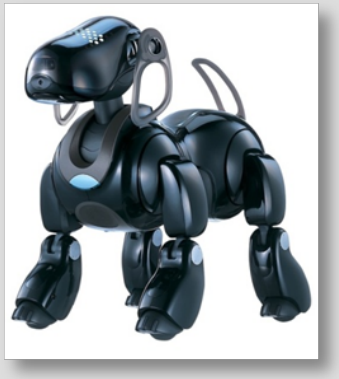
Markov chains – Part 2



Prof. Noah Snavely

CS1114

http://cs1114.cs.cornell.edu

Cornell University Computer Science

Administrivia

- Guest lecture on Thursday, Prof. Charles Van Loan

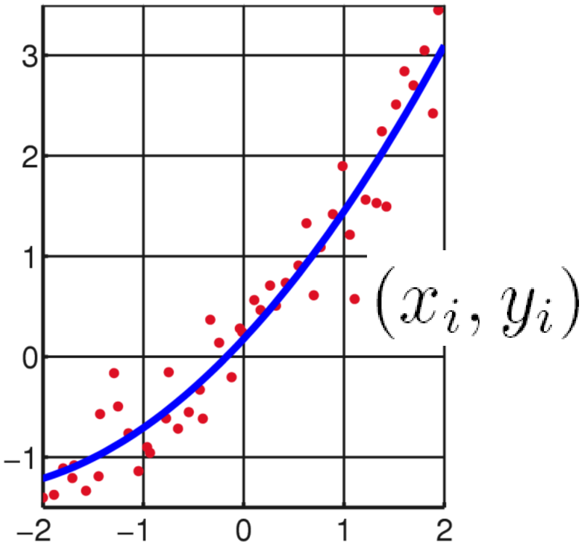
- Assignments:
  - A5P2 due on Friday by 5pm
  - A6 will be out on Friday

- Quiz next Thursday, 4/23

Administrivia

- Final projects
  - Due on Friday, May 8 (tentative)
  - You can work in groups of up to 3
  - Please form groups and send me a proposal for your final project by Wednesday
    - Not graded, but required

Least squares



Model: $y = ax^2 + bx + c$

Objective function: $\mathrm{Cost}(a, b, c) = \sum_{i=1}^{n} (y_i - (ax_i^2 + bx_i + c))^2$

Sorting and convex hull

- Suppose we had an $O(n)$ algorithm for sorting
- What does this tell us about the problem of computing a convex hull?

Incremental convex hull

- Starting from a triangle, we add *n* points, one at a time
- Given a convex hull with *k* points assume:
  - It takes *k* operations to test if a new point is inside the convex hull
  - It takes *k* operations to add a new point to the convex hull
  - Our convex hull never shrinks (in terms of number of points)

Incremental convex hull

- How long does it take if:
  - All of the points are inside the original triangle?

    Answer: 3 + 3 + 3 + … (*n* times) … + 3 = O(*n*)
  - None of the points are inside the prev. conv. hull?

    Answer: 6 + 8 + 10 + … (*n* times) … + 2*n* = O(n^2)
  - The first 10% of points are added, last 90% aren't

    Answer: 6 + 8 + … (0.1*n* times) … + 0.2*n* +
    0.1*n* + … (0.9*n* times) … + 0.1*n* = O(n^2)
  - *n* – sqrt(*n*) are not added, then sqrt(n) are

    Answer: 3 + 3 + … (*n* – sqrt(n) times) … + 3 +
    6 + … (sqrt(n) times) … + 2sqrt(*n*) = O(*n*)

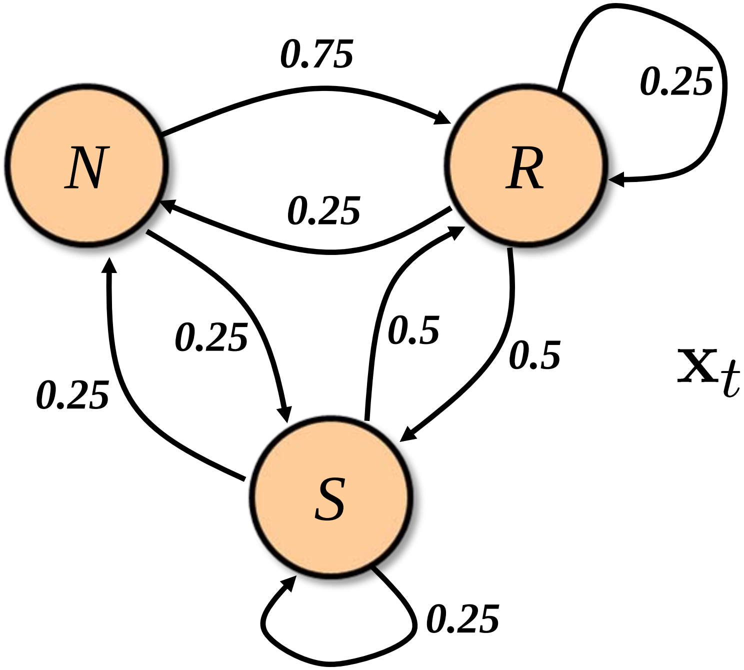
Computing nearest neighbors

- Suppose we have two sets of (*d*-dimensional) vectors, **A** and **B**

  A has *m* vectors, **B** has *n* vectors
- For each vector in **A**, we want to find the closest vector in **B**
- How do we do this quickly?
- The straightforward way (nested for loops) is horribly slow
- You learned a better way in section
- You'll learn another one tomorrow

Markov chains

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



$$
\mathbf{x}_{t-1}\;
\begin{array}{c|ccc|}
 & \multicolumn{3}{c}{\mathbf{x}_t} \\
 & \boldsymbol{N} & \boldsymbol{R} & \boldsymbol{S} \\
\boldsymbol{N} & 0.0 & 0.75 & 0.25 \\
\boldsymbol{R} & 0.25 & 0.25 & 0.5 \\
\boldsymbol{S} & 0.25 & 0.5 & 0.25
\end{array}
$$

Transition probabilities

Markov chains

- What's the weather 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
- Over a long time, we expect 20% of days to be nice, 44% to be rainy, and 36% to be snowy

Stationary distributions

- Constant columns are mapped to themselves:

$$\begin{bmatrix} 0.0 & 0.75 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

- So why do we converge to these particular columns?

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

Stationary distributions

- The vector [0.2 0.2 0.2]’ is unchanged by multiplication by ***P*** :

$$\begin{bmatrix} 0.0 & 0.75 & 0.25 \\ 0.25 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

- The vector [0.2 0.44 0.36]’ is unchanged by multiplication by $\boldsymbol{P^T}$:

$$\begin{bmatrix} 0.0 & 0.25 & 0.25 \\ 0.75 & 0.25 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.44 \\ 0.36 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.44 \\ 0.36 \end{bmatrix}$$

- Where have we seen this before?

Stationary distributions

- [0.2 0.2 0.2]' is called an *eigenvector* of ***P***
  - a vector **x** such that ***P*x** = **x**
  - (in linear algebra, you'll learn that the exact definition is a vector **x** such that ***P*x** = ***λ*x**)
- [0.2 0.44 0.36]' is an eigenvector of ***P′***
  - ***P′*x** = **x**
- If we look at a long sequence, the proportion of days we expect to be nice/rainy/snowy
- You won't have to remember this (at least for this class)

Markov Chain Example: Text

“A dog is a man’s best friend. It’s a dog eat dog world out there.”

$\mathbf{x}_{t-1}$

	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					

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

$\mathbf{x}_t$

Cornell University

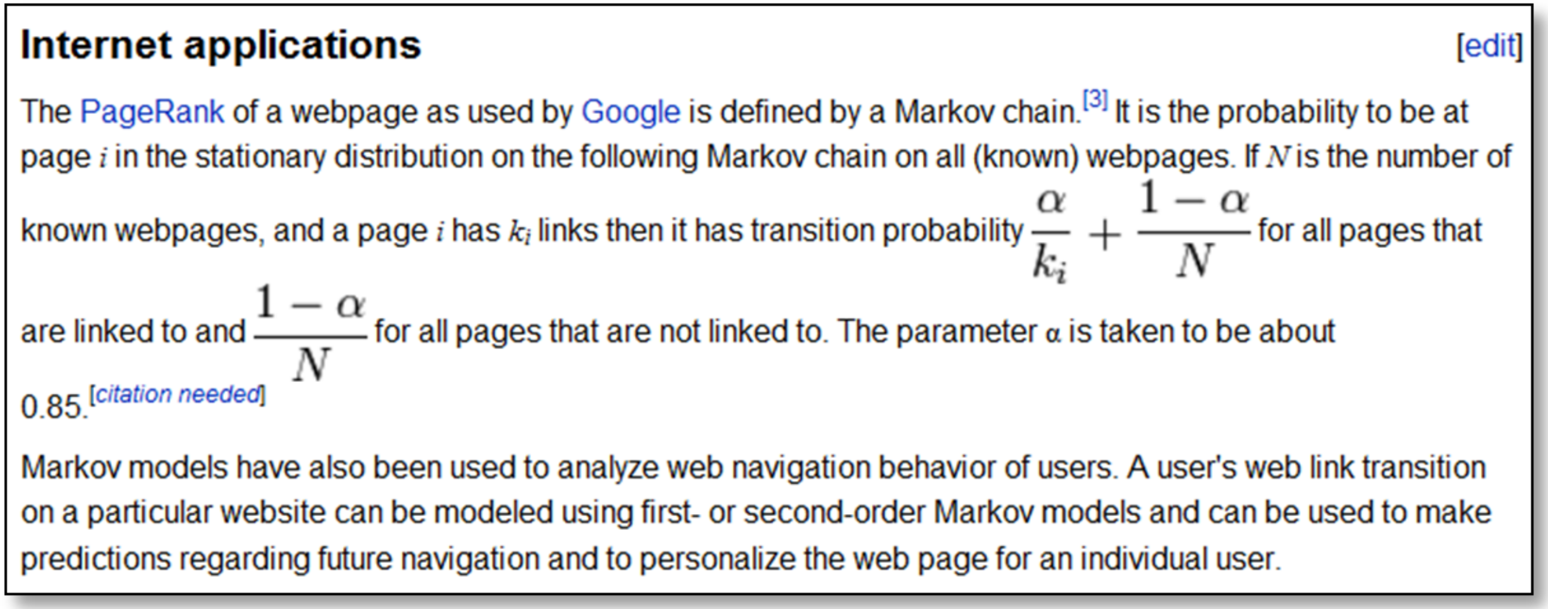
(Slide credit: Steve Seitz)

Text synthesis

- “It’s a dog is a man’s best friend. It’s a dog eat dog eat dog is a dog”
- This is called a *random walk*

- Questions about random walks:
  - What’s the probability of being at the word *dog* after generating *n* words?
  - How long does it take (on average) to generate all words at least once?
- These have important applications

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]

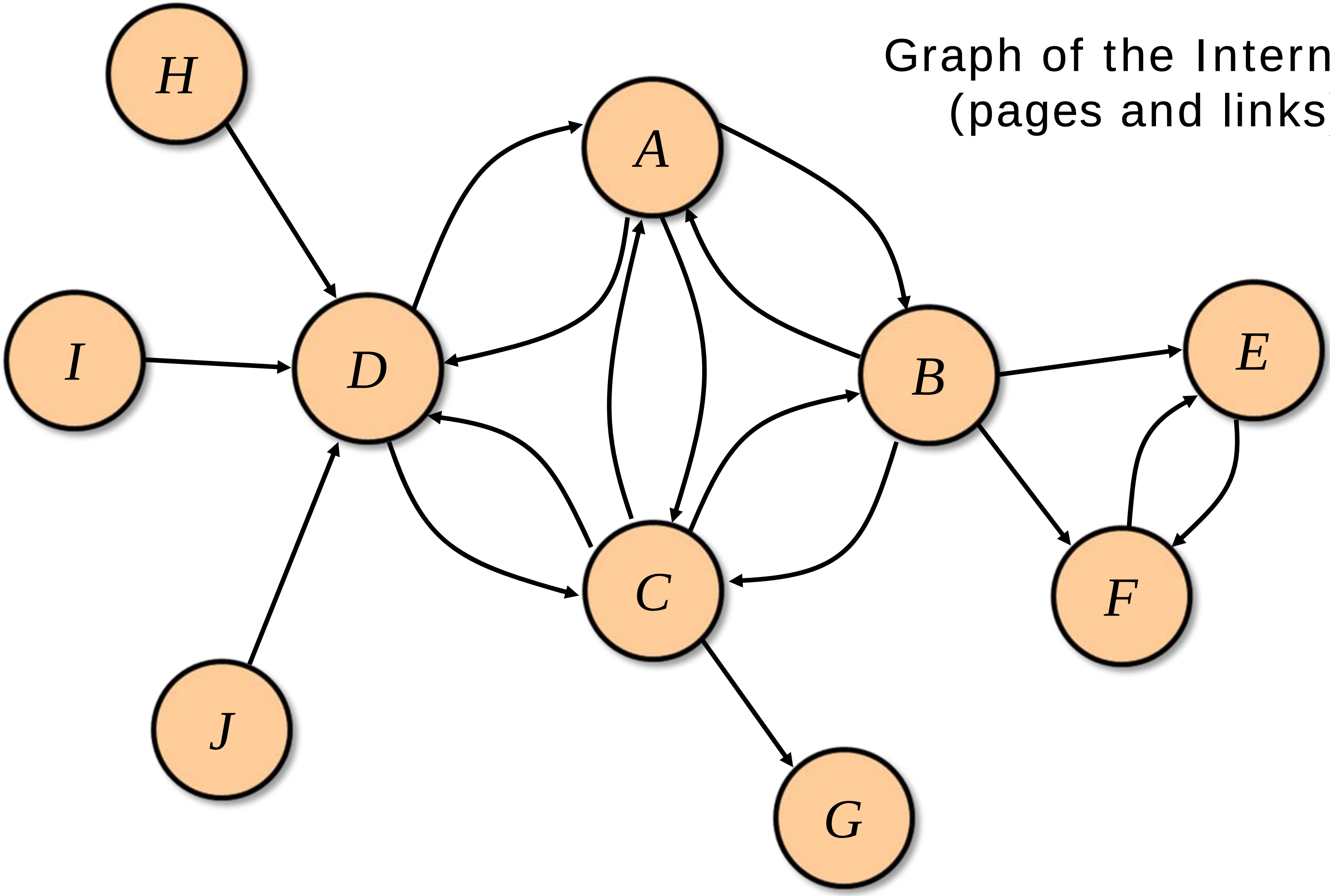
Markov models have also been used to analyze web navigation behavior of users. A user's web link transition on a particular website can be modeled using first- or second-order Markov models and can be used to make predictions regarding future navigation and to personalize the web page for an individual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

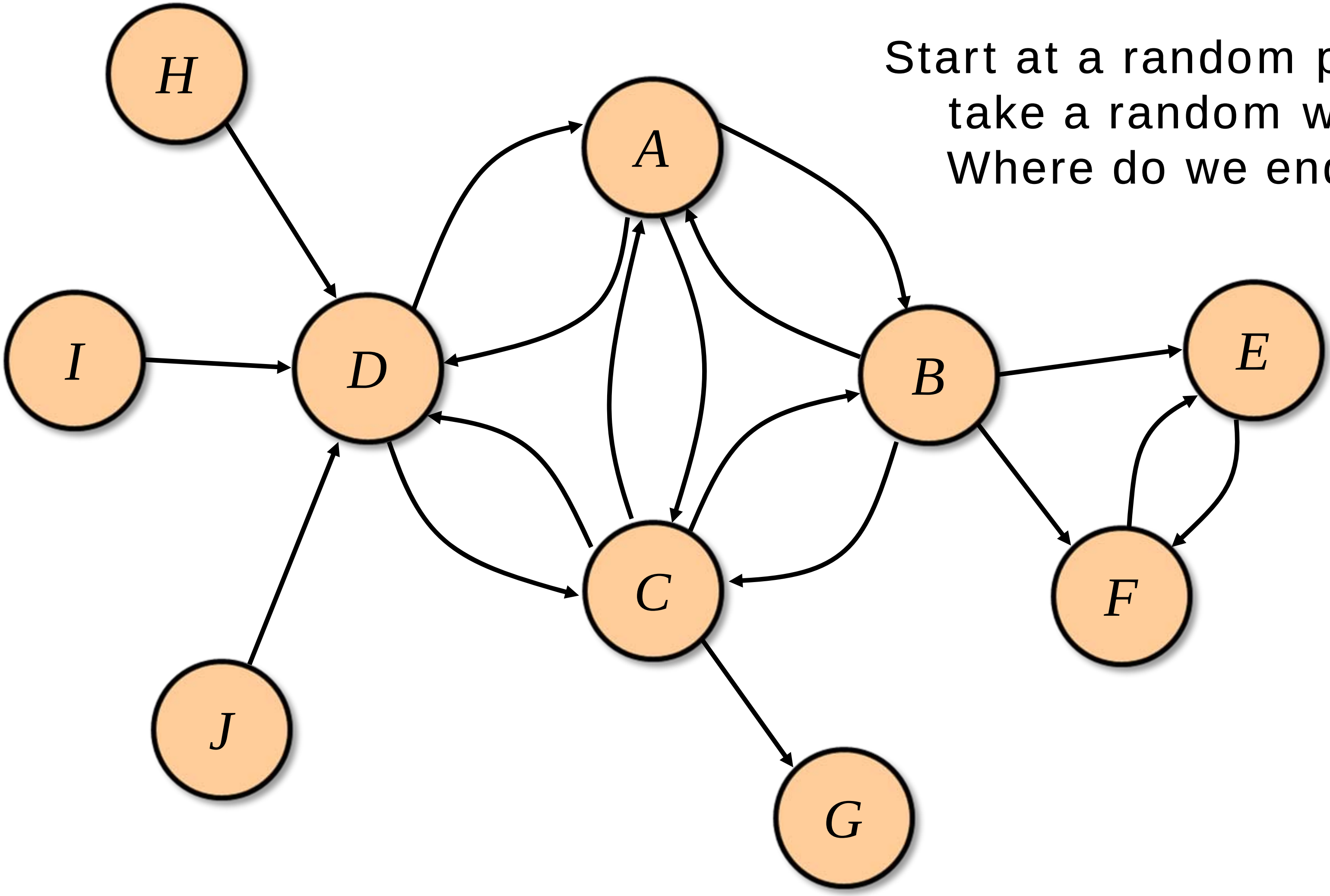
See also:
J. Kleinberg. Authoritative sources in a hyperlinked environment. Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.

Google's PageRank



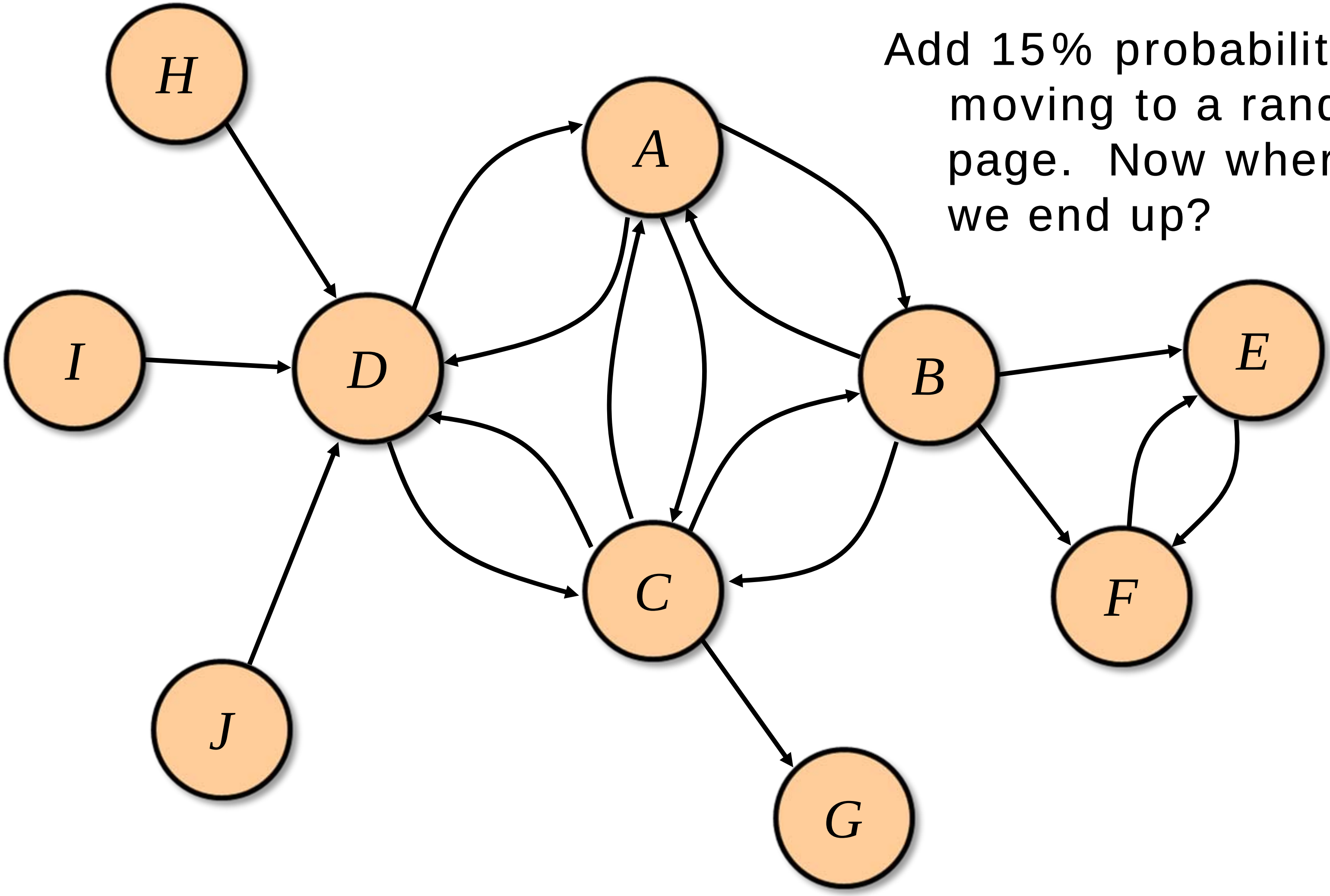
Graph of the Internet
(pages and links)

Google's PageRank



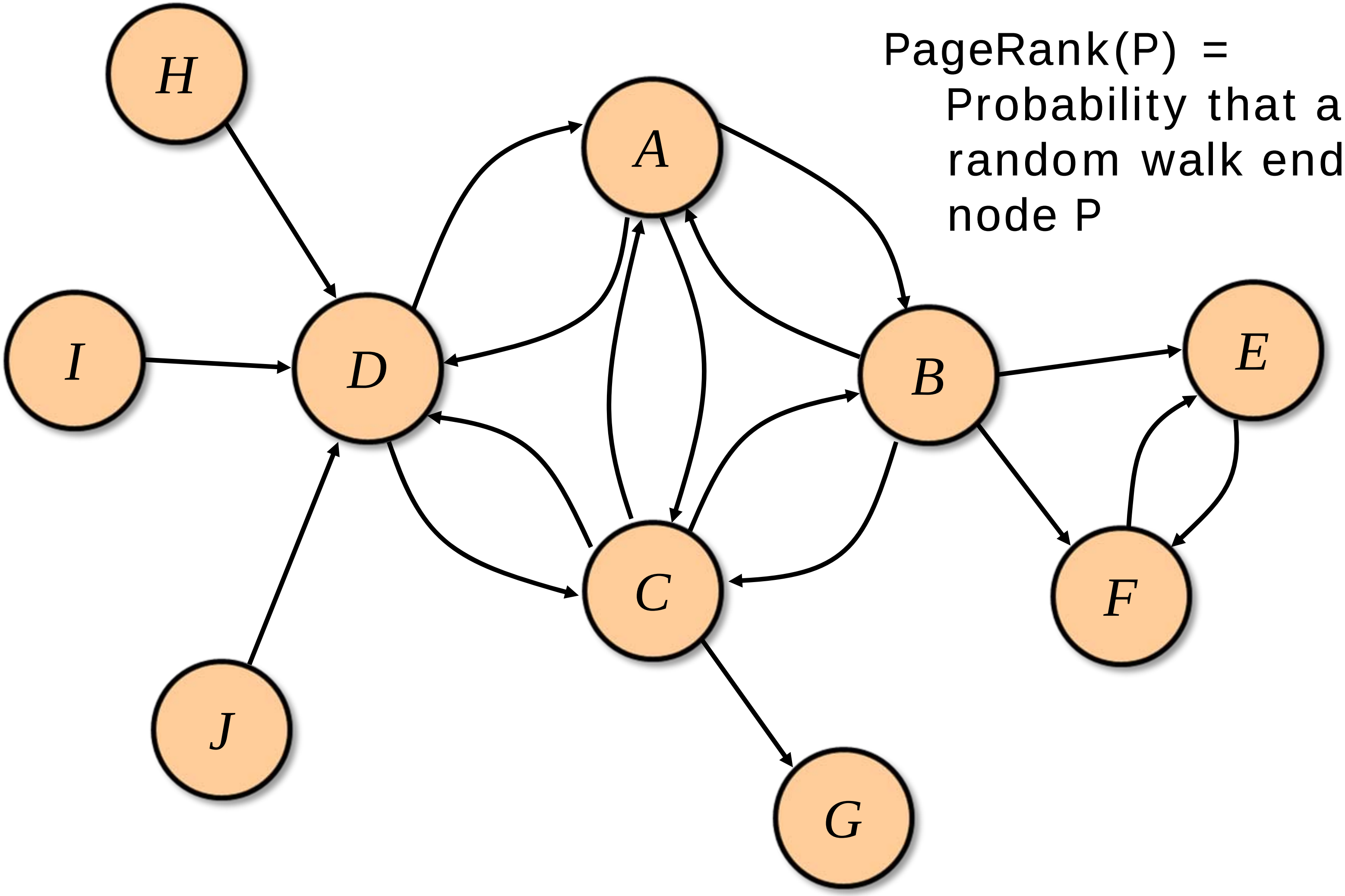
Start at a random page,
take a random walk.
Where do we end up?

Google's PageRank

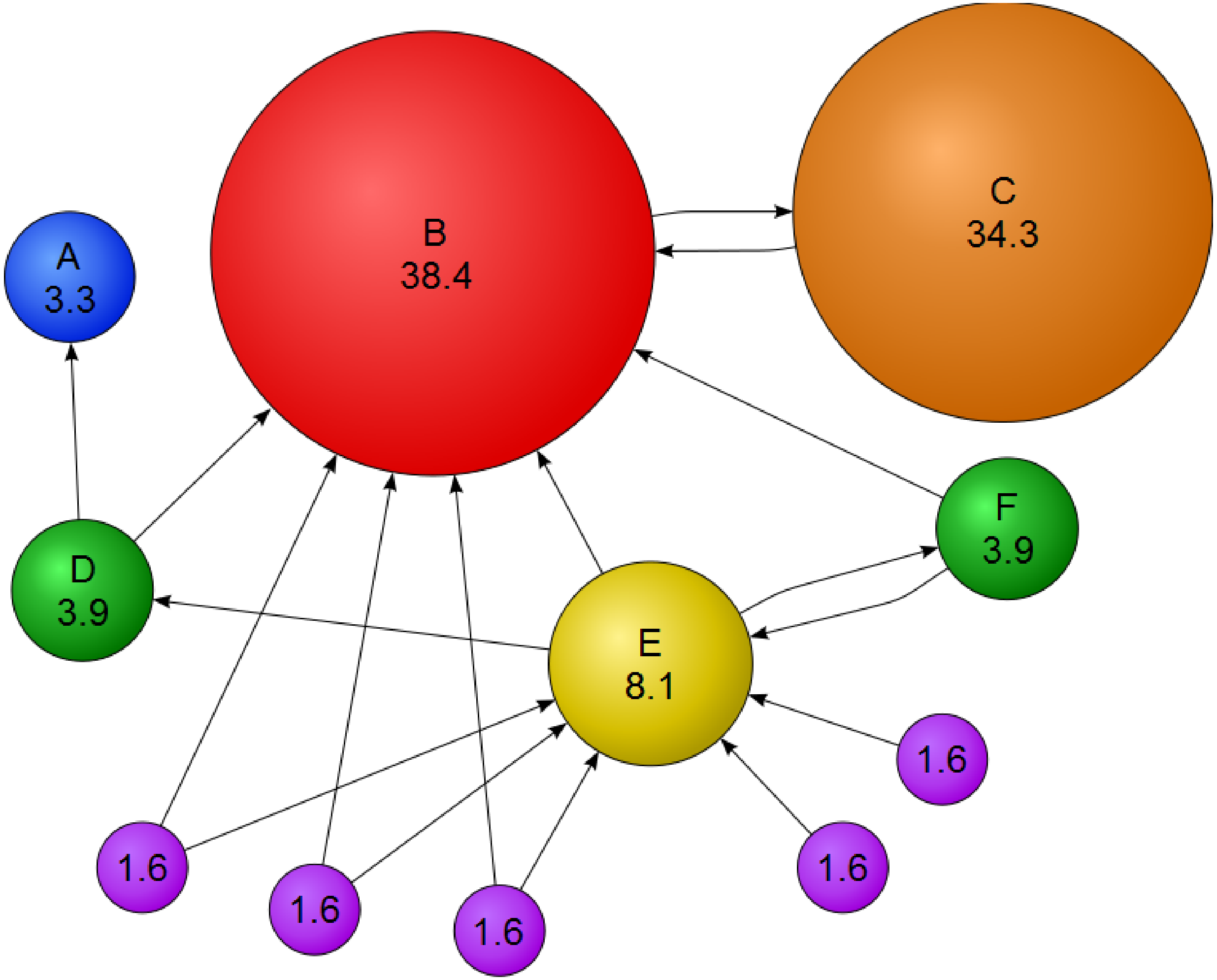


Add 15% probability of moving to a random page. Now where do we end up?

Google’s PageRank



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



(The ranks are an eigenvector of the transition matrix)

Back to text

- We can use Markov chains to generate new text
- Can they also help us recognize text?
  - In particular, the author?
  - Who wrote this paragraph?

    “Suppose that communal kitchen years to come perhaps. All trotting down with porringers and tommycans to be filled. Devour contents in the street. John Howard Parnell example the provost of Trinity every mother's son don't talk of your provosts and provost of Trinity women and children cabmen priests parsons fieldmarshals archbishops.”

Author recognition

- We can use Markov chains to generate new text
- Can they also help us recognize text?
  - How about this one?

    „Diess Alles schaute Zarathustra mit grosser Verwunderung; dann prüfte er jeden Einzelnen seiner Gäste mit leutseliger Neugierde, las ihre Seelen ab und wunderte sich von Neuem. Inzwischen hatten sich die Versammelten von ihren Sitzen erhoben und warteten mit Ehrfurcht, dass Zarathustra reden werde.“

The Federalist Papers

- 85 articles addressed to New York State, arguing for ratification of the Constitution (1787-1788)

- Written by “Publius” (?)
- Really written by three different authors:
  - John Jay, James Madison, Andrew Hamilton

- Who wrote which one?
  - 73 have known authors, 12 are in dispute

Can we help?

- Suppose we want to know who wrote a given paragraph/page/article of text
- Idea: for each suspected author:
  - Download all of their works
  - Compute the transition matrix
  - Find out which author’s transition matrix is the best fit to the paragraph

Author recognition

- Simple problem:

  Given two Markov chains, say *Austen* (*A*) and *Dickens* (*D*), and a string *s* (with *n* words), how do we decide whether *A* or *D* wrote *s*?

- Idea: For both *A* and *D*, compute the probability that a random walk of length *n* generates *s*

Probability of a sequence

- What is the probability of a given *n*-length sequence *s*?

$$s = s_1 \; s_2 \; s_3 \; \ldots \; s_n$$

- Probability of generating *s* = the product of transition probabilities:

$$\underbrace{\Pr(S_1 = s_1)}_{\text{Probability that a sequence starts with } s_1} \underbrace{\Pr(S_2 = s_2 | S_1 = s_1) \Pr(S_3 = s_3 | S_2 = s_2) \ldots \Pr(S_n = s_n | S_{n-1} = s_{n-1})}_{\text{Transition probabilities}}$$

Likelihood

- Compute this probability for *A* and *D*

$$\Pr(s|A)$$

“*likelihood*” of A

$$\Pr(s|A) > \Pr(s|D)$$

Jane Austen wrote *s*

$$\Pr(s|A) < \Pr(s|D)$$

Charles Dickens wrote *s*

$$\Pr(s|D)$$

“*likelihood*” of D

$$\Pr(s|A) = \Pr(s|D)$$

???

	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					

Pr(“is a dog is a dog”) = ?   *Pr*(“is dog man’s best friend”) = ?

Next time

- Section: more information on how to actually compute this
- Guest lecture by Charles Van Loan