Today

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

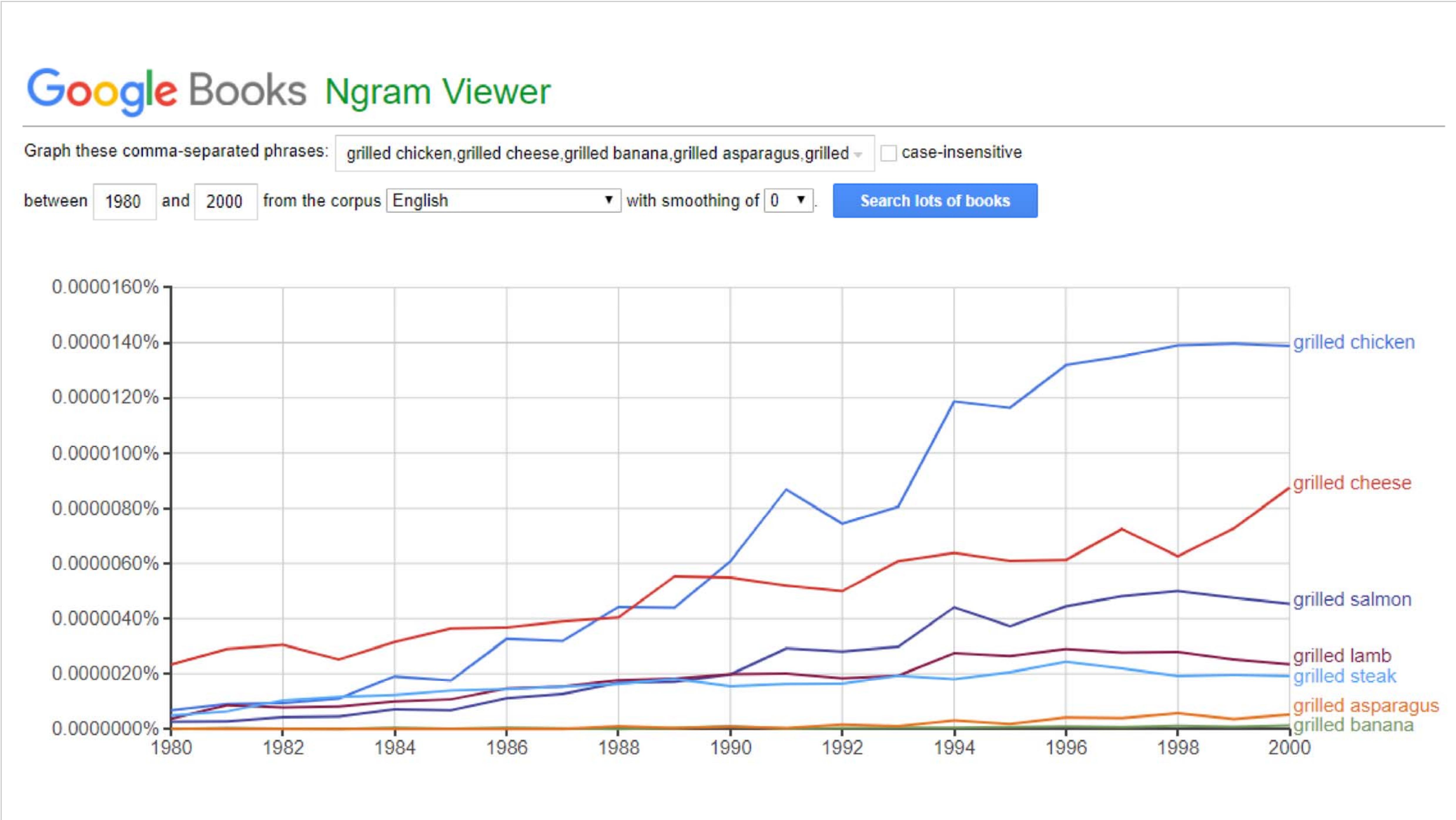
Next lecture

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

Complete the sentence

I had a glass of red wine with my grilled __tilapia__

steak
pork
cheese



Google Books Ngram Viewer
Graph these comma-separated phrases: grilled chicken,grilled cheese,grilled banana,grilled asparagus,grilled
case-insensitive
between 1980 and 2000 from the corpus English with smoothing of 0.
Search lots of books
0.0000160%
0.0000140%
0.0000120%
0.0000100%
0.0000080%
0.0000060%
0.0000040%
0.0000020%
0.0000000%
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
grilled chicken
grilled cheese
grilled salmon
grilled lamb
grilled steak
grilled asparagus
grilled banana

Describing a sequence with a Markov chain

- Let $X_0, X_1, \ldots$ be a sequence of discrete finite-valued random variables

- The sequence is a **Markov chain** if the probability distributions of the current value X_t only depend~~s~~ on the previous value X_{t-1}

$$\boxed{P(X_t | X_1, \ldots, X_{t-1}) = P(X_t | X_{t-1})}$$

- Also assume that the **transition probabilities** do not change with time

$$P(X_t | X_{t-1}) = P(X_{t-1} | X_{t-2}) = \cdots = P(X_1 | X_0)$$

Markov chain: coin example

Toss a fair coin until you see two heads in a row and then stop

- What is the probability of stopping after exactly 2 flips?

$P(HH) = 1/4$

- What is the probability of stopping after exactly 3 flips?

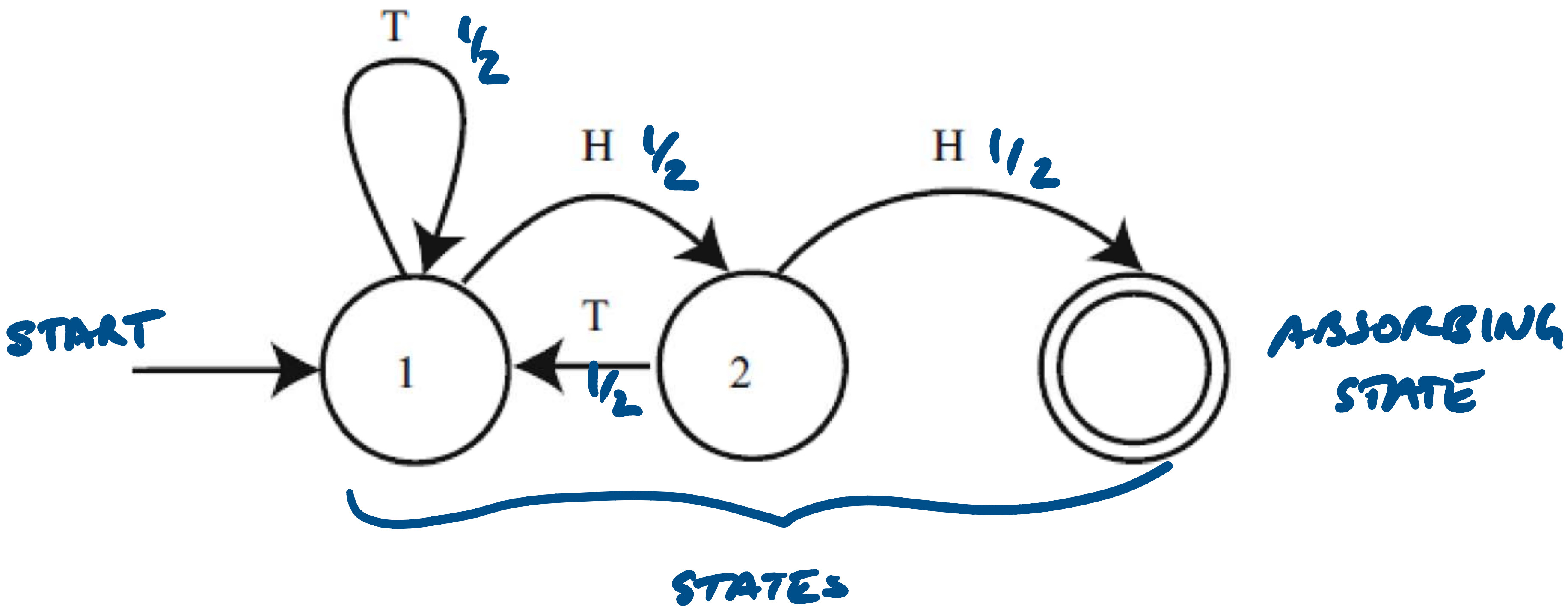
$P(THH) = 1/8$

- What is the probability of stopping after exactly 4 flips?

$P(HTHH \text{ or } TTHH) = \frac{2}{16} = \frac{1}{8}$

Markov chain as a graph: coin example

STATE DIAGRAM

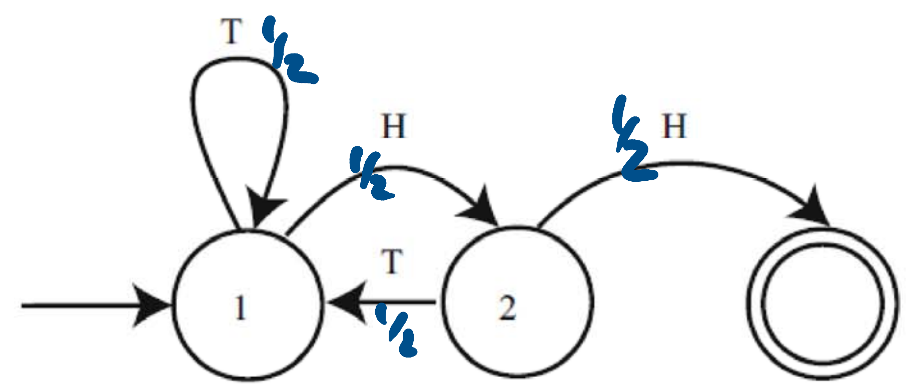


Forming a recurrence relation: coin example

- Let p_n be the probability of stopping after exactly n flips
  - We already know that $p_1 = 0,\ p_2 = {}^1/_4,\ p_3 = {}^1/_8,\ \ p_4 = {}^1/_8$

- If $n > 2$, there are two ways the sequence can start
  - Toss T and then end in exactly $n - 1$ tosses
  - Toss HT and then end in exactly $n - 2$ tosses

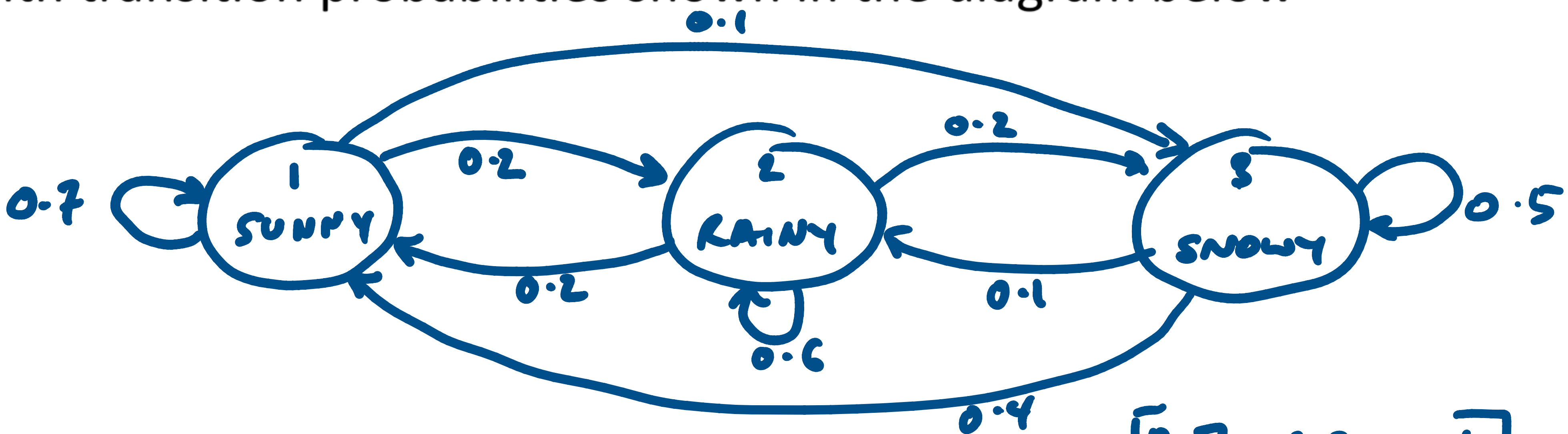
- So we derive a recurrence relation



$$p_n = \overset{P(T)}{\frac{1}{2}} p_{n-1} + \overset{P(HT)}{\frac{1}{4}} p_{n-2}$$

Transition probability matrix: weather example

- Let’s model daily weather as one of three states (sunny, rainy, snowy) with transition probabilities shown in the diagram below



- These probabilities can be represented in a **transition probability matrix**

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

Transition probability matrix properties

- The transition probability matrix P is a square matrix with entries p_{ij}

row column

- Since $p_{ij} = P(X_t = j | X_{t-1} = i)$

$$p_{ij} \geq 0 \quad \text{and} \quad \sum_j p_{ij} = 1$$

rows sum to 1

Probability distributions over states

- Let $\boldsymbol{\pi}$ be a row vector containing the probability distribution over states at $t = 0$

$$\pi_i = P(X_0 = i)$$

- Example: suppose that it is rainy today; then $\boldsymbol{\pi} = \begin{array}{c} \begin{array}{ccc} \text{SUNNY} & \text{RAINY} & \text{SNOWY} \end{array} \\ [0 \quad 1 \quad 0] \end{array}$

- Let $\mathbf{p}^{(t)}$ be a row vector containing the probability distribution over states at t

$$p_i^{(t)} = P(X_t = i)$$

$$p^{(0)} = \pi$$

Propagating the probability distribution

- Propagating from $t = 0$ to $t = 1$,

$$
\begin{aligned}
p_j^{(1)} &= P(X_1 = j) \\
&= \sum_i P(X_1 = j, X_0 = i) && \text{Total probability} \\
&= \sum_i P(X_1 = j \,|\, X_0 = i) P(X_0 = i) && \text{Bayes rule} \\
&= \sum_i p_{ij}\, \pi_i
\end{aligned}
$$

- In matrix notation, $\mathbf{p}^{(1)} = \boldsymbol{\pi} P$

Probability distributions: weather example

- Suppose that it is rainy today, so we know that $\boldsymbol{\pi} = [0 \;\; 1 \;\; 0]$

- What are the probability distributions for the weather tomorrow and the next day?

$$\mathbf{p}^{(1)} = \boldsymbol{\pi} P = [0 \;\; 1 \;\; 0] \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} = [0.2 \;\; 0.6 \;\; 0.2]$$

$$\mathbf{p}^{(2)} = \mathbf{p}^{(1)} P = [0.2 \;\; 0.6 \;\; 0.2] \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} = [\overset{0.34}{\cancel{0.28}} \;\; \overset{0.42}{\cancel{0.40}} \;\; \overset{0.24}{\cancel{0.32}}]$$

Propagating to $t = \infty$

- We have just seen that $\mathbf{p}^{(2)} = \mathbf{p}^{(1)}P = (\boldsymbol{\pi}P)P = \boldsymbol{\pi}P^2$

- In general $\mathbf{p}^{(t)} = \boldsymbol{\pi}P^t$

- If you can reach any state from any other state, the Markov chain is called **irreducible** and has the following property

$$\lim_{t\to\infty} \boldsymbol{\pi}P^t = \mathbf{s}$$

where is $\mathbf{s}$ called the stationary distribution of the Markov chain

$Ax = \lambda x$

Stationary distribution

- The stationary distribution $\mathbf{s}$ has the property that $\mathbf{s}P = \mathbf{s}$
- In other words, $\mathbf{s}$ is a row eigenvector of P with eigenvalue of 1
- Example: regardless of the initial probability distribution $\boldsymbol{\pi}$, the stationary distribution for our daily weather model is

$$\mathbf{s} = \lim_{t\to\infty} \boldsymbol{\pi} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}^t = \begin{bmatrix} \frac{18}{37} & \frac{11}{37} & \frac{8}{37} \end{bmatrix}$$

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.