Markov Chains
Two applications

Anna Karlin
Some slides by Alex Tsun

Agenda

- Discrete-Time Stochastic Processes
- Review Markov Chains
- Pagerank
- The metropolis algorithm

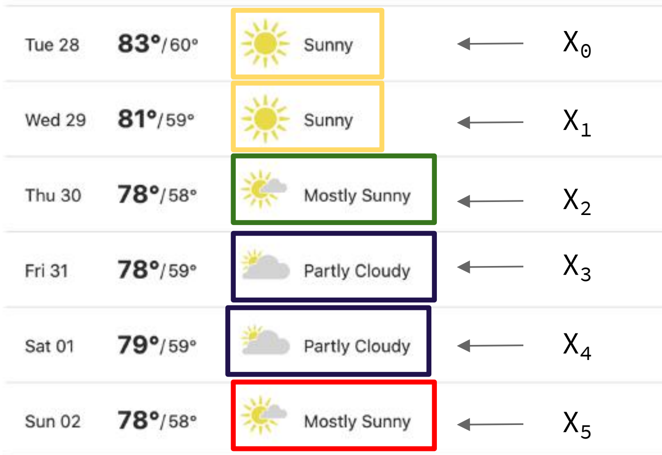
Discrete-Time Stochastic Processes

Discrete-Time Stochastic Process (DTSP): A discrete-time stochastic process is a sequence of random variables $X_0, X_1, X_2, \ldots$ where X_t is the value at time t.

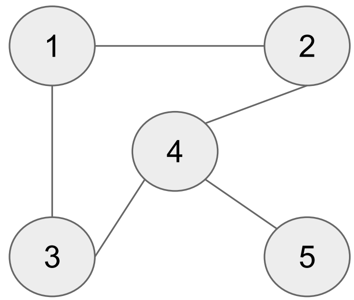
Examples:

- The temperature in Seattle each day.
- The stock price of TESLA each day.
- Your position on a one-dimensional number line during a "random walk".
- The Bernoulli process (at each time step, flip a coin independently).

Discrete-Time Stochastic Processes

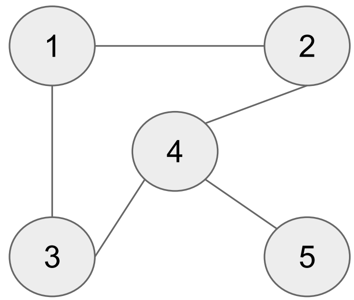


Random Walk on a Graph Example



Suppose we start at node 1, and at each time step, independently step to a neighboring node with equal probability.

Random Walk on a Graph Example

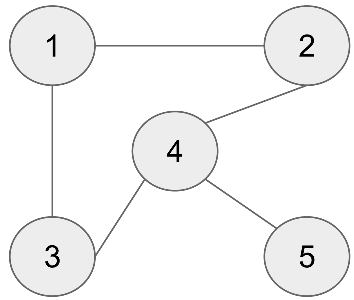


Suppose we start at node 1, and at each time step, independently step to a neighboring node with equal probability.

For example, $X_0 = 1$ since at time $t = 0$ we are at node 1. Then, X_1 is either 2 or 3 (but cannot be 4 or 5 since not neighbors of node 1). And so on.

This DTSP is actually a special type of DTSP called a **Markov Chain**.

Random Walk on a Graph Example



Three Key Properties:

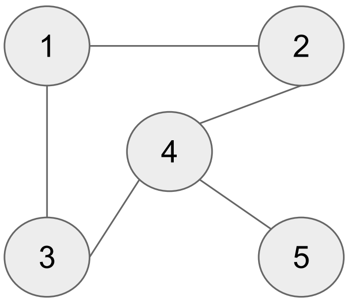
1. We only have finitely many states (we have 5 in this example: $\{1,2,3,4,5\}$).
2. We don't care about the past, given the present. That is, the distribution of where we *go* next **ONLY** depends on where we are ***currently***, and not any past history.
3. The transition probabilities are the **same** at each time step (e.g., if we are at node 1 at time $t = 0$ or $t = 152$, we always are equally likely to go to node 2 or 3).

Markov Chains

Discrete time stochastic process $X_1, X_2, \ldots, X_t, \ldots$
Such that

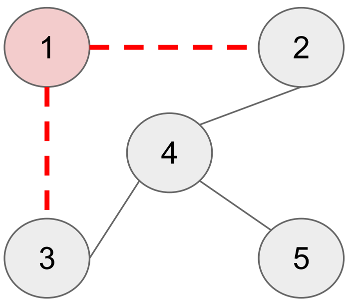
- There is a state space S, such that each X_t is in S.
- Pr $(X_{t+1}$ = j | X_t = i, history) = Pr $(X_{t+1}$ = j | X_t = i)
- p_{ij} = Pr $(X_{t+1}$ = j | X_t = i) for all i and j specified by transition probability matrix (TPM) P

Transition Probability Matrix Example



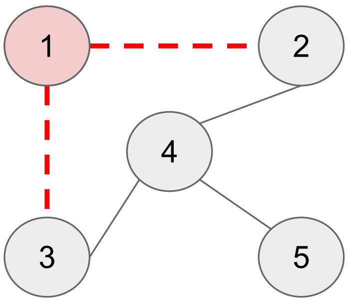
	To s_1	To s_2	To s_3	To s_4	To s_5
From s_1					
From s_2					
From s_3					
From s_4					
From s_5					

Transition Probability Matrix Example



	To s_1	To s_2	To s_3	To s_4	To s_5
From s_1					
From s_2					
From s_3					
From s_4					
From s_5					

Transition Probability Matrix Example

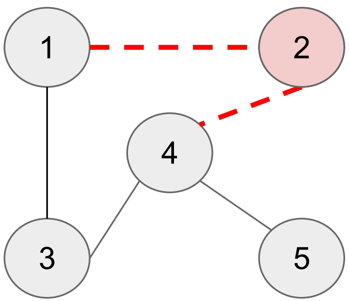


	To s_1	To s_2	To s_3	To s_4	To s_5
From s_1	0	1/2	1/2	0	0
From s_2					
From s_3					
From s_4					
From s_5					

$P(X_{t+1} = 3 \mid X_t = 1)$

$P(X_{t+1} = 2 \mid X_t = 1)$

Transition Probability Matrix Example

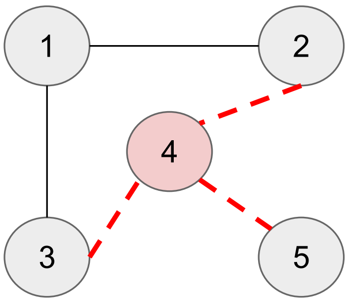


	To s_1	To s_2	To s_3	To s_4	To s_5
From s_1	0	1/2	1/2	0	0
From s_2	1/2	0	0	1/2	0
From s_3					
From s_4					
From s_5					

$P(X_{t+1} = 4 \mid X_t = 2)$

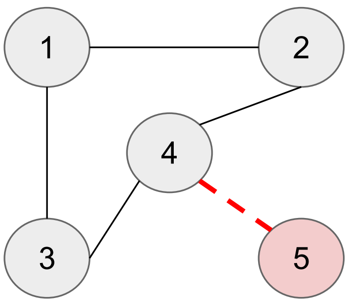
$P(X_{t+1} = 1 \mid X_t = 2)$

Transition Probability Matrix Example



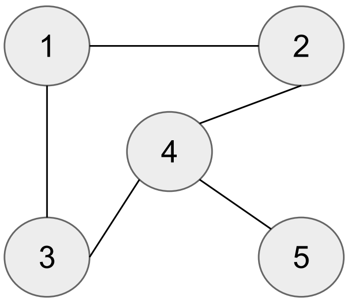
	To s_1	To s_2	To s_3	To s_4	To s_5
From s_1	0	1/2	1/2	0	0
From s_2	1/2	0	0	1/2	0
From s_3	1/2	0	0	1/2	0
From s_4	0	1/3	1/3	0	1/3
From s_5					

Transition Probability Matrix Example



	To s_1	To s_2	To s_3	To s_4	To s_5
From s_1	0	1/2	1/2	0	0
From s_2	1/2	0	0	1/2	0
From s_3	1/2	0	0	1/2	0
From s_4	0	1/3	1/3	0	1/3
From s_5	0	0	0	1	0

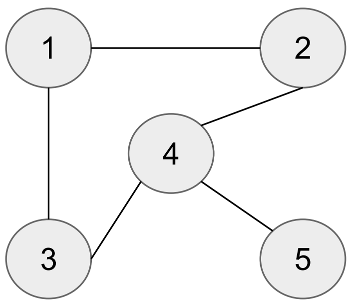
Transition Probability Matrix Example



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

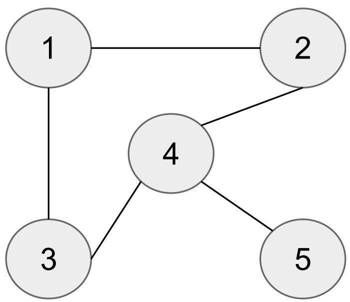
Computing Probability Example

$$P(X_2 = 3|X_0 = 2)$$



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

Computing Probability Example

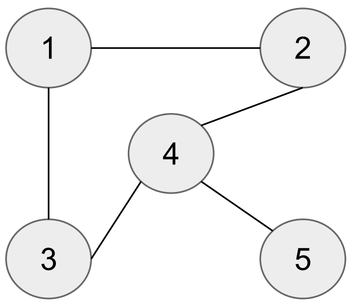


$$P(X_2 = 3 | X_0 = 2)$$

$$= \sum_{i=1}^{5} P(X_2 = 3 | X_0 = 2, X_1 = i) P(X_1 = i | X_0 = 2) \text{ [LTP]}$$

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

Computing Probability Example



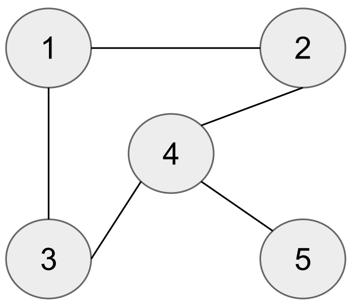
$$P(X_2 = 3|X_0 = 2)$$

$$= \sum_{i=1}^{5} P(X_2 = 3|X_0 = 2, X_1 = i)P(X_1 = i|X_0 = 2) \text{ [LTP]}$$

$$= \sum_{i=1}^{5} P(X_2 = 3|X_1 = i)P(X_1 = i|X_0 = 2) \text{ [Markov property]}$$

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

Computing Probability Example



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

$$P(X_2 = 3 | X_0 = 2)$$

$$= \sum_{i=1}^{5} P(X_2 = 3 | X_0 = 2, X_1 = i) P(X_1 = i | X_0 = 2) \text{ [LTP]}$$

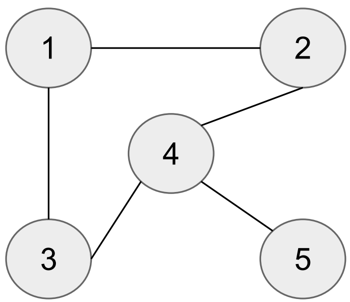
$$= \sum_{i=1}^{5} P(X_2 = 3 | X_1 = i) P(X_1 = i | X_0 = 2) \text{ [Markov property]}$$

$$= P(X_2 = 3 | X_1 = 1) P(X_1 = 1 | X_0 = 2) + P(X_2 = 3 | X_1 = 4) P(X_1 = 4 | X_0 = 2)$$

$$= P_{13}P_{21} + P_{43}P_{24} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{12}$$

Random Picture

Applying A Transition Matrix



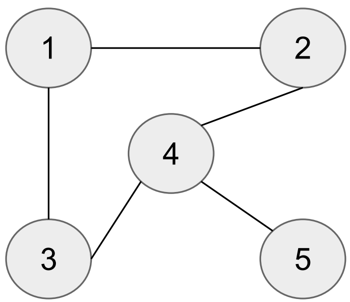
Suppose we start at a random state with these probabilities (sum to 1).

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

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

Applying a Transition Matrix



Suppose we start at a random state with these probabilities (sum to 1).

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

What is my belief distribution for where I am next?

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

Applying A Transition Matrix

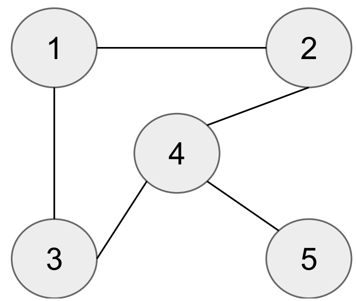
Let’s compute the matrix product vP (we’ll see why soon!).

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

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1} v_i, \quad \right]$$



Applying A Transition Matrix

Let's compute the matrix product vP (we'll see why soon!).

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

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1} v_i, \quad \right]$$

$P(X_1 = 1)$

$$\sum_{i=1}^{5} P_{i1} v_i = \sum_{i=1}^{5} P(X_1 = 1 | X_0 = i) P(X_0 = i) = P(X_1 = 1)$$

Def of TPM (works for any time) | Def of v | LTP

Applying A Transition Matrix

Let's compute the matrix product vP (we'll see why soon!).

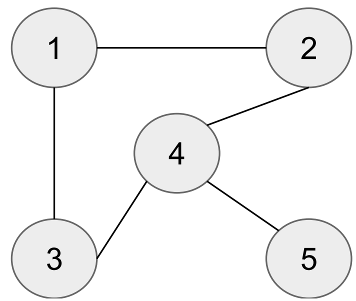
$$P = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1} v_i, \quad \sum_{i=1}^{5} P_{i2} v_i, \quad \right]$$

$P(X_1 = 1)$



Applying a Transition Matrix

Let's compute the matrix product vP (we'll see why soon!).

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

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1} v_i, \quad \sum_{i=1}^{5} P_{i2} v_i, \quad \right]$$

$P(X_1 = 1)$ $P(X_1 = 2)$

$$\sum_{i=1}^{5} P_{i2} v_i = \sum_{i=1}^{5} P(X_1 = 2 | X_0 = i) P(X_0 = i) = P(X_1 = 2)$$

Def of TPM (works for any time) — Def of v — LTP

Applying a Transition Matrix

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

Let's compute the matrix product vP (we'll see why soon!).

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 =$

$v = [0.25,\quad 0.45,\quad 0.15,$

HMMMM

$P_{i2}v_i,$ $]$

$= 2)$

$$\sum_{i=1}^{5} P_{i2}v_i = \sum_{i=1}^{5} P(X_1 = 2 | X_0 = i)P(X_0 = i) = P(X_1 = 2)$$

Def of TPM (works for any time) Def of v LTP

Applying A Transition Matrix

Let's compute the matrix product vP (we'll see why soon!).

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

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1}v_i, \quad \sum_{i=1}^{5} P_{i2}v_i, \quad \sum_{i=1}^{5} P_{i3}v_i, \quad \sum_{i=1}^{5} P_{i4}v_i, \quad \sum_{i=1}^{5} P_{i5}v_i\right]$$

$P(X_1 = 1)$ $P(X_1 = 2)$ $P(X_1 = 3)$ $P(X_1 = 4)$ $P(X_1 = 5)$

Right-multiplying by P gives the belief distribution at the next time step!

Applying a Transition Matrix

Let's compute the matrix product vP (we'll see why soon!).

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

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

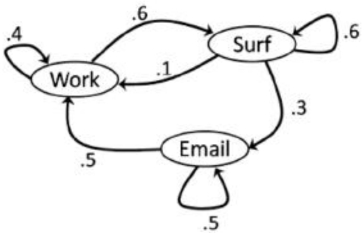
$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1} v_i, \quad \sum_{i=1}^{5} P_{i2} v_i, \quad \sum_{i=1}^{5} P_{i3} v_i, \quad \sum_{i=1}^{5} P_{i4} v_i, \quad \sum_{i=1}^{5} P_{i5} v_i\right]$$

$P(X_1 = 1)$ $P(X_1 = 2)$ $P(X_1 = 3)$ $P(X_1 = 4)$ $P(X_1 = 5)$

Right-multiplying by P gives the belief distribution at the next time step!

By induction (repeatedly applying this matrix P), vP^t is the distribution after t time steps.



$$P^2 = \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .22 & .6 & .18 \\ .25 & .42 & .33 \\ .45 & .3 & .25 \end{pmatrix} \end{array}$$

$$P^3 = \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .238 & .492 & .270 \\ .307 & .402 & .291 \\ .335 & .450 & .215 \end{pmatrix} \end{array}$$

$$P^{10} \approx \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .2940 & .4413 & .2648 \\ .2942 & .4411 & .2648 \\ .2942 & .4413 & .2648 \end{pmatrix} \end{array}$$

$$P^{30} \approx \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .29411764705 & .44117647059 & .26470588235 \\ .29411764706 & .44117647058 & .26470588235 \\ .29411764706 & .44117647059 & .26470588235 \end{pmatrix} \end{array}$$

$$P^{60} \approx \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

Applying A Transition Matrix

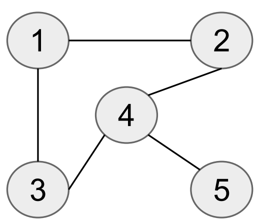
$$P^t = \begin{bmatrix} 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.3 & 0.1 \end{bmatrix}$$

$P(X_0 = 1)$ $P(X_0 = 2)$ $P(X_0 = 3)$ $P(X_0 = 4)$ $P(X_0 = 5)$

$$v = [0.25, \quad 0.45, \quad 0.15, \quad 0.05, \quad 0.10]$$

$$\left[\sum_{i=1}^{5} P_{i1} v_i, \quad \sum_{i=1}^{5} P_{i2} v_i, \quad \sum_{i=1}^{5} P_{i3} v_i, \quad \sum_{i=1}^{5} P_{i4} v_i, \quad \sum_{i=1}^{5} P_{i5} v_i\right]$$

$P(X_1 = 1)$ $P(X_1 = 2)$ $P(X_1 = 3)$ $P(X_1 = 4)$ $P(X_1 = 5)$



Applying A Transition Matrix

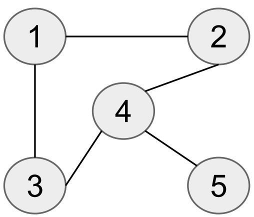
Let's compute the matrix product vP (we'll see why soon!).

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

$P(X_t = 1)$	$P(X_t = 2)$	$P(X_t = 3)$	$P(X_t = 4)$	$P(X_t = 5)$
(0.2	0.2	0.2	0.3	0.1.)

$$\left[\sum_{i=1}^{5} P_{i1} v_i , \quad \sum_{i=1}^{5} P_{i2} v_i , \quad \sum_{i=1}^{5} P_{i3} v_i , \quad \sum_{i=1}^{5} P_{i4} v_i , \quad \sum_{i=1}^{5} P_{i5} v_i \right]$$

$P(X_1 = 1)$ $P(X_1 = 2)$ $P(X_1 = 3)$ $P(X_1 = 4)$ $P(X_1 = 5)$



Stationary Distribution of a Markov Chain

<u>**Stationary Distribution**</u>: The stationary distribution of a Markov Chain with n states (which doesn't always exist), is the n-dimensional row vector π (which must be a probability distribution – nonnegative and sums to 1), such that

$$\pi P = \pi$$

Intuitively, it means that the distribution at the next time step is the same as the distribution at the current time step. This typically happens after a "long time" (called the **mixing time**) in the process (meaning after lots of transitions are taken).

Stationary Distribution of a Markov Chain

Fundamental Theorem of Markov Chains

Let P be the TPM for a Markov chain. Under minor technical conditions, including ”connectivity”

$$\lim_{t \to \infty} P^t_{ij} = \pi_j \qquad \forall i, j$$

where

$$\pi_j = \sum_{i=1}^{n} \pi_i p_{ij} \qquad j = 1, \ldots, n$$
$$\pi_1 + \pi_2 + \ldots + \pi_n = 1$$

Application 1: PageRank

1997

- Bill Clinton in White House
- Deep Blue beat world chess champion (Kasparov)

And the Internet kind of sucked

Nov ‘97: only one of the top 4 commercial search engines actually *found itself* when you search

The problem

Search engines worked by matching words

Top search for Bill Clinton

- `Bill Clinton Joke of the Day’ Website

Deeply susceptible to spammers and advertisers

How to fix?

Collect pages with decent textual match

Then **rank** them by some measure of ‘quality’ or ‘authority’.

Enter two groups:

- Jon Kleinberg (prof at Cornell)
- Larry Page and Sergey Brin (Ph.D. students at Stanford)

Both groups had the same brilliant idea

Larry Page and Sergey Brin (Ph.D. students at Stanford)

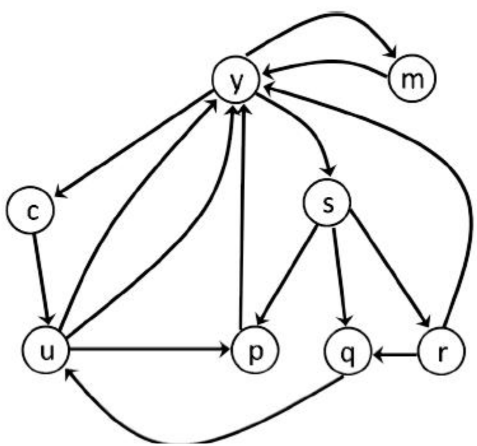
- Took the idea and founded Google, making billions

Jon Kleinberg (professor at Cornell)

- MacArthur Genius Prize, Nevanlina Prize and many other academic honors.

Pagerank

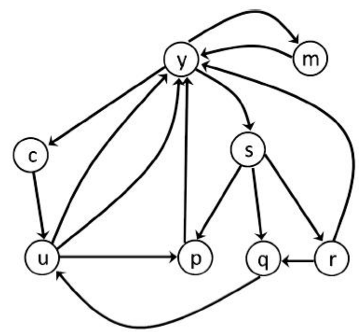
- Key idea is hyperlink analysis: take into account the directed graph structure of the web



Pagerank

Idea 1: “Citations”

As with academic publishing, it’s a good idea to think of each link to a page as a “citation” or “vote of quality”.



Rank pages by in-degree?

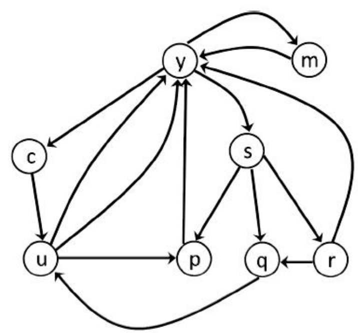
Problems with ranking pages by indegree

- Spamming
- Some linkers not discriminating
- Not all links created equal.

Perhaps we should weight the links somehow and then use the weights of the in-links to rank pages.

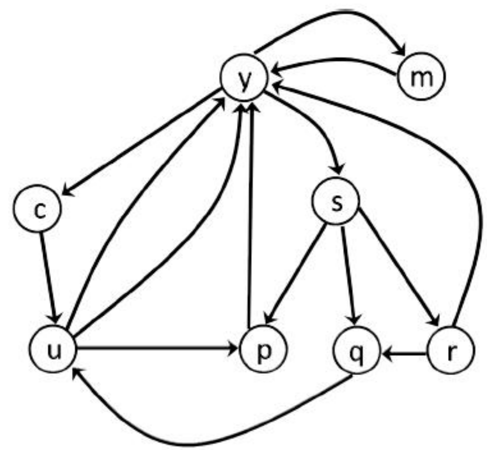
Inching towards Pagerank

- Web page has **high quality** if it's linked to by lots of **high quality** pages.
- A page is high quality if it links to lots of **high quality** pages.
- So kind of a recursive definition



Inching towards Pagerank

- If web page x has d outgoing links, one of which goes to y, then this contributes 1/d to the importance of y.
- But, we want to take importance/quality of x into account.
- Recursive definition.



These are the equations $\boldsymbol{qP = q}$

- Look familiar?

$$\boldsymbol{\pi P = \pi}$$

- Stationary distribution for what Markov chain?
- The Markov chain of a random surfer!!

Some issues

- Dangling nodes (dead ends)

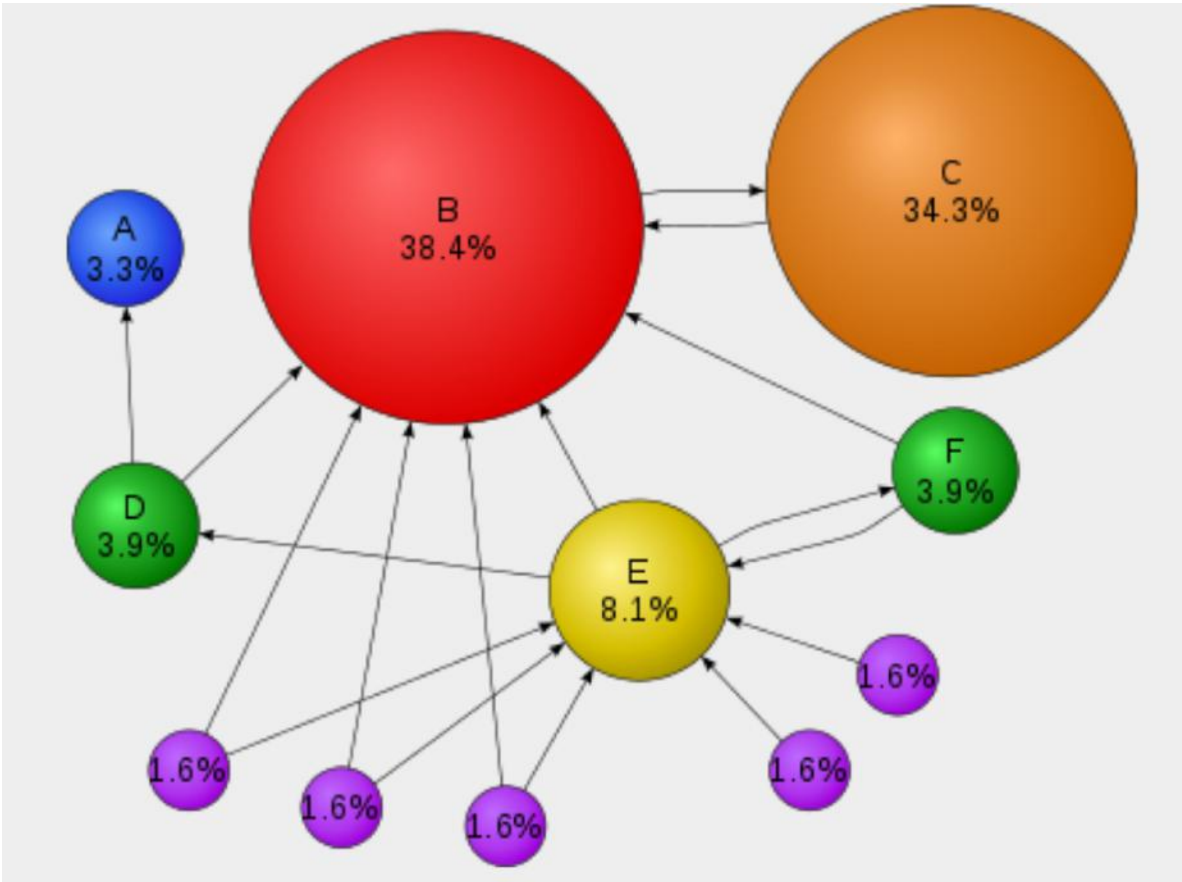
- Rank sinks - group of pages that only link to each other.

Final random surfer model

- Markov chain whose states are all the web pages in the world
- At each step, random surfer is looking at some web page. Next:
    - With probability p, follow a random link on that web page.
    - With probability 1-p, go to a uniformly random page on the web.

Compute stationary distribution for this Markov chain. Define Pagerank, i.e., ``quality" of a web page x to be its stationary probability π_x

On a query, return pages with "good textual match".
Rank them by their Pagerank.


A
3.3%
B
38.4%
C
34.3%
D
3.9%
E
8.1%
F
3.9%
1.6%
1.6%
1.6%
1.6%
1.6%
1.6%

Nowadays

- Tons more secret sauce to ranking search results…

Monte Carlo Methods

Collection of tools for estimating values through random sampling.

One of the most interesting tools in this toolbox:
Markov Chain Monte Carlo. (MCMC)

Example: The Independent set Problem

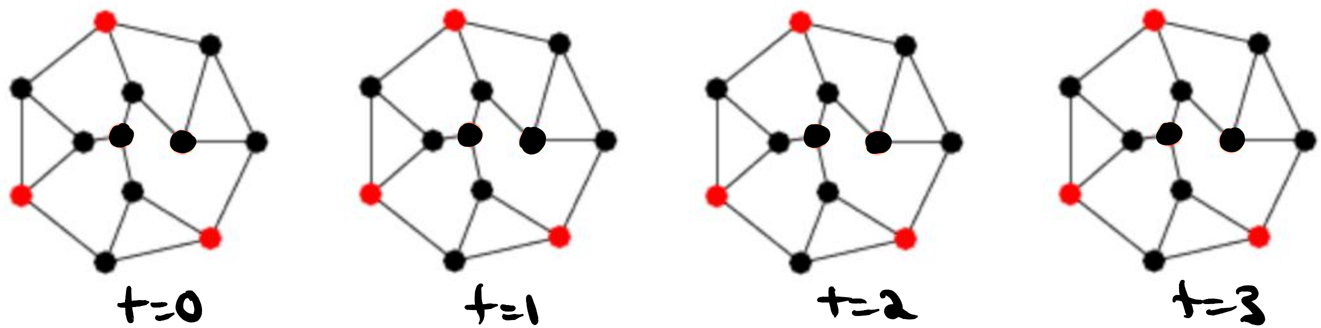
Given a graph G= (V,E), find the maximum independent set.

An independent set is a subset of vertices with no edges between them.

Finding a maximum independent set is a notoriously hard NP-complete problem.

Solving The Independent set Problem using MCMC?

Define a Markov chain on independent sets.

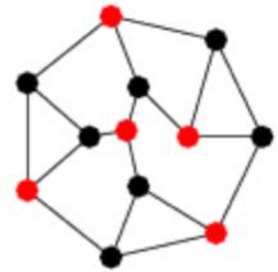


X_t is an independent set.

Choose v uniformly at random. If v $\in X_t$, remove it.

Otherwise, if X_t U v is independent, set $X_{t+1} := X_t$ U v.

What I really want



A Markov chain on independent sets such that:

* Given current state, can efficiently compute next state.

* Has a good stationary distribution (high stationary probability on independent sets that are large).

- Converges fast to its stationary distribution.

For example: might want π_I proportional to $2^{|I|}$

THIS SEEMS INSANELY DIFFICULT! Can't even compute π_I

The metropolis algorithm

Suppose you have a set of states S

And a desired stationary distribution π

Can you design a Markov Chain on S with stationary distribution π ?

This is what the Metropolis algorithm does.

The metropolis algorithm

Suppose you have a set of states S

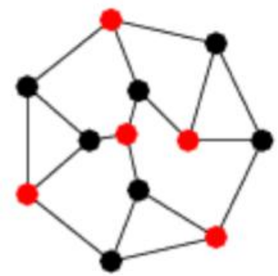
And a desired stationary distribution π

Can you design a Markov Chain on S with stationary distribution π ?

This is what the Metropolis algorithm does for us.

* Only needs to know ratios between stationary probabilities!

How it works



There is an underlying navigation graph.

The Metropolis Algorithm:

Start in some arbitrary state U_0.

for $t = 1, 2, 3, \ldots$

Pick a random neighbor v of U_{t-1} in the navigation graph.

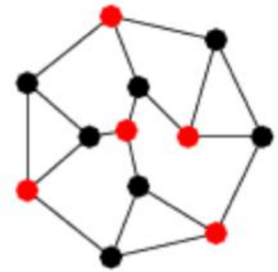
Compute the "acceptance ratio" $A = \pi[v]/\pi[U_{t-1}]$

If $A > 1$, set $U_t \leftarrow v$.

If $A \leq 1$, set $U_t \leftarrow v$ with probability A. Otherwise, set $U_t \leftarrow U_{t-1}$.

This Markov chain has stationary distribution $\pi[I]$ proportional to $2^{|I|}$

Metropolis Alorithm gets us partway



A Markov chain on independent sets such that:

* Given current state, can efficiently compute next state.

* Has a good stationary distribution: can set π_I proportional to $2^{|I|}$

- Converges fast to its stationary distribution????

Turns out to be of great importance to statistical physicists.