Recap

- (Ch 14) Markov chains and hidden Markov models
  - Graphical representation
  - Transition probability matrix
  - Propagating state distributions
  - The stationary distribution

Today

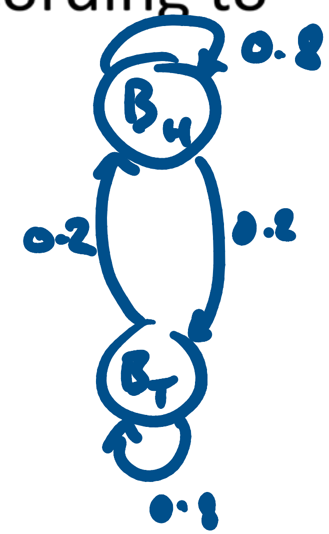
- (Ch 14) Markov chains and hidden Markov models
  - The PageRank algorithm
  - Hidden Markov models
  - Viterbi algorithm

Transition probability matrix: con example

- Let’s say a con artist has two biased coins that look the same
  - B_H is a coin that is biased towards heads
  - B_T is a coin that is biased towards tails

- The con artist makes bets with an unsuspecting victim using one coin at a time but can use sleight-of-hand to switch the coins according to the **transition probability matrix**

$$P = \begin{bmatrix} P(B_H|B_H) & P(B_T|B_H) \\ P(B_H|B_T) & P(B_T|B_T) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$



Stationary distribution: con example

- The stationary distribution $\mathbf{s}$ has the following properties
  - $\mathbf{s}P = \mathbf{s}$
  - $\mathbf{s}$ is a row eigenvector of P with eigenvalue of 1
  - Regardless of the initial probability distribution $\boldsymbol{\pi}$, $\lim_{t\to\infty} \boldsymbol{\pi}P^t = \mathbf{s}$ if the Markov chain is irreducible

- For the con example, we can find that $P = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$ leads to $\mathbf{s} = [0.5 \;\; 0.5]$ in two ways
  - Exactly, by finding the eigenvector with eigenvalue 1
  - Approximately, by finding $\boldsymbol{\pi}P^t$ for large values of t

The billion dollar eigenvector

The PageRank Citation Ranking: Bringing Order to the Web.

Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report. Stanford InfoLab.

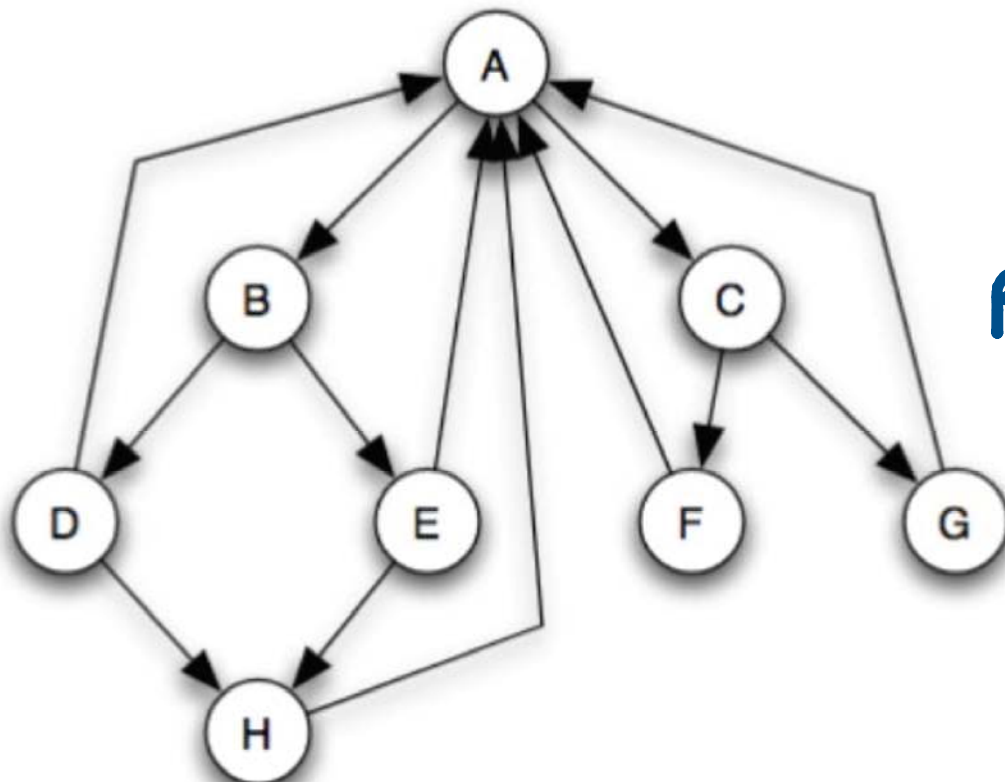
BibTeX DublinCore EndNote HTML

PDF
299Kb

Abstract

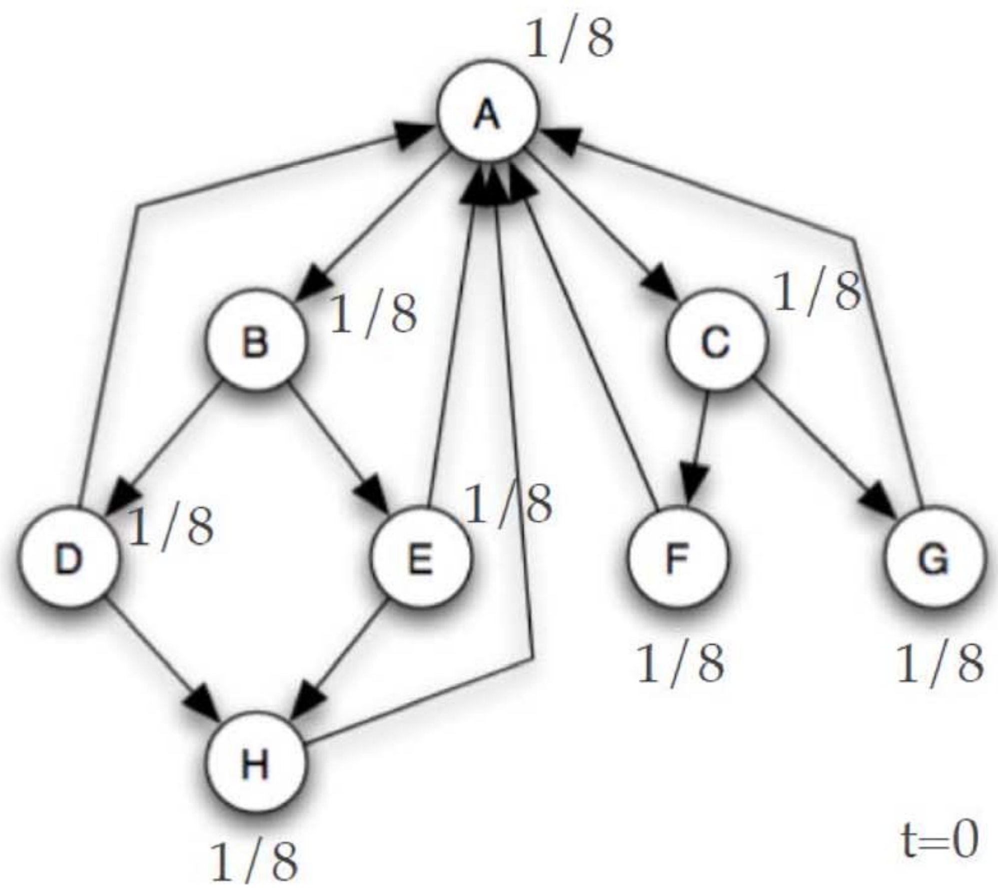
The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a mathod for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them. We compare PageRank to an idealized random Web surfer. We show how to efficiently compute PageRank for large numbers of pages. And, we show how to apply PageRank to search and to user navigation.

Randomly surfing a network of webpages



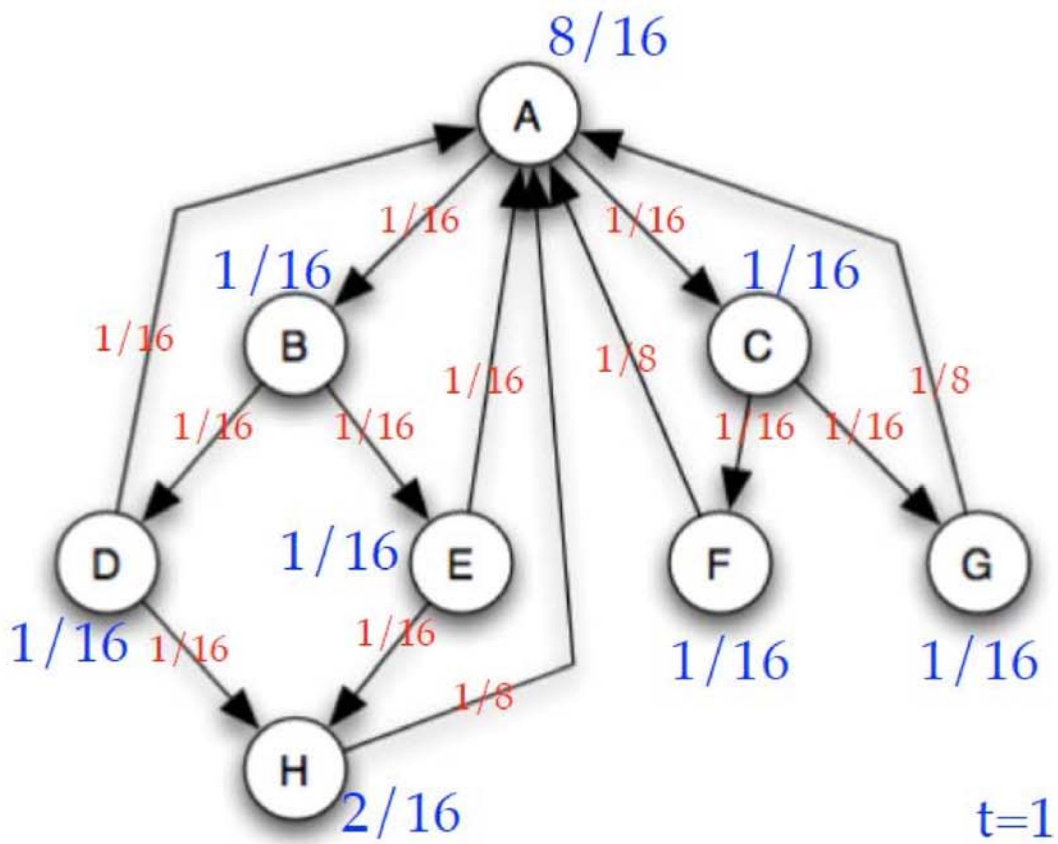
$$
P = \begin{array}{c} \\ A \\ B \\ C \\ \vdots \\ H \end{array}
\begin{array}{c}
\begin{array}{ccccc} A & B & C & \cdots & H \end{array} \\
\left[\begin{array}{ccccc}
0 & 0.5 & 0.5 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & \cdots & 0 \\
\vdots & & & & \\
1 & 0 & 0 & \cdots & 0
\end{array}\right]
\end{array}
$$

Initialize the distribution uniformly



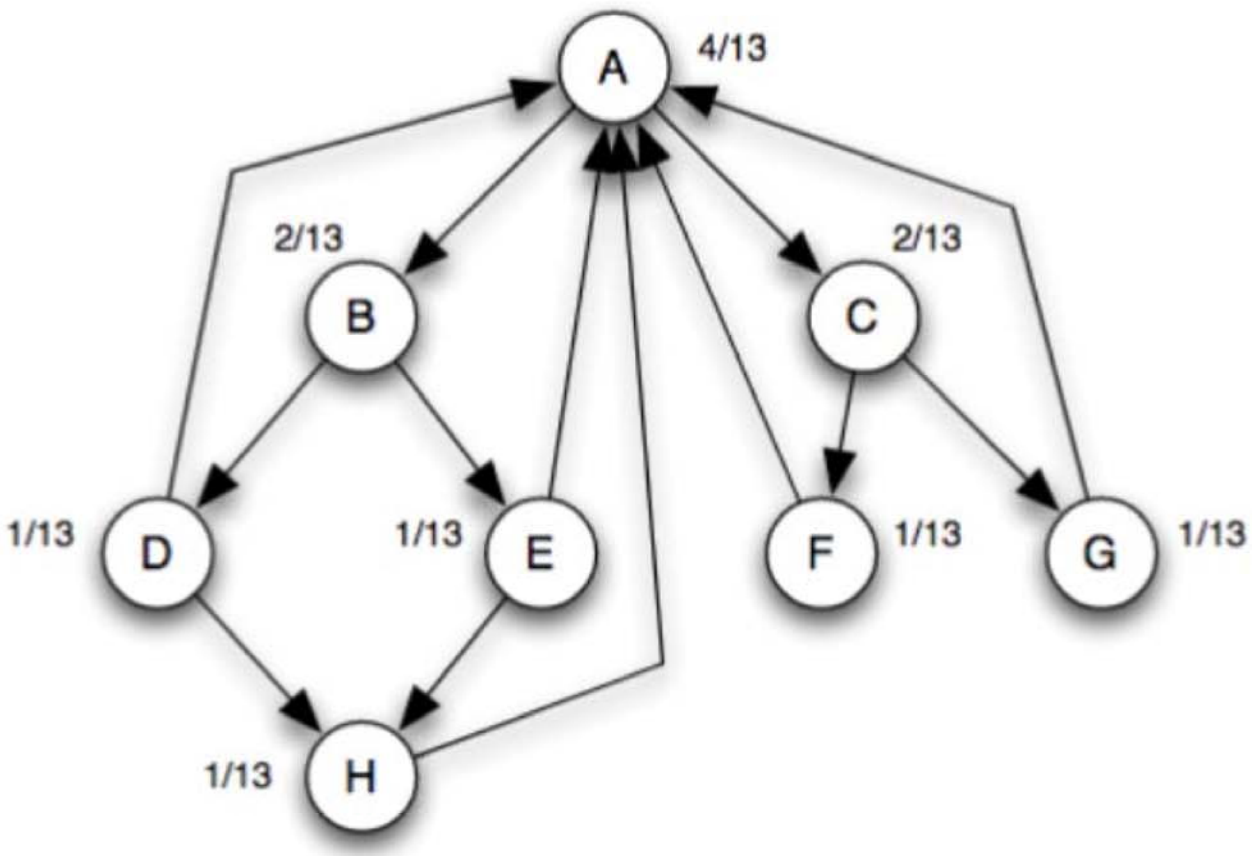
$$\begin{array}{cccccc} & A & B & C & \cdots & H \\ \pi = & \Big[\, \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \cdots\cdots & \frac{1}{8} \,\Big] \end{array}$$

Update the distribution iteratively ...



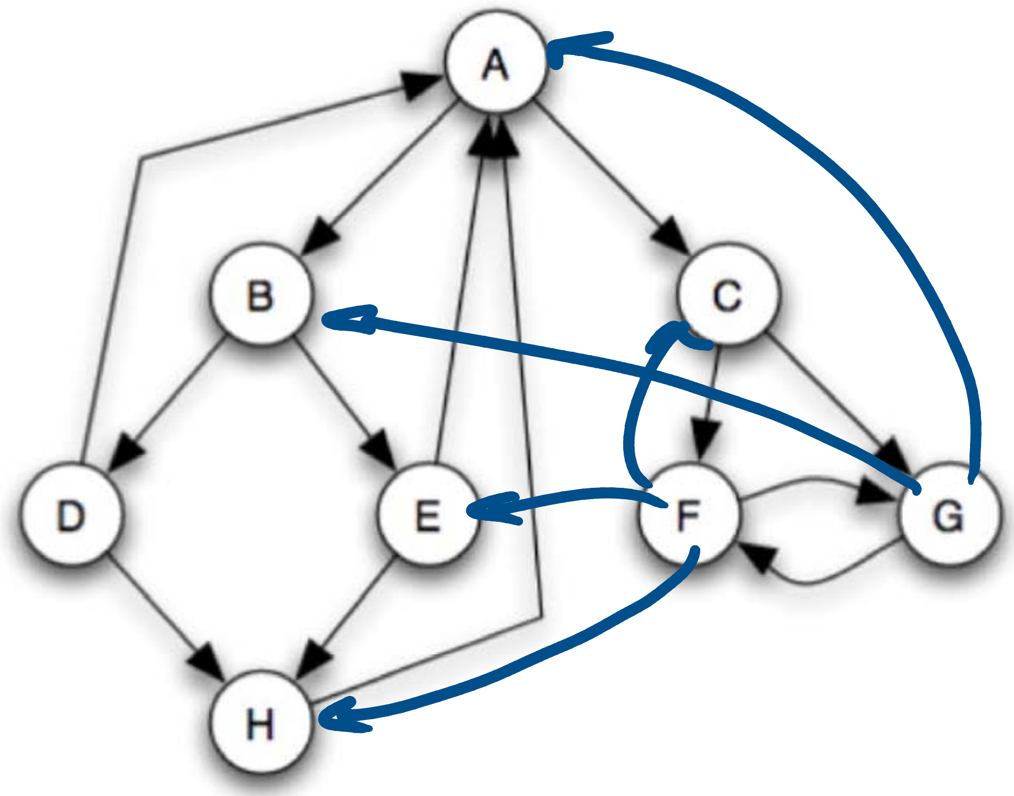
πP

… until the stationary distribution is reached



$$\lim_{t \to \infty} \pi P^t$$

Sometimes the surfer gets trapped



- The PageRank algorithm avoids the problem of getting trapped in a subnetwork of pages by allowing the web surfer to "teleport" from any webpage to any other with small probability
- Teleportation corresponds to entering a URL directly into the browser's address bar

Hidden Markov model: con example

- Let's add the following assumptions about the con artist's game
  - B_H has $P(H|B_H) = 0.6$ and B_T has $P(H|B_T) = 0.4$
  - The initial distribution $\boldsymbol{\pi} = [0.5 \;\; 0.5]$ (so that $\boldsymbol{\pi}P^t = [0.5 \;\; 0.5]$ for all t)

- You are an onlooker to the game
  - You know the properties of the coins, the initial distribution $\boldsymbol{\pi}$ over the coins and the transition probability matrix P
  - You do **not** know which coin the con artist is using at any time
  - If you observe the tosses HTH, what is the most likely sequence of coins used to generate the tosses?

Probability of sequence given observations

- For a given sequence of coins, we want to calculate

$$P(\text{sequence}|\text{observations}) = \frac{P(\text{sequence}, \text{observations})}{P(\text{observations})}$$

- Since $P(\text{observations})$ does not depend on the sequence, we will try to maximize $P(\text{sequence}, \text{observations})$

Comparing two sequences: con example

- What is $P(B_H B_T B_H, HTH)$?

$\pi(B_H)\, P(H|B_H)\, P(B_T|B_H)\, P(T|B_T)\, P(B_H|B_T)\, P(H|B_H)$

$(0.5)\quad(0.6)\quad\underbrace{(0.2)\quad(0.6)\quad(0.2)}_{0.024}\quad(0.6)$

- What is $P(B_H B_H B_H, HTH)$?

$\pi(B_H)\, P(H|B_H)\, P(B_H|B_H)\, P(T|B_H)\, P(B_H|B_H)\, P(H|B_H)$

$(0.5)\quad(0.6)\quad\underbrace{(0.8)\quad(0.4)\quad(0.8)}_{0.256}\quad(0.6)$

larger

The Viterbi algorithm

- The number of probabilities to evaluate grows exponentially with the length of the sequence
- The Viterbi algorithm is a dynamic programming technique that reduces the complexity to linear in the length of the sequence
- The key ideas are
  - Draw all state sequences on a trellis diagram
  - Work through the trellis step-by-step along the sequence retaining only the sequences that maximize the probability up to each step

[Not on final]

The Viterbi algorithm: con example