{ij} = v_j P_{ji}$$

Then $$\vec{\pi} = \underset{\text{normalized}}{\vec{v}} = \frac{\vec{v}}{v_1 + \cdots + v_n}$$

is the stationary distribution of P,

<u>and</u>

$$\pi_i P_{ij} = \pi_j P_{ji}$$

so P is reversible.

Pf: $\vec{v} P = \left\{ \begin{matrix} \text{claim} \\ \text{HW} \end{matrix} \right\}$ is just $\vec{v}$

Assume prop: Markov(G)

$$\text{state } i,\ v_i \ \underset{1/d_j}{\overset{1/d_i}{\rightleftarrows}}\ \text{state } j$$

So if $\vec{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}$

if $i \sim j$

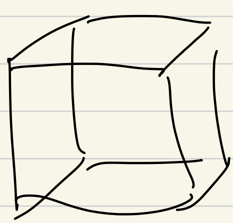
$$\begin{tikzcd} v_i \arrow[r, bend left, "1/d_i"] & v_j \arrow[l, bend left, "1/d_j"] \end{tikzcd}$$

$$d_i p_{ij} = d_i \frac{1}{d_i} = 1 = d_j p_{ji}$$

if $i \not\sim j$, $p_{ij} = 0 = p_{ji}$

Metropolis algorithm!

$\mathbb{B}_3$!



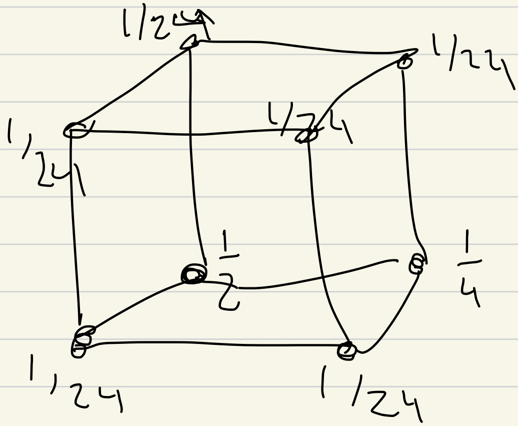
$\mathrm{Markov}(\mathbb{B}_3) = \frac{1}{3} A_G$

but we want to create version of

$P = \frac{1}{3} A_G$ with a given

stationary distrib, $\vec{\pi}$.

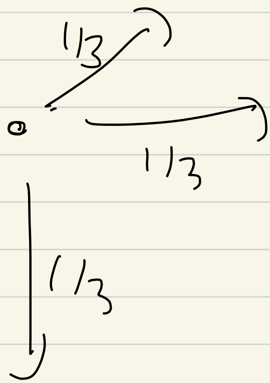
π:



$\mathbb{B}_{100}$ big

Metropolis chain:

$\mathbb{B}_3$

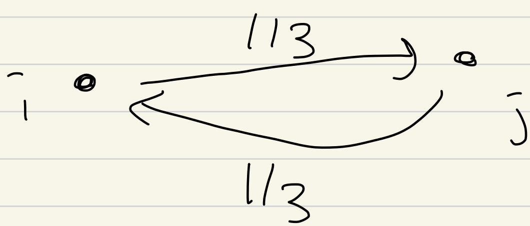


given π

d-regular

$\pi_i = \frac{1}{24}$ ——— $\frac{1}{24}$

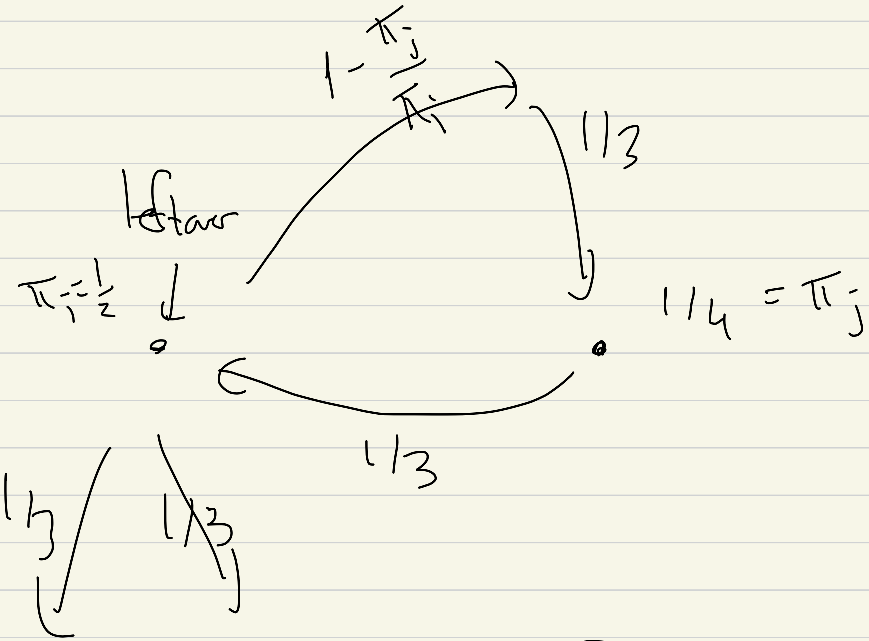
$\pi_i = \frac{1}{2}$ ——— $\pi_j = \frac{1}{4}$



rule: say $\pi_i \geq \pi_j$

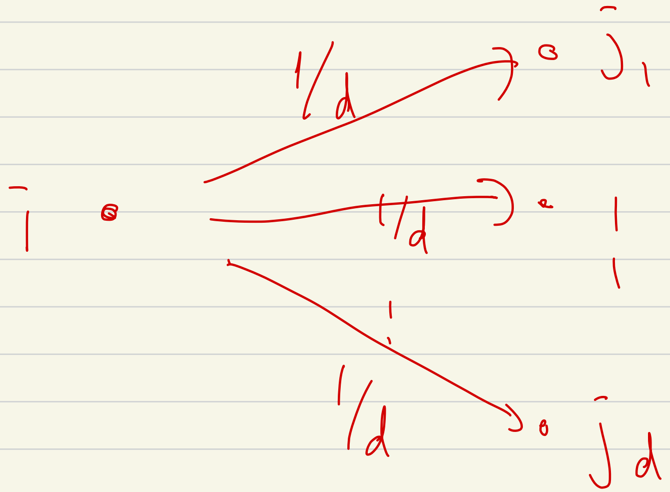
Set $p_{ji} = 1/d$

$$p_{ij} = \qquad \cdot \frac{\pi_j}{\pi_i} \quad \text{here} \quad \frac{1/4}{1/2} = \frac{1}{2}$$

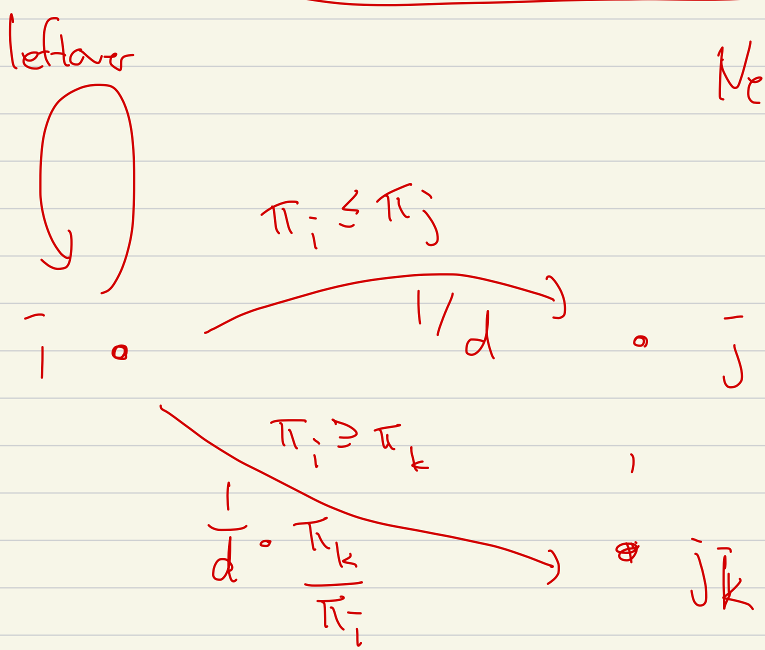


Claim: New chain, reversible, stationary dist $\vec{\pi}$

Old chain



New chain



Class ends ...

If Ψ is any reversible Markov mat,

such as $\Psi = \text{Markov}(G)$,

new Metropolis $i \neq j$

$$p_{ij} = \begin{cases} \psi_{ij} & \text{if } \pi_j \geq \pi_i \\ \psi_{ij} \dfrac{\pi_j}{\pi_i} & \pi_i \geq \pi_j \end{cases}$$

p_{ii} = leftover