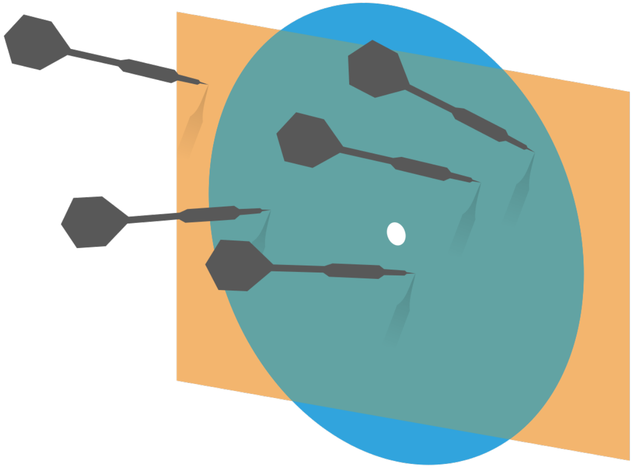
# Probability and Statistics for Computer Science



Credit: wikipedia

Conditional probability comes back in matrix!

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 12.08.2021

# Last time

- Markov Chain (I)

$$X_1 \quad X_2 \;-\;-\;-\; X_t$$

$$p(X_t \mid X_{t-1}) = f(t)$$

$$= c$$

$$p(X_1 \cap X_2 \cap \cdots X_t)$$

# Objective

- Markov Chain (II)
- Q/A
- Concept review

# 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)

$P(X_{n+1} \mid X_n)$

$X_{n-1} \cdots X_0$

$= f(n)$

# 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 $X_{t-1}$

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

Markov property

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

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

# Coin example

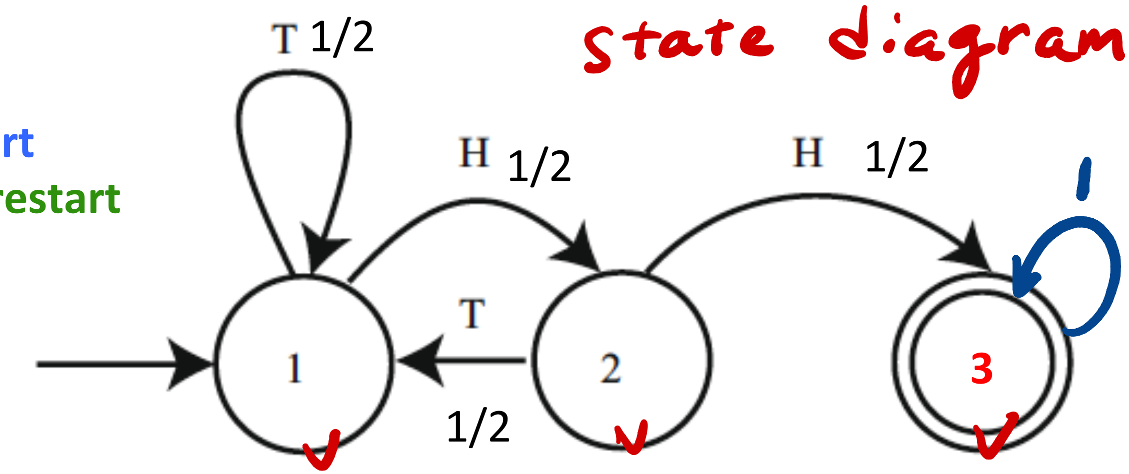
- Toss a fair coin until you see two heads in a row and then stop, what is the probability of stopping after exactly **n** flips?

$$\underbrace{* \; * * \; HH}_{n}$$

$$P(n = n_0) = ?$$

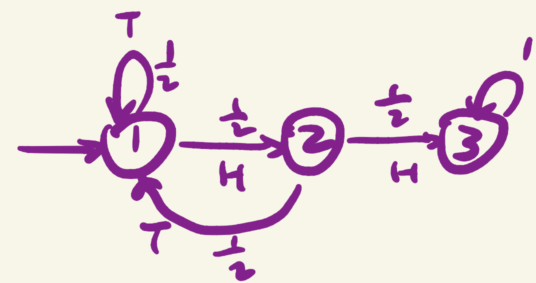
state diagram

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



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $N =$ | #1 | #2 | #3 | #4 | #5 | #6 | |
| Trials | T | T | H | T | H | H | $\theta$ |
| $X_N =$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | |
| State | 1 | 1 | 2 | 1 | 2 | 3 | |

Markov Property:




$$p_{ij} = P(X_{n+1}=j \mid X_n = i)$$

$$= P(X_{n+1}=j \mid X_n=i, \boxed{X_{n-1}=? \; - \; \cdots \; X_0=?})$$

This part can be any!!

# The model helps form recurrence formula

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

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

$P(n = n_0) = ?$

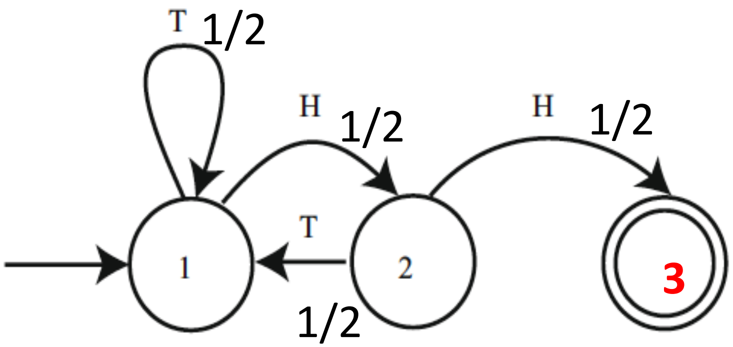
THH

$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$

HTHH

TTHH $= \frac{1}{8}$

$\frac{1}{16}$



# The model helps form recurrence formula

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

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

T (n-1)

- If $n > 2$, there are two ways the sequence starts

n

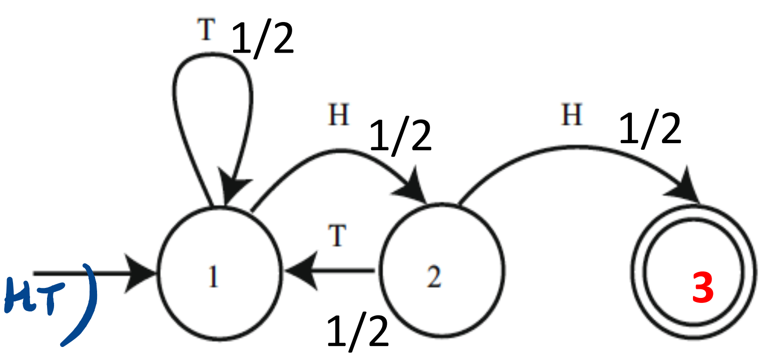
  - Toss T and finish in n-1 tosses
  - Or toss HT and finish in n-2 tosses

HT n-2

- So we can derive a recurrence relation

$$p_n = \frac{1}{2} p_{n-1} + \frac{1}{4} p_{n-2}$$

P(T)· P(n-1|T) P(HT)· P(n-2|HT)



$$r^n = \frac{1}{2} r^{n-1} + \frac{1}{4} r^{n-2}$$

$$r^2 = \frac{1}{2} r + \frac{1}{4}$$

$$4r^2 = 2r + 1$$

$$4r^2 - 2r - 1 = 0$$

$$r = \frac{2 \pm \sqrt{20}}{2 \times 4} = \frac{2 \pm 2\sqrt{5}}{8}$$

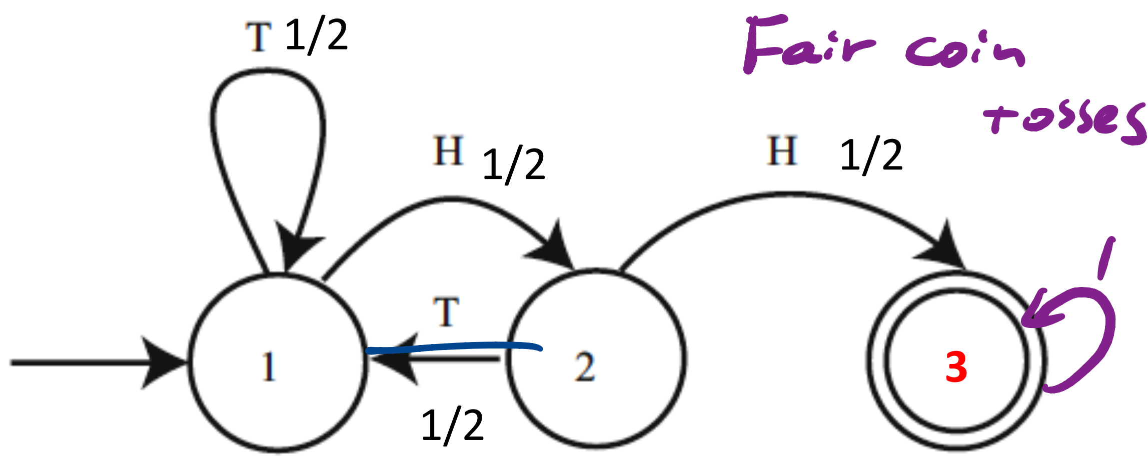
$$p_n = a r_1^n + b r_2^n$$

solve for a, b
using
$P_1 = 0$.
$P_2 = \frac{1}{4}$

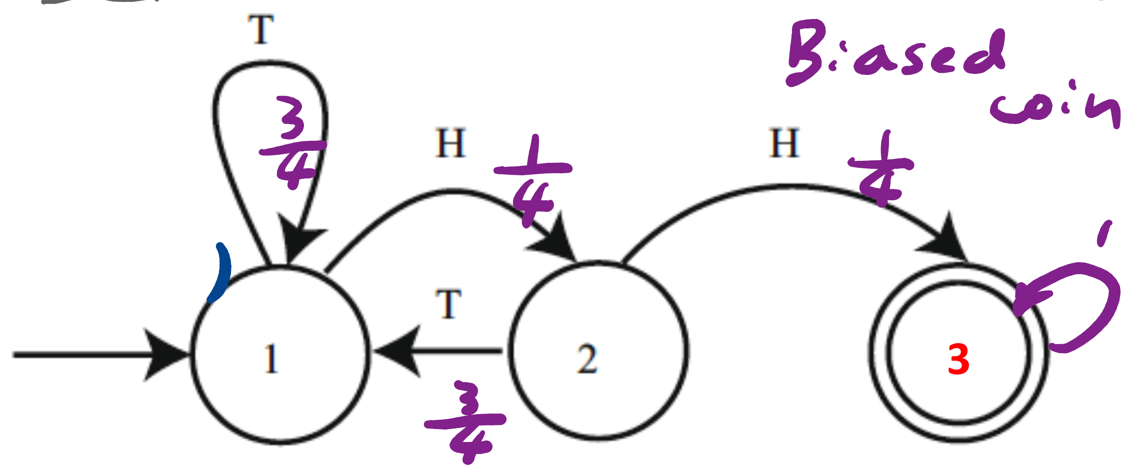
# Transition probability btw states

State

| t \ t+1 | 1 | 2 | 3 |
|---|---|---|---|
| 1 | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 |
| 2 | $\frac{1}{2}$ | 0 | $\frac{1}{2}$ |
| 3 | 0 | 0 | 1 |

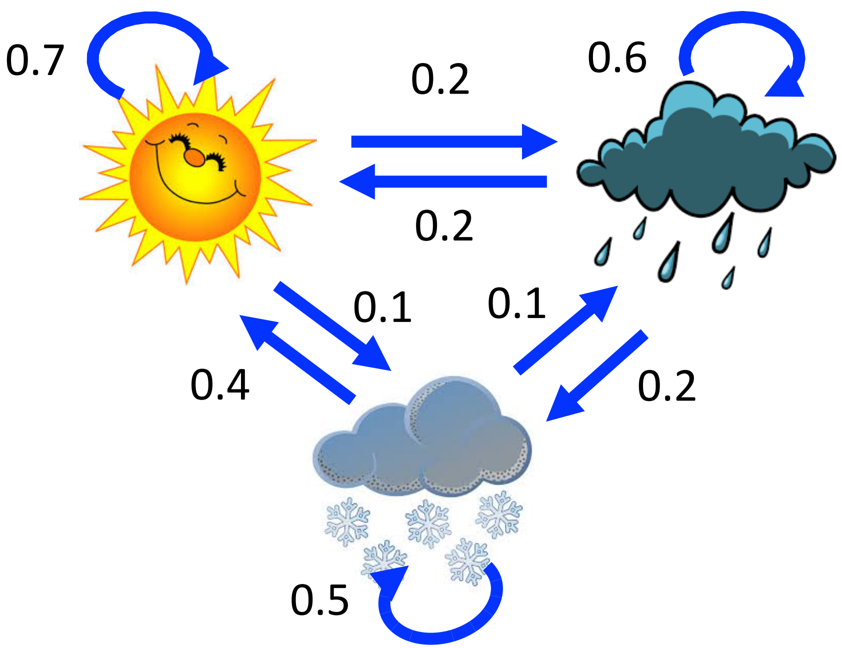



| | 1 | 2 | 3 |
|---|---|---|---|
| 1 | $\frac{3}{4}$ | $\frac{1}{4}$ | 0 |
| 2 | $\frac{3}{4}$ | 0 | $\frac{1}{4}$ |
| 3 | 0 | 0 | 1 |



# Transition probability matrix: weather model

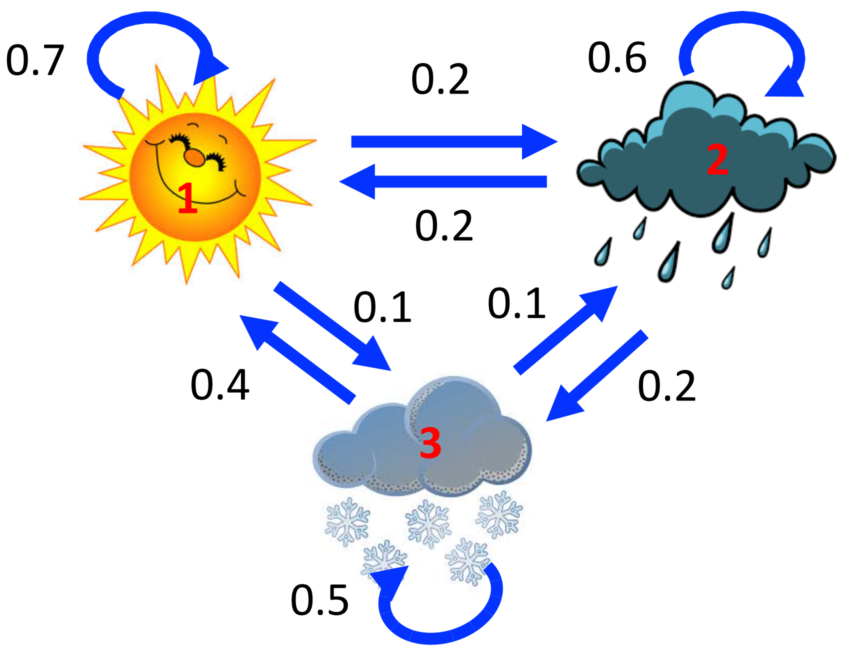
- Let's model daily weather as one of the three states (Sunny, Rainy, and Snowy) with Markov chain that has the transition probabilities as shown here.




| | Sunny | Rainy | Snowy |
|---|---|---|---|
| Sun | 0.7 | 0.2 | 0.1 |
| R | 0.2 | 0.6 | 0.2 |
| Snow | 0.4 | 0.1 | 0.5 |

# 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

$P(X_{t+1}=Sun \mid X_t=Sun)$

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

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.

$$P(X_{t+1} = s_a \mid X_t = s_b)$$
$$= C$$

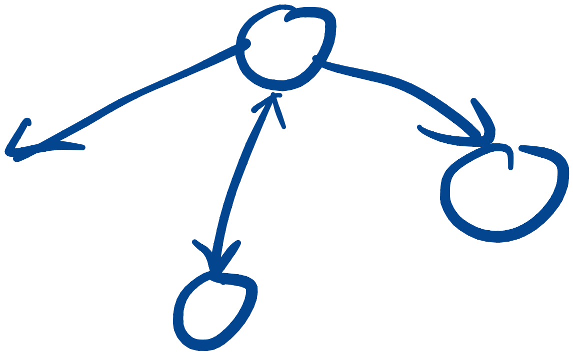
A. Yes.

$\boxed{\text{B.}}$ No.

# Q: The transition probabilities for a node sum to 1

A. Yes.

B. No.



# Transition probability matrix properties

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

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

1 $p_{ij} \geq 0$ and 2 $\sum_j p_{ij} = 1$

Stochastic Matrix

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

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

$$\boldsymbol{\pi} = \begin{array}{ccc} \text{Sunny} & \text{Rainy} & \text{Snowy} \\ \end{array}$$
$$\boldsymbol{\pi} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$P(A \cap B)$
$= P(A) P(B|A)$

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

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

$P(X_1 = 1)$
$\sum_j P(X_0 = S_j) P(X_1 | X_0)$

# Propagating the probability distribution

- Propagating from t=0 to t=1,

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

$\pi_0 = [0, 1, 0]$

- 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?

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

new prior

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

# 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: $\lim\limits_{t \to \infty} \boldsymbol{\pi} P^t = \boldsymbol{S}$

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

$\underset{t \to \infty}{P}(X_t)$

# Stationary distribution

- The stationary distribution $\boldsymbol{s}$ has the following property: $\boldsymbol{s}P = \boldsymbol{s}$

  $S_{t+1}$ $\quad S_t$

  $S_{\text{at } t} \cdot P = S_{\text{at } t+1}$

- $\boldsymbol{s}$ is a row eigenvector of **P** with eigenvalue 1

- In the example of the weather model, regardless of the initial distribution,

Sunny Rainy Snowy

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

$$sp = s$$

$$(sp)^T = s^T$$

$$p^T s^T = s^T$$

$$X = s^T$$

$$AX = X$$

$$\lambda = 1$$

# Chance of being up-to-date

In a class, students are either up-to-date or behind regarding progress. If a student is up-to-date, the student has 0.8 probability remaining up-to-date, if a student is behind, the student has 0.6 probability becoming up-to-date. Suppose the course is so long that it runs life long, what is the probability any student eventually gets up-to-date?

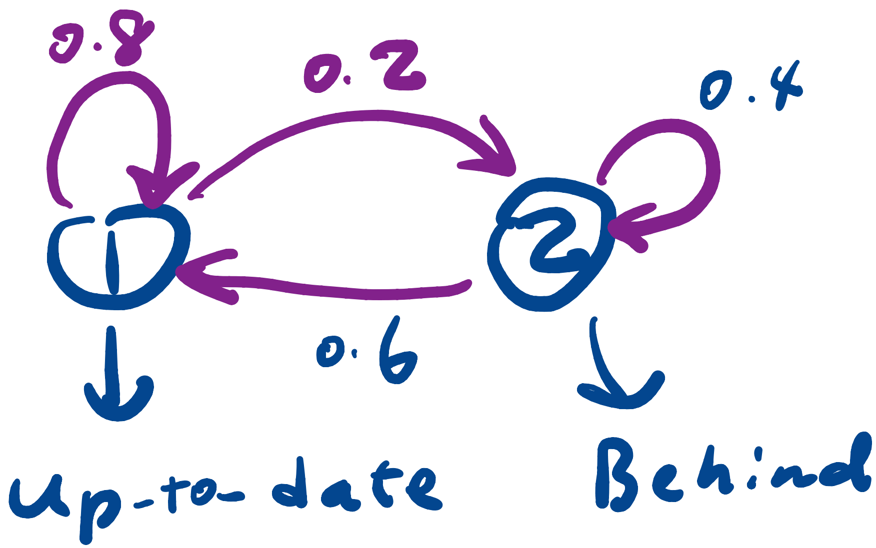
A) 25%

C) 75%

B) 50%

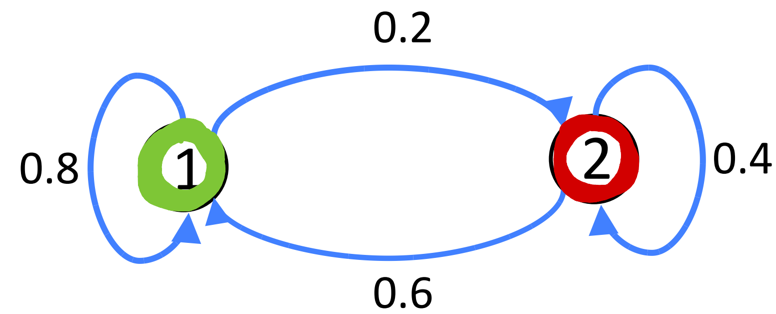
D) 95%

# The Markov Model



# Example: Up-to-date or behind model

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




**What's the transition matrix?**
If I start with $\pi$ = [ 0, 1] , what is my probability of being up-to-date eventually? 3/4

$$
P = \begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} (U)\,1 \\ (B)\,2 \end{array} & \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{array}
$$

$S = ?$

# Solving the stationary Markov

Given $SP = S$ What is $s$? $u^T$

$(SP)^T = S^T$ $s = [\frac{3}{4} \quad \frac{1}{4}]$

$P^T S^T = S^T$

$Au = u$ $(A = P^T,\ u = S^T)$

$Au = 1 \times u \Rightarrow Au = \lambda u$ $(\lambda = 1)$

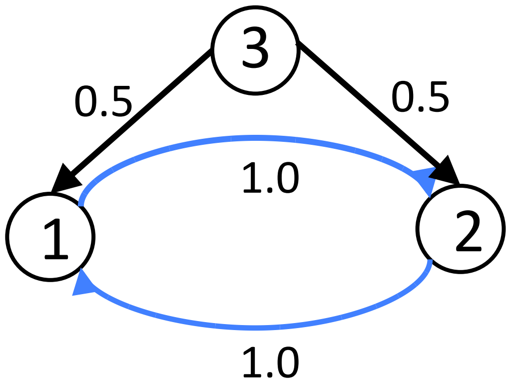
$[A - I]u = 0$ $\begin{bmatrix} 0.8-1 & 0.2-0 \\ 0.6-0 & 0.4-1 \end{bmatrix} u = 0$

$u = ?$ $u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ $u_1 + u_2 = 1$

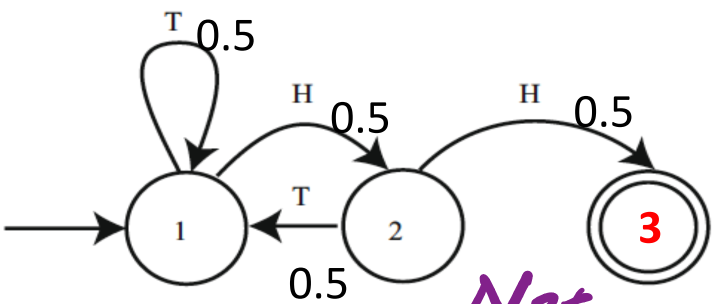
$u_1 = \frac{3}{4}$ $u_2 = \frac{1}{4}$

# Examples of non-stationary Markov chains

Not irreducible




Periodic




Not stationary

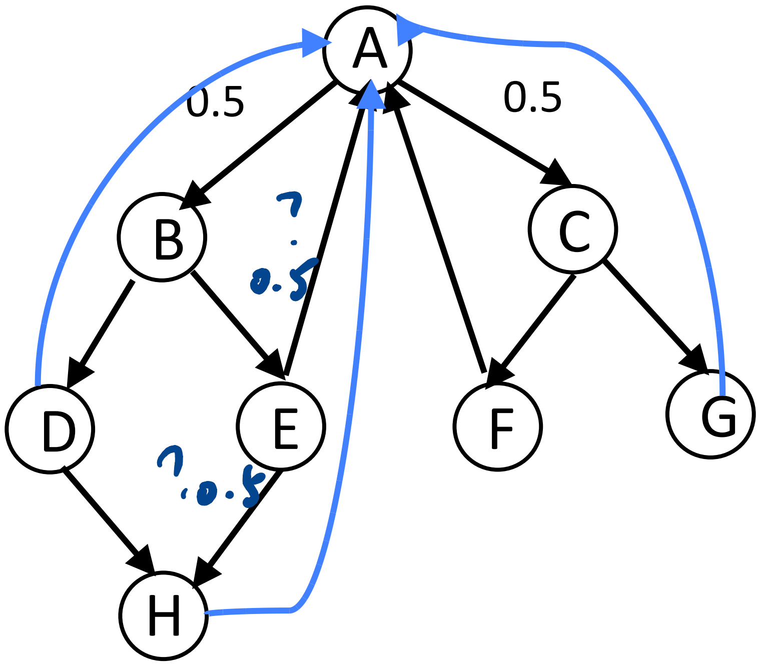
Absorbing

$SP = S$

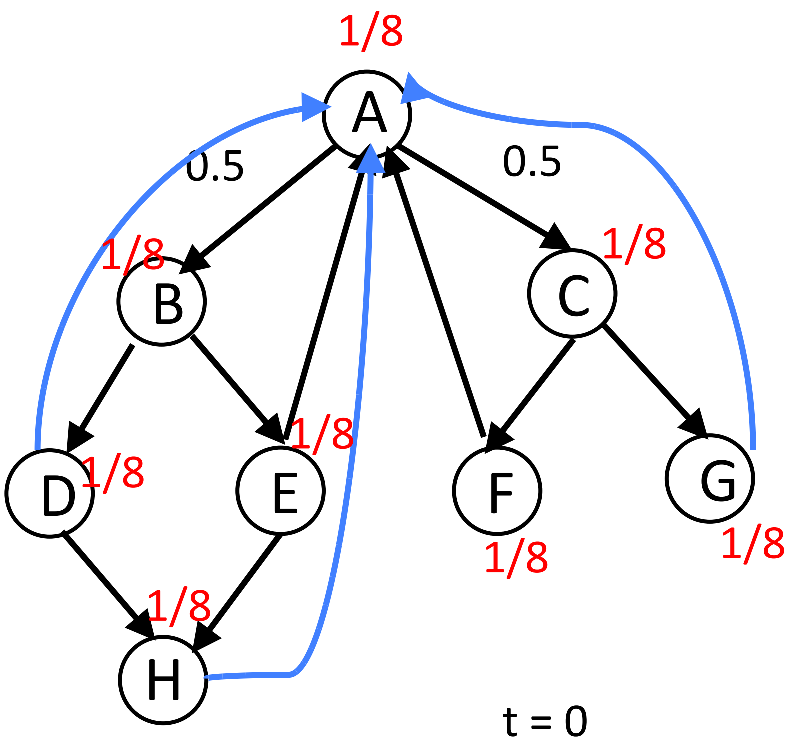
# PageRank Example

- 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



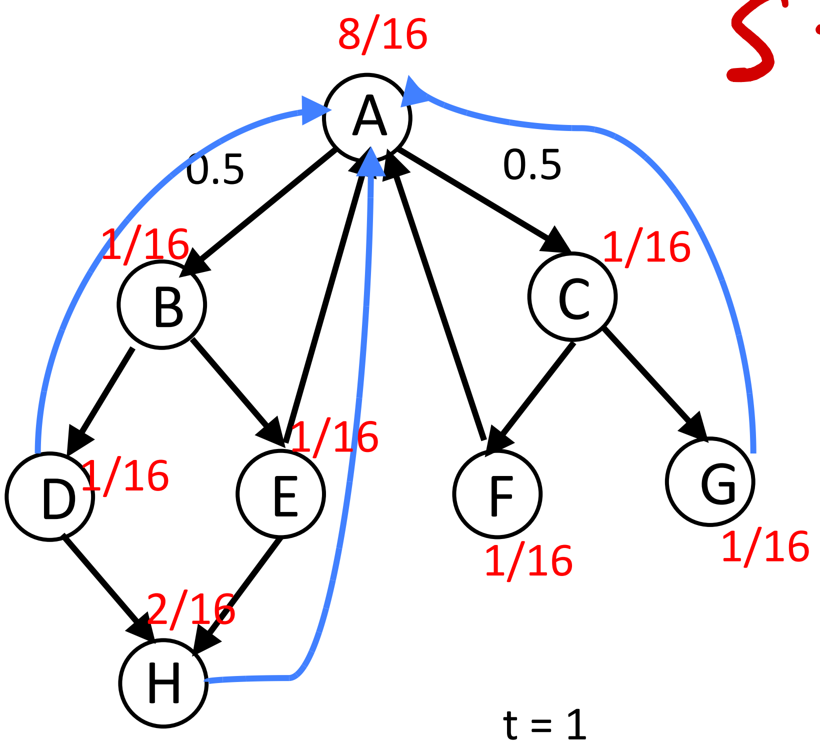

$$\boldsymbol{\pi} = \begin{bmatrix} 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 \end{bmatrix}$$

$$\begin{array}{cc} & \begin{array}{cccccccc} A & B & C & D & E & F & G & H \end{array} \\ P = & \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \end{array}$$

$\vdots$

H

# Update the distribution iteratively



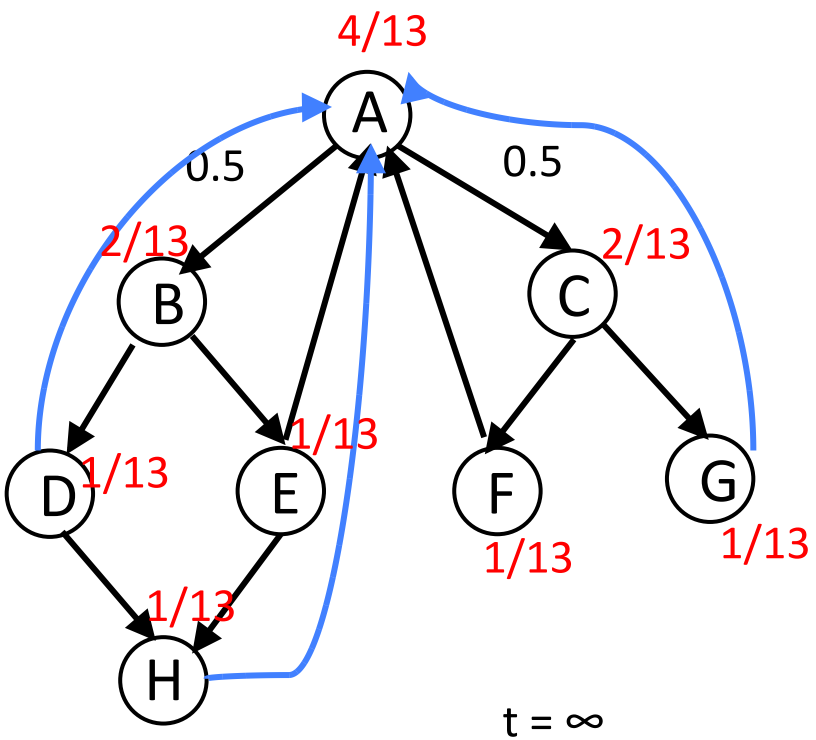

$$S = \lim_{t \to \infty} \pi P^t = ?$$

$$SP = S$$

$$(SP)^T = S^T$$

$$P^T S^T = S^T$$

# Until the stationary distribution

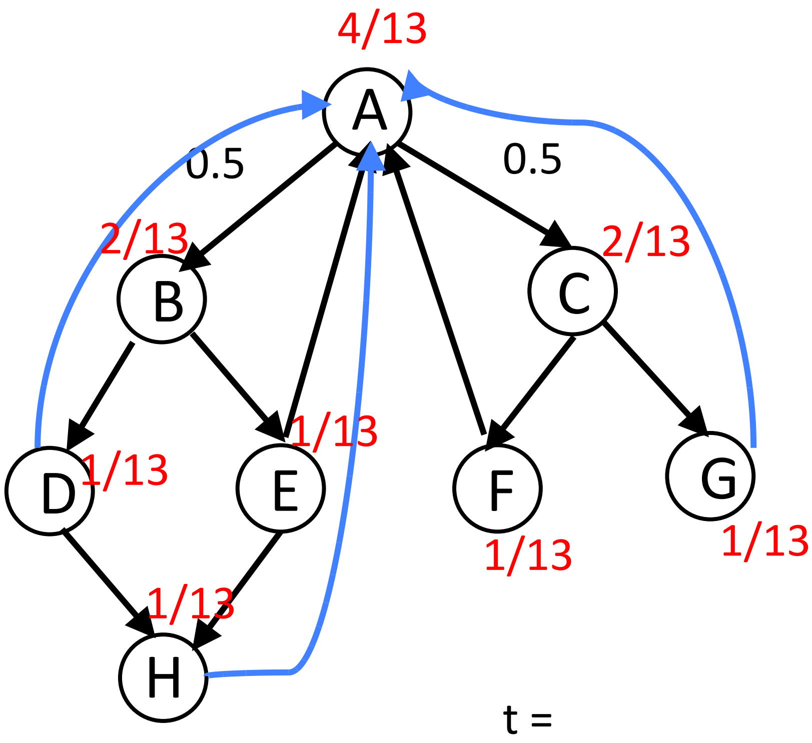



why don't we try $SP = S$ ?

# pages ~ $4 \times 10^9$

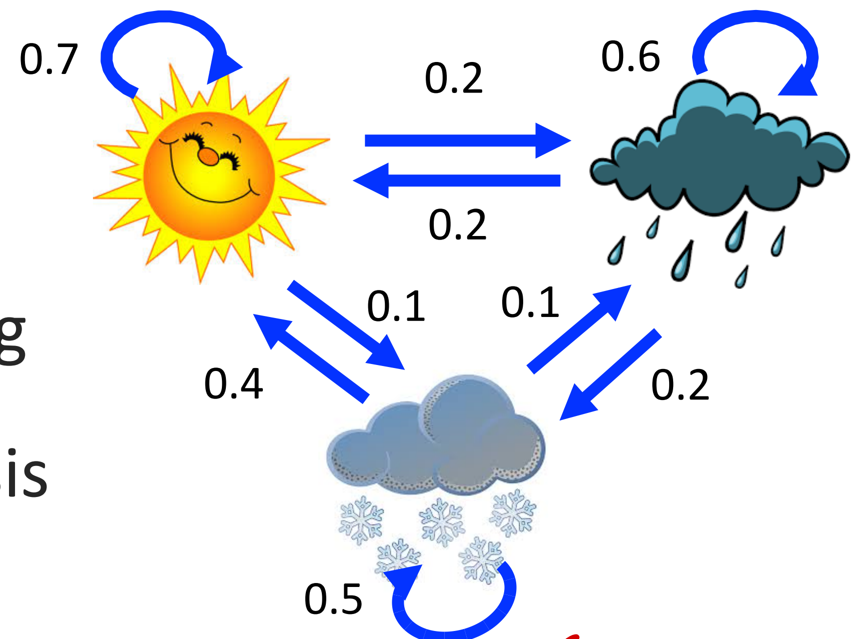
# If the surfer get trapped

- 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

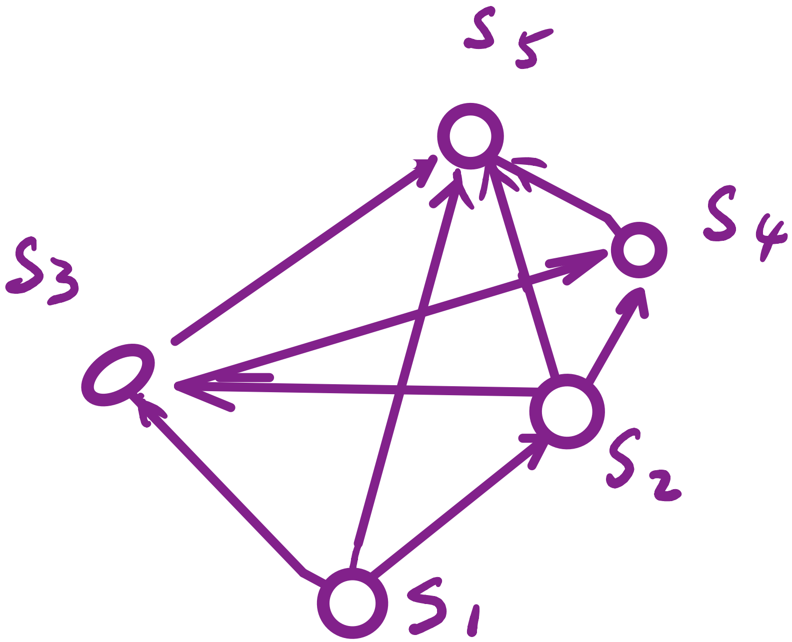



A T C G

C D E F G A B

octaves

# Communication Network example




$$
\begin{array}{c}
\begin{array}{ccccc} S_1 & S_2 & S_3 & S_4 & S_5 \end{array} \\
\begin{bmatrix}
0.1 & 0.2 & 0.3 & 0 & 0.3 \\
0 & 0.1 & 0.1 & 0.2 & 0.6 \\
0 & 0 & 0.1 & 0.1 & 0.8 \\
0 & 0 & 0 & 0.1 & 0.9 \\
0 & 0 & 0 & 0 & 1.0
\end{bmatrix}
\end{array}
$$

$$P_5^{(2)} = \pi_1 \cdot P(5)$$

$$\pi_0 = [1 \quad 0 \quad 0 \quad 0 \quad 0]$$

$$p_5^{(n)} = ? \qquad p_5^{(1)} = \qquad p_5^{(2)} = ?$$

$P_5^{(1)} = 0.3$

$P_5^{(2)} = 0.69$

$P_5^{(3)} = 0.844$

$P_5^{(4)} = 0.8806$

$P(X_{t-1} \cap X_t)$

$= P(X_{t-1}) \cdot P(X_t \mid X_{t-1})$

# Final Exam

- Time: 1:30pm 12/13 Mon. Central Time
- Duration: 3hrs
- Content coverage: Ch1-14, except 8, details are on Canvas
- Open book and lecture notes
- Format: 50 multiple choices, on PrairieLearn proctored by Staff on Zoom

# 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!*