Logica e Metodi Probabilistici per l’Informatica
AA 2021/2022
Sezione 3

Metodi Probabilistici per l’Informatica

Prof. Stefano Leonardi

Tutor: Dr. Federico Fusco

Programma

3. *Processi stocastici ed applicazioni*

- Le catene di Markov

- Random walks e l'algoritmo di Pagerank

- Random Walks for 2-SAT

- Random Walks in Undirected Graphs

Random Walks and Markov chains

- Random walks are important methods for fast random sampling and for designing fast algorithms
- Random walks can be modelled with Markov Chains
- We use it for several applications:
  - Pagerank
  - Solving 2-SAT Formulae
  - Connectivity graphs

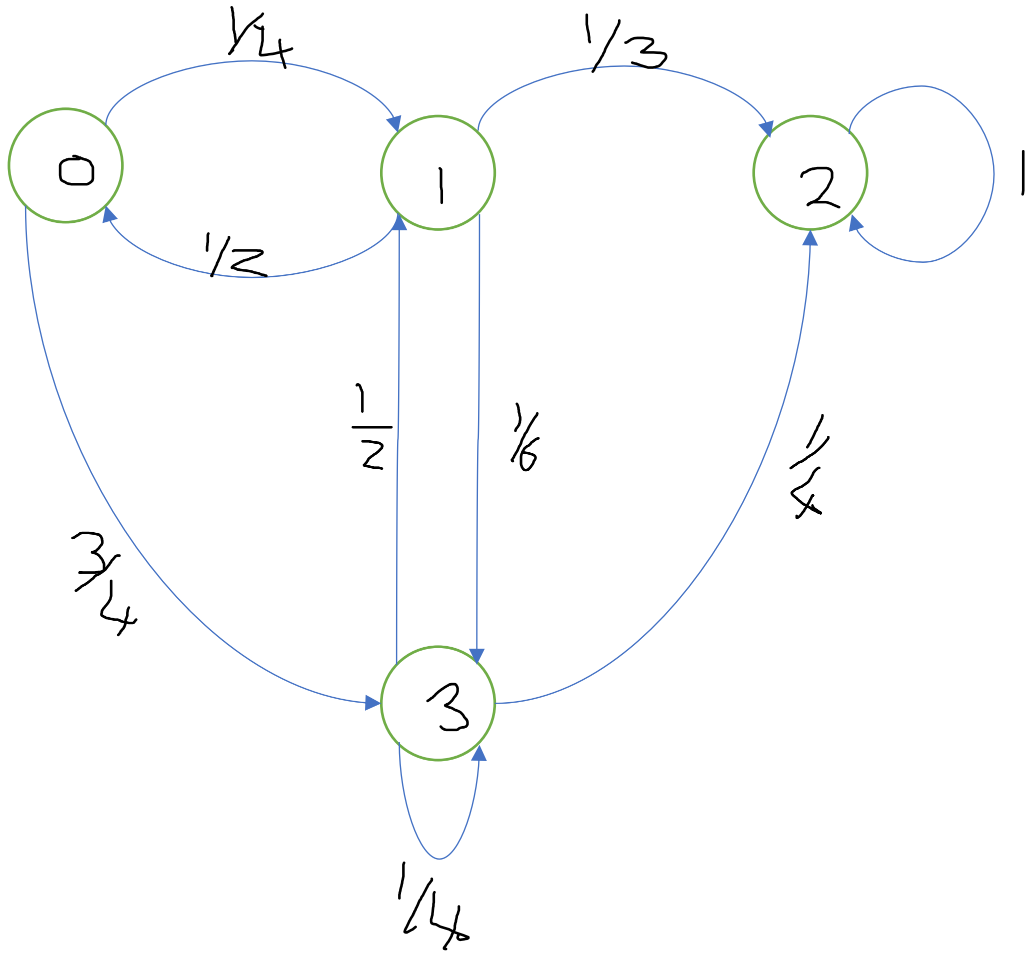
Random Walks and Markov Chains

- Given a graph, a random walk starts form a vertex, at each step the walk moves to an uniformly random neighbour
- Some of the basic mathematical questions are:

1. What is the limiting distribution (stationary distribution) of the random walk?
2. How long does it take before the walk approaches the limiting distribution? (mixing time)
3. Starting from a vertex s, what is the expected number of steps to first reach t? (hitting time)
4. How long does it take to reach every vertex at least once? (cover time)

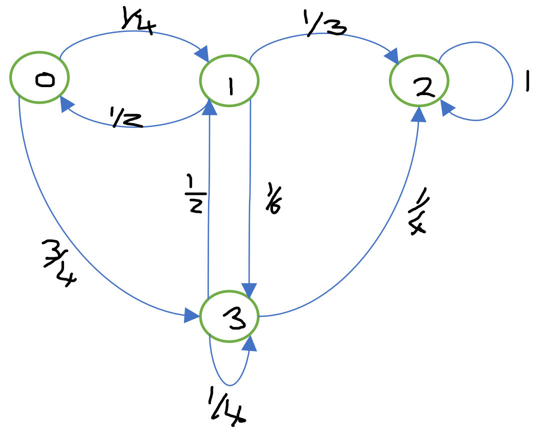
Markov Chains

- We model a random walk on a directed graph with a Markov chain:
  - Each vertex correspond to a node
  - Arc (i,j) corresponds to the transition probability from state i to state j



Formulate as a Matrix problem

- P_{ij}: transition probability from state i to state j
- Stochastic matrix: $\sum_j \mathrm{P_{ij}} = 1$
- X_t: state at time t
- $p_t(i)$: probability to be at state i at time t
- $\vec{p}_0 = (1,0,..,0)$: the walk is at state 0 at time 0
- $\vec{p}_0 = \left(\frac{1}{n},\frac{1}{n},..,\frac{1}{n}\right)$:the walk starts at a random node



$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{pmatrix} \end{array}$$

P

State transitions

- The state transition is defined as follows:

$$p_{t+1}(j) = \sum_{i=0}^{n-1} p_t(i) \cdot P_{ij}$$

- In compact form:

$$\vec{p}_{t+1} = \vec{p}_t \cdot P \text{ and generally } \vec{p}_{t+m} = \vec{p}_t \cdot P^m \text{ or } \vec{p}_t = \vec{p}_0 \cdot P^t$$

- Markov chains are memoryless stochastic processes:

$\Pr(X_t = a_t \mid X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots, X_0 = a_0) = \Pr(X_t = a_t \mid X_{t-1} = a_{t-1}) = P_{a_{t-1}a_t}$

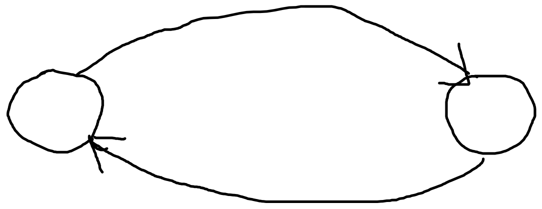
States of a Markov chain

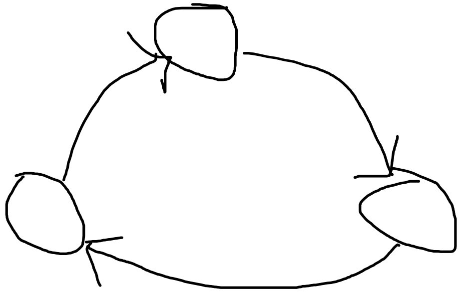
- A Markov chain is *irreducible* if for every pair of states $i, j \in V$ there exists a path from *i* to *j*
- If the Markov Chain is irreducible, for every $i, j \in V$ there exists a value *l* such that $\Pr(X_{t+l} = s_j \mid X_t = s_i) > 0$

- The period of a state s_i is defined as $d(s_i) = gcd\ \{ t \mid p_{ii}^{t} > 0 \}$
- A state *i* is aperiodic if $d(s_i) = 1$
- A Markov chain is aperiodic if all the states are aperiodic

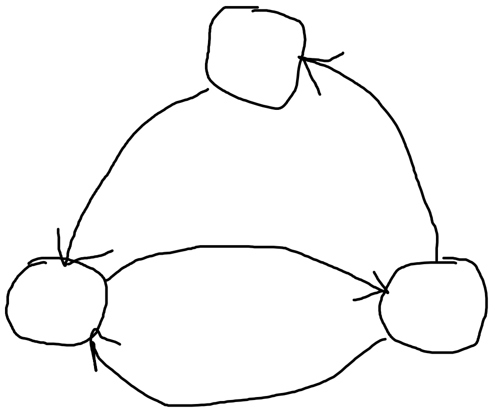
States of a Markov Chain

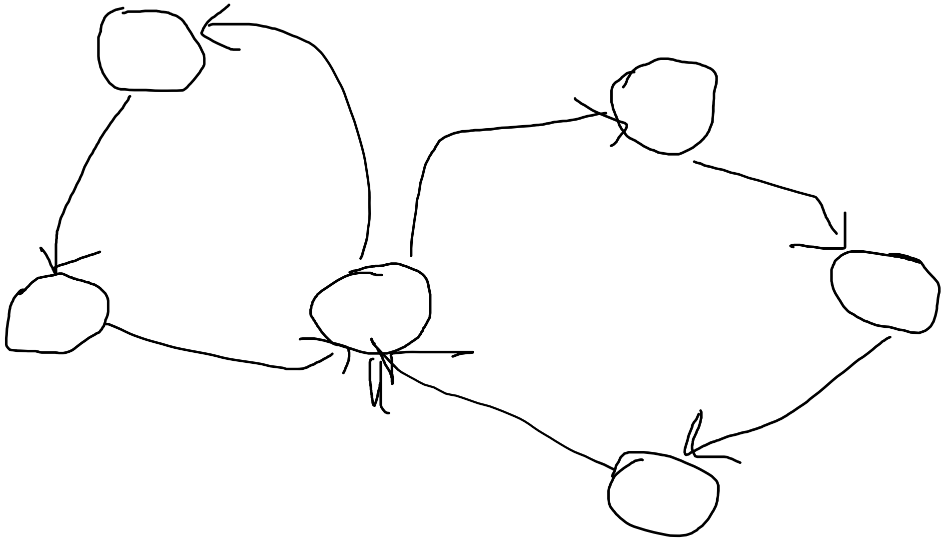
- Periodic Markov chains:





- Aperiodic Markov chains:





Ergodic Markov chains

- A Markov chain that is irreducible and aperiodic is called Ergodic

Theorem: For any finite, irreducible and aperiodic (ergodic) Markov chain, there exists a time T<∞ such that

$p_{ij}^{t}>0$ for any pair of states *i,j,* and for any *t>T*

Proof. If the MC is ergodic then there exists a path from i to j of length t, for any t>T.

Stationary distribution of Markov chain

Definition: A stationary distribution of a Markov chain is a probability distribution $\vec{\pi}$ such that $\vec{\pi} = \vec{\pi}P$

- Informally, $\vec{\pi}$ is a steady state/equilibrium/fixed point, as $\vec{\pi} = \vec{\pi}P^{t}$ for any $t \geq 0$
- Given a Markov chain, after we run long enough, then we will forget about the history and converge to the same distribution.
- Two questions:
  - Does a steady state always exist?
  - Is it unique?

Stationary distribution of Markov chain

Fundamental theorem of Markov chains:

Theorem. For any finite, irreducible, and aperiodic Markov chain, the following holds:

1. There exists a stationary distribution $\vec{\pi}$
2. The distribution $\vec{p}_t$ will converge to $\vec{\pi}$ as $t \to \infty$ no matter which is the initial distribution $\vec{p}_0$
3. There is a unique stationary distribution

Pagerank

- Consider a directed graph describing the relationships between a set of web pages: there is an arc from i to j if there is a link from i to j
- We like to rank pages according to importance
- Intuitevely, a page linked from many other pages is important
- This motivate the following random walk algorithm

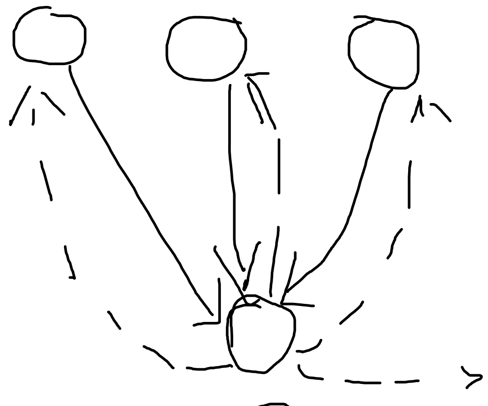
Pagerank Algorithm

1. Initially, the random walk is at some page of the graph
2. In each step, the random walk follows one random outgoing link

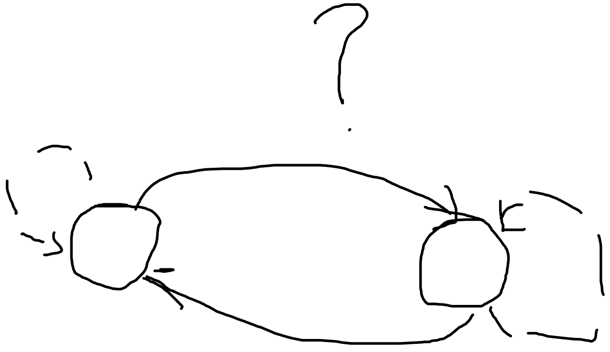
Pagerank

Problems:

1. Dead-ends. The random walk stops at the dead ends
   Solution: At a dead-end, jump to a random page



2. Aperiodicity. If the graph is aperiodic, there is no stationary distribution
   Solution: with probability α jump to a random page
   with probability 1-α follow a random outgoing link (teleporting)



Pagerank

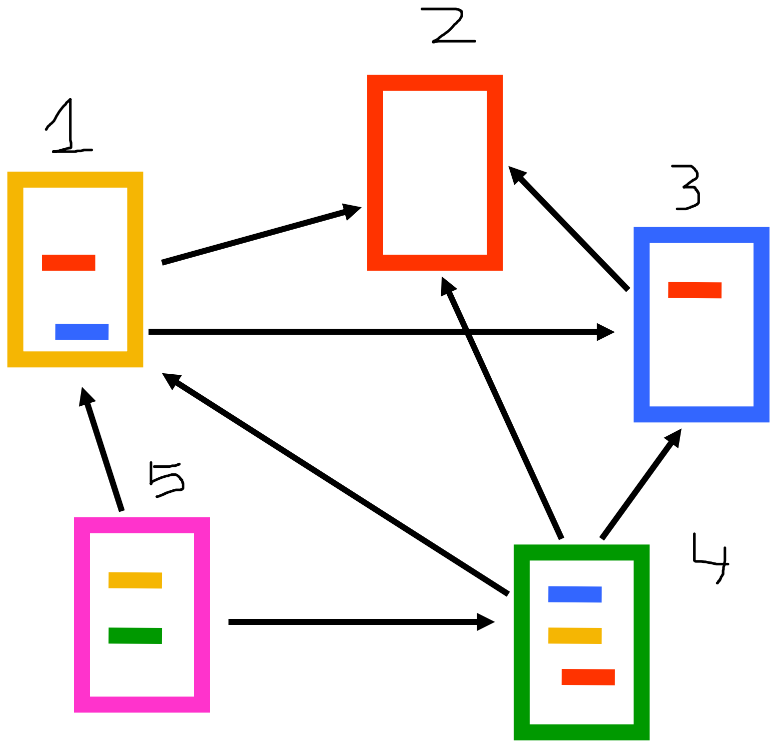
- The Pagerank Markov chain with teleporting and random jump is irreducible and aperiodic.
- There exists a stationary distribution of the Markov chain that is unique
- The equilibrium Pagerank values are equal to the stationary distribution of the random walk!
- Pagerank Markov chain:

$$P_{ij} = (1 - \alpha)\frac{1}{deg^{out}(i)} + \alpha\frac{1}{n}$$

The PageRank Markov chain

- Adiacency matrix of the graph
- Vertex 2 has no outgoing node

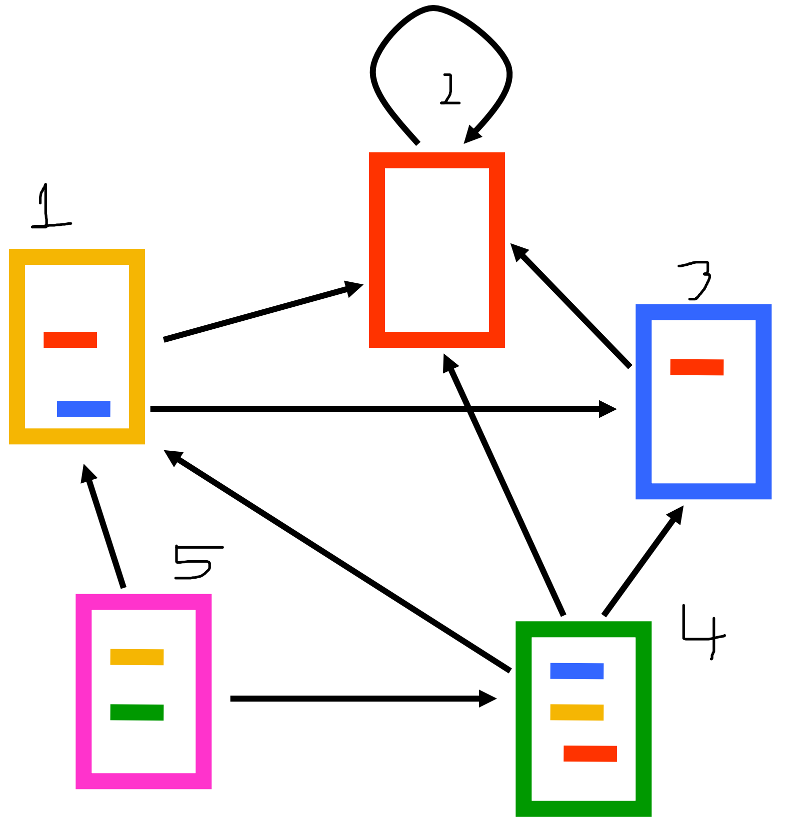
$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$



The PageRank Markov chain

- What about sink nodes?
  - what happens when the random walk moves to a node without any outgoing inks?

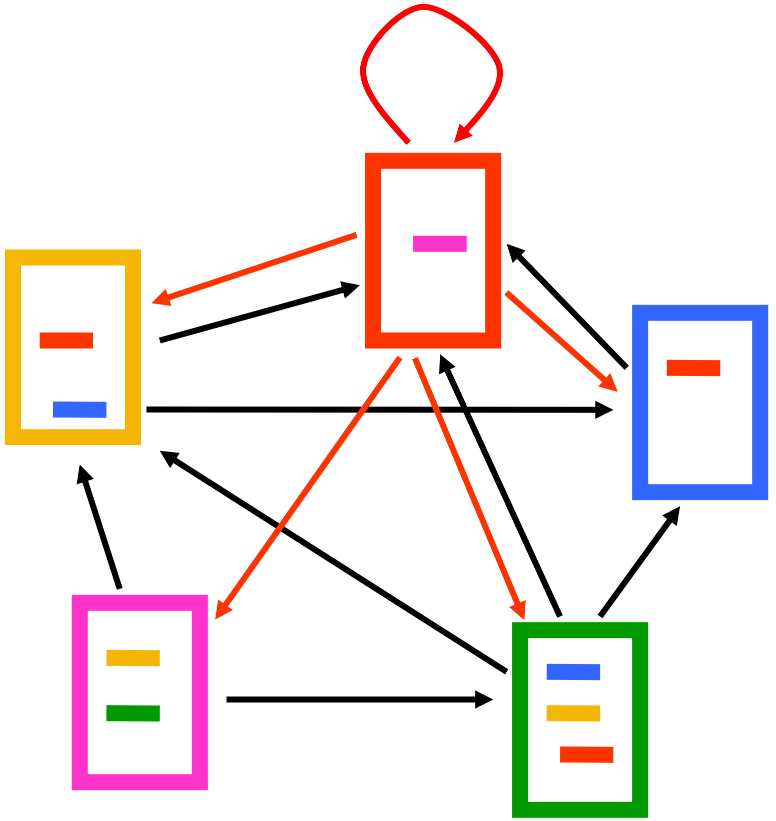
$$P_{RW} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$



The PageRank Markov chain

- Replace these row vectors with a vector the uniform vector

$$P_{RW} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ \boxed{1/5 \quad 1/5 \quad 1/5 \quad 1/5 \quad 1/5} \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$



The PageRank Markov chain

- Add a random jump to a random node with prob α

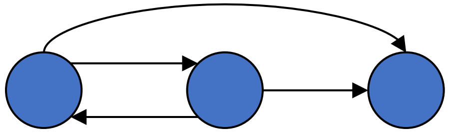
$$
\mathrm{P}_{\mathrm{PR}} = (1-\alpha)\begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix} + \alpha \begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}
$$

$P_{PR} = (1-\alpha)P_{RW} + \alpha U$, where U is the uniform matrix with rows summing to 1

Transition matrix for pagerank

- Take the adjacency matrix A
- If a line i has no 1s set P_{ij} = 1/N
- For the rest of the rows:
  - Set:

$$P_{ij} = (1-\alpha)P_{RW} + \frac{\alpha}{N} = (1-\alpha)\frac{A_{ij}}{(\#\ \text{1s in line } i)} + \frac{\alpha}{N}$$



$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad P_{RW} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P = \begin{bmatrix} \frac{\alpha}{3} & \frac{1}{2} - \frac{\alpha}{6} & \frac{1}{2} - \frac{\alpha}{6} \\ \frac{1}{2} - \frac{\alpha}{6} & \frac{\alpha}{3} & \frac{1}{2} - \frac{\alpha}{6} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

The stationary distribution of a Markov chain

Theorem:

For any finite, irreducible, and aperiodic Markov chain, the following holds:

1. There exists a stationary distribution $\vec{\pi}$
2. The distribution $\vec{p}_t$ will converge to $\vec{\pi}$ as $t \to \infty$ no matter which is the initial distribution $\vec{p}_0$
3. There is a unique stationary distribution

The stationary distribution of a Markov chain

- A distribution X is stationary if $X = X P$
- We start from distribution $X_0 = (X_0^1, X_0^2, \ldots, X_0^n)$
- After 1 step the position of the random walk is given by the distribution $X_1 = X_0 P$
- After t steps we have $X_t = X_0 P^t$

The stationary distribution of a Markov chain

- Theorem 1: Let P be the transition matrix of a markov chain that is aperiodic and strongly connected. Then, $\lim_{t\to\infty} P^t = P^\infty$ where

$$P^\infty = \begin{bmatrix} \pi_1\, \pi_2 & \cdots & \pi_n \\ \vdots & \ddots & \vdots \\ \pi_1 \pi_2 & \cdots & \pi_n \end{bmatrix}$$

- Let us draw the consequences of the above theorem where $\pi = (\pi_1, \pi_2, \dots., \pi_n)$

The stationary distribution of a Markov chain

Corollary 1: $\pi = (\pi_1, \pi_2, \ldots., \pi_n)$ is a stationary distribution

Proof.

$X_0 P^{\infty} = \pi$ since X_0 is a distribution. Let us now prove that $\pi P = \pi$.

$$P^{\infty} P = (\lim_{t \to \infty} P^t) P$$
$$= (\lim_{t \to \infty} P^{t+1})$$
$$= P^{\infty}$$

And therefore $\pi P = X_0 P^{\infty} P = X_0 P^{\infty} = \pi$.

The stationary distribution of a Markov chain

- Corollary 2: For any initial distribution X_0, the sequence $X_t = X_{t-1} P = X_0 P^t$ converges to π, i.e.,. $\lim_{t\to\infty} X_t = \pi$, i.e., the stationary distribution is unique.

Proof:
$$\begin{aligned}\lim_{t\to\infty} X_t &= \lim_{t\to\infty}(X_0 P^t) \\ &= X_0 \lim_{t\to\infty}(P^t) \\ &= X_0 P^{\infty} \\ &= \sum_{i=1}^{n} X_0^i \pi \\ &= \pi \sum_{i=1}^{n} X_0^i \\ &= \pi\end{aligned}$$

The stationary distribution of a Markov chain

Proof of Theorem 1:

- Keep track of the smallest and largest values, resp. m_t and M_t, of a column j as t goes to infinity.
- We assume $p_{ij}^{t} > 0$ and $\delta = \min_{ij} p_{ij}$
- Since n>1, $\delta \leq \frac{1}{2}$
- We'll prove:

(i.) The sequence $\{m_t\}$ is non decreasing

(ii.) The sequence $\{M_t\}$ is non increasing

$(iii.)\Delta_t = M_t - m_t$ goes to 0 exponentially fast!

The stationary distribution of a Markov chain

Proof of Theorem 1 (contd): We prove *(i)*

- $$m_{t+1} = \min_i p_{ij}^{t+1}$$
$$= \min_i \sum_{k=1}^{n} p_{ik} p_{kj}^{t}$$
$$\geq \min_i \sum_{k=1}^{n} p_{ik} m_t$$
$$= \left(\min_i \sum_{k=1}^{n} p_{ik}\right) m_t$$
$$= m_t$$

Analogously, we prove *(ii.)*

The stationary distribution of a Markov chain

Proof of Theorem 1 (contd): We prove *(iii)*

- Let l be the row where M_t lies
- $m_{t+1} = \min_i p_{ij}^{t+1}$

$$= \min_i \sum_{k=1}^{n} p_{ik} p_{kj}^{t} = \min_i p_{il} M_t + \sum_{k \neq l} p_{ik} p_{kj}^{t}$$

$$\geq \min_i p_{il} M_t + \sum_{k \neq l} p_{ik} m_t = p_{il} M_t + (1 - p_{il}) m_t$$

$$\geq \delta\, M_t + (1 - \delta)\, m_t \quad \textit{(a.)}$$

Analogously, we prove $M_{t+1} \leq \delta\, m_t + (1 - \delta)\, M_t$ *(b.)*

The stationary distribution of a Markov chain

Proof of Theorem 1 (contd): We prove *(iii)*

- Taking *(a.)* and *(b.)* together
- $\Delta_{t+1} = M_{t+1} - m_{t+1}$

  $\leq \delta\, m_t + (1-\delta)\, M_t - (\delta\, M_t + (1-\delta)\, m_t)$

  $= (1 - 2\,\delta)(M_t - m_t)$

  $= (1 - 2\,\delta)\Delta_t$

  $\leq (1 - 2\,\delta)^t$
- It goes to 0 exponentially fast for $0 < \delta \leq ½$
- For Pagerank, $\alpha > 0$ ensures $p_{ij} > 0$, $e^{-2\delta t} < 1/n$ if $t > \frac{1}{2\delta} \ln n$

Random walks for 2-SAT

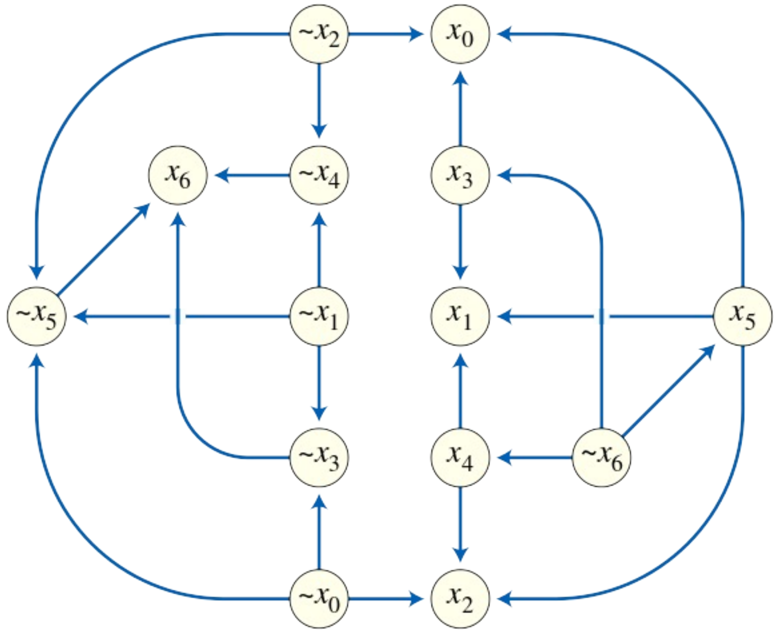
- DNF 2-SAT Formula:

$$(x_1 \lor \overline{x_2}) \land (x_1 \lor x_3) \land (x_1 \lor x_2) \land (x_4 \lor \overline{x_3}) \land (x_4 \lor \overline{x_1})$$

- Given a formula with two variables for each clause, there exists a polynomial time algorithm that finds a true assignment If exists.
- The algorithm builds a graph with a node for each x and $\bar{x}$ literals, and for each clause $(x \lor y)$ an edge $(\bar{x}, y)$ and $(\bar{y}, x)$.
- The Formula is NOT satisfiable iff for a variable *x* there exists a path from $\bar{x}$ to x and vice-versa.
- This can be tested in linear time using an SCC algorithm.

Random walks for 2-SAT

$$(x_0 \cup x_2) \cap (x_0 \cup \neg x_3) \cap (x_1 \cup \neg x_3) \cap (x_1 \cup \neg x_4) \cap$$
$$(x_2 \cup \neg x_4) \cap (x_0 \cup \neg x_5) \cap (x_1 \cup \neg x_5) \cap (x_2 \cup \neg x_5) \cap$$
$$(x_3 \cup x_6) \cap (x_4 \cup x_6) \cap (x_5 \cup x_6).$$

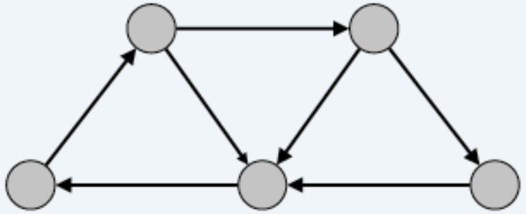


Random Walks for 2-SAT

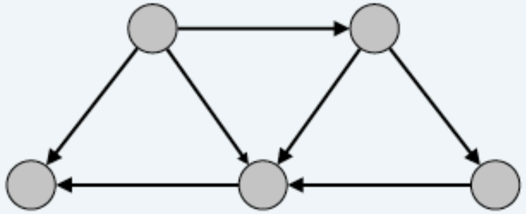
Theorem. Can determine if G is strongly connected in $O(m + n)$ time.

Pf.

- Pick any node s.
- Run BFS from s in G.
- Run BFS from s in $G^{reverse}$. reverse orientation of every edge in G
- Return true iff all nodes reached in both BFS executions.
- Correctness follows immediately from previous lemma. ▪



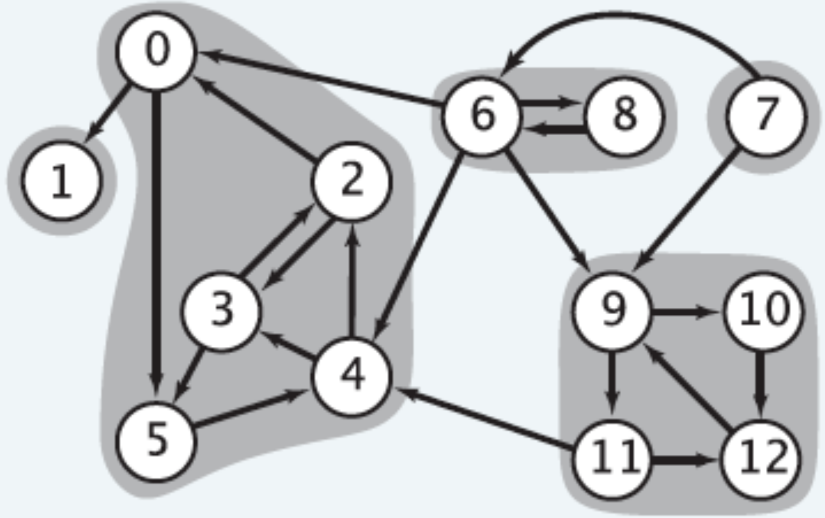
strongly connected



not strongly connected

Random Walks for 2-SAT

Def. A strong component is a maximal subset of mutually reachable nodes.



Theorem. [Tarjan 1972] Can find all strong components in $O(m+n)$ time.

Random walks for 2-SAT

- We give an algorithm for 2-SAT based on random walks

1. Start from an arbitrary assignment
2. Repeat up to $2Cn^2$ times or terminate if all clauses are satisfied
   I. Choose an arbitrary clause that is not satisfied
   II. Choose a random variable in the clause and switch its value
3. Return a satisfying assignment if it is found, otherwise return "unsatisfiable"

Let us consider the case in which there exists a satisfying assignment but the algorithm did not find it

Random walks for 2-SAT

- How do we analyse the algorithm?
- It is hard to measure the number of satisfied clauses because switching a variable can change many clauses
- Let us try to measure how close do we are to a satisfying assignment assuming that this exists.

1. S: satisfying assignment
2. A_i: assignment made in the i-th step of the algorithm
3. X_i: number of variables that have the same value in A_i and S

Random walks for 2-SAT

- If X_i=n, then A_i=S and we have the formula satisfied
- Keep track of X_i when a satisfying assignment is not found yet
- If X_i=0 then $\Pr(X_i=1 \mid X_i=0) = 1$
- Suppose $1 \leq X_i \leq n-1$. In an unsatisfied clause, since S is a satisfying assignment, there must be a variable in the class that has different values in A_i and S.
- Since we pick a random variable in that class, with prob. ½ we pick such variable and correct. Therefore:

$\Pr(X_{i+1}=j+1 \mid X_i=j) \geq 1/2$ and $\Pr(X_{i+1}=j-1 \mid X_i=j) \leq 1/2$

Random walks for 2-SAT

- To model as a random walk on a line, we consider a pessimistic version:

- $\Pr(Y_{i+1}=1 \mid Y_i=0) = 1$
- $\Pr(Y_{i+1}=j+1 \mid Y_i=j) = ½$
- $\Pr(Y_{i+1}=j-1 \mid Y_i=j) = ½$

- The expected time for Y is not smaller than the expected time for X

Random walks for 2-SAT

- This can be seen as a random walk on the line
- Let us give an upper bound on the expected time for Y to reach n

Let h_j be the expected number of steps to reach n when starting at j

Then $h_j = \frac{1}{2}(h_{j-1} + 1) + \frac{1}{2}(h_{j+1} + 1) = \frac{h_{j-1} + h_{j+1}}{2} + 1 \Leftrightarrow h_j - h_{j+1} = h_{j-1} - h_j + 2$

Random walks for 2-SAT

- To compute h_j we need to solve a system of linear equations:

- $h_n=0$
- $h_j - h_{j+1} = h_{j-1} - h_j + 2$
- $h_0 - h_1 = 1$

- By induction, we have that $h_j - h_{j+1} = 2j+1$
- We'd like to determine

$$h_0 - h_n = \sum_{i=0}^{n-1} h_i - h_{i+1} = \sum_{i=0}^{n-1} 2i + 1 = 2\left(\frac{(n-1)n}{2}\right) + n = n^2$$

Random walks for 2-SAT

- So, if we run the algorithm for $2\ n^2$ steps, by Markov's inequality, the probability of not finding a satisfying assignment is at most ½
- If we run for $2C\ n^2$ steps, the failure probability is at most 2^{-C} .
- The algorithm can also be implemented in polynomial time.

Random walks in undirected graph

- A random walk on an undirected graph can be seen as the movement of a particle between the vertices (states)of an undirected graph G=(V,E)
- If the particle is at vertex *i,* then the probability to move to vertex *j* on edge *(i,j)* is $\frac{1}{\mathrm{d(i)}}$ where d(i) is the degree of vertex *i*
- A random walk on an undirected graph is aperiodic if and only if the graph is not bipartite (it does not have odd cycles)
- Since the graph is connected, there exists a stationary distribution and it is unique.

Stationary distribution of a random walks in an undirected graph

Theorem: A random walk on G converges to a stationary distribution $\vec{\pi}$ where $\pi_i = \frac{d(i)}{2E}$.

Proof: Since $\sum_i \pi_i = \sum_i \frac{d(i)}{2E} = 1, \vec{\pi}$ is a probability distribution.

We also need to prove that $\vec{\pi}$ P= $\vec{\pi}$:

$$\pi_i = \sum_{j \in N(i)} \frac{d(j)}{2|E|} \frac{1}{d(j)} = \frac{d(i)}{2|E|}$$

Hitting time

$h_{i,j}$: expected number of steps to reach j from i

Corollary: For any vertex i in G: $h_{i,i} = \frac{2\,|E|}{d(i)}$.

Lemma: For every edge (i,j)$\in E$, $h_{i,j} < 2\,|E|$.

We compute $h_{i,i}$ in two different ways:

$$\frac{2\,|E|}{d(i)} = h_{i,i} = \frac{1}{d(i)} \sum_{j \in N(i)} (1 + h_{j,i}) \Longrightarrow 2\,|E| = \sum_{j \in N(i)} (1 + h_{j,i})$$

and therefore $h_{j,i} < 2\,|E|$.

Cover time of a random walk

- Definition: The cover time of a graph G=(V,E) is the maximum over all vertices *i* in V of the expected time needed to visit all the nodes of the random walk starting at vertex *i.*

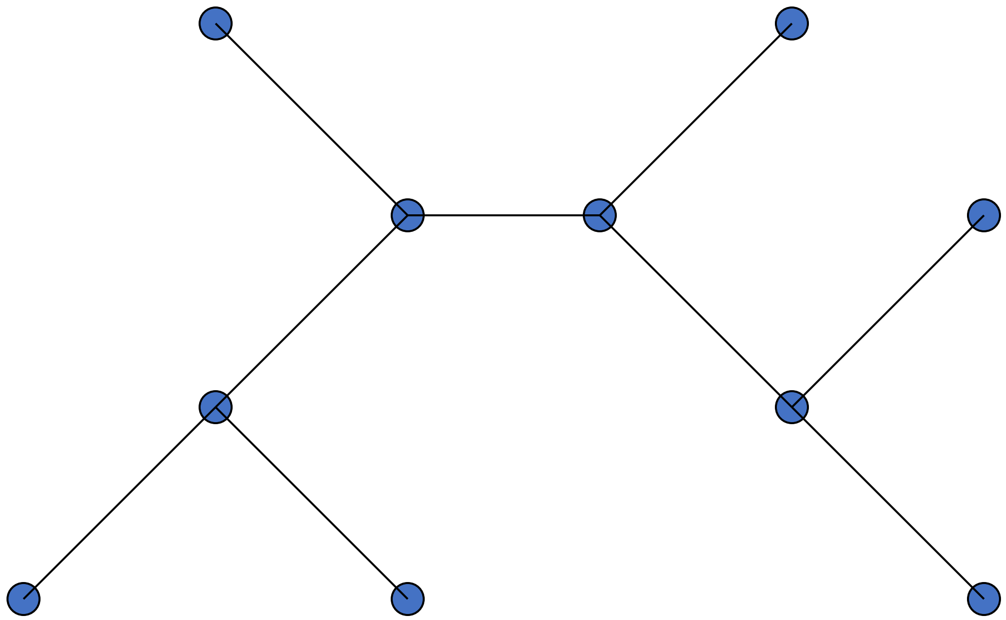
- Lemma: the cover time of G=(V,E) is bounded by 4 |V| |E|.

Proof: Find a spanning tree of the graph and duplicate all the edges in order to construct an Eulerian tour that visits all the 2(|V|-1) vertices v_0, v_1, …., $v_{2|V|-2}$. The expected time to go through all the vertices in the Eulerian tour is an upper bound on the cover time

$$\sum_{i=0}^{2|V|-3} h_{v_i v_{i+1}} < (2|V|-2)(2|E|) < 4\,|V||E|$$

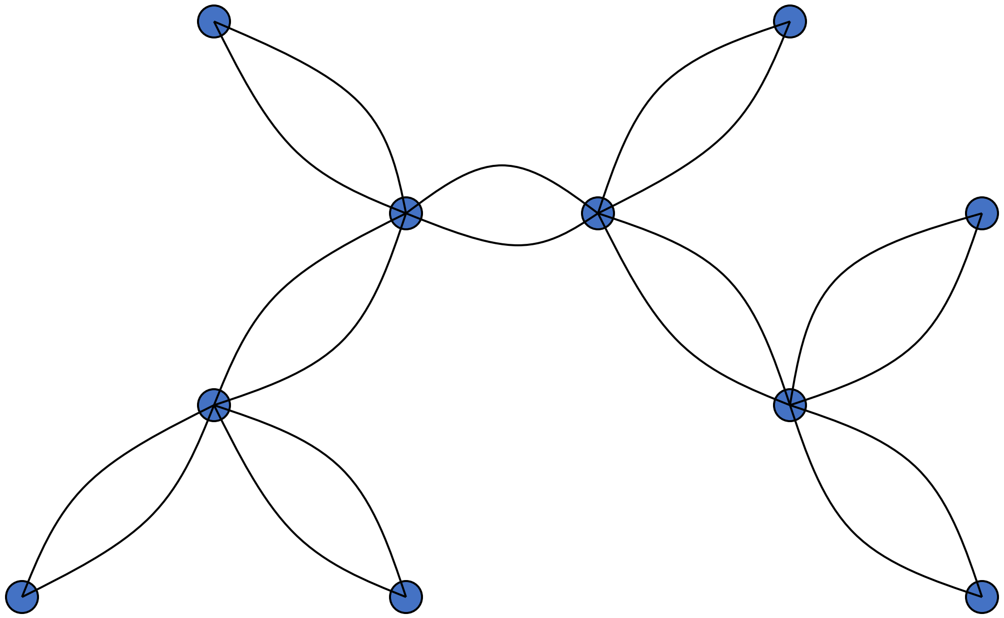
Cover time of a random walk

Find a spanning tree of the graph



Cover time of a random walk

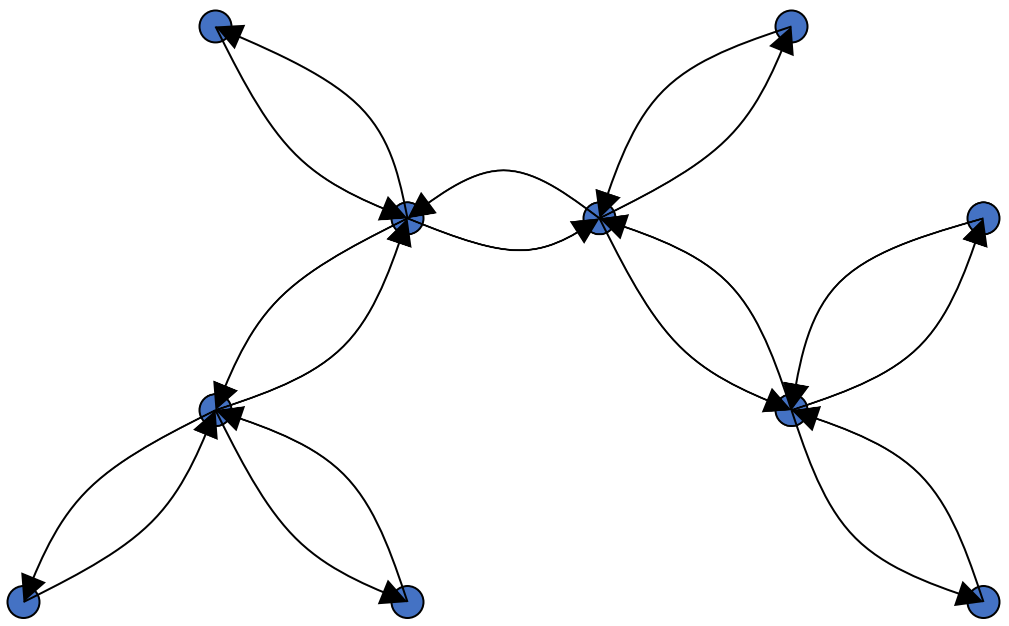
How to formalize the idea of "following" a minimum spanning tree?



Key idea: double all the edges and find an Eulerian tour.

Spanning Tree and TSP

How to formalize the idea of "following" a minimum spanning tree?



Key idea: double all the edges and find an Eulerian tour.

There always exists an Eulerian in Graph of vertices with even degree

s-t connectivity in a Graph

- The standard BFS algorithm for s-t connectivity finds a path from *s* to *t* *if* it *exists* using $\Omega(n)$ space
- We give a algorithms based on random walk that uses only O(log n) space.
- s-t Connectivity Algorithm:

  1. Start a random walk from s.

  2. If the walk reaches t within $4n^3$ steps, return that there is a path. Otherwise return that there is no path.

s-t connectivity in a Graph

- Theorem: The s-t connectivity algorithm returns the correct answer with probability 1/2, and it only errs by returning that there is no path from *s* to *t* when there is such a path.
- Proof: The expected time to reach t from s (if there is a path) is bounded from above by the cover time of their shared component, which is at most $4nm < 2n^3$.

By Markov's inequality, the probability that a walk takes more than $4n^3$ steps to reach s from t is at most 1/2.

s-t connectivity in a Graph

- The algorithm must keep track of its current position, which takes O(log n) bits.

- The algorithm also needs to track the number of steps taken in the random walk, which also takes O(log n) bits;

- This is all the memory required as long there exists a mechanism for selecting a random neighbour in the random walk.