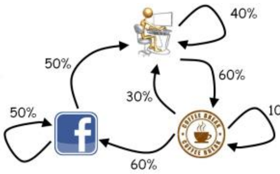
Great Theoretical Ideas In Computer Science

Victor Adamchik

CS 15-251

Carnegie Mellon University

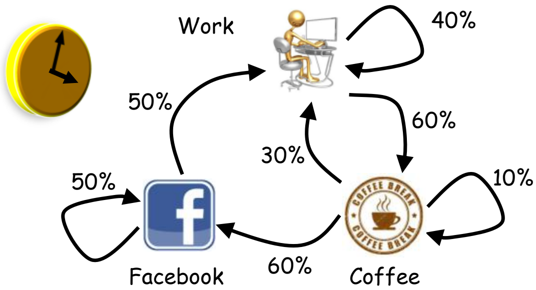
Markov Chains, Random Walks



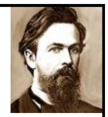
Outline

Markov Chains
Transition matrix
Invariant distribution
PageRank
Random walk on graphs
Randomized Algorithm

A day of my life

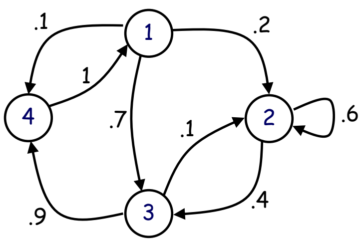


Markov Chain – Definition



- Directed graph, self-loops OK
- Always assumed strongly connected in 251
- Each edge labeled by a positive probability
- At each node ("state"), the probabilities on outgoing edges sum up to 1.
- The process starts in one of these states and moves successively from one state to another

Markov Chain



In 1907, A. A. Markov began the study of chance process. In this process, all of the past outcomes could influence our predictions for the next experiment. This type of process is called a Markov chain

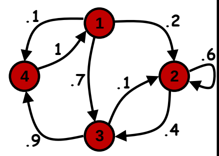
Markov Chain – Notation

Suppose there are n states.

n×n *transition* matrix M:

$M_{i,j}$= Pr [i → j in 1 step]

$$M = \begin{pmatrix} 0 & .2 & .7 & .1 \\ 0 & .6 & .4 & 0 \\ 0 & .1 & 0 & .9 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$



Rows sum to 1
("stochastic matrix")

Markov Chain – Notation

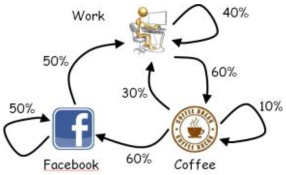
For time t = 0, 1, 2, 3, …

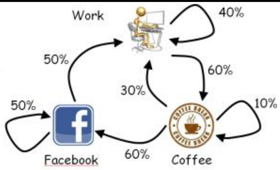
X_t denotes the state (node) at time t.

Somebody decides on X_0.

Then $X_1, X_2, X_3, \dots$ are random variables.

$X_0 = W$
$X_1 = C$
$X_2 = C$
$X_3 = W$
and so on





$$M = \begin{array}{c} \\ W \\ C \\ F \end{array} \begin{array}{c} \begin{array}{ccc} W & C & F \end{array} \\ \begin{pmatrix} .4 & .6 & 0 \\ .3 & .1 & .6 \\ .5 & 0 & .5 \end{pmatrix} \end{array}$$

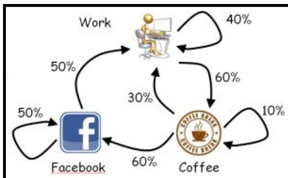
What is Pr $[X_1 = W \mid X_0 = C]$?

Pr $[X_1 = C \mid X_0 = W]$ = .6

Pr $[X_1 = C \mid X_0 = F]$ = 0

Pr $[X_6 = W \mid X_5 = C]$ = .3

Pr $[X_{t+1} = j \mid X_t = i]$ = $M[i,j]$



$$M = \begin{array}{c} \\ W \\ C \\ F \end{array} \begin{array}{c} \begin{array}{ccc} W & C & F \end{array} \\ \begin{pmatrix} .4 & .6 & 0 \\ .3 & .1 & .6 \\ .5 & 0 & .5 \end{pmatrix} \end{array}$$

What is Pr $[X_2 = W \mid X_0 = C]$?

Conditioning on X_1, using Law of Total Probability

Pr $[X_2 = W \mid X_0 = C]$ =

Pr $[X_1 = W \mid X_0 = C] \cdot$ Pr $[X_2 = W \mid X_1 = W]$

+ Pr $[X_1 = C \mid X_0 = C] \cdot$ Pr $[X_2 = W \mid X_1 = C]$

+ Pr $[X_1 = F \mid X_0 = C] \cdot$ Pr $[X_2 = W \mid X_1 = F]$

$= .3 \cdot .4 + .1 \cdot .3 + .6 \cdot .5 = .45$

In general, what is Pr $[X_2 = j \mid X_0 = i]$?

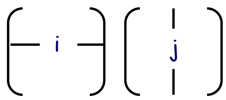
Conditioning on X_1, using Law of Total Prob…

$$\Pr[X_2=j|X_0=i] = \sum_{k=1}^{n} \Pr[X_1 = k \mid X_0 = i]\Pr[X_2 = j \mid X_1 = k] =$$

$$= \sum_{k=1}^{n} M[i,k]\,M[k,j]$$

$$= M^2[i,j]$$

Matrix multiplication



i's row j's column

What is $\Pr[X_3 = j \mid X_0 = i]$?

Conditioning on X_2, using Law of Total Prob…

$$\sum_{k=1}^{n} \Pr[X_2 = k \mid X_0 = i]\Pr[X_3 = j \mid X_2 = k] =$$

$$= \sum_{k=1}^{n} M^2[i,k]\,M[k,j] = M^3[i,j]$$

In general, Pr $[X_n = j \mid X_0 = i] = M^n[i, j]$.

M^n gives the probability that the Markov chain starting at state i will be in state j after n steps.

A random initial state

Often assume the initial state X_0 is also chosen randomly in some way…

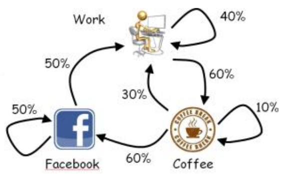
e.g., $X_0 \sim$

W	C	F
50%	20%	30%

a distribution vector (nonnegative, adds to 1)

distribution vector for X_0 usually denoted π_0

$X_0 \sim \pi_0 = \begin{bmatrix} W & C & F \\ 50\% & 20\% & 30\% \end{bmatrix}$



$$M = \begin{array}{c|ccc} & W & C & F \\ \hline W & .4 & .6 & 0 \\ C & .3 & .1 & .6 \\ F & .5 & 0 & .5 \end{array}$$

$\Pr[X_1 = W] = .5 \cdot .4 + .2 \cdot .3 + .3 \cdot .5 = .41$

Conditioning on X_0, using Law of Total Prob…

In general, if $X_0 \sim \pi_0$, what is $\Pr[X_1 = j]$?

Conditioning on X_0, using Law of Total Prob…

$$\sum_{k=1}^{n} \Pr[X_0 = k] \Pr[X_1 = j \mid X_0 = k] =$$

$$= \sum_{k=1}^{n} \pi_0[k] M[k,j] = (\pi_0 \bullet M)[j]$$

$(\ \pi_0\ )$ vector $\begin{pmatrix} M \end{pmatrix}$ matrix

I.e., the distribution vector for X_1 is $\pi_1 = \pi_0 \cdot M$

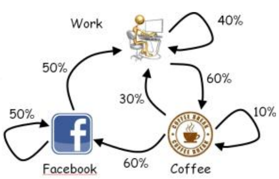
And, the distribution vector for X_n is $\pi_n = \pi_0 \cdot M^n$

Let M be the transition matrix of a Markov chain, and let is π_0 be the probability vector which represents the starting distribution. Then the probability that the chain is in state i after n steps is the i-th entry in the vector

$$\pi_n = \pi_0 \cdot M^n$$

The Invariant Distribution

(aka the Stationary Distribution)



$$M = \begin{array}{c|ccc} & W & C & F \\ \hline W & .4 & .6 & 0 \\ C & .3 & .1 & .6 \\ F & .5 & 0 & .5 \end{array}$$

Recall: $M^n[i,j] = \Pr[i \to j \text{ in exactly } n \text{ steps}]$

$$M^2 = \begin{pmatrix} .34 & .3 & .36 \\ .45 & .19 & .36 \\ .45 & .3 & .25 \end{pmatrix} \qquad M^7 = \begin{pmatrix} .405413 & .269831 & .324756 \\ .405546 & .270497 & .323957 \\ .40528 & .27063 & .32409 \end{pmatrix}$$

$$M^{15} = \begin{pmatrix} .405405 & .27027 & .324324 \\ .405405 & .27027 & .324324 \\ .405405 & .27027 & .324324 \end{pmatrix}$$

What's up with this?

$$M^{15} = \begin{pmatrix} .405405 & .27027 & .324324 \\ .405405 & .27027 & .324324 \\ .405405 & .27027 & .324324 \end{pmatrix}$$

This limiting row (assuming the limit exists) is called the invariant distribution π.

"In the long run,
40.6% of the time I'm working,
27.0% of the time I'm on coffee break,
32.4% of the time I'm on Facebook."

Invariant Distribution Calculation

Raising M to a large power is annoying.

"π is invariant": if you start in this distribution and you take one more step, you're still in the distribution.

i.e., $\pi = \pi M$

For fixed M, this yields a system of equations.

$\pi = \pi M$

$$\begin{pmatrix} \pi[W] & \pi[C] & \pi[F] \end{pmatrix} = \begin{pmatrix} \pi[W] & \pi[C] & \pi[F] \end{pmatrix} \begin{pmatrix} .4 & .6 & 0 \\ .3 & .1 & .6 \\ .5 & 0 & .5 \end{pmatrix}$$

$\pi[W] = .4\ \pi[W] + .3\ \pi[C] + .5\ \pi[F]$
$\pi[C] = .6\ \pi[W] + .1\ \pi[C] + 0\ \pi[F]$
$\pi[F] = 0\ \pi[W] + .6\ \pi[C] + .5\ \pi[F]$

And we need to add another equation

(in order to get a unique solution)

$1 = \pi[W] + \pi[C] + \pi[F]$

$\pi = \pi M$

Solving the system in *Mathematica*, yields

```
Solve[{w == 0.3 c + 0.5 f + 0.4 w,
       c == 0.1 c + 0.6 w,
       f == 0.6 c + 0.5 f,
       1 == w + c + f}, {w, c, f}]
```

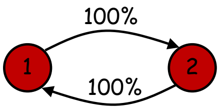
$\pi[W] = 0.405405$, $\pi[C] = 0.27027$, $\pi[F] = 0.324324$

Fundamental Theorem

Given a (finite, strongly connected) Markov Chain with transition matrix M, there is a unique invariant distribution π satisfying $\pi = \pi M$.

Fundamental Theorem

… unless the chain has some stupid "periodicity"



No limiting dist., but $\pi = (\frac{1}{2}\ \frac{1}{2})$ is still invariant.

Expected Time from u to u

In a Markov Chain with invariant distribution π, suppose $\pi[u] = 1/3$.

If you walked for N steps, you would expect to be at state u about N/3 times.

The average time between *successive visits* to u would be about $\frac{N}{N/3} = 3$.

Not hard to turn this into a theorem.

Mean First Recurrence Theorem

In a Markov Chain with invariant distribution π,

$$E[\#\text{ steps to from u to u}] = \frac{1}{\pi[u]}$$

Markov Chain Summary

M[i,j] = Pr[i → j in 1 step], transition matrix

$M^n[i,j]$ = Pr[i → j in exactly n steps]

If π_t is distribution at time t, $\pi_t = \pi_0 M^t$

∃ a unique *invariant distribution* π s.t. $\pi = \pi M$

$$\pi = \lim_{t \to \infty} \pi_t$$

E[# steps to go from u to u] =

Interlude: Altavista

1997: Web search was horrible. You search for "CMU", it finds all the pages containing "CMU" & sorts by # occurrences.

Interlude: PageRank

Sites should be considered important not only if they are linked to by many others, but also if they link to many others.

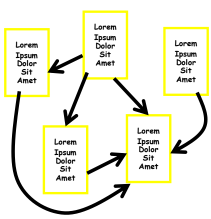
Page and Brin

Billionaires

Jon Kleinberg

Nevanlinna Prize, 10k euro

Interlude: PageRank



Measure importance with Random Surfer model:

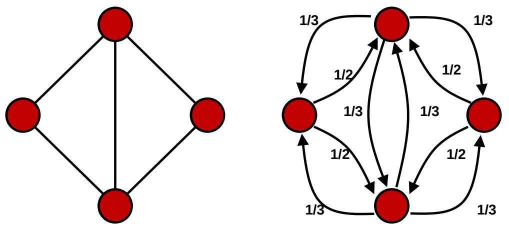
- Follows a random outgoing link with prob. α
- Jumps to a completely random page with prob. 1–α
- α is a parameter (≈ 85%)

PageRank: compute the invariant distribution π, rank pages u by highest $\pi[u]$ value!

Random walks on undirected graphs

Connected undirected graph.

Each step: go to a random neighbor.

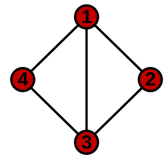


What is the transition matrix M?

What is the transition matrix M?

Adjacency matrix: degrees Transition matrix:

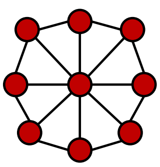
$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \begin{matrix} \div d_1 \\ \div d_2 \\ \div d_3 \\ \div d_4 \end{matrix} \qquad M = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$



What is the invariant distribution π?

Assuming no "stupid periodicity", same as the limiting distribution.

(periodicity iff bipartite, actually)



Higher degree
≡ higher limiting prob?

Could $\pi[u]$ just be proportional to degree d_u?

Theorem: In random walk on undirected graph $G=(n,m)$, inv. distribution $\pi = \begin{pmatrix} \frac{d_1}{2m} & \frac{d_2}{2m} & \cdots & \frac{d_n}{2m} \end{pmatrix}$

$(\Sigma d_i = 2m)$

Proof: We need to verify $\pi M = \pi$.

$$\pi M = \begin{pmatrix} \frac{d_1}{2m} & \frac{d_2}{2m} & \cdots & \frac{d_n}{2m} \end{pmatrix} \begin{pmatrix} a_{11}/d_1 & a_{12}/d_1 & \cdots & a_{1n}/d_1 \\ a_{21}/d_2 & a_{22}/d_2 & \cdots & a_{2n}/d_2 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}/d_n & a_{n2}/d_n & \cdots & a_{nn}/d_n \end{pmatrix}$$

Consider j's row:

$$\frac{d_1}{2m}\frac{a_{1j}}{d_1} + \frac{d_2}{2m}\frac{a_{2j}}{d_2} + \ldots + \frac{d_n}{2m}\frac{a_{nj}}{d_n} = \frac{1}{2m}\sum_{k=1}^{n} a_{kj} = \frac{d_j}{2m}$$

Thus,

$$\pi M = \begin{pmatrix} \frac{d_1}{2m} & \frac{d_2}{2m} & \cdots & \frac{d_n}{2m} \end{pmatrix} = \pi$$

Corollary

In random walk on undirected (connected) graph G,

$$E[\#\text{ steps to go from u to u}] = \frac{2m}{d_u}$$

Proof:

Mean first recurrence theorem:

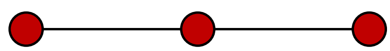
In a Markov Chain with invariant distribution π,

$$E[\#\text{ steps to go from u to u}] = \frac{1}{\pi[u]}$$

Examples

m=#edges

$$\pi = \begin{pmatrix} \frac{d_1}{2m} & \frac{d_2}{2m} & \cdots & \frac{d_n}{2m} \end{pmatrix}$$

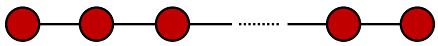


π:	1/4	1/2	1/4
E[v->v]:	4	2	4

Examples

$\pi = \left(\frac{d_1}{2m} \; \frac{d_2}{2m} \; \cdots \; \frac{d_n}{2m} \right)$

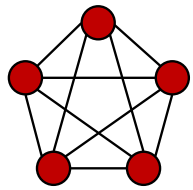
P_{n+1}, the path on n+1 nodes:



π:	$\frac{1}{2n}$	$\frac{2}{2n}$	$\frac{2}{2n}$	$\frac{2}{2n}$	$\frac{1}{2n}$
E[v->v]:	2n	n	n	n	2n

Examples

The clique on n nodes:



m = n(n-1)/2

d_v=n-1

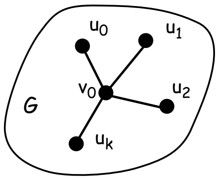
π = (1/n 1/n 1/n ··· 1/n)

E[v->v] = n

<u>Proposition:</u> Let (u_0,v_0) be an edge in G=(n,m).

E [# steps to go $u_0 \to v_0$] ≤ 2m–1

Proof: Suppose v_0 is connected to $u_0, u_1, \ldots, u_k$.



$\frac{2m}{d_{v0}} = E[\#\,steps\ v_0 \to u_0] =$

Use conditioning on the first step.

$= \sum_{i=0}^{k} Pr[v_0 \to u_i] \cdot E[\#\,steps\ v_0 \to u_0 \mid v_0 \to u_i]$

Drop all terms but i=0

$= \sum_{i=0}^{k} \frac{1}{d_{v0}} \cdot (1 + E[\#\,steps\ u_i \to v_0]) \geq \frac{1}{d_{v0}} \cdot (1 + E[\#\,steps\ u_0 \to v_0])$

<u>Theorem:</u> Let G=(n,m) be a connected graph. Let u and v be *any* two vertices. Then

E[# steps u->v] ≤ 2 m n ≤ n^3

Proof:

Pick a path u, w_1, w_2, …, w_r, v. At most n nodes.

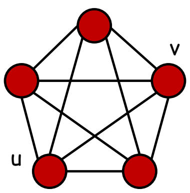
E[# of steps u→v] ≤ E[# of steps u→w_1→···→w_r→v]

= E[u→w_1] + E[w_1→w_2] + ··· + E[w_r→v]

≤ 2m + 2m + ··· + 2m ≤ 2 m n.

Examples

The clique on n nodes:



Thm: E [# steps to hit v starting from u] ≤ 2mn ≤ n^3

Actually: # steps to hit v starting from u ~ Geom, so expectation is n–1.

Connectivity problem

Given graph G, possibly disconnected, and two vertices u and v. Are u and v connected?

Easily solved in O(m) time using DFS/BFS.

Requires 'marking' nodes, hence ≥ n bits of memory need to be allocated.

Do it with O(1) memory.

A randomized algorithm

[Aleliunas,Karp,Lipton,Lovász,Rackoff in 1979]

one variable — four variables

z := u
for t = 1 ... $1000n^3$
  z := random-neighbor(z)
  if z = v, return "YES"
end for
return "NO"

```
for t0 = 1…1000
  for t1 = 1…n
    for t2 = 1…n
      for t3 = 1…n
```

couple more variables

Assume a variable can hold a number between 1 & n.

A randomized algorithm for CONN:

z := u
for t = 1 ... $1000n^3$
  z := random-neighbor(z)
  if z = v, return "YES"
end for
return "NO"

True answer is NO: alg. always says NO

True answer is YES: alg. says YES w/prob ≥ 99.9%

Why?

PROOF. Suppose u and v are indeed in the same connected component.

We do a random walk from u until we hit v.

Let T = # steps it takes, a random variable.

Then, $E[T] \le n^3$, by our theorem.

$$\Pr[T > 1000n^3] = ???$$

Markov's Inequality: $\Pr[X \ge c] \le \frac{E[X]}{c}$

$$\Pr[T > 1000n^3] \le \frac{n^3}{1000n^3} \le 0.1\%$$

Here's What You Need to Know…

Markov Chains
Transition matrix
Invariant distribution
PageRank
Random walk on graphs
Randomized Algorithm