Today

Random walks

Stationary Distributions

Cover Times

UST Conn

Random walks

Markov chains:

Ω = set of "states" (or nodes) (here always FINITE)

$X_0 \dots X_t \in \Omega$ sequence of visited states

Markovian property:

$$\Pr[X_{t+1} = y \mid X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_t = x_t]$$
$$= \Pr[X_{t+1} = y \mid X_t = x_t]$$

Next step depends only on where you are. Not how you got there.

Wlog, assume transitions independent of time:

i.e. $P(x,y) = \Pr[X_{t+1} = y \mid X_t = x]$

so can use "transition matrix" to represent it

Important special case:

transitions uniform on subset corresponding to neighbors of node

def. random walk on $G = (V, E)$

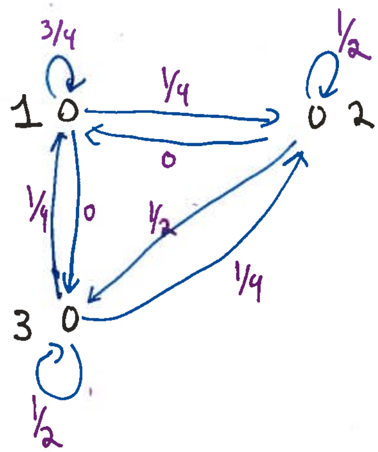
is a sequence $S_0 S_1 \dots$ of nodes

where S_0 is a start node.

At each step i, S_{i+1} picked uniformly from $\underbrace{N(S_i)}_{\text{outedges}}$

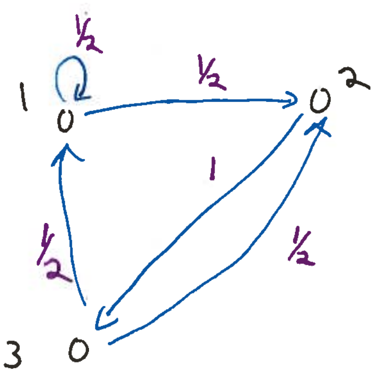
examples

Markov chain



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

random walk on digraph



$$
p: \begin{array}{c|ccc} & 1 & 2 & 3 \\ \hline 1 & 1/2 & 1/2 & 0 \\ 2 & 0 & 0 & 1 \\ 3 & 1/2 & 1/2 & 0 \end{array}
$$

$d(i) = \#$ outedges of node i

$$
p(i,j) = \begin{cases} \frac{1}{d(i)} & \text{if } (i,j) \in E \\ 0 & \text{o.w.} \end{cases}
$$

$$
\forall i \quad \sum_j P(i,j) = 1
$$

Distributions after t steps

Transition probabilities for t steps: $P^t(x,y) = \begin{cases} P(x,y) & t=1 \\ \sum_z P(x,z)P^{t-1}(z,y) & t>1 \end{cases}$ matrix multiplication $P^t = \underbrace{P \times P \times \dots \times P}_{t \text{ times}}$

Initial distribution: $\pi^{(0)} = (\pi_1^{(0)} \dots \pi_n^{(0)})$ where $\pi_i^0 = \Pr[\text{start at node } i]$

distribution after one step:

$$\pi^{(1)} = \pi^{(0)} \cdot P = \left(\sum_z P(z,1) \cdot \pi(z), \sum_z P(z,2)\pi(z), \dots \right)$$

⋮

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

Finite Markov Chain Properties

stochastic matrix: rows of P sum to 1

all M.C.'s have this property

doubly stochastic matrix: rows & columns sum to 1

not even all interesting M.C.'s satisfy this

e.g. random walk on undirected graph or digraph in which indegree = outdegree = const for all nodes

irreducible: ("strongly connected")

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

↕ change of quantifier order

ergodic: $\exists t_0$ s.t. $\forall t > t_0 \; \forall x,y \; P^t(x,y) > 0$ ← stronger than irreducible! why?

Aperiodic:

gcd of "possible" cycle length = 1

$\forall x \quad \gcd \{t : P^t(x,x) > 0\} = 1$ not bipartite, k-partite...

<u>Thm</u> Ergodic $\Leftrightarrow$ Irreducible + Aperiodic

Stationary Distributions

does it depend on π_0? { stationary distribution π } so $\pi^{(t)} = \pi^{(t-1)}$

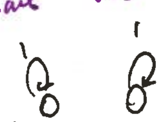
$\forall y \quad \pi(y) = \sum_x \pi(x) P(x,y)$

Will consider P s.t. π is unique & exists } i.e. doesn't depend on π_0

if periodic: could have no stat. dist. or several

if $\pi_0 = (0,1)$ then $\pi_{2i} = (0,1)$, $\pi_{2i+1} = (1,0)$

if reducible: could have lots of stat. dist.



Some stat dist's: $(\frac{1}{2}, \frac{1}{2})$ $(0,1)$ $(1,0)$...

<u>Important Thm</u> every ergodic M.C. has unique stationary distribution

Stationary dist. of undirected graph:

$$\pi = \left(\frac{\deg(x_1)}{2|E|}, \frac{\deg(x_2)}{2|E|}, \dots \right)$$

- so d-regular graphs have π = uniform (also indegree = outdegree = d digraphs & $\underbrace{\text{doubly stochastic}}_{\text{this implies the others!}}$ P M.C.'s)
- not true in general for digraphs
- bipartite, periodic graphs may have other stat. dists.

Hitting times

h_{ii} = E [time starting at i to return to i]

$= \frac{1}{\pi(i)}$ ← Very useful theorem!

h_{ij} = E [time starting at i to reach j]

Cover time of undirected graph

$C_u(G)$ = E [# steps to reach all nodes in G on walk starting from u]

$$C(G) = \max_u C_u(G)$$

Cover time Examples:

- $\mathcal{C}(K_n^*)$ where K_n^* = complete graph with self-loops at each node } so aperiodic

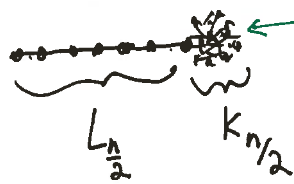
  $= \Theta(n \ln n)$ by coupon collector argument

- $\mathcal{C}(L_n)$ where L_n = n node line

  $= \Theta(n^2)$

- $\mathcal{C}$(lollipop)

  $= \Theta(n^3)$



← worst place to start

Thm $\mathcal{C}(G) \le 8m(n-1)$

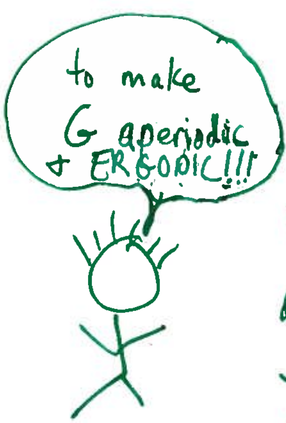
Proof

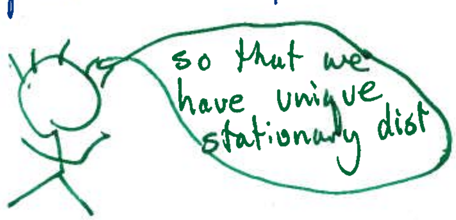
First - transform G (undirected) into G' (directed) (see example on pg 8)

to make G aperiodic, add d_u self loops to each u (ie. take self-loop with prob $1/2$)

Claim: $\mathcal{C}(G') = 2\,\mathcal{C}(G)$

why? can transform paths in G' by removing self-loops, expected # self-loops = $1/2$ (length of path)





Next, commute times + a lemma:

<u>def.</u> $C_{ij} = E$[#steps for r.w. starting at i to hit j + return to i] "commute time"

<u>Claim</u> $C_{ij} = h_{ij} + h_{ji}$ (linearity of expectation)

<u>Lemma</u> $\forall (u,v) \in E \quad C_{uv} \leq O(m)$

<u>Pf of lemma</u>

Key idea: (actually will show $C_{vu} \leq O(m)$ but it's symmetric)

if traverse (u,v) twice

$(u,v) \ldots (u,v)$

have performed commute from $v \rightsquigarrow u \rightsquigarrow v$

Plan: show E[time between visits to (u,v)] is $O(m)$

$\Longrightarrow C_{uv}$ is $O(m)$

Given $G' = (V, E)$ (G with added self loops)

Construct (directed) G'' representing walks on directed edges of G'

line graph

$E \to V''$ new nodes (u,v) are edges (u,v) in G'

$\underbrace{(u,v)(v,w)}_{\text{consecutive edges}} \to E''$ new edges are length 2 paths in G'

visit edge in G' twice $\Leftrightarrow$ visit node in G'' twice

example

G

$\Longrightarrow$

G'

$1 \to 2 \to 1$

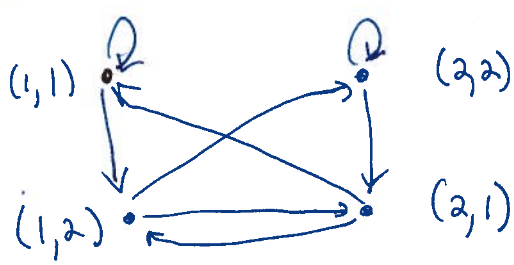
$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 0 & 1 \\ 2 & 1 & 0 \end{array}$$

$1 \to 1 \to 2 \to 2 \to 2 \to 1$

$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1/2 & 1/2 \\ 2 & 1/2 & 1/2 \end{array}$$

$\Downarrow$

G''

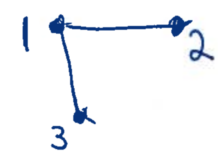


	(1,1)	(1,2)	(2,2)	(2,1)
(1,1)	1/2	1/2	0	0
(1,2)	0	0	1/2	1/2
(2,2)	0	0	1/2	1/2
(2,1)	1/2	1/2	0	0

(more complicated example) rw 8

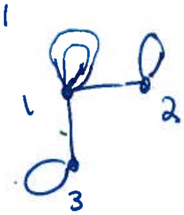
example

G



$\Rightarrow$

G'



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

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

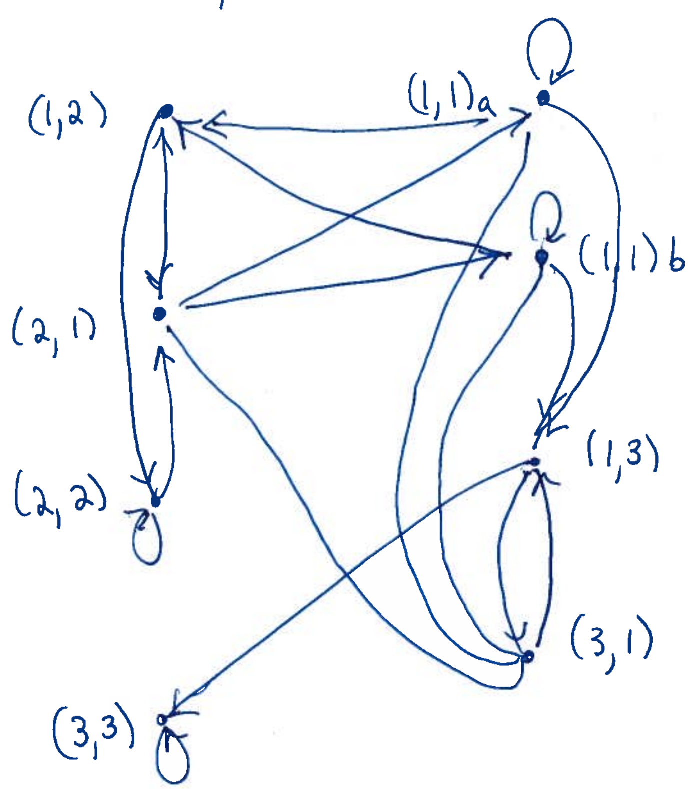
$1 \to 2 \to 1$

$1 \to 1 \to 2 \to 1$

$\Downarrow$

$(1,1) \to (1,2) \to (2,1)$

G''



Note: G'' is doubly stochastic:

$$Q_{(u,v)(v,w)} = P_{vw} = \frac{1}{d(v)} \quad \text{if} \quad (u,v)\,(v,w) \in E$$

$$\forall\ (v,w) \in \underset{V''}{\overset{\shortparallel}{E}} \qquad \underbrace{\sum_{\substack{(u,v) \text{ s.t.} \\ (u,v)(v,w) \in E'}}}_{\text{column sum}} Q_{(u,v)(v,w)} = \sum_{(u,v)\in E} \frac{1}{d(v)} = 1$$

$\therefore\ \pi$ of G'' is uniform

we need that walk on G'' is ergodic. Irreducible follows from G' irreducible. Aperiodic comes from self-loops.

so $\pi_u = \frac{1}{|V''|} = \frac{1}{4m}$

u: edge in G'

$4m$: # edges in G' ($(u,w) \to (u,v), (v,u)$ + 2 self loops)

$$h_{uu} = \frac{1}{\pi_u} = 4m \qquad \text{for all nodes } u \text{ in } G''$$

uu: edge in G'

$u \shortparallel (a,b)$ in G'

$\blacksquare$

if start at v conditioned on coming from edge (u,v) (proof of lemma) expect $\leq 4m$ steps to see (u,v) again.

But, its an M.C! so conditioning doesn't affect.

$\Rightarrow$ if start at v, expect to see (u,v) in $\leq 4m$ steps.

$\Rightarrow$ $C(v,u) = C(u,v) = O(m)$

Note: valid only for $(u,v) \in E$

$\blacksquare$

Wrapping it up:

Lemma $C(G) = O(nm) = O(n^3)$

Pf.

start vertex V_0

$T \leftarrow$ spanning tree rooted at V_0

edges in $T = n-1$

$V_0\ V_1 \cdots V_{2n-2}$ is depth 1st traversal st. each edge appears twice, once in each direction (a,b) + (b,a)

$V_{2n-2} = V_0$

$$C(G) \le \sum_{j=0}^{2n-3} h_{V_j V_{j+1}}$$

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

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

$$= O(nm)$$

■

S-T connectivity (UST-Conn)

Input: Undirected G, nodes s,t

Output: "Yes" if s & t connected

"No" o.w.

Can solve in poly time, in many ways.

What about small space?

RL = class of problems solvable by randomized log-space computations

[no charge for input space (read only), but can only have const # ptrs ...]

Thm UST-Conn $\in$ RL

Algorithm:

start at s

take random walk for n^3 steps

if ever see t, output "Yes"

o.w. output "No"

Complexity:

Keep track of # steps so far

edges at each node & toss coin to pick one randomly

logspace

Behavior:

If s,t not connected, never output "yes"

If s,t connected

$$h_{st} \le C_s(G_s) \le n^3$$

connected component of s

$$\Pr[\text{output "no"}] = \Pr[\text{start at } s, \text{ walk} \ge c \cdot E[C_s(G_s)] \text{ steps} + \text{don't see } t]$$

$$\le \frac{1}{c} \quad \text{by Markov's} \neq$$

Comments

- Actually USTCONN $\in L$!!!
- Open is $RL = L$?

  we know $RL \in L^{3/2}$