6.5420 Lec 4

Markov Chains
+ Random walks

- Stationary Dist.
- Cover Times

Markov Chain

set of states : Ω

$x_1 \cdots x_t \in \Omega^t$: sequence of visited states

Markovian Property :

$$\mathbb{P}[X_{t+1} = y \mid X_0 = x_0, X_1 = x_1, \cdots, X_t = x_t]$$

$$= \mathbb{P}[X_{t+1} = y \mid X_t = x_t]$$

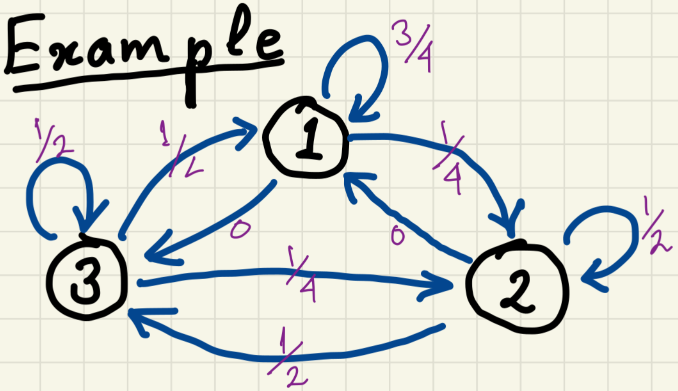
Only current state matters
NOT how we get there

Transitions independent of time

def: $P(x,y) = \mathbb{P}[X_{t+1} = y \mid X_t = x]$

Represent w/ "transition matrix"

Example



$$P = \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 3/4 & 1/4 & 0 \\ 2 & 0 & 1/2 & 1/2 \\ 3 & 1/4 & 1/4 & 1/2 \end{array}$$

Important special case:

Transition to uniformly random neighbor

def: Random Walk on $G=(V,E)$ is a sequence $S_0 S_1 \cdots$ of nodes

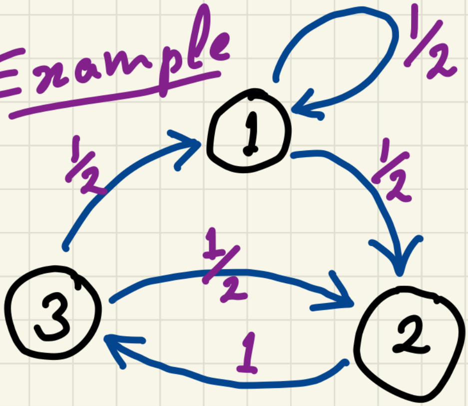
S_0 ← start node

S_{i+1} chosen uniformly from $\underbrace{N(S_i)}_{\text{outedges}}$

Let d_v = # outedges of v

$$P(x,y) = \begin{cases} \dfrac{1}{d_x} & \text{if } (x,y) \in E \\ 0 & \text{o.w} \end{cases}$$

Example

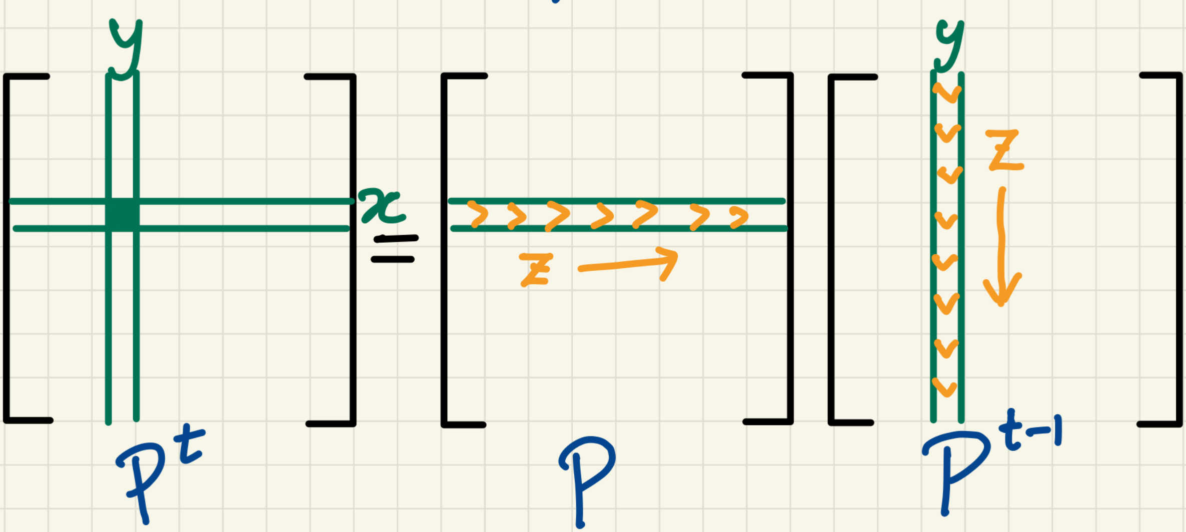


$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

Distribution after t steps

Recursively

$$P^t(x,y) = \begin{cases} P(x,y) & \text{if } t=1 \\ \sum_z P(x,z)\,P^{t-1}(z,y) & \text{if } t>1 \end{cases}$$



Initial dist. $\pi^{(0)} = \pi_1^{(0)} \pi_2^{(0)} \cdots$

$$\pi^{(0)} \xrightarrow{\text{one step}} \pi^{(1)} \longrightarrow \pi^{(2)} \longrightarrow \cdots$$

$\pi^{(1)} = \pi^{(0)} P$

$\pi^{(2)} = \pi^{(1)} P = \pi^{(0)} P^2$

t-step distribution : $\pi^{(0)} P^t$

Does this converge ?

Properties

Irreducible (strongly connected)

$$\forall x,y \quad \exists t(x,y) \text{ s.t } P^{t(x,y)}_{(x,y)} > 0$$

Aperiodic : $\forall x \ \gcd\{t : P^t(x,x) > 0\} = 1$

(gcd of possible cycle lengths = 1)

Ergodic : $\exists t^* \text{ s.t. } \forall t > t^* \ P^t(x,y) > 0$

Ergodic $\Longleftrightarrow$ Irreducible + Aperiodic

Stationary Distribution

π s.t $\forall_x \; \pi(x) = \sum_y \pi(y) P(y,x)$

or $\pi = \pi P$

(consider P s.t. π^* exists + unique
i.e does not depend on $\pi^{(0)}$)

Periodic  | Reducible 

$(0,1) \to (1,0) \to (0,1) \to \cdots$ | $(\frac{1}{2}\ \frac{1}{2})$ $(0,1)$ $(1,0)$

Thm: Irreducible M.C. $\Rightarrow$ Unique π^*

Undirected Graph $G=(V,E)$

$$\pi^* = \left(\frac{d_{v_1}}{2|E|}, \frac{d_{v_2}}{2|E|}, \cdots \right)$$

- π^* uniform for d-reg graphs

  Also for digraphs when indeg = outdeg = d
- Not true for general digraphs

Hitting Time

def: $h_{xy} = \mathbb{E}[\#\text{ steps to go } x \rightsquigarrow y]$

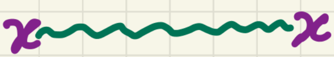
h_{xx} : Recurrence time

Thm: $h_{xx} = \dfrac{1}{\pi^*(x)}$

Pf Consider a very long walk

$\pi^*(x)$ fraction of the positions are x

$\Rightarrow$ Averge gap between occurences



$h_{xx} = \pi^*(x)^{-1}$

Cover Time

$C_v(G) = \mathbb{E}[\#\text{ steps to visit all nodes in } G \text{ starting at } v]$

$C(G) = \max_v C_v(G)$

Cover Time Examples

- $C(K_n)$ K_n is the complete graph on n vertices w/ self loops at each node

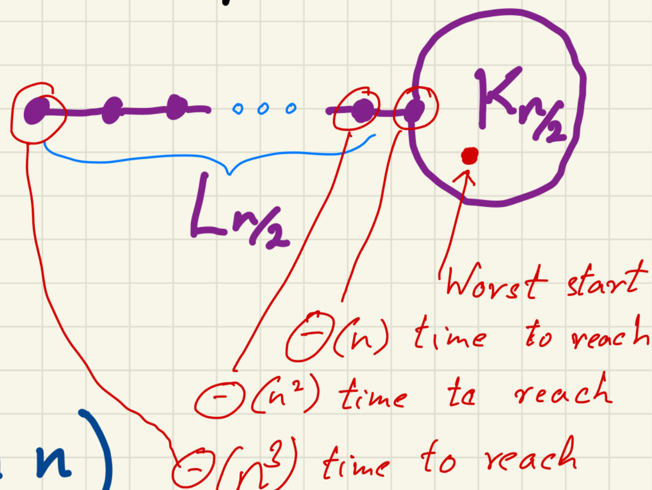
  $= \Theta(n \log n)$ ← coupon collector

- $C(L_n)$ L_n is the line graph w/ self loops at each node

  $= \Theta(n^2)$

- $C(\text{lollipop})$

  $= \Theta(n^3)$



Thm:

$C(G) \le O(m\,n)$ for undirected connected G

Pf.

Commute Time

def $C_{xy} = \mathbb{E}[\#\text{ steps for } x \rightsquigarrow y \rightsquigarrow x]$

$= h_{xy} + h_{yx}$ (linearity of expectation)

Lemma: $\forall (x,y) \in E \quad C_{xy} \le O(m)$

pf Consider a long walk

$u_1, u_2, u_3, \ldots$

where $u_i \in V$ and $(u_i\ u_{i+1}) \in E\ \forall i$

We look for commutes of the following form $\underbrace{\boxed{x \to y} \to \circ \circ \circ \to \boxed{x \to y}}_{x \rightsquigarrow y \rightsquigarrow x \text{ commute}}$

Prob of finding (x, y)

$$\mathbb{P}[(u_i, u_{i+1}) = (x, y)]$$

$$= \mathbb{P}[u_i = x] \cdot \mathbb{P}[u_{i+1} = y \mid u_i = x]$$

$$= \pi^*(x) \cdot \frac{1}{d_x}$$

$$= \frac{d_x}{2m} \cdot \frac{1}{d_x} = \boxed{\frac{1}{2m}}$$

$$\sim x-y \sim\sim\sim\sim x-y \sim\sim\sim x-y \sim$$

$\therefore$ $\frac{1}{2m}$ fraction of the edges are $x-y$

So, expected gap between consecutive occurences of $x-y$ is $2m$

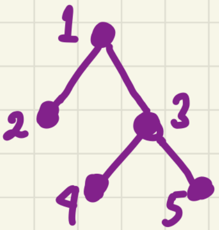
$\therefore$ $C_{xy} \leq O(m)$

$$\boxed{x-y \sim\sim\sim\sim\sim\sim x}-y$$

Finally, consider $T \subseteq G'$
where T is a spanning tree (n-1 edges)

$$\underbrace{v_0 \; v_1 \; v_2 \; \circ \circ \circ \; v_{2n-2}}_{}$$

DFS traversal of T



$\Rightarrow$ 1 2 1 3 4 3 5 3 1

Each edge (u,v) appears twice, as (u,v) & (v,u)

Using the DFS traversal sequence

$$C(G) \leq \sum_{j=0}^{2n-3} h_{v_j v_{j+1}}$$

$$= \sum_{(u,v)\in T} C_{u,v} \quad \left(C_{uv} = h_{uv} + h_{vu}\right)$$

$$= \sum_{(u,v \in T)} O(m) = O(mn)$$