Seoul
National
University

Markov Chains & Random Walks

Name: Chong-kwon Kim

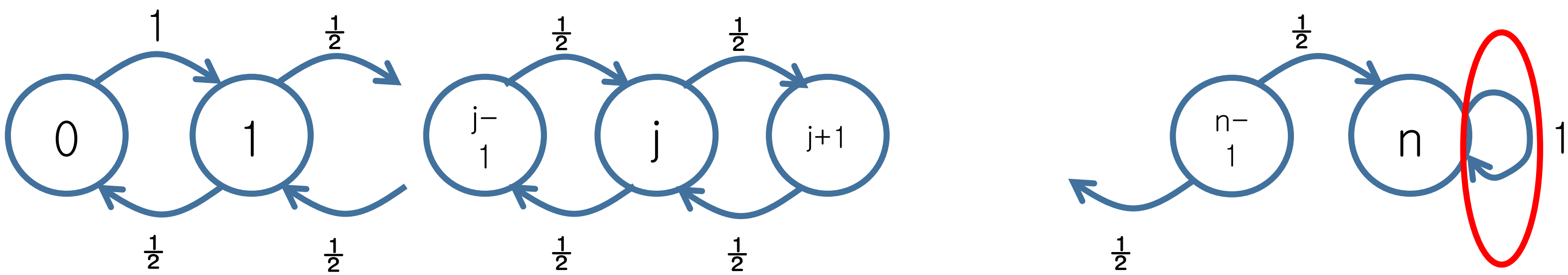
SCONE
Lab.

Stationary Distribution

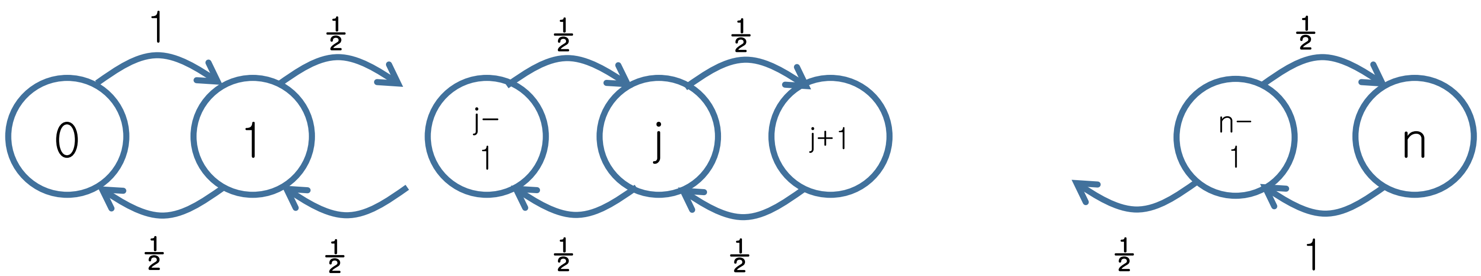
- Previously, we consider short-term behavior of an MC
- To examine the performance of systems, we need to analyze long-term behaviors
  - ➔ What is the expected wait time in a bank?
  - ➔ Probability that Kia Tigers tickets are sold out
- We need to know state probabilities many transitions after
- Some Markov Chains converge to a certain unique state probability distribution as many transitions occur
  - For all i and j, $\lim_{n\to\infty} P_{i,j}^{n} \to p_j$
- Some do not converge
- What properties make an MC to converge to a stationary distribution?

Example

- Revisit the 2-SAT problem

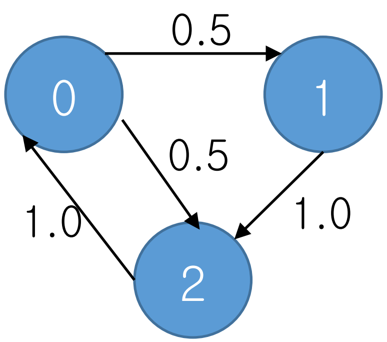


- Let Pj be the stationary probability of state j after ∞ transitions
  - If satisfiable, then Pn = 1
- Modified MC



Stationary Distribution

- Consider an MC

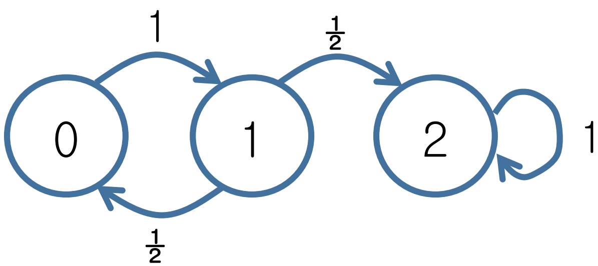


$$P=\begin{bmatrix}0 & 0.5 & 0.5\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}\quad P^2=\begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{2}\\ 1 & 0 & 0\\ 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}\quad P^4=\begin{bmatrix}\frac{1}{4} & \frac{1}{4} & \frac{1}{2}\\ \frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4}\end{bmatrix}\quad P^8=\begin{bmatrix}\frac{7}{16} & \frac{3}{16} & \frac{6}{16}\\ \frac{6}{16} & \frac{4}{16} & \frac{6}{16}\\ \frac{6}{16} & \frac{3}{16} & \frac{7}{16}\end{bmatrix}$$

$$P^\infty=\begin{bmatrix}0.4 & 0.2 & 0.4\\ 0.4 & 0.2 & 0.4\\ 0.4 & 0.2 & 0.4\end{bmatrix}$$

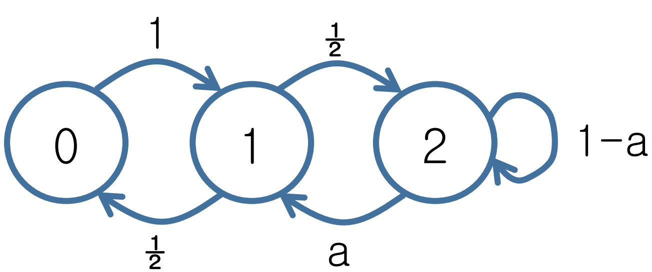
- State probabilities do not change once it converges to a stationary distribution
  - At time t, $\bar{p}(t) = (0.4, 0.2, 0.4)$
    From $\bar{p}(t+1) = \bar{p}(t)\cdot P$, $\bar{p}(t+1) = (0.4, 0.2, 0.4)$ also
- Also, an MC converge to the same stationary distribution regardless of the initial state
  - Let $\bar{p}(0) = (p_o, p_1, p_2)$
  - $\bar{p}(\infty) = \bar{p}(0)\cdot P^\infty = (0.4, 0.2, 0.4)$

Properties of MC



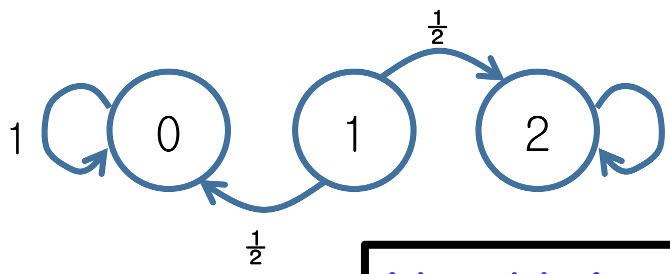
2-SAT with 2 variables

$$P=\begin{bmatrix}0 & 1 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 1\end{bmatrix} \quad P^2=\begin{bmatrix}\frac{1}{2} & 1 & \frac{1}{2}\\ 0 & \frac{1}{2} & \frac{1}{2}\\ 0 & 0 & 1\end{bmatrix} \quad P^4=\begin{bmatrix}1/4 & 0 & 3/4\\ 0 & 1/4 & 3/4\\ 0 & 0 & 1\end{bmatrix} \quad P^n=\begin{bmatrix}0 & 0 & 1\\ 0 & 0 & 1\\ 0 & 0 & 1\end{bmatrix} \text{ as } n \to \infty$$



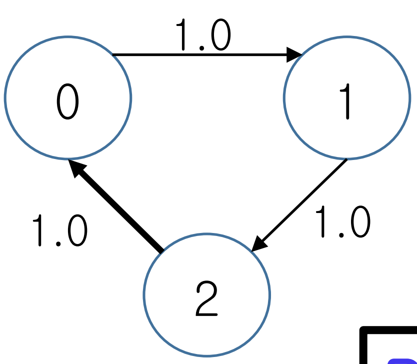
$$P=\begin{bmatrix}0 & 1 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & a & 1-a\end{bmatrix} \quad P^n=\begin{bmatrix}a/(3a+1) & 2a/(3a+1) & 1/(3a+1)\\ a/(3a+1) & 2a/(3a+1) & 1/(3a+1)\\ a/(3a+1) & 2a/(3a+1) & 1/(3a+1)\end{bmatrix}$$

Properties of MC



$$P=\begin{bmatrix}1 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 1\end{bmatrix} \quad P^2=\begin{bmatrix}1 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 1\end{bmatrix} \quad P^\infty=\begin{bmatrix}1 & 0 & 0\\ \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 1\end{bmatrix}$$

Not Unique
Long-term state probabilities vary depending on the initial state



$$P^{3n+1}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix} \quad P^{3n+2}=\begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix} \quad P^{3n}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

Periodic
Long-term state probabilities do not converge

Concepts & Definitions

- Definitions
  - Accessible: State j is accessible form i iff $P_{i,j}^{n} > \mathbf{0}$ for some n > 0

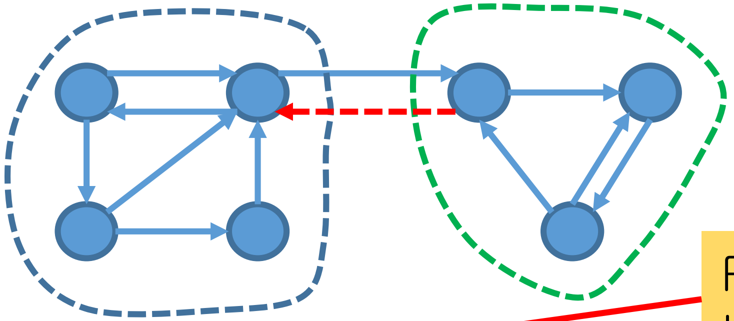
    ➔ There is a path from i to j in the directed graph
  - Communicating: i and j communicate if they are accessible from each other

    ➔ There is a path from i to j and vice versa

    i ↔ j
- Properties of Communicating relation
  1. Reflexive: $\mathrm{i} \leftrightarrow \mathrm{i}$
  2. Symmetric: If i ↔ j, then j ↔ i
  3. Transitive: If i ↔ j and j ↔ k, then i ↔ k

Concepts & Definitions

- More definitions and concepts
  - Communicating class: A set of nodes communicating each other
- Irreducible MC: Consists of a single communicating class

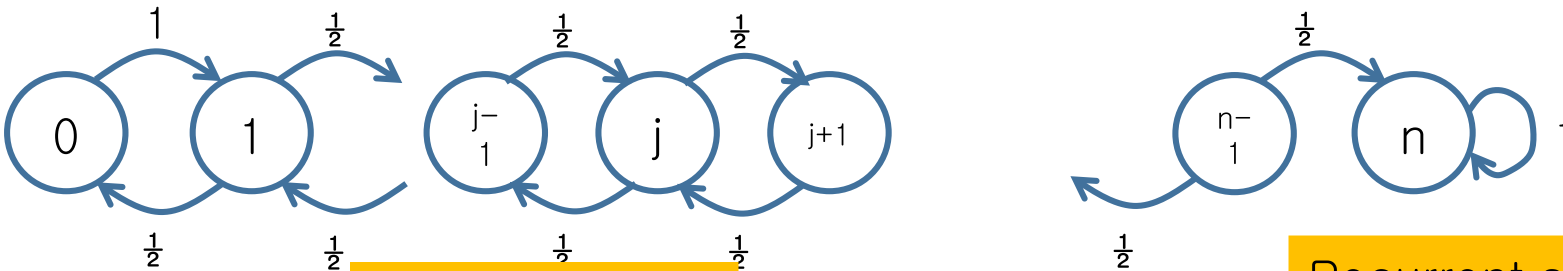


Reducible
But, irreducible if RED link is added

Prob. of visiting state j first time at the t-th transitions from i

- Let $r_{i,j}^{t} = \Pr(X_t = j \text{ and for } 1 \le s \le t-1, X_s \ne j \mid X_o = i)$
- Recurrent & Transient state:
  - State i is recurrent if $\sum_{t\geq1}^{\infty} r_{i,i}^{t} = 1$
  - State i is transient if $\sum_{t\geq1}^{\infty} r_{i,i}^{t} < 1$
- Recurrent MC: An MC all of its states are recurrent

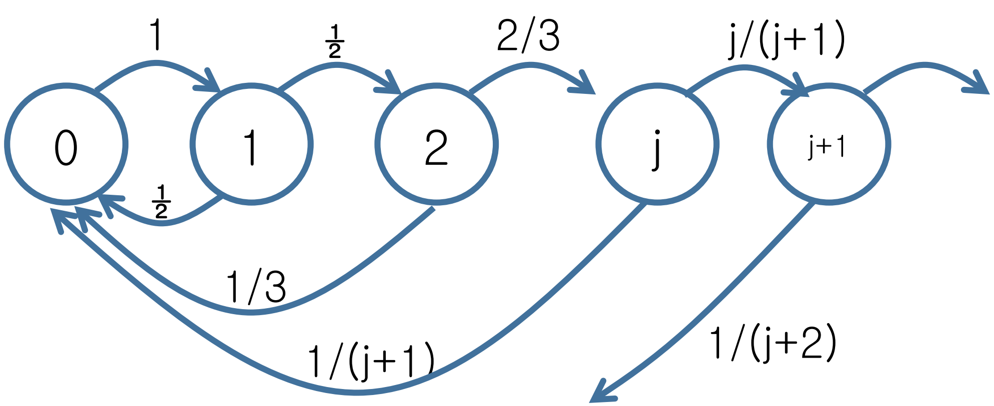
Concepts & Definitions



Transient states

Recurrent state
Absorbing state

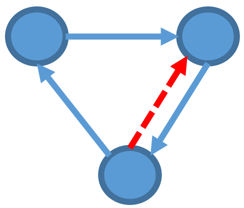
- **Positive recurrent**: A state is positive recurrent if the expected time to return to itself ($h_{i,i}$) is finite
  - $h_{i,i} = \sum_{t\geq 1}^{\infty} t \cdot r_{i,i}^{t} < \infty$



State 0 is Recurrent state
But, null recurrent

Concepts & Definitions

- **Periodic** state: A state is periodic if $r_{i,i}^{t} = 0$, for only t's that are divisible by Δ



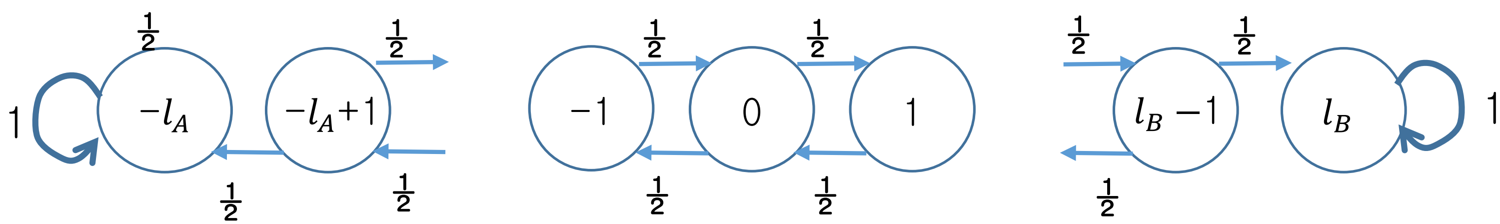
Periodic
Becomes Aperiodic if RED edge is added

- **Ergodic** state: A state is ergodic if it is positive recurrent and aperiodic

- **Ergodic MC**: An MC all of its states are ergodic

MC Example: Gambler's Ruin

- **Gambler's Ruin**
  - Two players A and B with l_A, l_B dollars, respectively
  - If one wins, take one dollar from the opponent
    - Pr(Win) = Pr (Lose)= 1/2
  - From the viewpoint of player A, quit if he loses all (l_A) or wins l_B dolloars
  - What is the probability (q) that player A wins?
  - State j: Dollar that A wins



  - Two recurrent states, $-l_A,\ \ and\ \ l_B$ and transient states
  - Let q be a probability that the MC terminates at state l_B

  ➔ Pr(Terminates at state $-l_A$) = 1−q

MC–Example

- Note, because win/lose probabilities are the same, $\mathrm{E}[W^t] = 0$ where is W^t is gain of player A after t plays

  $$\lim_{t\to\infty} \mathrm{E}[W^t] = 0 = l_B \cdot q + l_A \cdot (1-q)$$

  ➔ $q = \frac{l_A}{l_A + l_B}$

- Another approach
  - q_j: Probability that Player A wins l_B from the state j
  - On the condition of the result of the first play

    $q_j = \frac{1}{2} \cdot q_{j+1} + \frac{1}{2} \cdot q_{j-1}$, $\quad j = -l_A + 1, \cdots, l_B - 1$

  - Solve the equations

Stationary Distribution of MC

- Recall $\bar{p}(t+1) = \bar{p}(t) \cdot P$
  - When t is small, $\bar{p}(t)$ changes as t changes
  - For some MC, $\bar{p}(t)$ becomes stationary at large t

Definition: Stationary Distribution

$\bar{\pi}$ is the stationary distribution if $\bar{\pi} = \bar{\pi} \cdot P$ & $\bar{\pi} \cdot \bar{1} = 1$

Renewal Theory

Assume an irreducible, ergodic MC.

➔ For all i, $\lim_{t\to\infty} P_{i,i}^t$ exists and $\lim_{t\to\infty} P_{i,i}^t = 1/h_{i,i}$

- Interpretation
  - Expected number of transitions to revisit i is $h_{i,i}$
  - Among many transitions(N), the expected number of visits to i is N/$h_{i,i}$+o(1)

  ➔ Probability that the state is i among N observations
  = (N/$h_{i,i}$+o(1)) /N = $1/h_{i,i}$

Stationary Distribution of MC

Theorem

- A finite, irreducible, ergodic MC has the following properties
  - ① Has a unique stationary distribution $\bar{\pi} = (\pi_0, \pi_1, \cdots, \pi_n)$
  - ② For all j and i $\lim_{t\to\infty} P_{j,i}^t$ exists and is independent of j
  - ③ $\pi_i = \lim_{t\to\infty} P_{j,i}^t = 1/h_{i,i}$

Proof

- Using the renewal theory, $\pi_i = \lim_{t\to\infty} P_{i,i}^t = 1/h_{i,i}$

  Will prove that $\lim_{t\to\infty} P_{j,i}^t = \lim_{t\to\infty} P_{i,i}^t = 1/h_{i,i}$ for all j

  irreducible

  $\sum_{k=1}^{\infty} r_{j,i}^k = 1$ and for any ε > 0, we can find t_1 such that $\sum_{k=1}^{t_1} r_{j,i}^k \geq 1 - \varepsilon$

  - $\lim_{t\to\infty} P_{j,i}^t = lim \sum_{k=1}^{t} r_{j,i}^k \cdot P_{i,i}^{t-k} \geq \lim_{t\to\infty} \sum_{k=1}^{t_1} r_{j,i}^k \cdot P_{i,i}^{t-k} \geq (1-\varepsilon) \lim_{t\to\infty} P_{i,i}^t$

Stationary Distribution of MC

- Also
  - $P_{j,i}^{t} = \sum_{k=1}^{t} r_{j,i}^{k} \cdot P_{i,i}^{t-k}$

    $\leq \sum_{k=1}^{t_1} r_{j,i}^{k} \cdot P_{i,i}^{t-k} + \varepsilon$

    $\sum_{k=1}^{t_1} r_{j,i}^{k} \geq 1 - \varepsilon$

    ➔ $\lim_{t\to\infty} P_{j,i}^{t} \leq \lim_{t\to\infty} (\sum_{k=1}^{t_1} r_{j,i}^{k} \cdot P_{i,i}^{t-k} + \varepsilon)$

    $\leq \lim_{t\to\infty} P_{i,i}^{t} + \varepsilon$
- Therefore, $\lim_{t\to\infty} \sum_{i=1}^{n} P_{j,i}^{t} = \lim_{t\to\infty} P_{i,i}^{t} = 1/h_{i,i}$

- Let $\pi_i = \lim_{t\to\infty} P_{j,i}^{t} = 1/h_{i,i}$
- Now, prove that π_i is a stationary distribution
  - $\sum_{i=1}^{n} \pi_i = \sum_{i=1}^{n} \lim_{t\to\infty} P_{j,i}^{t} = \lim_{t\to\infty} \sum_{i=1}^{n} P_{j,i}^{t} = 1$
  - $\pi_i = \lim_{t\to\infty} P_{j,i}^{t} = \sum_{l=1}^{n} \lim_{t\to\infty} P_{j,l}^{t-1} \cdot P_{l,i} = \sum_{l=1}^{n} \pi_l \cdot P_{l,i}$

    ➔ $\bar{\pi} = \bar{\pi} \cdot P$
- Finally, prove that $\bar{\pi}$ is unique (Easy & left as a HW)

Stationary Distribution Computation

- Solve the set of linear equation
  - $\bar{\pi} = \bar{\pi} \cdot P$
  - $\sum_{i=1}^{n} \pi_i = 1$

- Example

$$P = \begin{pmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/6 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{pmatrix}$$

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$
$$\frac{1}{2}\pi_1 + \frac{1}{4}\pi_2 = \pi_0$$
$$\frac{1}{4}\pi_0 + \frac{1}{4}\pi_2 + \frac{1}{2}\pi_3 = \pi_1$$
$$\frac{1}{3}\pi_1 + \frac{1}{2}\pi_2 + \frac{1}{4}\pi_3 = \pi_2$$

Cut Set Method

- Theorem
  - Consider an irreducible, ergodic MC. Let S be the subset of the MC. In the stationary distribution, the probability that leaves from the set equals the probability that enters to the set.

- Proof
  - For a subset of single state (Assume it is state j)
    - $\pi_j = \sum_{i=1}^{n} \pi_i \cdot P_{i,j}$

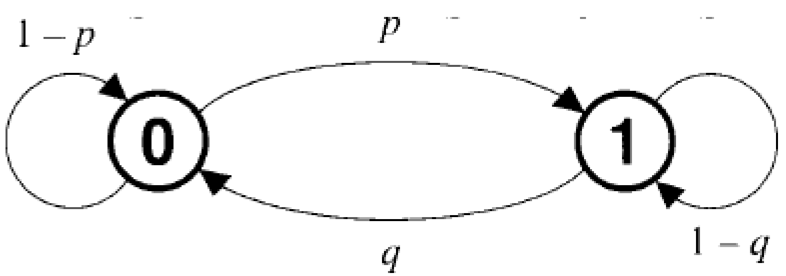
      Probability into state j
    - Also, $\pi_j = \pi_j \sum_{i=1}^{n} P_{j,i}$ = Probability from state j

  ➔ generalize to arbitrary subset

Cut Set Method

- Cut set method
  - Apply the Theorem to (n−1) subsets and obtain (n−1) equations
  - Define subsets that simplify the computation

- Example



MC that models the bursty occurrences
Bit error, collisions, ⋯

$$P=\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Based on $\bar{\boldsymbol{\pi}} = \bar{\boldsymbol{\pi}} \cdot \boldsymbol{P}$

$$\pi_0(1-p) + \pi_1 q = \pi_0$$
$$\pi_0 p + \pi_1(1-q) = \pi_1$$
$$\pi_0 + \pi_1 = 1$$

Via cut set method

$$\pi_0 p = \pi_1 q$$
$$\pi_0 + \pi_1 = 1$$

Time Reversibility

- Theorem
  - Consider an irreducible, ergodic MC with transition matrix P. If $\bar{\pi}$ is a distribution and if it satisfies

$$\pi_i \cdot P_{i,j} = \pi_j \cdot P_{j,i}$$

    Then $\bar{\pi}$ is the unique stationary distribution of the MC.

- Proof
  - From $\pi_i \cdot P_{i,j} = \pi_j \cdot P_{j,i}$
  - $\sum_{j=1}^{n} \pi_i \cdot P_{i,j} = \sum_{j=1}^{n} \pi_j \cdot P_{j,i}$

    $\equiv \pi_i$

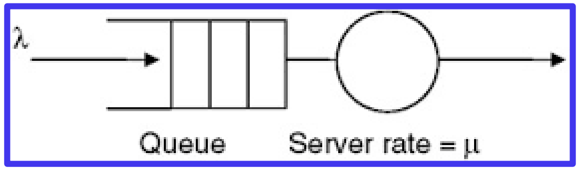
    ➔ $\bar{\pi} = \bar{\pi} \cdot P$

    ➔ $\bar{\pi}$ is unique stationary distribution of the MC

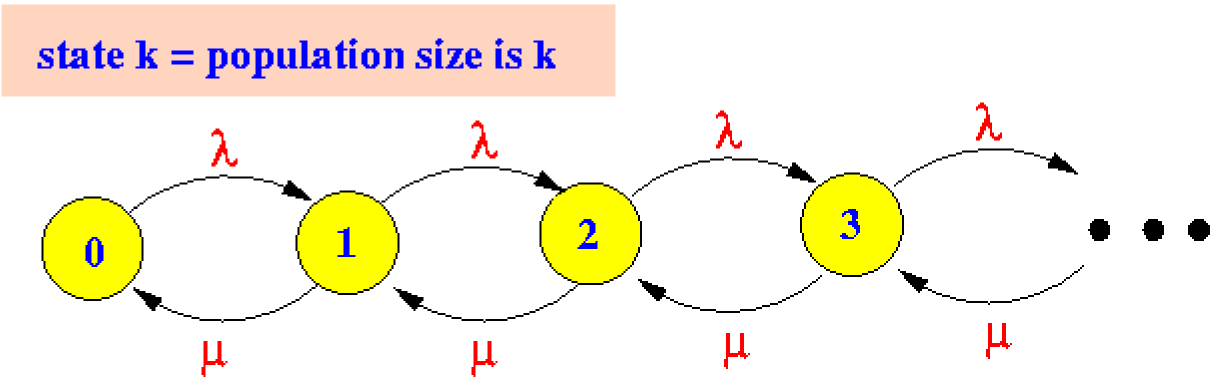
Example – Simple Queue

- Queue
  - Most (virtually all) resources are shared
    ➔ Sometimes, must wait until other customers in service (or of higher priority) finish their services
  - Bank teller, Dining hall, Bus, …

- Queueing theory
  - Study the performance of queues
- Simple queue
  - The system has one server and it contains up to n customers (jobs)
  - The server will serve a customer(job) whenever there are in the system
  - Scheduling is FCFS(First Come First Serve)
  - Once started, the service is finishes with probability of μ
  - A new customer arrives with probability of λ



Example – Simple Queue



Add transition probability

$$P_{i,i} = 1 - \lambda, \quad i = 0$$
$$P_{i,i} = 1 - \mu, \quad i = n$$
$$P_{i,i} = 1 - \lambda - \mu, \quad o.w.$$

$\lambda \cdot \pi_i = \mu \cdot \pi_{i+1}$ ➔ $\pi_{i+1} = (\frac{\lambda}{\mu}) \cdot \pi_i,$ for $i = 0,1, \ldots n-1$

➔ $\pi_i = (\frac{\lambda}{\mu})^i \cdot \pi_0,$ for $i = 1,2, \ldots, n$

From $\sum_{i=0}^{n} \pi_i$ ➔ $\pi_0 = \frac{1}{\sum_{i=0}^{n} (\frac{\lambda}{\mu})^i}$

$$\pi_i = (\frac{\lambda}{\mu})^i / \sum_{i=0}^{n} (\frac{\lambda}{\mu})^i$$

Random Walk

- Can observe the evidence of transitions between states, but cannot measure the transition probabilities

  ➔ Assume that transitions to neighbor states are equally probable
  - Transition probability from node u to v = 1/d(u) where d(u) is # u's neighbors
- Random walk model can be applied to both directed and undirected graphs
  - The most flourishing one is Google's PageRank

Random walk

From Wikipedia, the free encyclopedia

A **random walk** is a mathematical object, known as a stochastic or random process, that describes a p
some mathematical space such as the integers. For example, the path traced by a molecule as it travel
animal, the price of a fluctuating stock and the financial status of a gambler can all be approximated by
random in reality. As illustrated by those examples, random walks have applications to many scientific f
physics, chemistry, biology as well as economics. Random walks explain the observed behaviors of ma
fundamental model for the recorded stochastic activity. As a more mathematical application, the value c
in agent-based modelling environment.[1][2] The term *random walk* was first introduced by Karl Pearson

Stochastic process

From Wikipedia, the free encyclopedia
(Redirected from Random process)

In probability theory and related fields, a **stochastic** or r
random variables were associated with or indexed by a
numerical values of some system randomly changing ov
movement of a gas molecule.[1][4][5] Stochastic processe

Karl Pearson

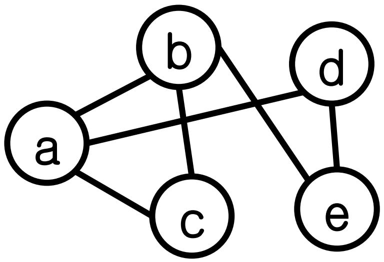
From Wikipedia, the free encyclopedia

For the English cricketer, see Karl Pearson (cricketer).

Karl Pearson FRS[1] (/ˈpɪərsən/; originally named **Carl**; 27 Marc
credited with establishing the discipline of mathematical statistic
1911, and contributed significantly to the field of biometrics, met

Quick Int. to Graph Theory

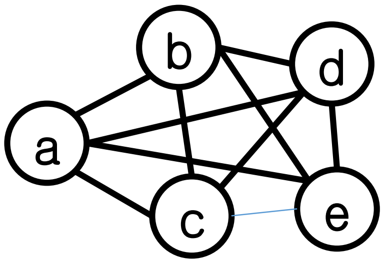
- A graph, G(V,E), consists of two sets
  - Node (Vertex)
  - Edge (Link, Arc)



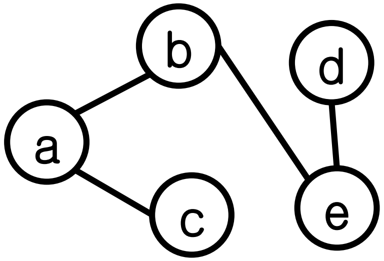
- An edge can be directed or undirected
- An edge may have weight
  - Meaning: distance, transition probability, ..
- Neighbor of a node
  - Nodes that are directly connected to the node
- Path
  - Sequence of connected links from source node to destination node
  - Path length (cost): Sum of weights of edges in a path
- Cycle
  - A path to and from a same node

Quick Int. to Graph Theory

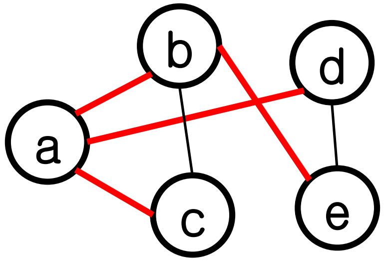
- Complete Graph
  - All node pairs are directly connected
    - # edges in a complete graph of n nodes = ?



- Tree
  - A graph without cycle



- Spanning tree
  - A sub-tree that spans all nodes
    - # edges in a spanning tree = ?



Computation of S. P.

- Claim:
  - A random walk on G has a stationary distribution $\bar{\pi}$, where $\boldsymbol{\pi_v}$ = d(v) / 2|E|
- Proof
  - First, show that $\sum_{v=1}^{n} \pi_v = 1$
  - Now, show $\pi_v \cdot P_{v,u} = \pi_u \cdot P_{u,v}$ for any neighbor pair (u, v)

    OR show $\pi_v = \sum_{u=1}^{n} \pi_u \cdot P_{u,v}$
- For any vertex u in G,
  - $h_{u,u} = \frac{2|E|}{d(u)}$

Cover Time of RW

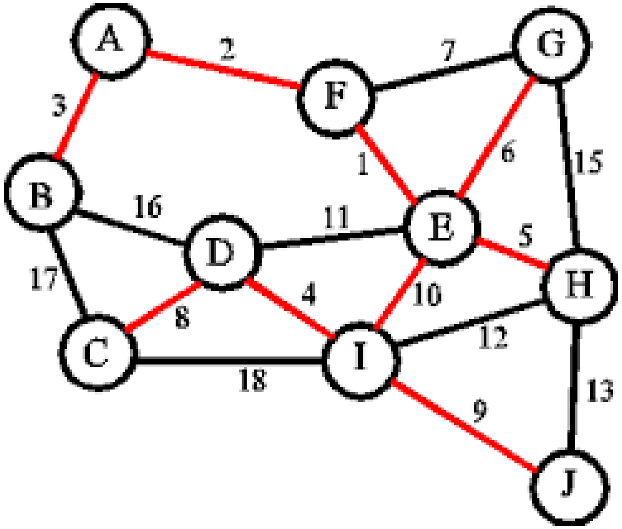
- Definition: Cover Time
  - Expected time to visit all vertices started from node v
  - Max. over v

- Claim
  - The cover time of G=(V, E) is bounded by 4|V|•|E|

- Proof
  - First, show that for (u,v) ∈ E, $h_{v,u} < 2|E|$
    - Express $h_{u,u}$ conditioned on the first random walk
    - $h_{u,u} = \frac{1}{d(u)} \sum_{v \in N(u)} (1 + h_{v,u})$ ➔ $h_{v,u} < 2|E|$

      $\boxed{\equiv 2|E|/d(u)}$
  - Let v_0 be the starting vertex
  - Create a Spanning Tree (SP) from v_0

Cover Time of RW



Spanning Tree: A Tree that include all vertices
|V|-1 edges

Let G be the start node
G-E-F-A-B-A-F-E-I-D-C-D-I-J-I-E-H-E-G
How many edges in the tour?

Walk w/o repeated edges
Forward & return directions

Assume a worst case tour along the SP

We traverse each edge at most two times to start from and return to the origin

- Let $v_0 - v_1 - v_2 - \cdots - v_0$ be a cyclic tour
- Cover time = $\sum_{i=0}^{2|V|-3} h_{v_i, v_{i+1}} < 4|V| \cdot |E|$

Why?

s–t Connectivity

- Problem
  - Determine if nodes s and t are connected
- Aim to minimize the *space* (not time) required to run a algorithm
- There is a randomized algorithm that solve the problem using O(log n) bits

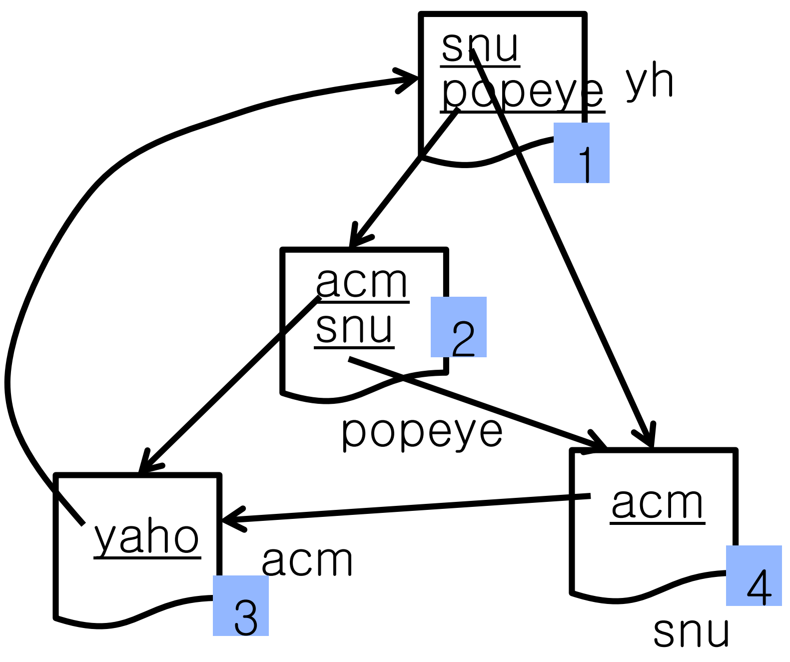
> Algorithm
> 1. Start a random walk from s up to $4n^3$ steps, terminate if t is reached
> 2. Report Success or Fail

- Claim: The algorithm is right with probability $\frac{1}{2}$
- Proof:
  - Let X be the random variable of time from s to t
  - E[X] < ??
  - Compute Pr(X $\geq 4n^3$) using Markov's Inequality

PageRank

S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", WWW, 1998.

- Model page navigation as a random walk along URLs



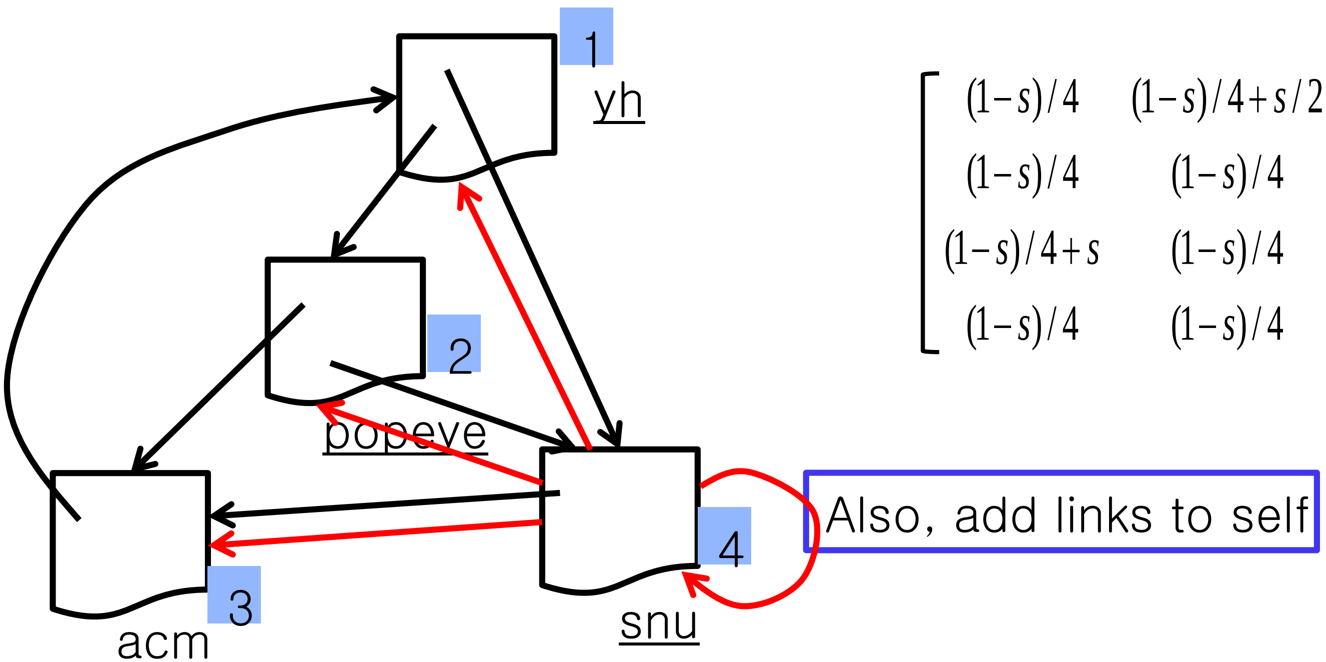
If there are n out-links then the prob. to follow a certain link is 1/n

$P_{ij} = 1/k_i$, k_i: the out-degree of node i

$$\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

PageRank

- Add Random Jump (Transport) to make the MC irreducible, ergodic



$$\begin{bmatrix} (1-s)/4 & (1-s)/4+s/2 & (1-s)/4 & (1-s)/4+s/2 \\ (1-s)/4 & (1-s)/4 & (1-s)/4+s/2 & (1-s)/4+s/2 \\ (1-s)/4+s & (1-s)/4 & (1-s)/4 & (1-s)/4 \\ (1-s)/4 & (1-s)/4 & (1-s)/4+s & (1-s)/4 \end{bmatrix}$$

Main Idea: An important pages get many references
Importance of a page = Sum of weighted references