# Markov Chains

# Statistical Problem

. We may have an underlying evolving system

(new state) = f(old state, noise)

- Input data: series of observations $X_1, X_2 \dots X_t$

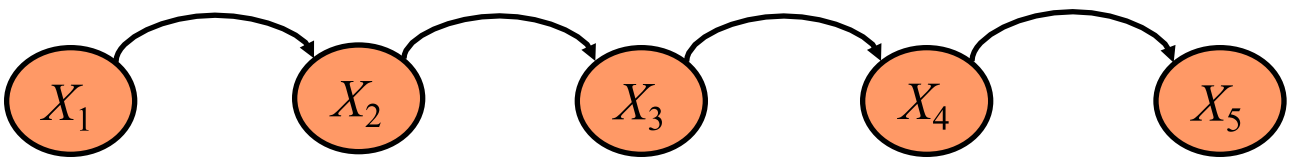
- Consecutive speech feature vectors are related to each other.

- We cannot assume that observations are i.i.d.

# Markov Process

- **Markov Property:** The state of the system at time $t+1$ depends only on the state of the system at time $t$

$$\Pr\left[X_{t+1} = x_{t+1} \mid X_1 \cdots X_t = x_1 \cdots x_t\right] = \Pr\left[X_{t+1} = x_{t+1} \mid X_t = x_t\right]$$




- **Stationary Assumption:** Transition probabilities are independent of time ($t$)

$$\Pr\left[X_{t+1} = b \mid X_t = a\right] = p_{ab}$$

**Bounded memory transition model**

# Markov Process

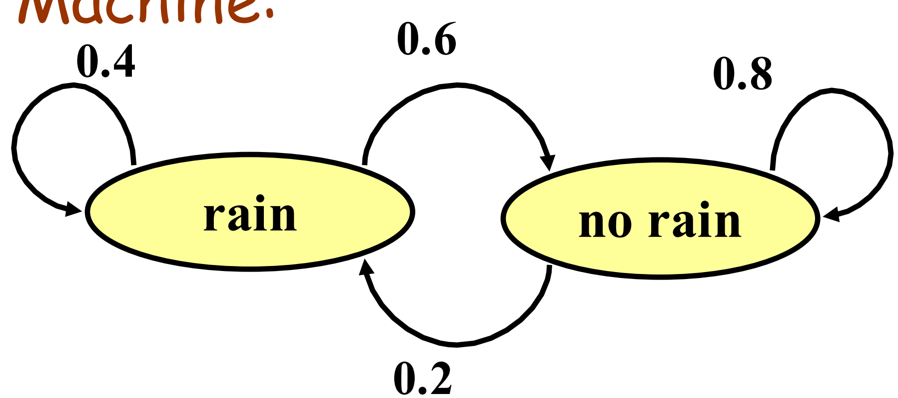
## Simple Example

**Weather:**

- raining today ⇒ 40% rain tomorrow
  ⇒ 60% no rain tomorrow

- not raining today ⇒ 20% rain tomorrow
  ⇒ 80% no rain tomorrow

**Stochastic Finite State Machine:**




$Pr(X_{T+1}=norain|X_T=rain)=0.6$

$Pr(X_{T+1}=rain|X_T=rain)=0.4$

# Markov Process

## Simple Example

**Weather:**

- **raining today** ⇒ **40% rain tomorrow**
  ⇒ **60% no rain tomorrow**

- **not raining today** ⇒ **20% rain tomorrow**
  ⇒ **80% no rain tomorrow**

The transition matrix:

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

- Stochastic matrix:
  Rows sum up to 1

# Markov Process

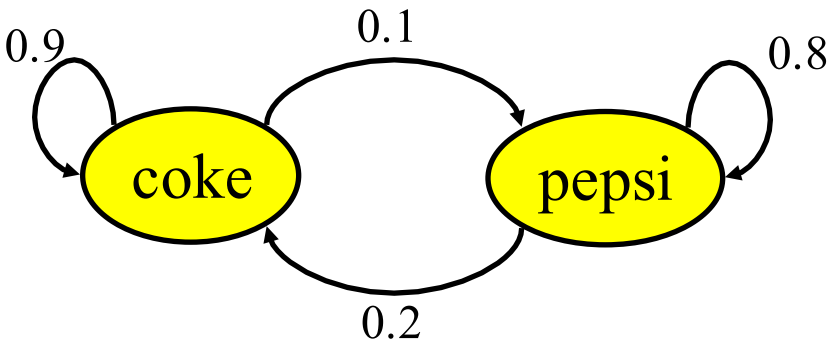
## Coke vs. Pepsi Example

- Given that a person's last cola purchase was Coke, there is a 90% chance that his next cola purchase will also be Coke.

- If a person's last cola purchase was Pepsi, there is an 80% chance that his next cola purchase will also be Pepsi.

transition matrix:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$



# Markov Process

## Coke vs. Pepsi Example (cont)

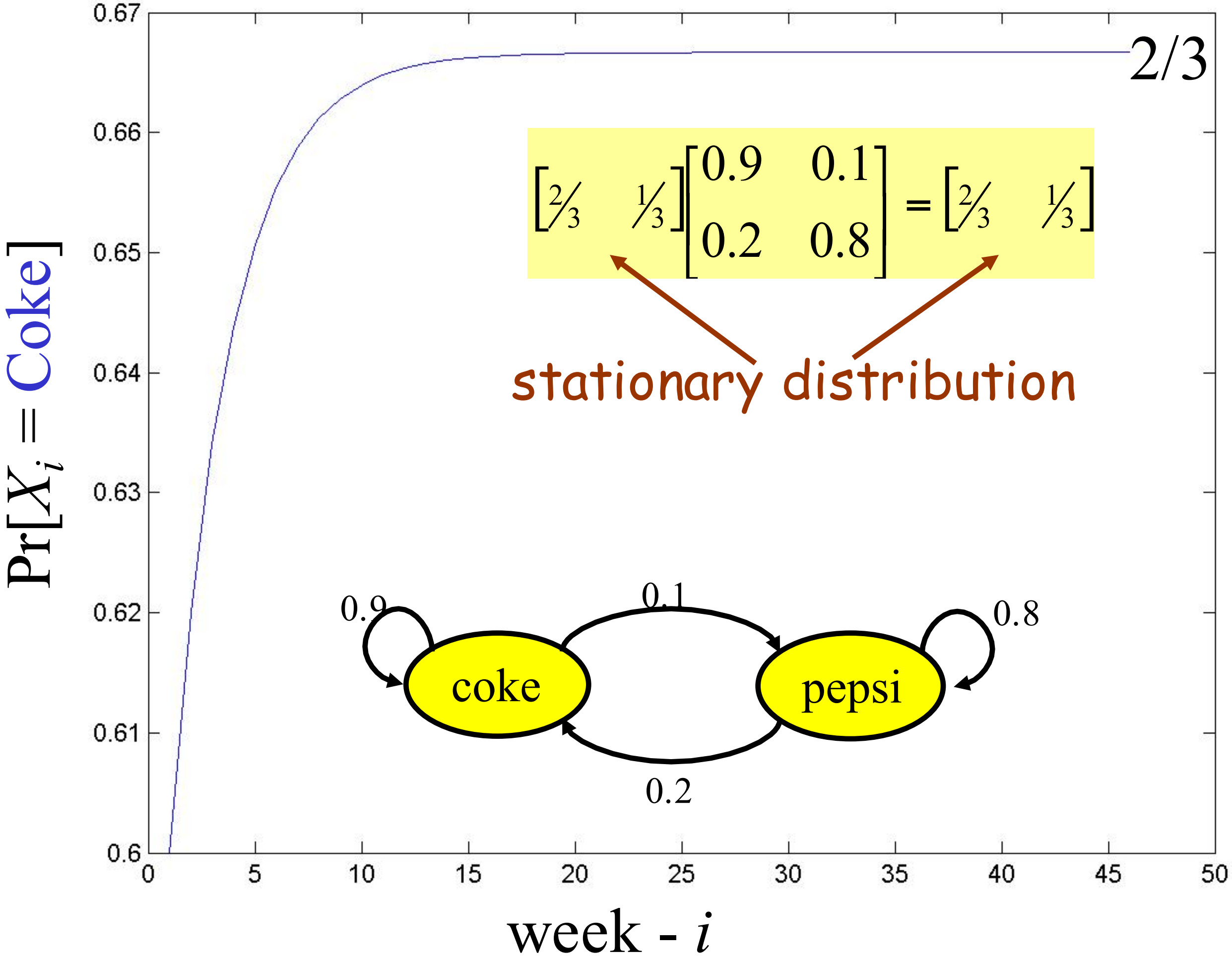
Given that a person is currently a Coke purchaser, what is the probability that he will purchase Pepsi **three** purchases from now?

$$P^3 = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix} = \begin{bmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{bmatrix}$$

# Markov Process

Coke vs. Pepsi Example (cont)

Simulation:




Eigen-Value problem

# Steady-State Probabilities

Property 2: Let $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_m)$ is the $m$-dimensional row vector of steady-state (unconditional) probabilities for the state space $\boldsymbol{S} = \{1,\ldots,m\}$. To find steady-state probabilities, solve linear system:

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}, \ \Sigma_{j=1,m}\, \pi_j = 1, \ \pi_j \geq 0, \ j = 1,\ldots,m$$

Brand switching example:

$$(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \begin{bmatrix} 0.90 & 0.07 & 0.03 \\ 0.02 & 0.82 & 0.16 \\ 0.20 & 0.12 & 0.68 \end{bmatrix}$$

$$\pi_1 + \pi_2 + \pi_2 = 1, \ \pi_1 \geq 0, \ \pi_2 \geq 0, \ \pi_3 \geq 0$$

# Steady-State Equations for Brand Switching Example

$$\pi_1 = 0.90\pi_1 + 0.02\pi_2 + 0.20\pi_3$$

$$\pi_2 = 0.07\pi_1 + 0.82\pi_2 + 0.12\pi_3$$

$$\pi_3 = 0.03\pi_1 + 0.16\pi_2 + 0.68\pi_3$$

$$\pi_1 + \pi_2 + \pi_3 = 1$$

$$\pi_1 \geq 0, \ \pi_2 \geq 0, \ \pi_3 \geq 0$$

Total of 4 equations in 3 unknowns

➔ Discard 3rd equation and solve the remaining system to get :

$$\pi_1 = 0.474, \ \pi_2 = 0.321, \ \pi_3 = 0.205$$

➔ $$q_1(0) = 0.25, \ q_2(0) = 0.46, \ q_3(0) = 0.29$$

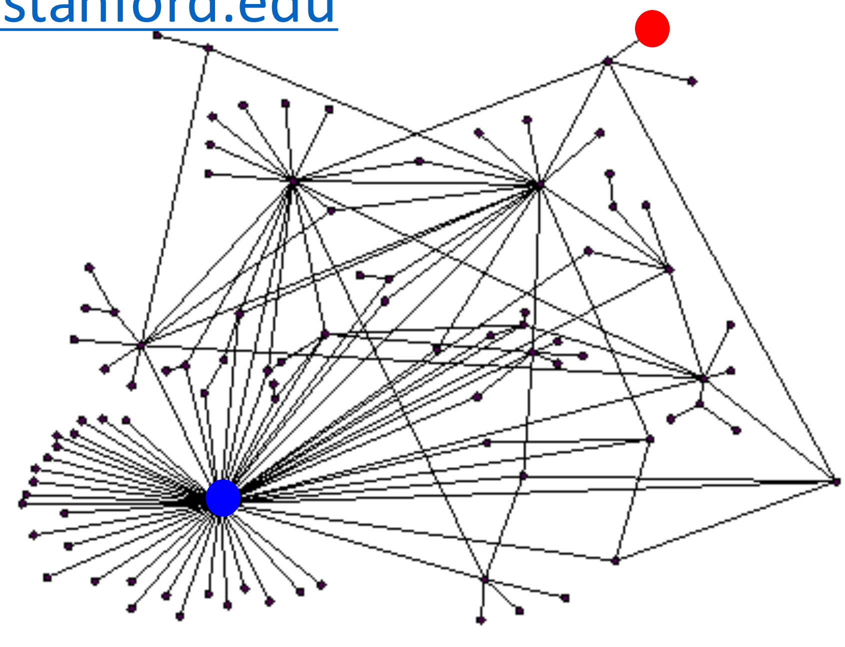
Steady-state probabilities may not exist for some Markov chains

# Ranking Nodes on the Graph: PageRank (Google)

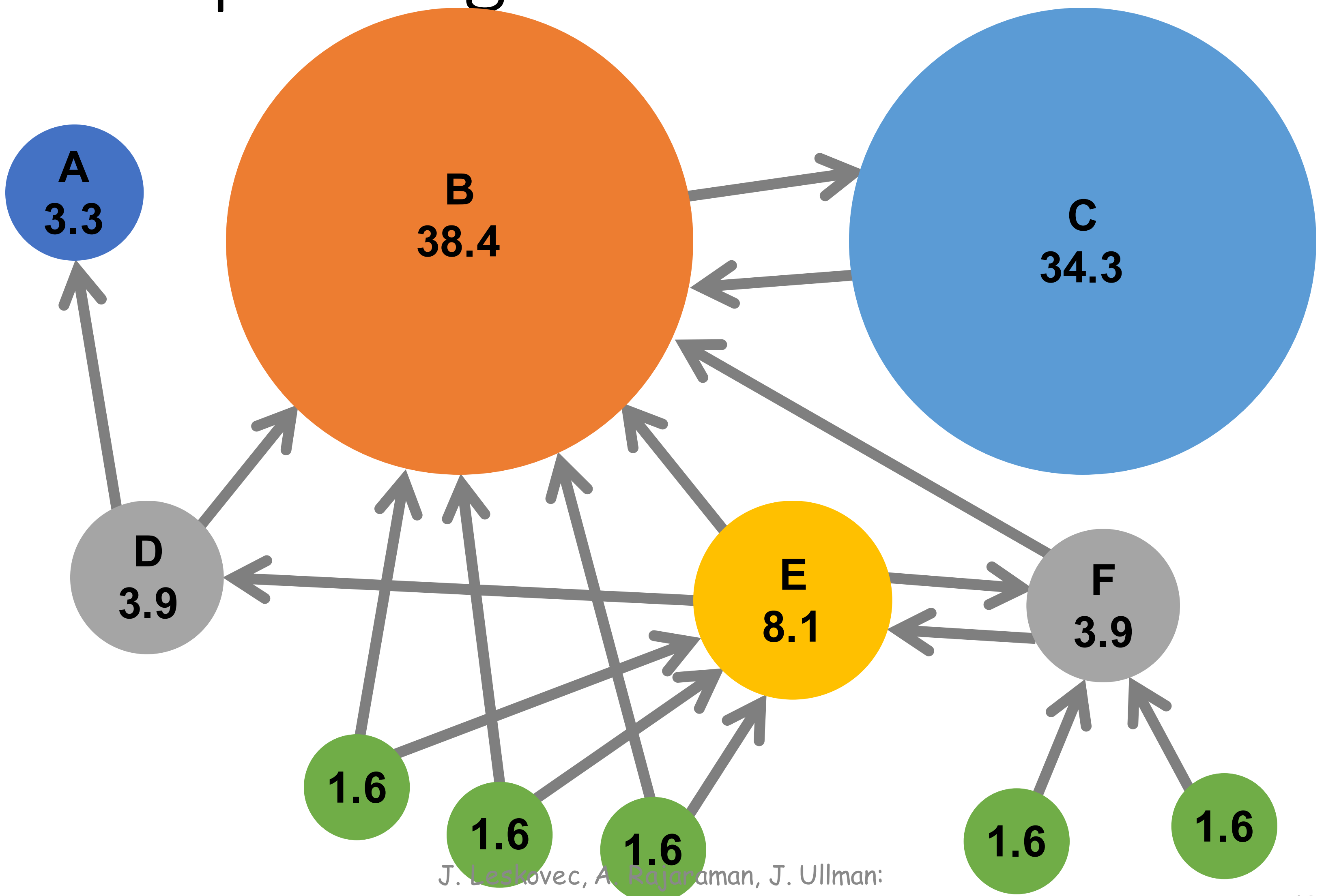
- **All Internet web pages are not equally "important"**

  www.joe-schmoe.com vs. www.stanford.edu

- There is large diversity in the web-graph node connectivity. **Let's rank the pages by the link structure!**



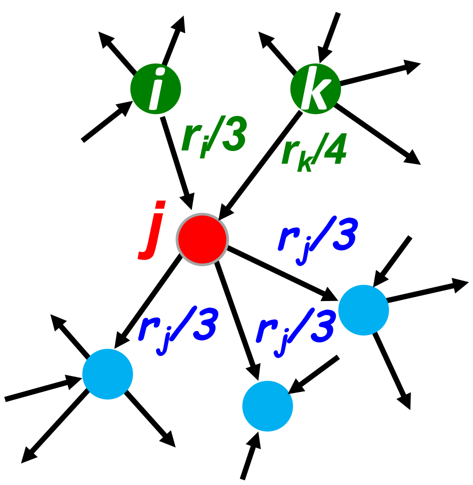
# Example: PageRank Scores



# Simple Recursive Formulation

- Each link's vote is proportional to the **importance** of its source page
- If page $j$ with importance $r_j$ has $n$ out-links, each link gets $r_j / n$ votes
- Page $j$'s own importance is the sum of the votes on its in-links

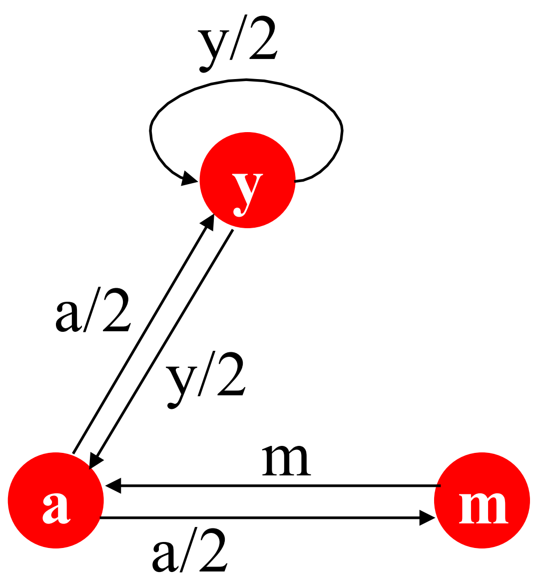
$$r_j = r_i/3 + r_k/4$$



# PageRank: The Markov Model

- **A "vote" from an important page is worth more**
- **A page is important if it is pointed to by other important pages**
- **Define a "rank" $r_j$ for page $j$**

[redacted]



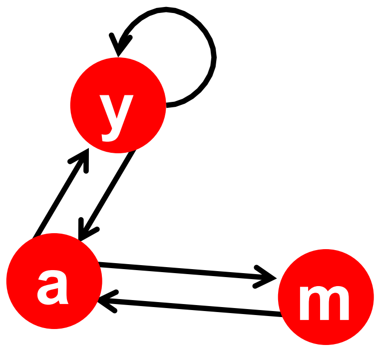

$$r_j = \sum_{i \to j} \frac{r_i}{\mathrm{d}_i}$$

**Equations:**

$$r_y = r_y/2 + r_a/2$$
$$r_a = r_y/2 + r_m$$
$$r_m = r_a/2$$

$d_i$ **… out-degree of r**

# Example: Web Equations




| | y | a | m |
|---|---|---|---|
| y | ½ | ½ | 0 |
| a | ½ | 0 | 1 |
| m | 0 | ½ | 0 |

$\boldsymbol{r = M \cdot r}$

$$r_y = r_y/2 + r_a/2$$
$$r_a = r_y/2 + r_m$$
$$r_m = r_a/2$$

$$\begin{bmatrix} y \\ a \\ m \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} y \\ a \\ m \end{bmatrix}$$

Notice that the web transition matrix $M = P^T$

# Solving the steady-state Equations

**Equations:**

$r_y = r_y/2 + r_a/2$

$r_a = r_y/2 + r_m$

$r_m = r_a/2$

- **3 equations, 3 unknowns, no constants**
  - No unique solution
  - All solutions equivalent modulo the scale factor
- **Additional constraint forces uniqueness:**
  - $\boldsymbol{r_y + r_a + r_m = 1}$
  - **Solution:** $\boldsymbol{r_y = \frac{2}{5},\ r_a = \frac{2}{5},\ r_m = \frac{1}{5}}$
- **Gaussian elimination method works for small examples, but we need a better method for large web-size graphs**
- **We need a new formulation!**

# Eigenvector Formulation

- **The web equations can be written**

$$\boldsymbol{r} = \boldsymbol{M} \cdot \boldsymbol{r}$$

- So the **rank vector** ***r*** is an **eigenvector** of the stochastic web matrix ***M***
  - In fact, its first or principal eigenvector, with corresponding eigenvalue ***1***
    - Largest eigenvalue of ***M*** is **1** since ***M*** is column stochastic (with non-negative entries)
      - *We know **r** is unit length and each column of **M** sums to one, so* $\boldsymbol{Mr} \leq \mathbf{1}$

**NOTE: *x*** is an eigenvector with the corresponding eigenvalue **λ** if:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{\lambda}\boldsymbol{x}$$

- **We can now efficiently solve for *r*!**
  **The method is called Power iteration**

# PageRank: Power Iteration Method

- **Given a web graph with *n* nodes, where the nodes are pages and edges are hyperlinks**
- **Power iteration:** a simple iterative scheme
  - Suppose there are *N* web pages
  - Initialize: $\mathbf{r}^{(0)} = [1/N,\ldots.,1/N]^T$
  - Iterate: $\mathbf{r}^{(t+1)} = \mathbf{M} \cdot \mathbf{r}^{(t)}$
  - Stop when $|\mathbf{r}^{(t+1)} - \mathbf{r}^{(t)}|_1 < \varepsilon$

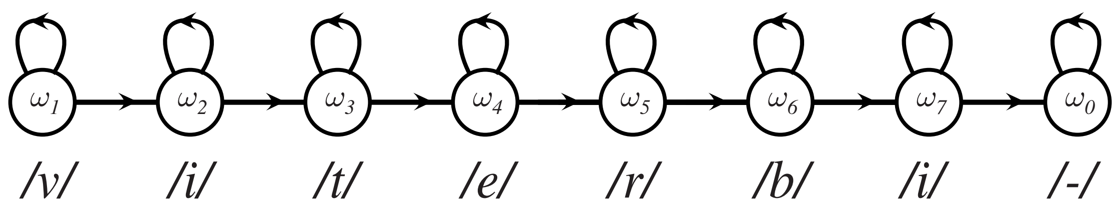
$$r_j^{(t+1)} = \sum_{i \to j} \frac{r_i^{(t)}}{\mathrm{d}_i}$$

$d_i$ …. out-degree of node i

$|\mathbf{x}|_1 = \sum_{1 \le i \le N} |x_i|$ is the **L1** norm
Can use any other vector norm, e.g., Euclidean

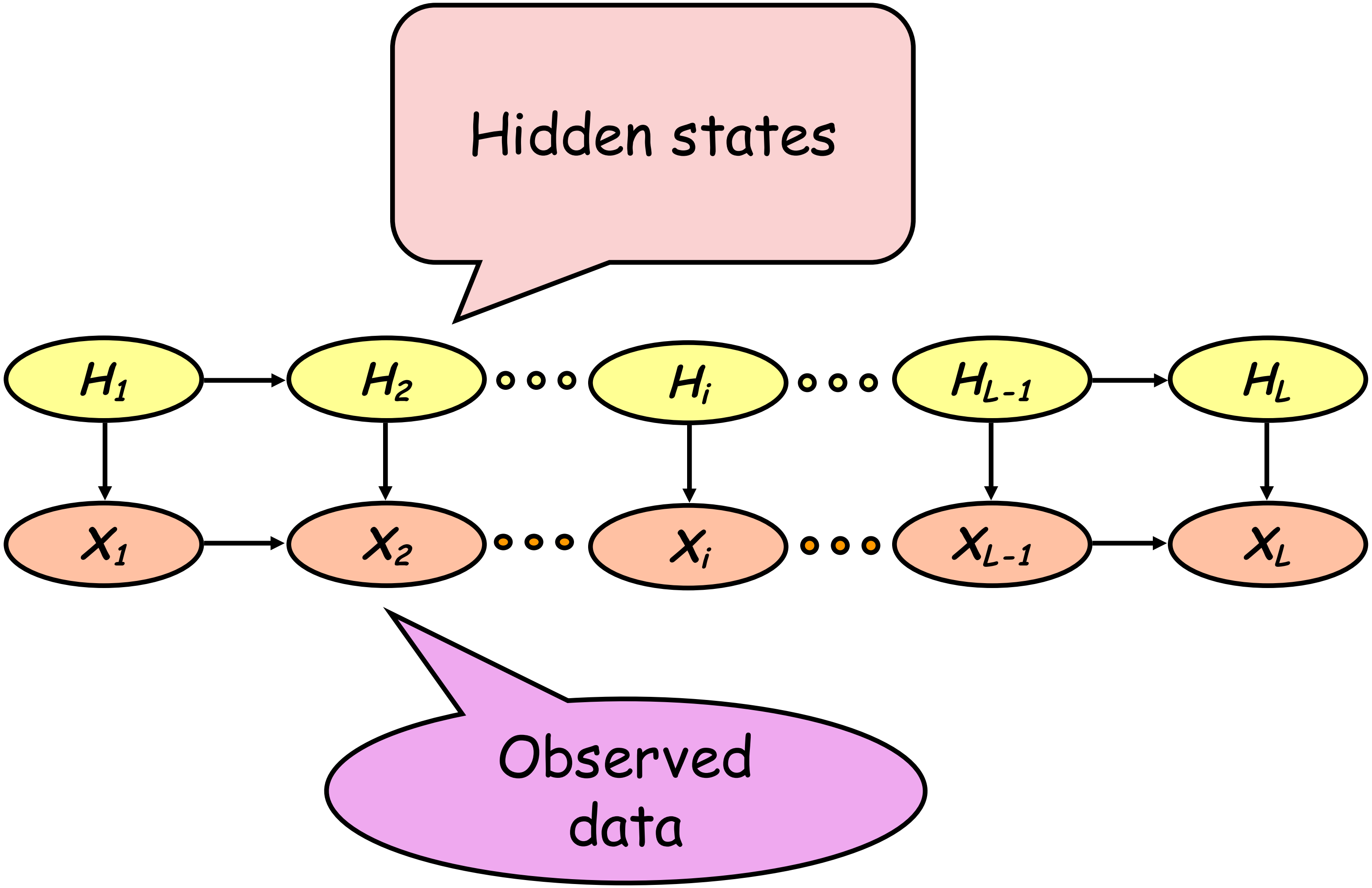
# Markov Chain Structure in Speech

- Left-right model



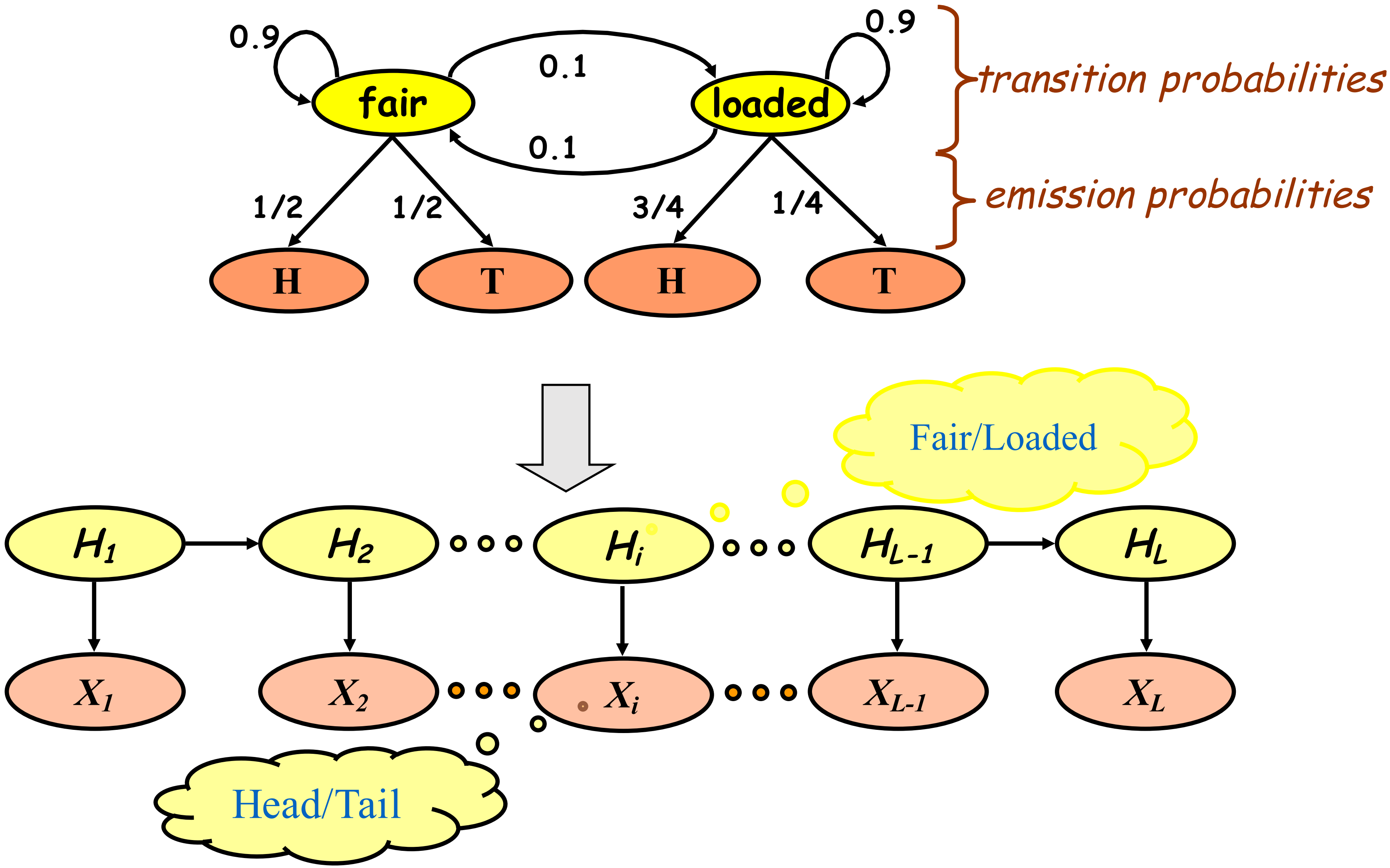

- Ideally each phoneme corresponds to a state but it may not be the case in practice!

# Hidden Markov Models - HMM



# Hidden Markov Models - HMM

## Coin-Tossing Example



# HMM

- Doubly embedded random process
- One of the process: Sequence of states is not observable (hidden)
- The state sequence may not be unique, even if we know that we begin in state one.
- However, some state sequences may be more likely than others.

- Learning: Given the HMM structure (number of visible and hidden states) and a training set of visible state sequences, determine the transition probabilities for hidden and visible states
- Evaluation: Computing the probability that a sequence of visible states was generated by a given HMM
- Decoding: Determine the most likely sequence of hidden states that produced a sequence of visible states

# References

- We will follow the following paper:

A tutorial on hidden Markov models and selected applications in speech recognition

LR **Rabiner** - Proceedings of the IEEE, 1989 - ieeexplore.ieee.org

A short version of the above paper

- An introduction to hidden Markov models

LR **Rabiner**, BH Juang - ASSP Magazine, IEEE, 1986 - ieeexplore.ieee.org

Longer version of the paper:

- **Fundamentals of Speech Recognition** 1st Edition

by Lawrence Rabiner (Author), Biing-Hwang Juang (Author)