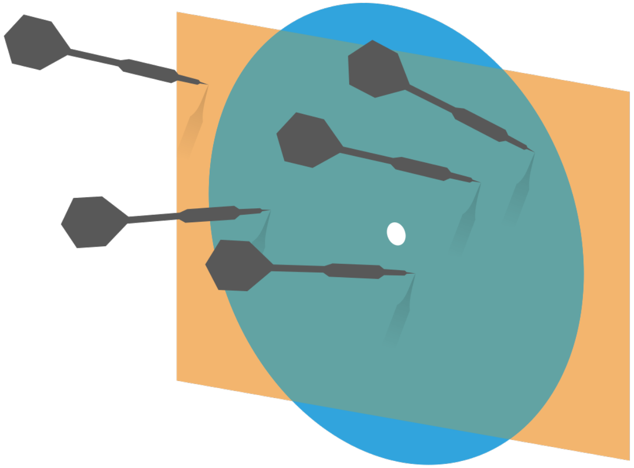
Probability and Statistics for Computer Science



Credit: wikipedia

Conditional probability comes back in matrix!

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 05.04.2021

Last Time

- Application of Clustering

  Cluster Center Histogram

- Markov Chain (I)

Objectives

- Markov Chain (II)

- Application of Stationary Markov chain

  PageRank Algo.

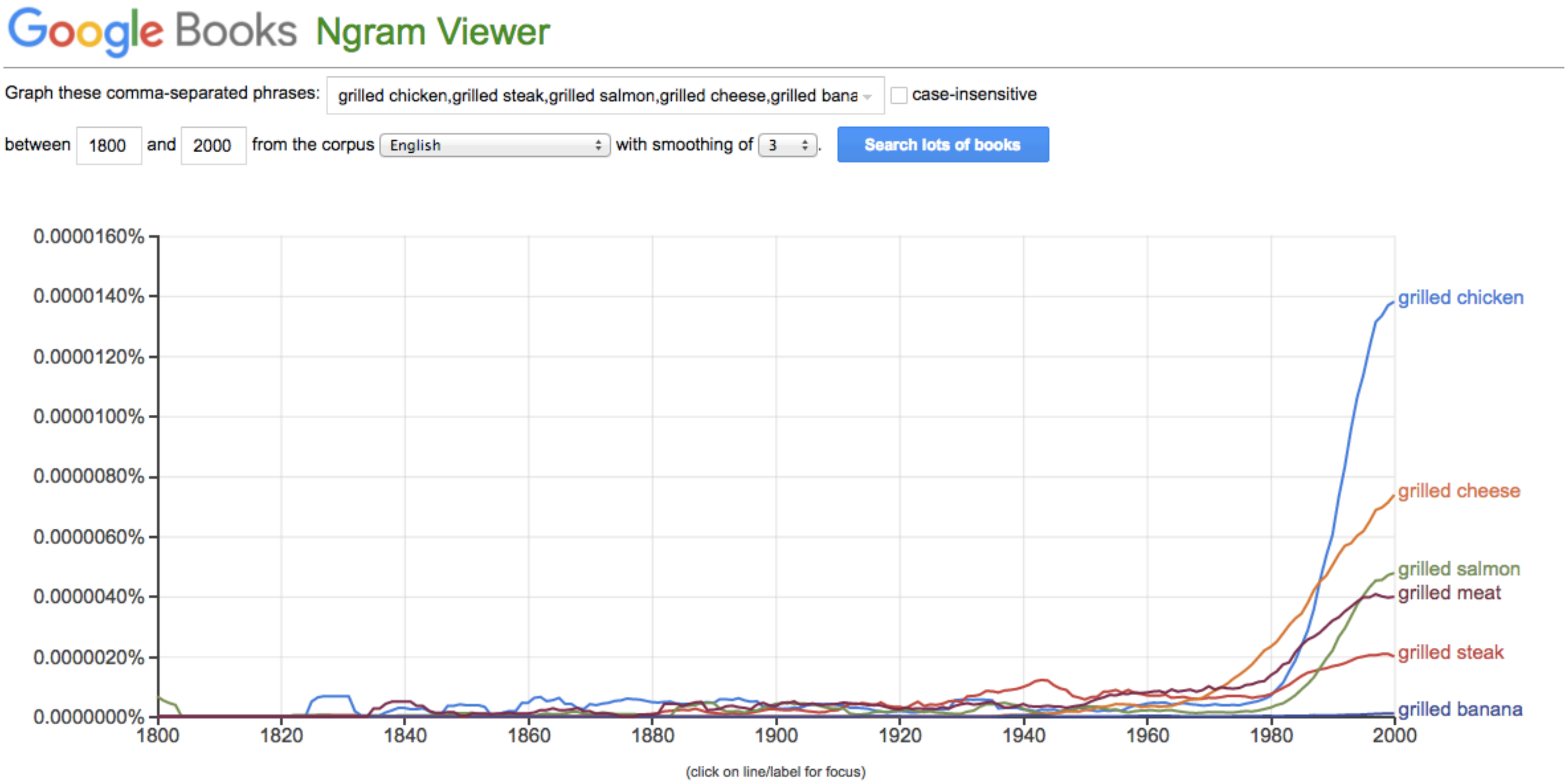
Motivation

- So far, the processes we learned such as **Bernoulli and Poisson** process are sequences of **independent** trials.
- There are a lot of real world situations where sequences of events are **Not independent** In comparison.
- Markov chain is one type of characterization of a series of **dependent** trials.

An example of dependent events in a sequence

I had a glass of wine with my grilled __________

An example of dependent events in a sequence

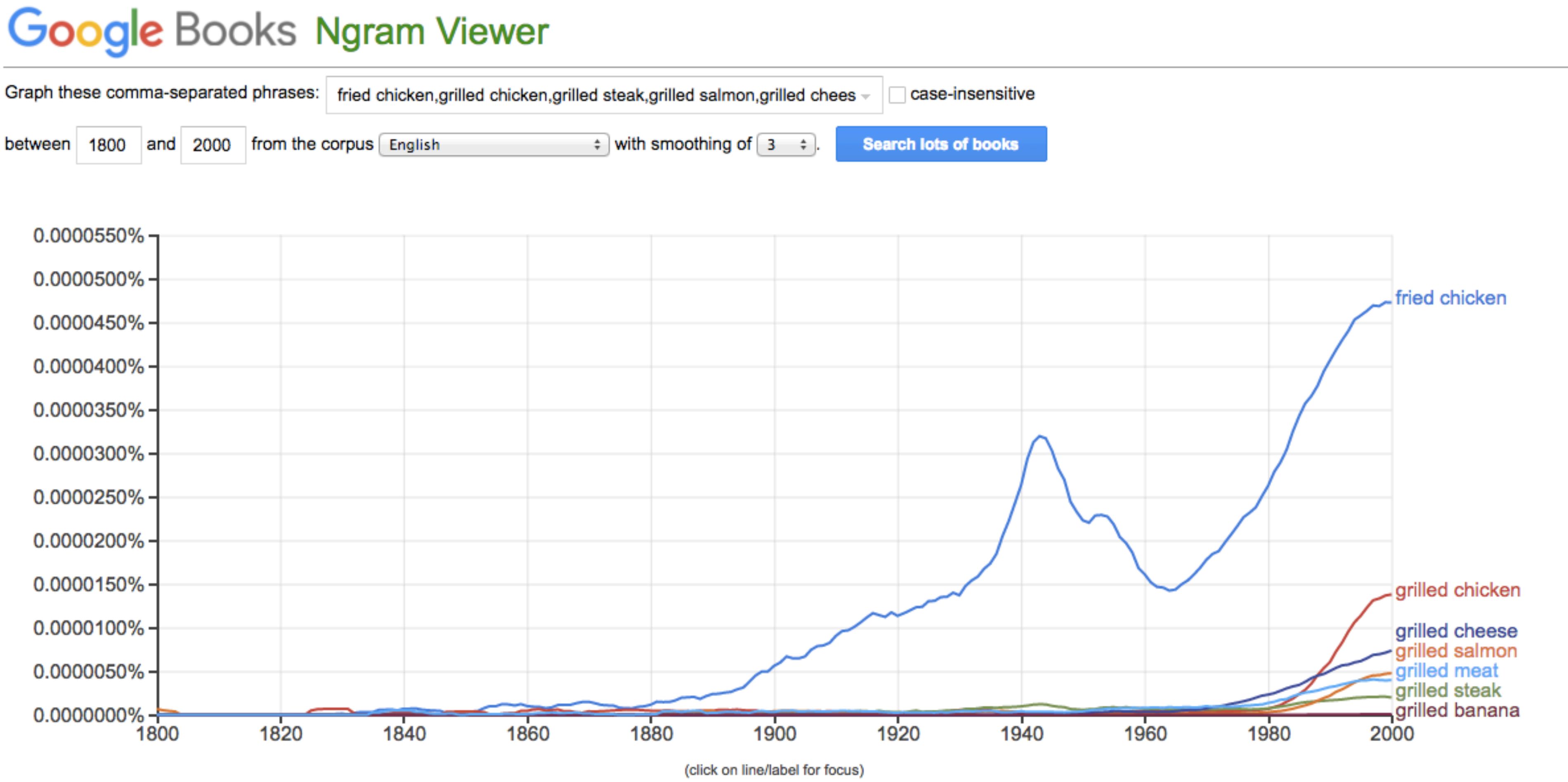


Graph these comma-separated phrases: grilled chicken,grilled steak,grilled salmon,grilled cheese,grilled bana

☐ case-insensitive

between 1800 and 2000 from the corpus English with smoothing of 3.

An example of dependent events in a sequence



Graph these comma-separated phrases: fried chicken,grilled chicken,grilled steak,grilled salmon,grilled chees ☐ case-insensitive

between 1800 and 2000 from the corpus English with smoothing of 3.

Markov chain

- Markov chain is a process in which outcome of any trial in a sequence is **conditioned by the outcome of the trial immediately preceding, but not by earlier ones.**
- Such dependence is called **chain dependence**

Andrey Markov (1856-1922)

Markov chain in terms of probability

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

- The sequence is a Markov chain if the probability distribution X_t only depends on the distribution of the immediately preceding random variable X_{t-1}

$$\boxed{P(X_t | X_0 \ldots, X_{t-1}) = P(X_t | X_{t-1})}$$

- If the conditional probabilities (transition probabilities) do **NOT change with time**, it's called **constant Markov chain**.

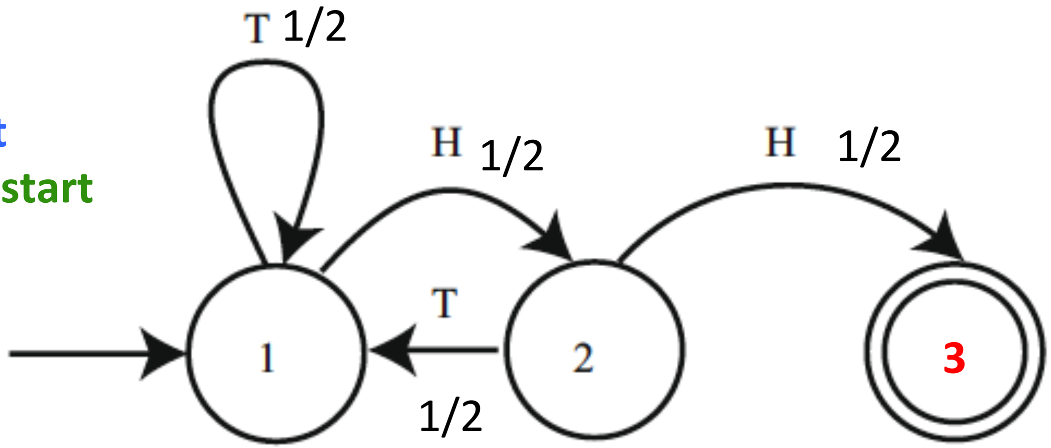
$$P(X_t | X_{t-1}) = P(X_{t-1} | X_{t-2}) = \ldots = P(X_1 | X_0)$$

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 **n** flips?

- Use a state diagram, which is a **directed graph**. Circles are the states of likely outcomes. Arrow directions show the direction of transitions. Numbers over the arrows show transition probabilities.

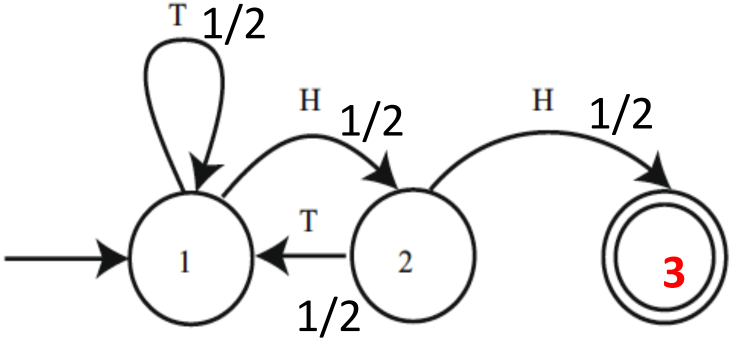
1 -> Start or just had tail/restart
2 -> had one head after start/restart
3 -> 2heads in a row/Stop



The model helps form recurrence formula

- Let p_n be the probability of stopping after **n** flips

$$p_1 = 0 \quad p_2 = 1/4 \quad p_3 = 1/8 \quad p_4 = 1/8 \quad \ldots$$



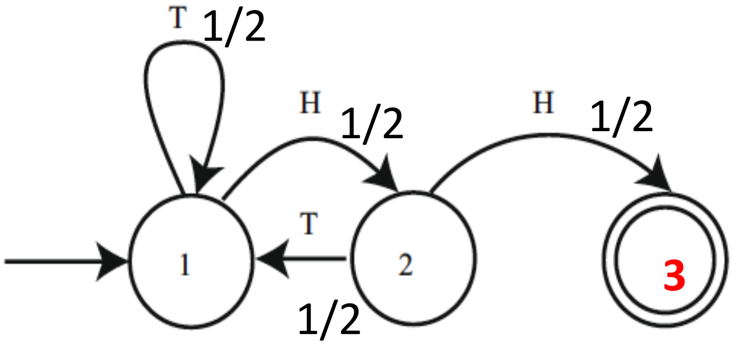
The model helps form recurrence formula

- Let p_n be the probability of stopping after **n** flips

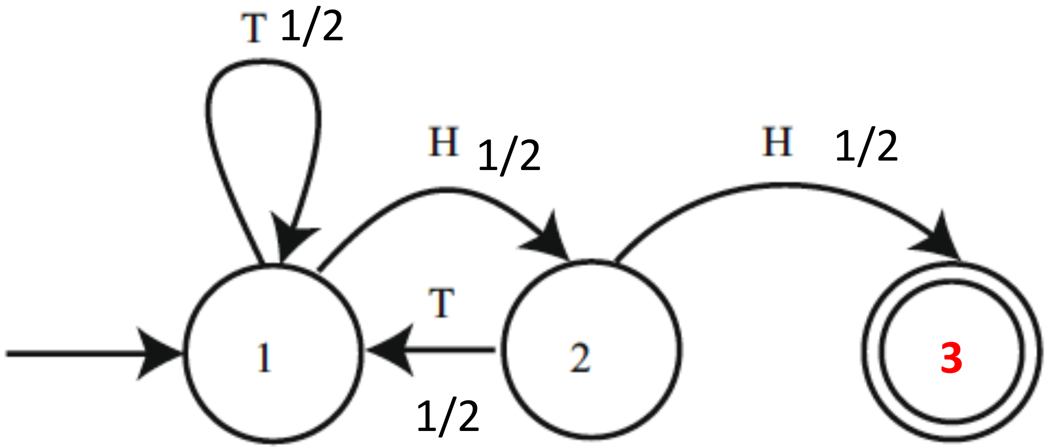
$$p_1 = 0 \quad p_2 = 1/4 \quad p_3 = 1/8 \quad p_4 = 1/8 \quad \ldots$$

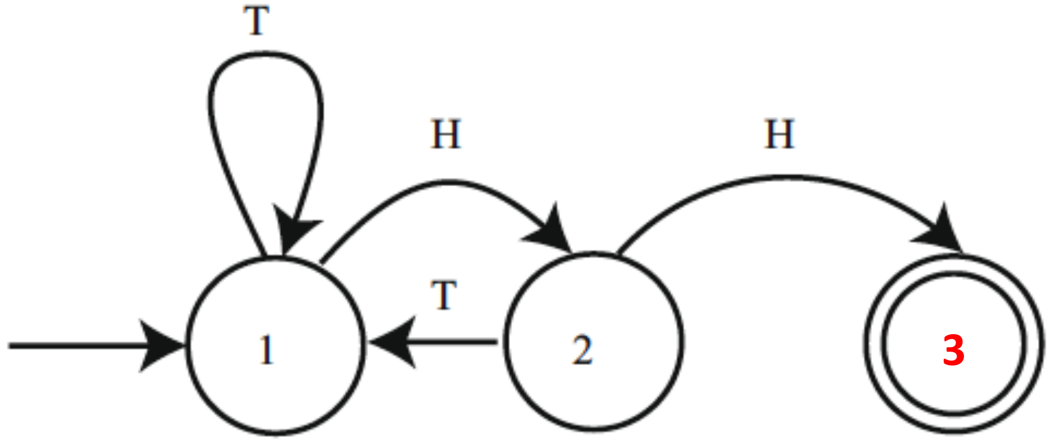
- If $n > 2$, there are two ways the sequence starts
  - Toss T and finish in n-1 tosses
  - Or toss HT and finish in n-2 tosses
- So we can derive a recurrence relation

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



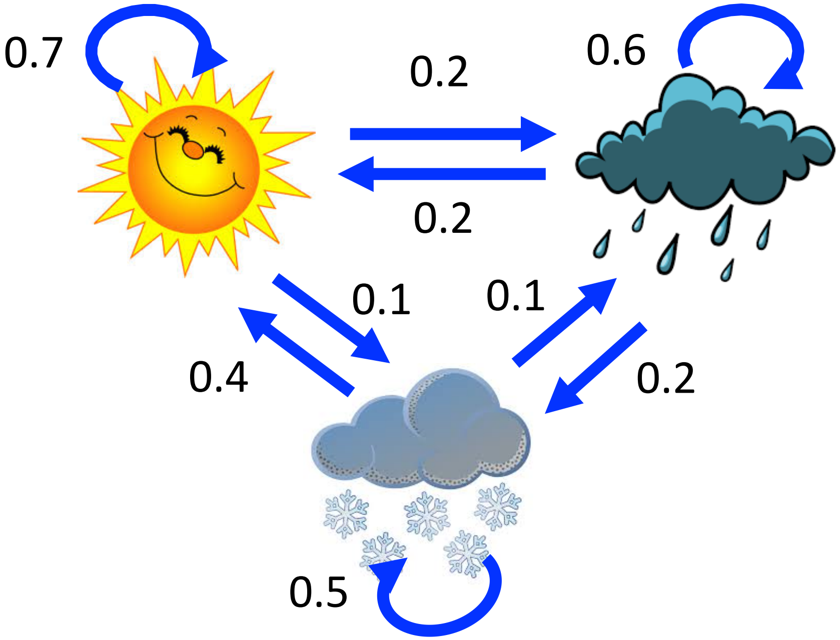
Transition probability btw states





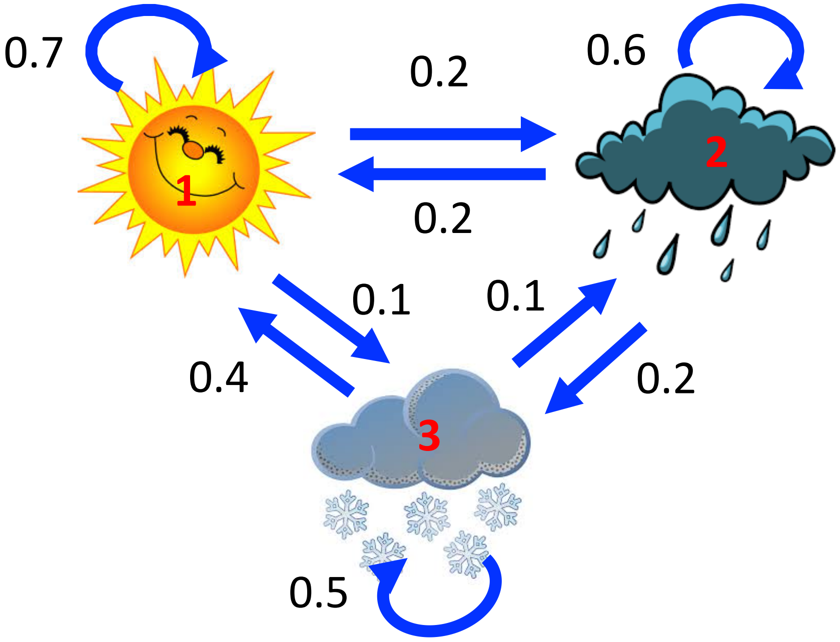
Transition probability matrix: weather model

- Let’s model daily weather as one of the three states (Sunny, Rainy, and Snowy) with Markov chain that has the transition probabilities as shown here.



Transition probability matrix: weather model

- Let's model daily weather as one of the three states (Sunny, Rainy, and Snowy) with Markov chain that has the transition probabilities as shown here.



i, the current state at time point t

j, the next state at time point t+1

$$\boldsymbol{P} = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{cccc} \text{Sunny} & \text{Rainy} & \text{Snowy} & \\ \left[\begin{matrix} 0.7 \\ 0.2 \\ 0.4 \end{matrix}\right. & \begin{matrix} 0.2 \\ 0.6 \\ 0.1 \end{matrix} & \left.\begin{matrix} 0.1 \\ 0.2 \\ 0.5 \end{matrix}\right] & \begin{matrix} \text{Sunny} \\ \text{Rainy} \\ \text{Snowy} \end{matrix} \end{array}$$

The transition probability matrix

Q: Is this TRUE?

For a constant Markov Chain, at any step **t**, the probability distribution among the states remain the same.

A. Yes.

B. No.

Q: The transition probabilities for a node sum to 1

A. Yes.

B. No.

Only the row sum is 1, that is: the probabilities associated with outgoing arrows sum to 1.

Transition probability matrix properties

- The transition probability matrix $\boldsymbol{P}$ is a square matrix with entries p_{ij}
- Since $p_{ij} = P(X_t = j | X_{t-1} = i)$

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

$$\boldsymbol{P} = \begin{array}{cc} & \begin{array}{ccc} \text{Sunny} & \text{Rainy} & \text{Snowy} \end{array} \\ & \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} \begin{array}{l} \text{Sunny} \\ \text{Rainy} \\ \text{Snowy} \end{array} \end{array}$$

↑

The transition probability matrix

Probability distributions over states

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

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

- For example: if it is rainy today, and today is t=0, then

$$\boldsymbol{\pi} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

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

$$\mathrm{p}_i^{(t)} = P(X_t = i)$$

Propagating the probability distribution

- Propagating from t=0 to t=1,

$$\begin{aligned} P_j^{(1)} &= P(X_1 = j) \\ &= \sum_i P(X_1 = j, X_0 = i) \\ &= \sum_i P(X_1 = j | X_0 = i) P(X_0 = i) \\ &= \sum_i p_{ij} \pi_i \end{aligned}$$

- In matrix notation,

$$\boxed{\boldsymbol{p}^{(1)} = \boldsymbol{\pi} P}$$

Probability distributions:

- Suppose that it is rainy, we have the initial probability distribution. $\boldsymbol{\pi} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$

- What are the probability distributions for tomorrow and the day after tomorrow?

$$\boldsymbol{p}^{(1)} = \boldsymbol{\pi} P$$

$$\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(1)} P$$

Propagating to t= ∞

- We have just seen that

$$\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(1)}P = (\boldsymbol{\pi}P)P = \boldsymbol{\pi}P^2$$

- So in general $\boldsymbol{p}^{(t)} = \boldsymbol{\pi}P^t$

- If one state can be reached from any other state in the graph, the Markov chain is called **irreducible** **(single chain)**.

- Furthermore , if it satisfies: $\lim_{t\to\infty} \boldsymbol{\pi}P^t = \boldsymbol{S}$

  then the Markov chain is stationary and **S** is the stationary distribution.

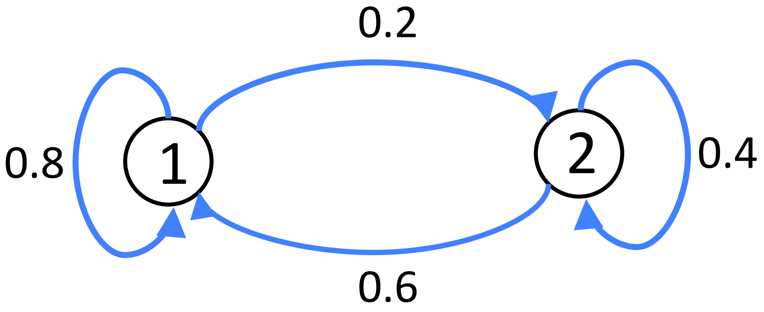
Stationary distribution

- The stationary distribution $\boldsymbol{s}$ has the following property: $\boldsymbol{s}P = \boldsymbol{s}$
- $\boldsymbol{s}$ is a row eigenvector of **P** with eigenvalue 1
- In the example of the weather model, regardless of the initial distribution,

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

Example: Up-to-date or behind model

State 1: Up-to-date
State 2: Behind



What's the transition matrix?

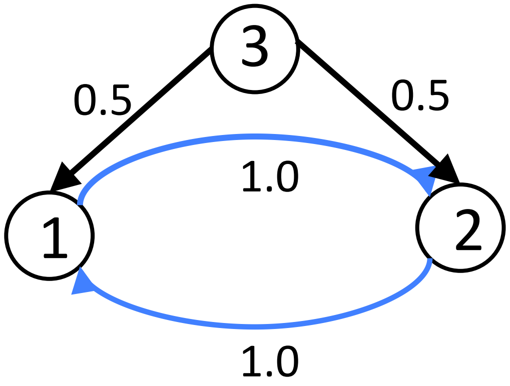
If I start with π = [0, 1] , what is my probability of being up-to-date eventually? 3/4

$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix}$$

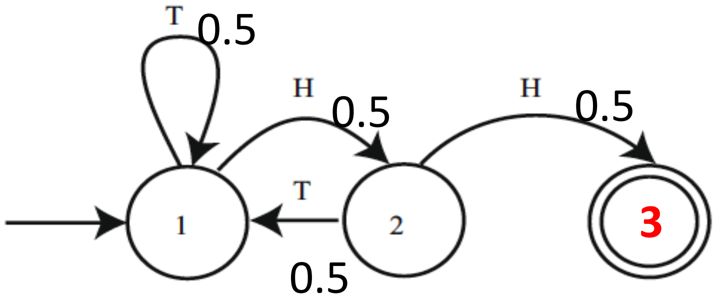
Example: Up-to-date or behind model

$$SP = S \Rightarrow (SP)^T = S^T \Rightarrow P^T S^T = S^T$$
$$(P^T - I)S^T = 0$$

Examples of non-stationary Markov chains



Periodic

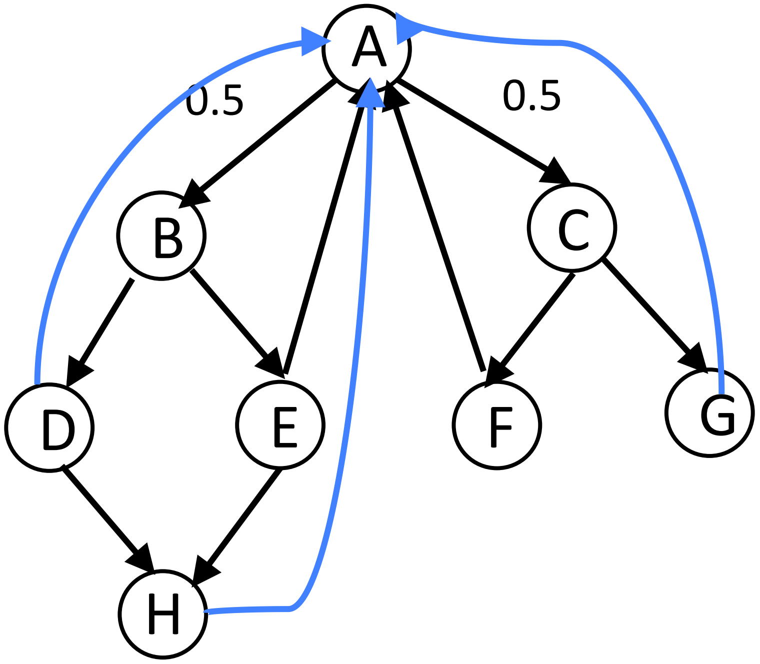


Absorbing

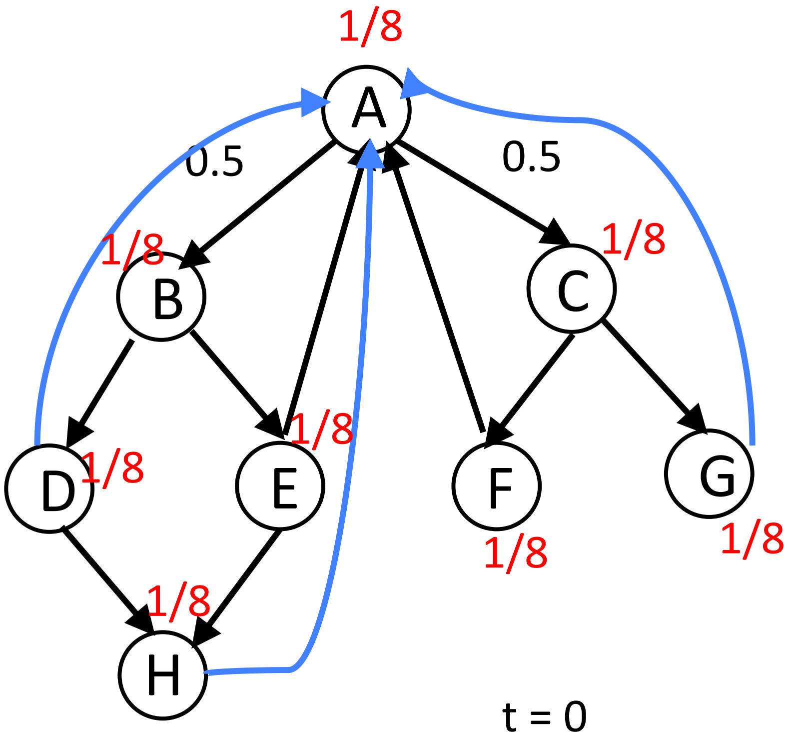
PageRank Example

- How to rate web pages objectively?
- The PageRank algorithm by Page et al. made **Google** successful
- The method utilized **Markov chain model** and applied it to the large list of webpages.
- To illustrate the point, we use a small-size example and assume a simple **stationary model**.

Suppose we are randomly surfing a network of webpages



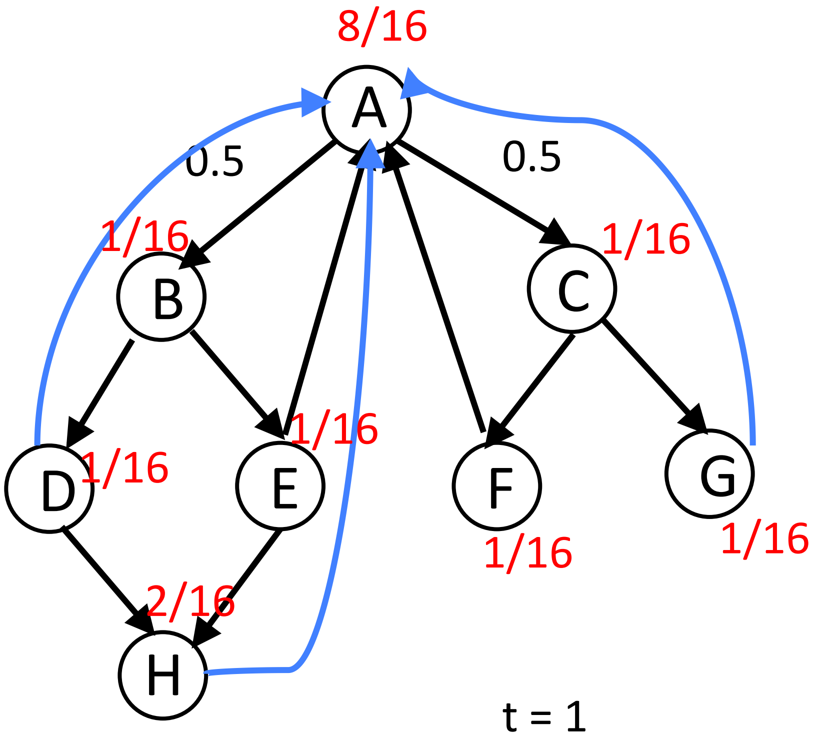
Initialize the distribution uniformly



t = 0

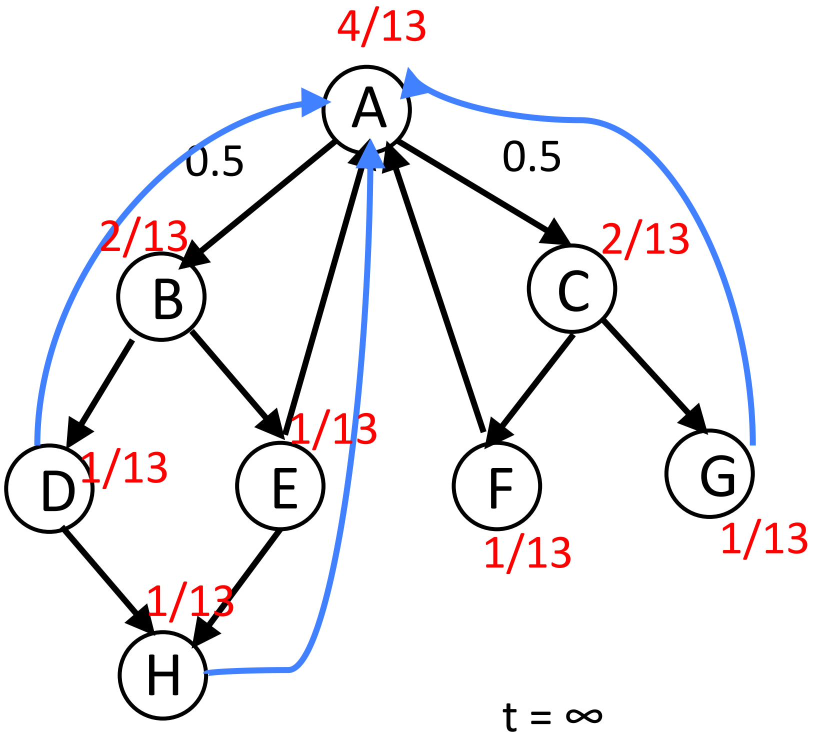
π =

Update the distribution iteratively



t = 1

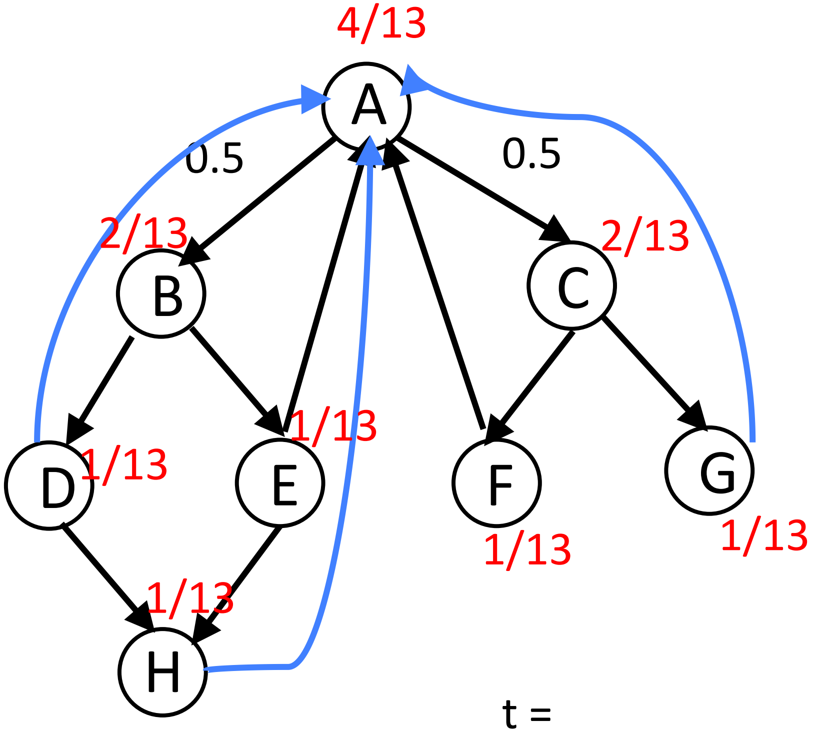
Until the stationary distribution



If the surfer get trapped

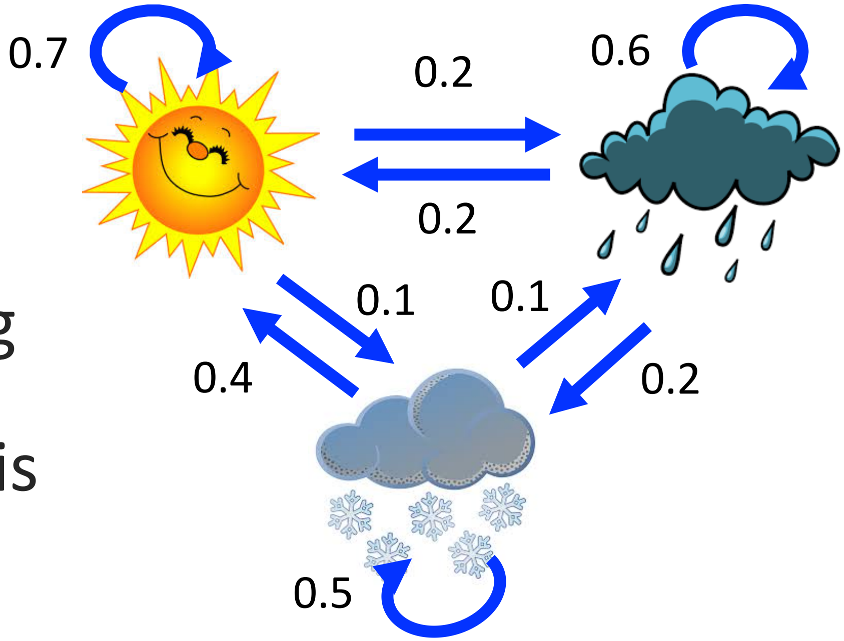
- Allow "teleport" with small probability from any page to another

- Or allow "teleport" with user input of URL



Diverse applications of Markov Model

- Communication network
- Queue modeling
- DNA sequence modeling
- Natural language processing
- Single-cell large data analysis
- Financial/Economic model
- Music



Final Exam

- Time: 8am-11am 5/10 Mon. Central Time
- Conflicts need to be requested 2 days ahead.
- Duration: 3hrs
- Content coverage: Ch1-14, except 8, details are on Compass
- Open book and lecture notes
- Format: 50 multiple choices, on PrairieLearn proctored by CBTF

Additional References

- Robert V. Hogg, Elliot A. Tanis and Dale L. Zimmerman. “Probability and Statistical Inference”
- Kelvin Murphy, “Machine learning, A Probabilistic perspective”

Acknowledgement

Thank You!