Sequences I

Prof. Noah Snavely
CS1114
http://cs1114.cs.cornell.edu

Cornell University
Computer Science

Administrivia

- Assignment 5, due Friday, April 20th, 5pm
- Assignment 6 will be released early next week

Administrivia

- Final projects
  - Due on Tuesday, May 15 (tentative) via demo
  - Group project (groups of two)
  - Please form groups and send me a proposal for your final project by next Thursday, 4/19
  - Proposal should include:
    - Your group members
    - The problem you are going to solve
    - Any special equipment you need from us

Final project suggestions

- Find and follow moving objects in the world (or other robots)
- Coordinate robots to do something interesting (e.g., dance)
- Robot maze
- Build a musical instrument using robots
- Recognize a Sudoku puzzle from an image
- Automatic image colorization
- Anything else you want to do that involves implementing a non-trivial algorithm

- We'll have a demo session on the due date

New topic: modeling sequences

- Lots of interesting things in the world can be thought of as *sequences*

- Ordering of heads/tails in multiple coin flips
- Ordering of moves in rock/paper/scissors
- Text
- Music
- Closing stock prices
- Web pages you visit on Wikipedia

How are sequences generated?

- For some sequences, each element is generated *independently*
  - Coin flips

- For others, the next element is generated deterministically
  - 1, 2, 3, 4, 5, … ?

- For others, the next element depends on previous elements, but exhibits some randomness
  - The sequence of web pages you visit on Wikipedia
  - We'll focus on these (many interesting sequences can be modeled this way)

Markov chains


Andrei Markov

- A *sequence* of discrete random variables $\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n$

$$\boxed{\mathrm{x}_1} \rightarrow \boxed{\mathrm{x}_2} \rightarrow \boxed{\mathrm{x}_3} \rightarrow \boxed{\mathrm{x}_4} \rightarrow \boxed{\mathrm{x}_5}$$

  - x_t is the **state** of the model at time t
  - **Markov assumption**: each state is dependent only on the previous one
    - dependency given by a **conditional probability**:

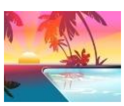
$$p(\mathrm{x}_t | \mathrm{x}_{t-1})$$

  - This is actually a *first-order* Markov chain
  - An *N'th-order* Markov chain: $p(\mathrm{x}_t | \mathrm{x}_{t-1}, \ldots, \mathrm{x}_{t-N})$

Markov chains

- Example: Springtime in Ithaca

  Three possible conditions: nice, rainy, snowy



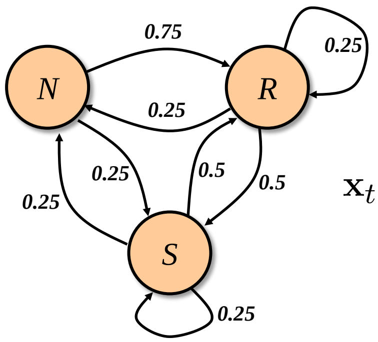
If it's nice today, then tomorrow it will be:
- rainy 75% of the time
- snowy 25% of the time

If it's rainy today, then tomorrow it will be:
- rainy 25% of the time
- nice 25% of the time
- snowy 50% of the time

If it's snowy today, then tomorrow it will be:
- rainy 50% of the time
- nice 25% of the time
- snowy 25% of the time

Markov chains

- Example: Springtime in Ithaca
- We can represent this as a kind of graph
- (N = Nice, S = Snowy, R = Rainy)



$\mathbf{x}_{t-1}$ \ $\mathbf{x}_t$	***N***	***R***	***S***
N	0.0	0.75	0.25
R	0.25	0.25	0.5
S	0.25	0.5	0.25

Transition probabilities

Markov chains

- Example: Springtime in Ithaca
- We can represent this as a kind of graph
- (N = Nice, S = Snowy, R = Rainy)

$\mathbf{x}_{t-1}$ \ $\mathbf{x}_t$	***N***	***R***	***S***
N	0.0	0.75	0.25
R	0.25	0.25	0.5
S	0.25	0.5	0.25

Transition probabilities

If it’s nice today, what’s the probability that it will be nice tomorrow?

If it’s nice today, what’s the probability that it will be nice the day after tomorrow?

Markov chains

$$\mathbf{P} = \begin{array}{c} \\ N \\ R \\ S \end{array}\begin{array}{c} \begin{array}{ccc} N & R & S \end{array} \\ \begin{bmatrix} 0.0 & 0.75 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} \end{array}$$

- The transition matrix at time *t*+2 is $\mathbf{P}^2$

$$\begin{bmatrix} 0.0 & 0.75 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}\begin{bmatrix} 0.0 & 0.75 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} = \begin{bmatrix} 0.2500 & 0.3125 & 0.4375 \\ 0.1875 & 0.5000 & 0.3125 \\ 0.1875 & 0.4375 & 0.3750 \end{bmatrix}$$

- The transition matrix at time *t*+n is $\mathbf{P}^n$

Markov chains

- What's will the weather be like in 20 days?

$$P^{20} = \begin{bmatrix} 0.2 & 0.44 & 0.36 \\ 0.2 & 0.44 & 0.36 \\ 0.2 & 0.44 & 0.36 \end{bmatrix}$$

- Almost completely independent of the weather today
- The row [0.2 0.44 0.36] is called the stationary distribution of the Markov chain

Markov chains

- Where do we get the transition matrix from?
- One answer: we can learn it from lots of data (e.g., 20 years of weather data)

Markov Chain Example: Text

"A dog is a man's best friend. It's a dog eat dog world out there."

Rows: $\mathbf{x}_{t-1}$; columns: $\mathbf{x}_t$; entries: $p(\mathbf{x}_t|\mathbf{x}_{t-1})$

	a	dog	is	man's	best	friend	it's	eat	world	out	there	.
a		2/3		1/3								
dog			1/3					1/3	1/3			
is	1											
man's					1							
best						1						
friend												1
it's	1											
eat		1										
world										1		
out											1	
there												1
.							1					

(Slide credit: Steve Seitz)

Text synthesis

- Create plausible looking poetry, love letters, term papers, etc.
- Most basic algorithm:
  1. Build transition matrix
     - find all blocks of N consecutive words/letters in training documents
     - compute probability of occurance $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-(n-1)})$
  2. Given words $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$
     - compute $\mathbf{x}_k$ by sampling from $p(\mathbf{x}_t | \mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-(n-1)})$

- Example on board...

[Scientific American, June 1989, Dewdney]

"I Spent an Interesting Evening Recently with a Grain of Salt"

- Mark V. Shaney

(computer-generated contributor to UseNet News group called net.singles)

- You can try it online here: http://www.yisongyue.com/shaney/

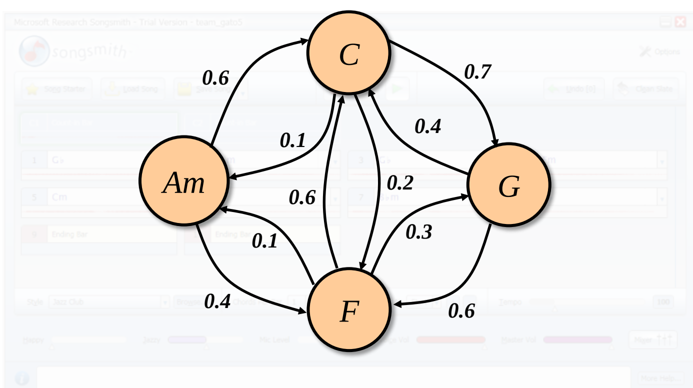
- Output of 2nd order word-level Markov Chain after training on 90,000 word philosophical essay:
- *"Perhaps only the allegory of simulation is unendurable--more cruel than Artaud's Theatre of Cruelty, which was the first to practice deterrence, abstraction, disconnection, deterritorialisation, etc.; and if it were our own past. We are witnessing the end of the negative form. But nothing separates one pole from the very swing of voting "rights" to electoral..."*

Cornell University

Text synthesis

- Jane Austen's *Pride and Prejudice*:
  - 121,549 words
  - 8,828 unique words (most common: 'the')
  - 7,800,000 possible pairs of words
  - 58,786 pairs (0.75%) actually appeared
  - most common pair?
  - Given a model learned from this text, we can
    - generate more "Jane Austen"-like novels
    - estimate the likelihood that a snippet of text was written by Jane Austen

Music synthesis



- Chord progressions learned from large database of guitar tablature

Google’s PageRank

Internet applications [edit]

The PageRank of a webpage as used by Google is defined by a Markov chain.[3] It is the probability to be at page i in the stationary distribution on the following Markov chain on all (known) webpages. If N is the number of known webpages, and a page i has k_i links then it has transition probability $\frac{\alpha}{k_i} + \frac{1-\alpha}{N}$ for all pages that are linked to and $\frac{1-\alpha}{N}$ for all pages that are not linked to. The parameter α is taken to be about 0.85.[citation needed]

Markov models have also been used to analyze web navigation behavior of users. A user's web link transition on a particular website can be modeled using first- or second-order Markov models and can be used to make predictions regarding future navigation and to personalize the web page for an individual user.

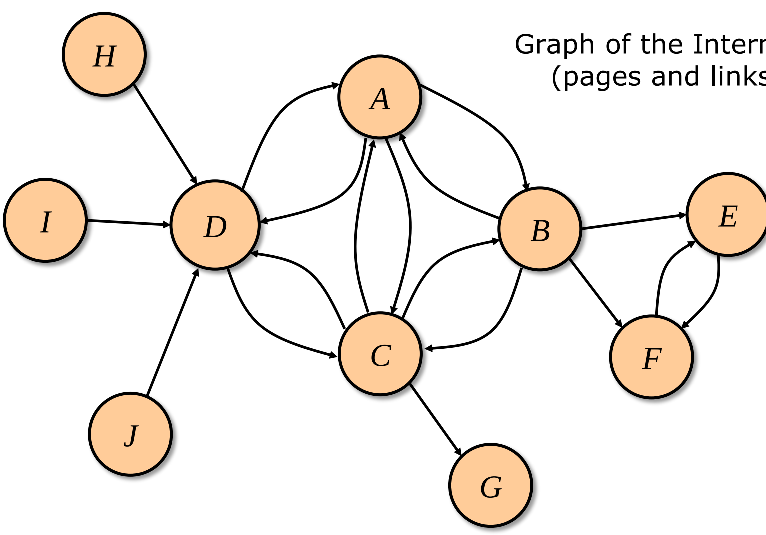
http://en.wikipedia.org/wiki/Markov_chain

Page, Lawrence; Brin, Sergey; Motwani, Rajeev and Winograd, Terry (1999). *The PageRank citation ranking: Bringing order to the Web*.

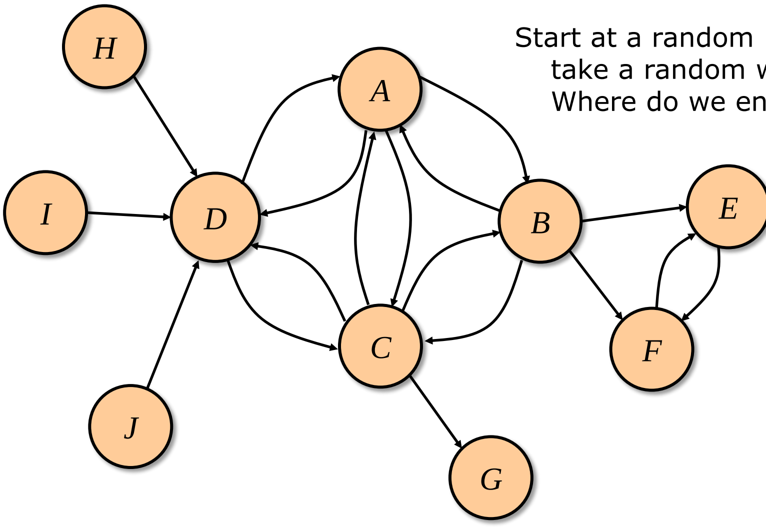
See also:
J. Kleinberg. Authoritative sources in a hyperlinked environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.

Google’s PageRank

Graph of the Internet (pages and links)

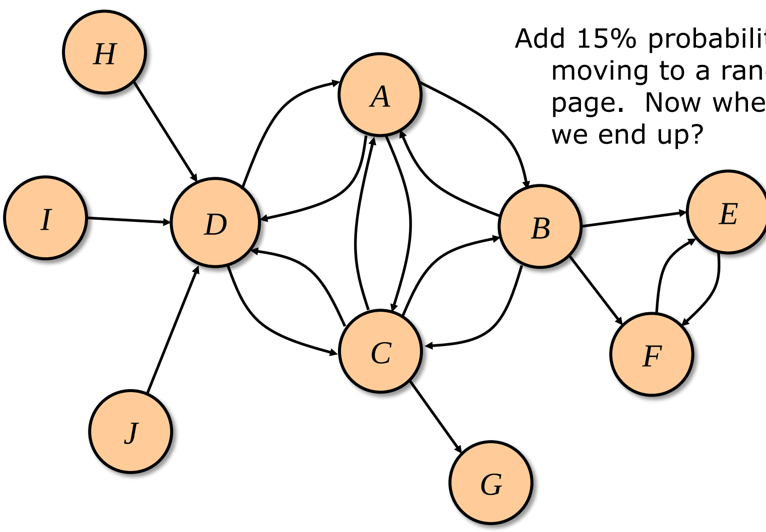


Google's PageRank



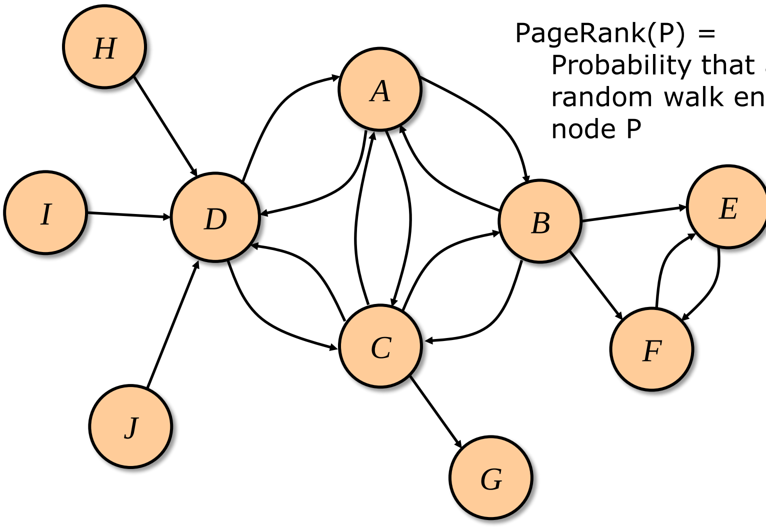
Start at a random page, take a random walk. Where do we end up?

Google's PageRank

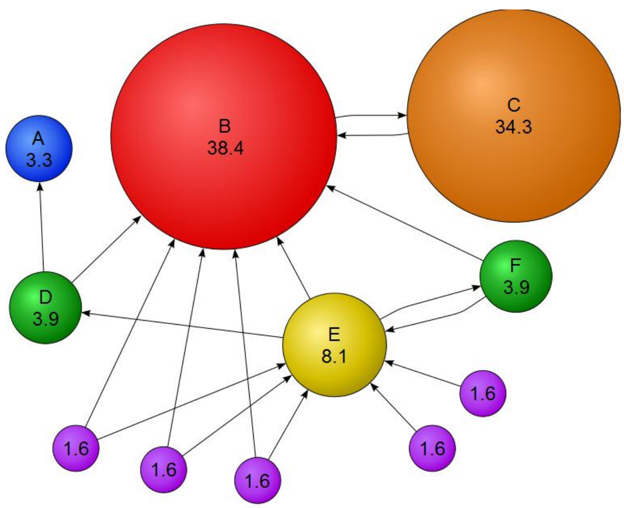


Add 15% probability of moving to a random page. Now where do we end up?

Google's PageRank



PageRank(P) =
Probability that a long random walk ends at node P



Questions?