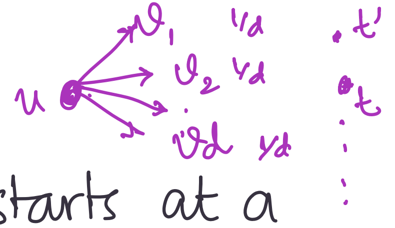
Random Walks and Markov chains.



Given a graph $G=(V,E)$, a random walk starts at a node and at each step moves to a uniformly random neighbour.

(Stationary distr.)

(mixing time)

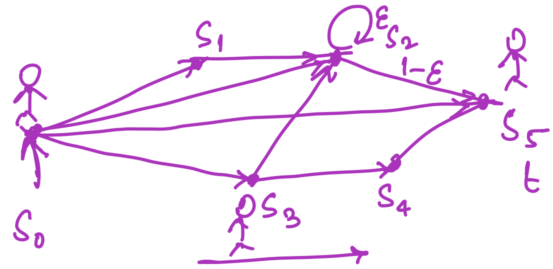
Qn: 1. What is the limiting distribution?

2. How many steps before reaching limiting distribution

3. Starting from node s, how long does it take to reach t? ← (hitting time)

4. How long does it take to visit each node at least once?

(cover time).



Markov chain

↳ More generalized random walk on a directed graph.

- $G=(V,E)$ $\quad |V|=n$ and $|E|=m$.
- Transition probability matrix $P_{n\times n}$

P_{ij} = prob that j is the next state given i is the current state.

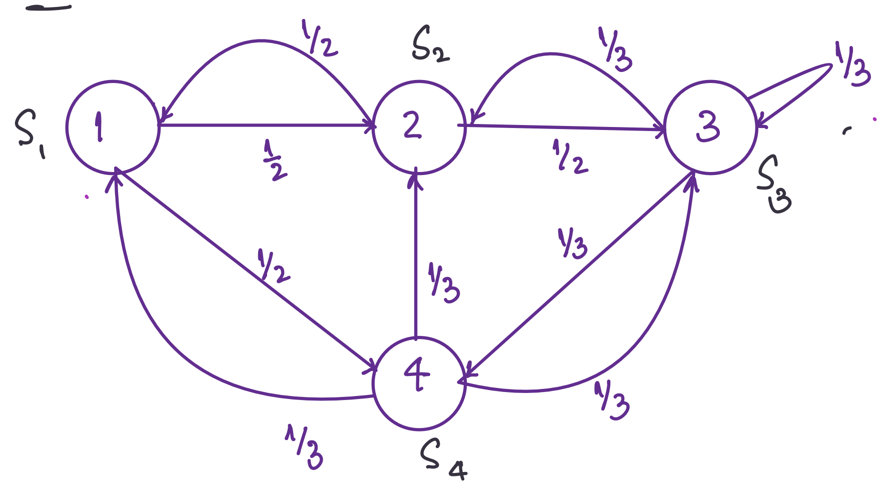
- X_t : State of the Markov chain at time t.

$$\Pr[X_t=a_t \mid X_{t-1}=a_{t-1},\dots,X_0=a_0] = \Pr[X_t=a_t \mid X_{t-1}=a_{t-1}] = P_{a_{t-1},a_t}$$

"Future behaviour of Markov chain depends only on its current state".

Ex:

$(0, 1/2, 0, 1/2) \leftarrow P_0 = (1, 0, 0, 0)$



$$P = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ \frac{1}{2} & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix}$$

$P_1 = P_0 \cdot P \cdot \underline{(1,0,0,0)}[P]$

Let $p_i(t)$ be the prob of being at state s_i at time t.

$P_t = (P_1(t), P_2(t), \dots, P_n(t))$ // If we start randomly, $P_0 = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$.

Else if we start from node 1, $P_0 = (1, 0, \dots, 0)$.

Observation: $P_{t+1} = P_t \cdot P$.

$$(p_t \cdot P)_j = \sum_{i=1}^{n} p_t(i) \cdot P_{ij}.$$

$$\Rightarrow \quad P_{t+m} = P_t \cdot P^m.$$

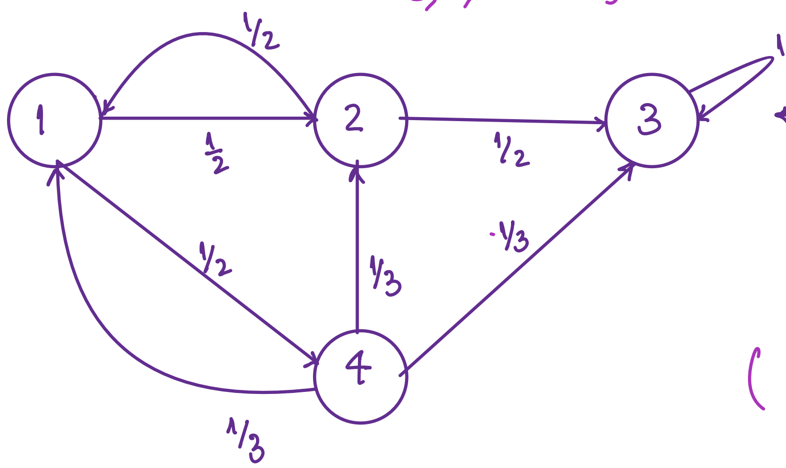
$$p_t = p_0 \cdot P^t$$

We call the vertices/nodes in a Markov Chain as states.

Defn: A MC is irreducible if G is strongly connected.
(That is $\forall s_i \neq s_j$, $\exists$ a directed path from $s_i \rightsquigarrow s_j$).

$$\forall i \neq j, \ \Pr[\exists m > 0, X_{t+m} = s_j \mid X_t = s_i] > 0.$$

$\gcd\left\{\begin{matrix}2, 4, \dots, \\ 3, 6, \dots\end{matrix}\right\} = \underline{1}$ $\qquad$ $\underline{1, 2, 3, \dots}, \dots$
$= 1$



← Not irreducible.

2, 4, 3

$(P^t)_{i,i} :=$ Prob of reaching i after t time steps >0 starting from i

Defn: Period of a state s_i in MC is defined as

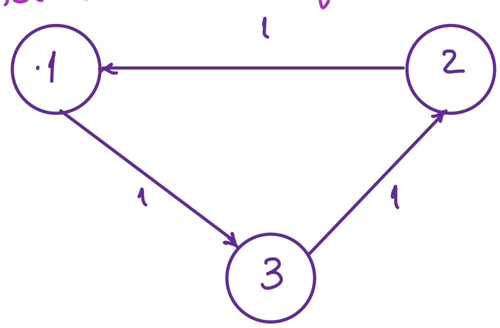
$$d(s_i) = \gcd\left(\{\underline{t \in \mathbb{N}} \mid (P^t)_{ii} > 0, t > 0\}\right).$$

In other words, $d(s_i) = T$ then if at time t, $P_t(i) > 0$ then t belongs to an arithmetic progression

$$\{a + i \cdot T \mid i \geq 0\}$$

positive int that depends on initial distr.

MC is said to be aperiodic if $d(s_i) = 1$ for all states $s_i \in \{s_1, \dots, s_n\}$.



Periodic w/ period 3 for all states.

- Let $P_t(i,j) = \Pr[X_t = s_j \mid X_0 = s_i]$.
- Let $r_t(i,j) = \Pr[X_t = s_j \text{ and } \forall\ 1 \leq u \leq t-1,\ X_u \neq s_j \mid X_0 = s_i]$

"First visit to s_j starting from s_i"

- Starting from s_i, prob of a transition into state s_j is given by

$$f_{ij} = \sum_{t>0} r_t(i,j).$$

- Expected no. of time steps to reach s_j starting from s_i:

$$h_{ij} = \sum_{t>0} t \cdot r_t(i,j)$$

Theorem: If a MC is finite, irreducible and aperiodic then $\exists T < \infty$ s.t $(P^t)_{ij} > 0 \quad \forall t \geq T, \forall i,j \in V.$

$P_{t+1} = P_t$

Defn: A stationary distribution for MC w/ transition matrix P is a probability distribution π s.t

$$\pi = \pi P.$$

$\exists T$ s.t $\forall t \geq T$

$P_{t+1} = P_t.$

Fundamental theorem of Markov chains:

For any finite, irreducible and aperiodic MC,

1. There exists a unique stationary distribution π s.t for all $1 \leq i \leq n$, $\pi_i > 0$.

2. For $1 \leq i \leq n$, $f_{ii} = 1$ and $h_{ii} = 1/\pi_i$.

3. Let $N(i,t)$ be the no. of times MC visits state i in t steps. Then $\lim_{t \to \infty} \frac{N(i,t)}{t} = \pi_i$.

Revisiting random walks on graphs:

- $G = (V, E)$
- non-bipartite
- undirected
- Connected

$$P_{u,v} = \begin{cases} \frac{1}{\deg(u)} & \text{if } (u,v) \in E, \\ 0 & \text{o/w.} \end{cases}$$

$|V| = n$ and $|E| = m$.

$\Rightarrow$ directed version of G
$\vec{G}$: it is irreducible, finite.

Claim: G is aperiodic.

Pf: • Every vertex has a closed walk of even lengths given by traversing the same edge twice in succession.

G is non-bipartite $\Rightarrow$ odd cycles. $\Rightarrow$ GCD of closed walks $= 1$.

$\Rightarrow$ GCDs of closed walks $= 1$ for every vertex.

a, $b =$ odd, $(a+b+a) =$ odd.

If G is non-bipartite $\Rightarrow \exists$ a closed walk of odd length for every vertex.

Lemma: For all $v \in V$, $\pi_v = \frac{\deg(v)}{2m}$.

Pf: π is stationary if $\underset{1\times n}{\pi} = \underbrace{\underset{1\times n}{\pi}\ \underset{n\times n}{P}}_{1\times n}$. That is for all $v \in V$, $\pi_v = (\pi P)_v = \sum_u \pi_u \cdot P_{uv}$.

$|V| = n$

$\pi = (\pi_1, \ldots \pi_v \ldots \pi_n)$

System of n linear equations in n. vars.

Claim: $\pi_u = \frac{\deg(u)}{2m}$ is a solution.

$m = \#$ Edges.

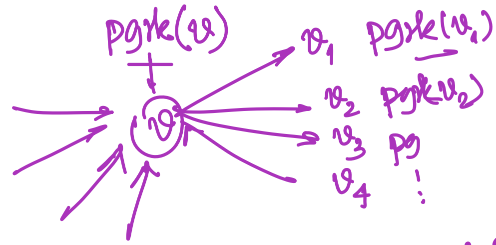
$\boxed{\sum_u \pi_u = 1}$

$\sum \frac{\deg(u)}{2m}$

$$\sum_u \pi_u \cdot P_{uv} = \sum_{\substack{u \\ (u,v)\in E}} \frac{\deg(u)}{2m} \times \frac{1}{\deg(u)} = \frac{\deg(v)}{2m} = \pi_v$$

Given by no. of non zero entries in vth col of P.

From Fundamental Theorem, there is a unique solution to $\pi = \pi P$. Thus, $\pi_v = \frac{\deg(v)}{2m}$ is a stationary distr.



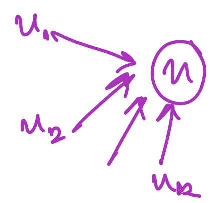
(Simple) Pagerank algorithm:

- Assign a pagerank of $\frac{1}{n}$ to each node.

$\frac{\text{pgrk}(v)}{\deg(v)}$

- For each node, divide the current page rank value by the out degree of the node and send to each "neighbour", out-link equally.

- Pagerank of each node is updated to sum of all values recieved.



$\sum \frac{\text{pgrk}(u_i)}{\deg(u_i)}$

Can be modeled as a MC.

$$P_{ij} = \begin{cases} \frac{1}{d_{out}(i)} & \text{if page } i \text{ links to } j. \\ 0 & \text{otherwise.} \end{cases}$$

$$\text{pagerank}^{(t)}(j) = \sum_{i:(i,j)\in E} \frac{\text{pagerank}^{(t-1)}(i)}{d_{out}(i)}.$$

// Apply the fundamental theorem if the web is finite, irreducible and aperiodic.

$$(P_1^{(0)}, \dots, P_n^{(0)}) \cdot P$$

$$'' \quad (P_1^{(1)}, \dots, P_n^{(1)})$$

$$P_j^{(1)} = \sum_i P_i^{(0)} \cdot P_{ij} = \sum_i P_i^{(0)} \cdot \frac{1}{d_{out}(i)} = \sum \frac{\text{pagerk}^{(0)}(i)}{d_{out}(i)}$$

$$\Pi_{1\times n} = \Pi_{1\times n} P_{n\times n}$$

$$(\pi_1, \dots, \pi_n) = (\pi_1, \dots, \pi_n)\begin{bmatrix} & & \\ & P_{ij} & \\ & & \end{bmatrix}$$

$$(\pi_1, \dots, \pi_n) = \left(\sum_{i=1}^{n} \pi_i \cdot P_{i1}, \sum_{i=1}^{n} \pi_i \cdot P_{i2}, \dots \right)$$

$$\left.\pi_j = \sum_{i=1}^{n} \pi_i \cdot P_{ij} \quad \forall j \in [n] \right\} S$$

Stationary distr. is a solution to the system of equations S.