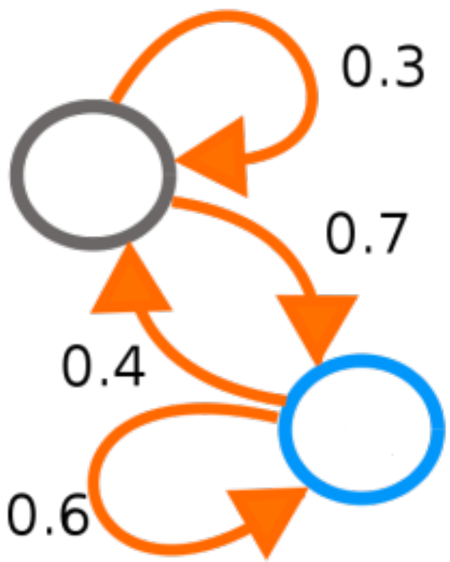
15-251
Great Theoretical Ideas in Computer Science

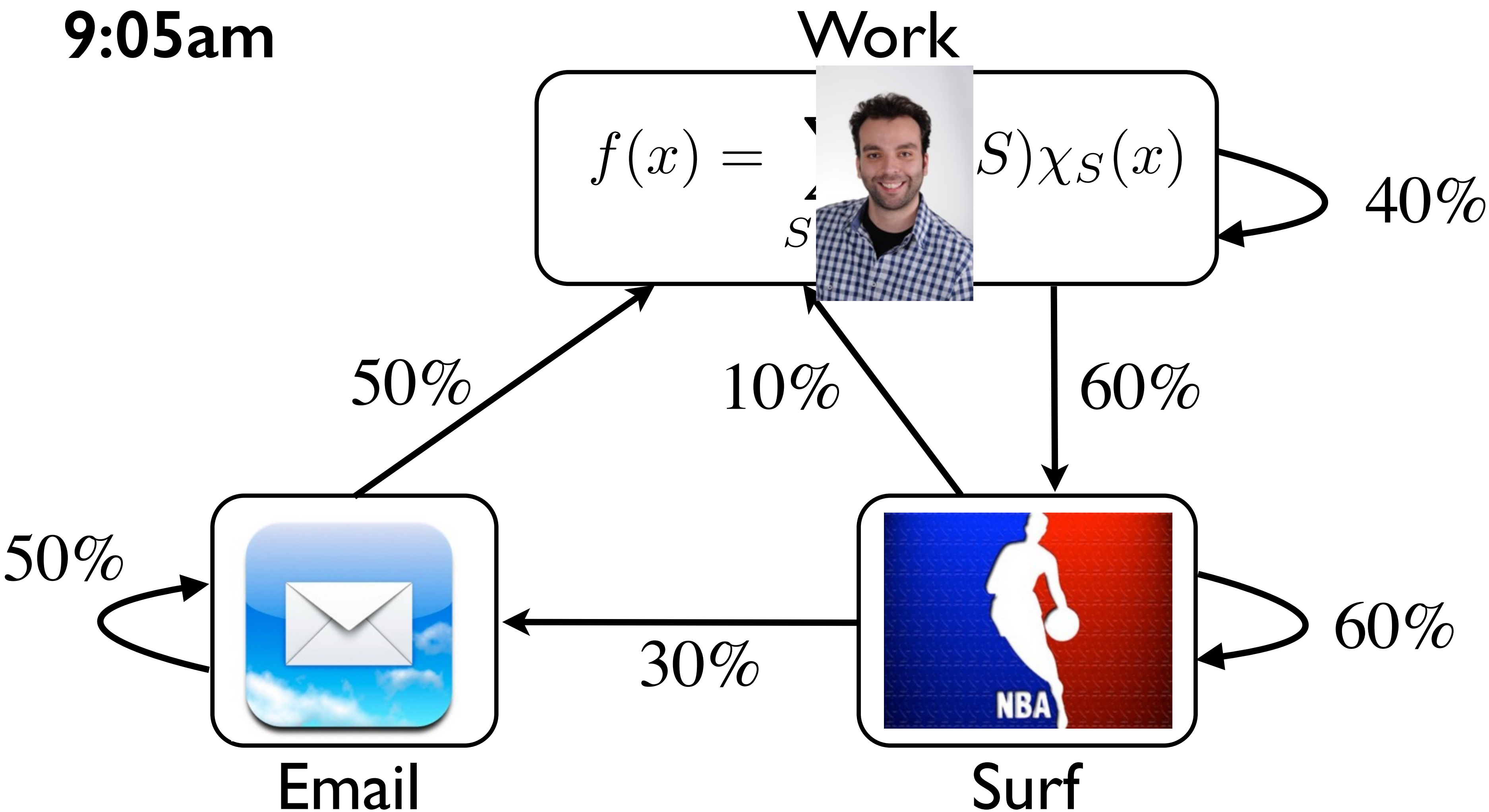
Lecture 20:
Markov Chains

November 3rd, 2016



My typical day (when I was a student)

9:05am



Markov **Model**

Andrey Markov (1856 - 1922)

Russian mathematician.

Famous for his work on <u>random processes</u>.

($\Pr[X \geq c \cdot \mathbf{E}[X]] \leq 1/c$ is Markov's Inequality.)

A model for the evolution of a random system.

The future is independent of the past, given the present.

Cool things about the Markov model

- It is a very general and natural model.

  Extraordinary number of applications in many different disciplines:

    computer science, mathematics, biology, physics, chemistry, economics, psychology, music, baseball,...

- The model is simple and neat.

- A beautiful mathematical theory behind it.

  Starts simple, goes deep.

The plan

Motivating examples and applications

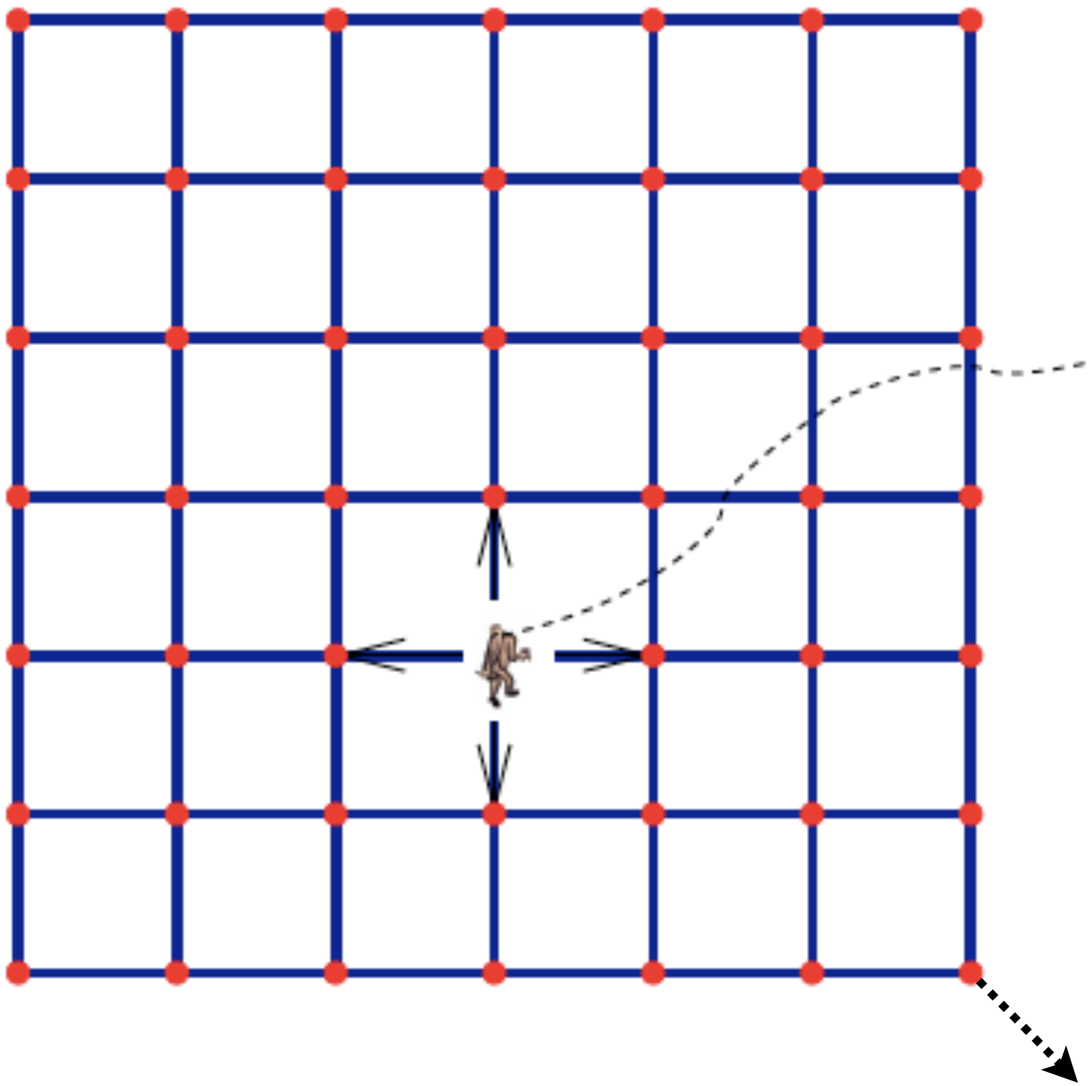
Basic mathematical representation and properties

A bit more on applications

The future is independent of the past, given the present.

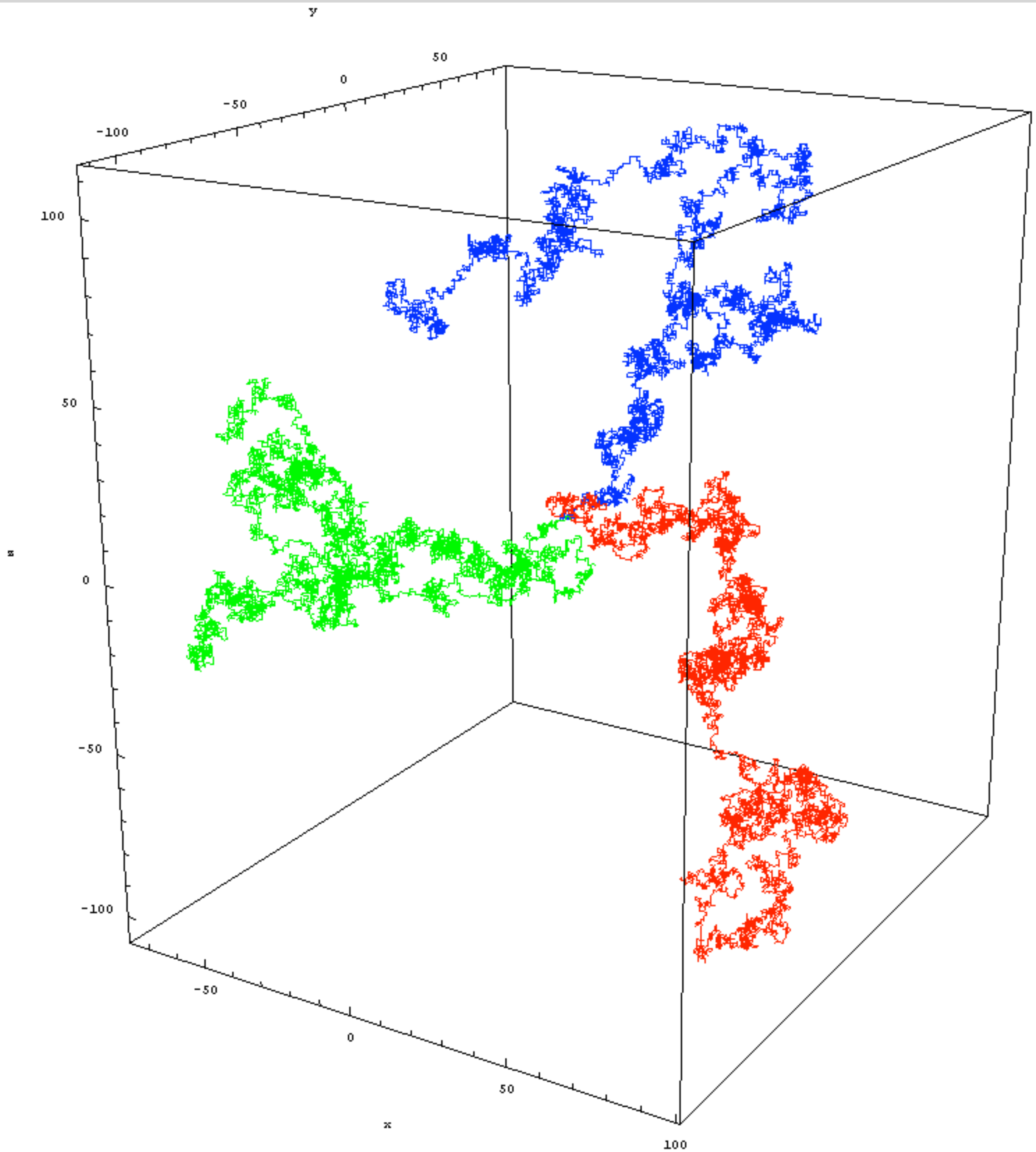
Some Examples of Markov Models

Example: Drunkard Walk



Salvador Dali (1922)
The Drunkard

Example: Diffusion Process



Example: Weather

A very(!!) simplified model for the weather.

Probabilities on a daily basis:

S = sunny
R = rainy

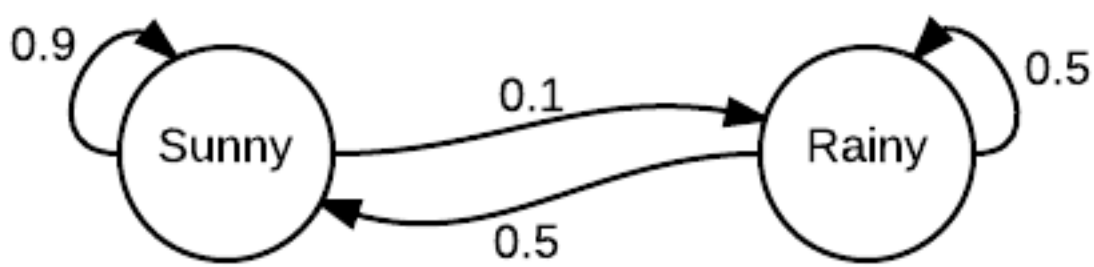
Pr[sunny to rainy] = 0.1

Pr[sunny to sunny] = 0.9

Pr[rainy to rainy] = 0.5

Pr[rainy to sunny] = 0.5

$$\begin{array}{cc} & \begin{array}{cc} S & R \end{array} \\ \begin{array}{c} S \\ R \end{array} & \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \end{array}$$



Encode more information about current state for a more accurate model.

Example: Life Insurance

Goal of life insurance company:

figure out how much to charge the clients.

Find a model for how long a client will live.

Probabilistic model of health on a monthly basis:

Pr[healthy to sick] = 0.3
Pr[sick to healthy] = 0.8
Pr[sick to death] = 0.1
Pr[healthy to death] = 0.01
Pr[healthy to healthy] = 0.69
Pr[sick to sick] = 0.1
Pr[death to death] = 1

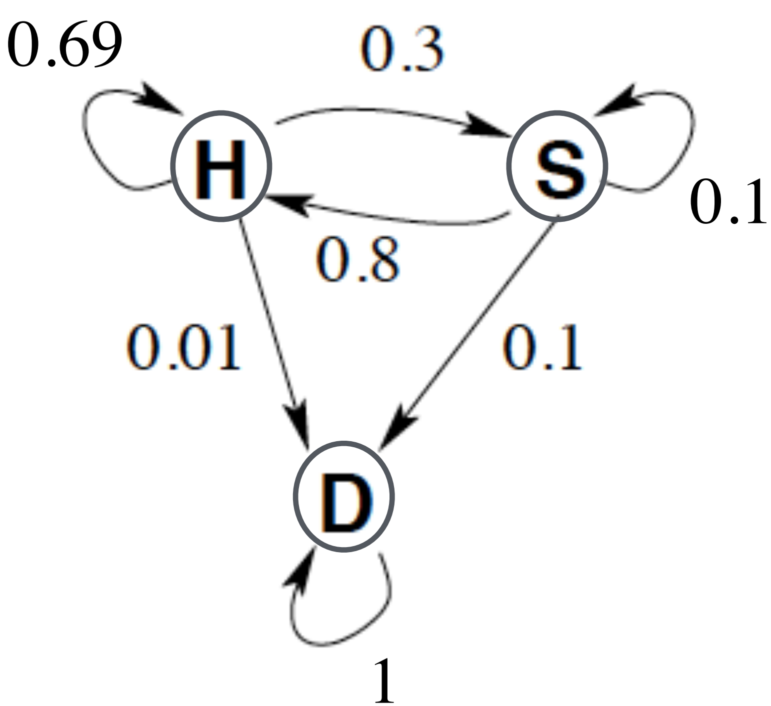
Example: Life Insurance

Goal of life insurance company:

figure out how much to charge the clients.

Find a model for how long a client will live.

Probabilistic model of health on a monthly basis:

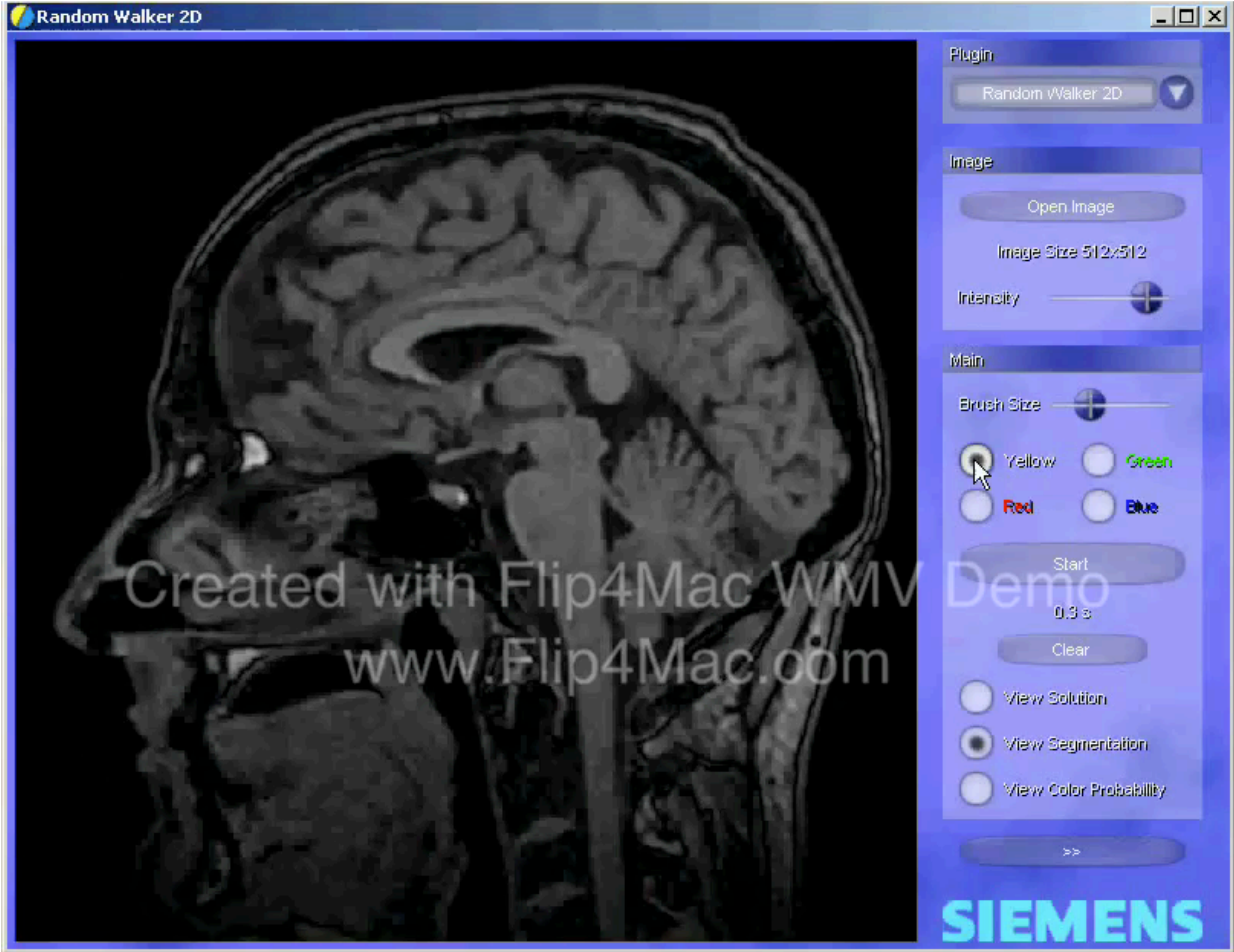


$$\begin{array}{c} \\ H \\ S \\ D \end{array} \begin{array}{c} \begin{array}{ccc} H & S & D \end{array} \\ \begin{bmatrix} 0.69 & 0.3 & 0.01 \\ 0.8 & 0.1 & 0.1 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Some Applications of Markov Models

Application: Algorithmic Music Composition

Application: Image Segmentation



Application: Automatic Text Generation

Random text generated by a computer
(putting random words together):

"While at a conference a few weeks back, I spent an interesting evening with a grain of salt."

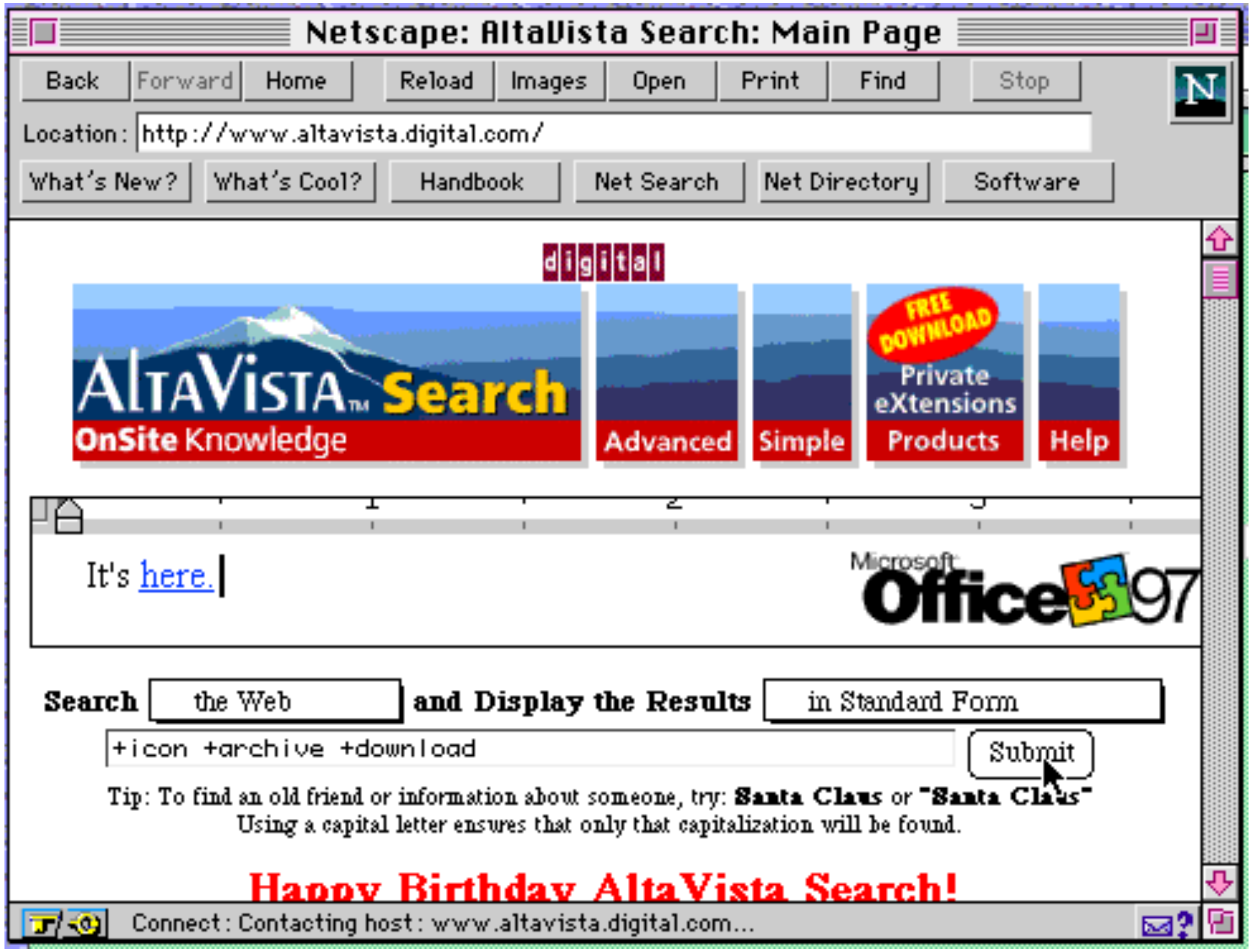
Google: Mark V Shaney

Application: Speech Recognition

Speech recognition software programs use Markov models to listen to the sound of your voice and convert it into text.

Application: Google PageRank

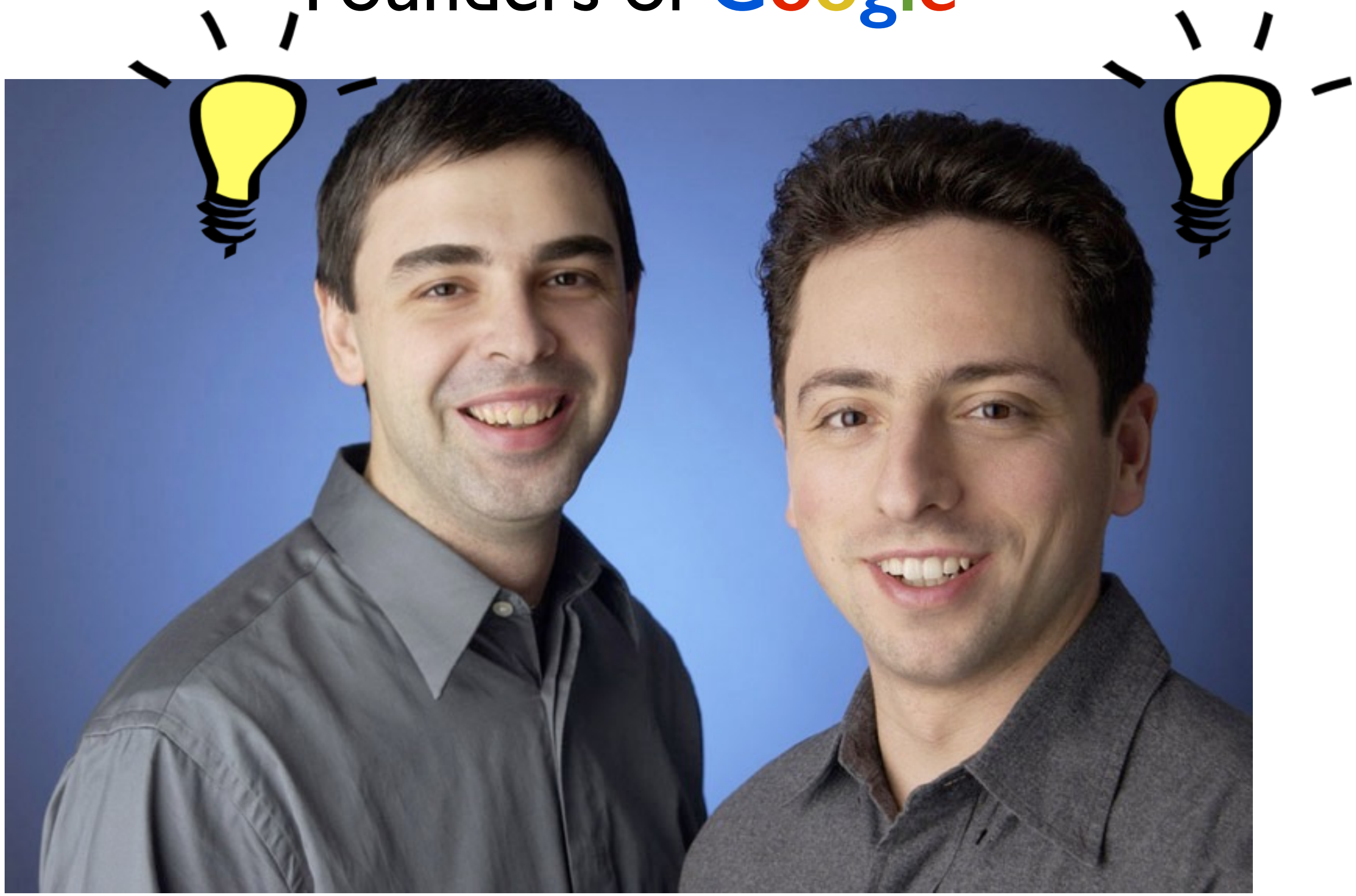
1997: Web search was horrible



Sorts webpages by number of occurrences of keyword(s).

Application: Google PageRank

Founders of Google



Larry Page

Sergey Brin

$20Billionaires

Application: Google PageRank

Jon Kleinberg

Nevanlinna Prize

Application: Google PageRank

How does Google order the webpages displayed after a search?

<u>**2 important factors:**</u>

- Relevance of the page.

- <u>Reputation</u> of the page.

  ↓

  The number and reputation of links pointing to that page.

  Reputation is measured using PageRank.

  PageRank is calculated using a Markov Chain.

The plan

Motivating examples and applications

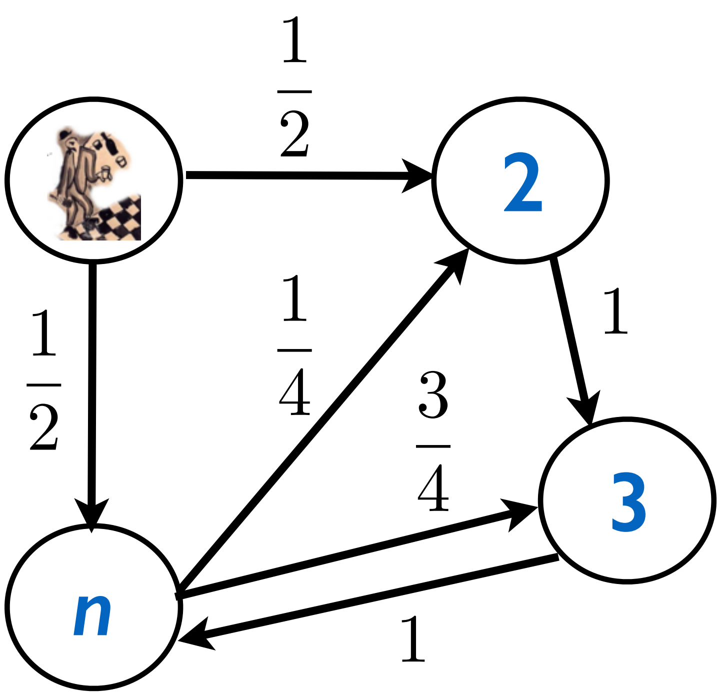
Basic mathematical representation and properties

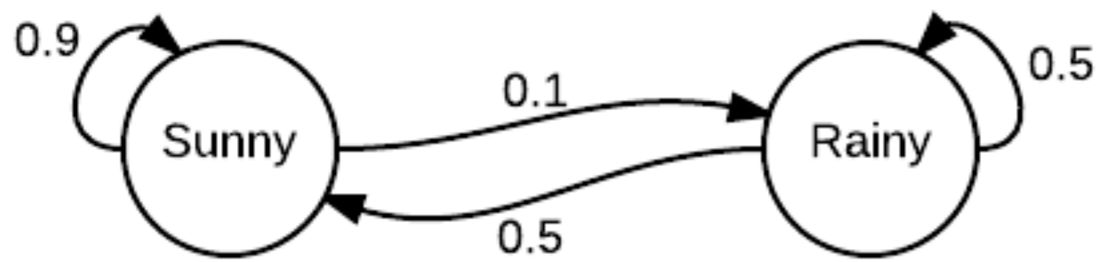
A bit more on applications

The Setting

There is a system with *n* possible states/values {1, 2, …, n}.

At each time step, the state changes probabilistically.





Memoryless

The next state only depends on the current state.

Evolution of the system: random walk on the graph.

The Definition

A **Markov Chain** is a directed graph with $V = \{1, 2, \ldots, n\}$ such that:

- Each edge is labeled with a value in $(0, 1]$ (a positive probability).
  self-loops allowed
- At each vertex, the probabilities on outgoing edges sum to 1.

(We usually assume the graph is strongly connected. i.e. there is a directed path from *i* to *j* for any *i* and *j*.)

The vertices of the graph are called states.

The edges are called transitions.

The label of an edge is a transition probability.

Notation

Given some Markov Chain with n states:

For each $t = 0, 1, 2, 3, \ldots$ we have a random variable:

$X_t =$ the state we are in after t steps.

Define $\pi_t[i] = \Pr[X_t = i]$.

$$\pi_t = \begin{array}{cccc} 1 & 2 & & n \\ \left[p_1\right. & p_2 & \cdots & \left.p_n\right] \end{array}$$

$$\sum_i p_i = 1$$

$\pi_t[i] =$ probability of being in state *i* after exactly **t** steps.

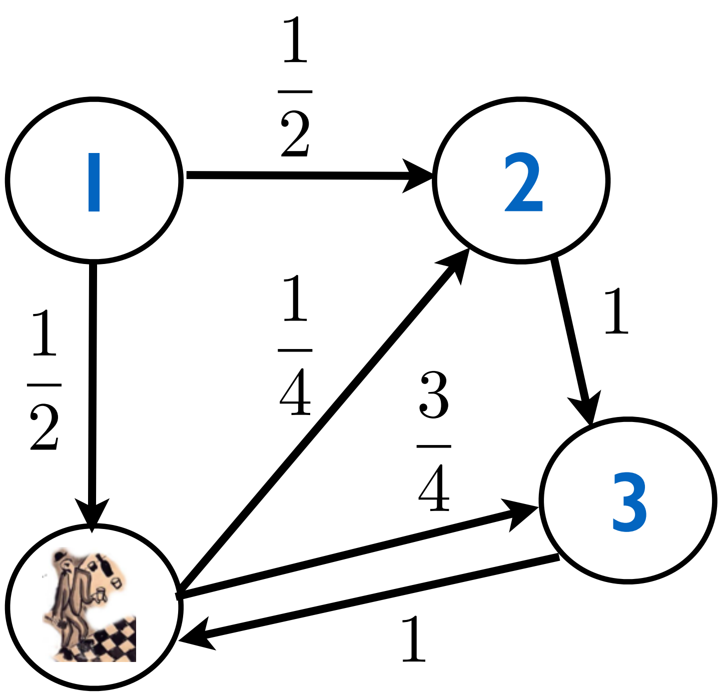
We write $X_t \sim \pi_t$. (X_t has distribution π_t)

Note that someone has to provide π_0.

Once this is known, we get the distributions $\pi_1, \pi_2, \ldots$

Notation

Let's say we start at state 1, i.e., $X_0 \sim \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \end{array}$ $[1 \quad 0 \quad 0 \quad 0] = \pi_0$



$$X_0 = 1 \qquad X_0 \sim \pi_0$$

$$X_1 = 4 \qquad X_1 \sim \pi_1$$

$$X_2 = 3 \qquad X_2 \sim \pi_2$$

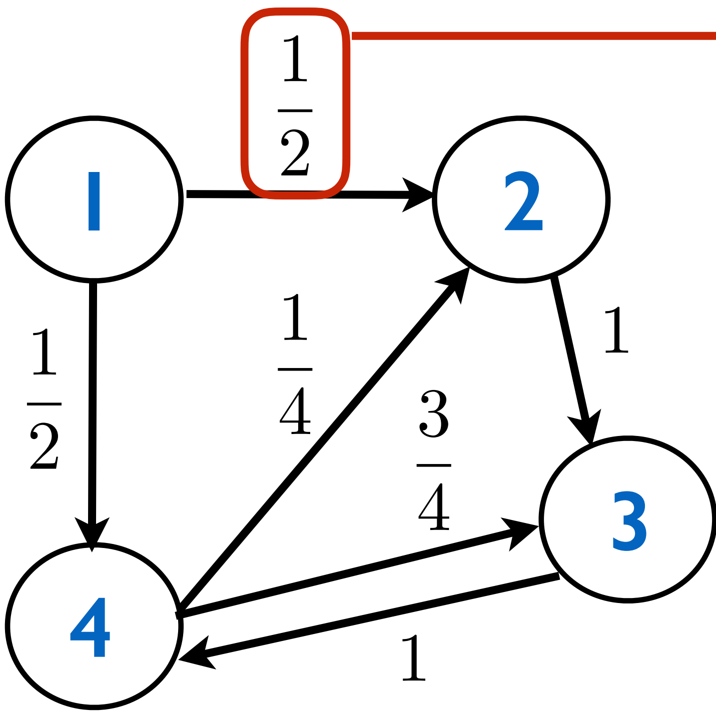
$$X_3 = 4 \qquad X_3 \sim \pi_3$$

$$X_4 = 2 \qquad X_4 \sim \pi_4$$

$$\vdots$$

Notation

Let's say we start at state 1, i.e., $X_0 \sim \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \end{array}$ $[1 \quad 0 \quad 0 \quad 0] = \pi_0$



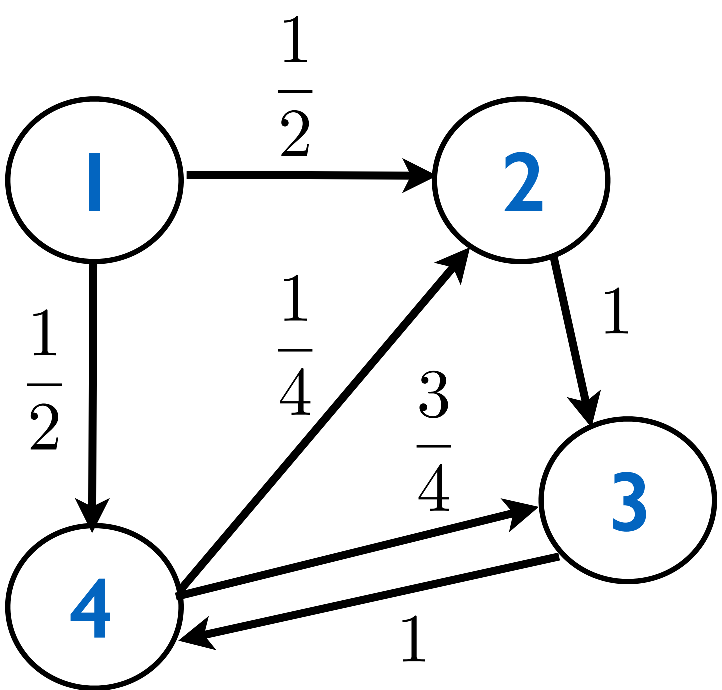
$$\Pr[1 \to 2 \text{ in one step}]$$

$$= \Pr[X_1 = 2 \mid X_0 = 1]$$

$$= \Pr[X_t = 2 \mid X_{t-1} = 1]$$

Notation

Let's say we start at state 1, i.e., $X_0 \sim \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \end{array} = \pi_0$



$$\Pr[X_1 = 2 | X_0 = 1] = \frac{1}{2}$$

$$\Pr[X_1 = 3 | X_0 = 1] = 0$$

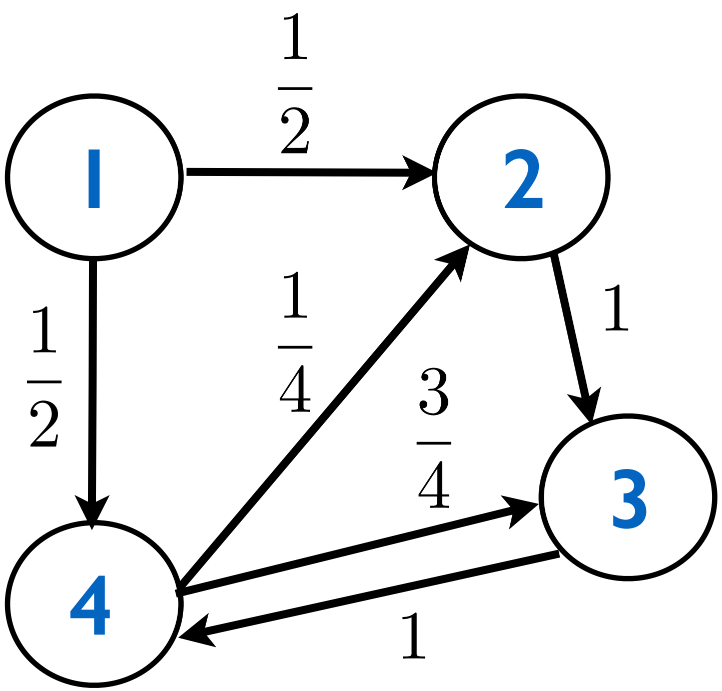
$$\Pr[X_1 = 4 | X_0 = 1] = \frac{1}{2}$$

$$\Pr[X_1 = 1 | X_0 = 1] = 0$$

$$\forall t \quad \Pr[X_t = 2 | X_{t-1} = 4] = \frac{1}{4}$$

$$\forall t \quad \Pr[X_t = 3 | X_{t-1} = 2] = 1$$

Notation



$$\begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix} \end{array}$$

Transition Matrix

A Markov Chain with **n** states
can be characterized by the **n x n** **transition matrix** K :

$$\forall i, j \in \{1, 2, \ldots, n\} \quad K[i,j] = \Pr[X_t = j \mid X_{t-1} = i]$$

$$= \Pr[i \to j \text{ in one step}]$$

Note: rows of K sum to 1.

Some Fundamental and Natural Questions

What is the probability of being in state ***i*** after ***t*** steps (given some initial state)?

$$\pi_t[i] = ?$$

What is the expected time of reaching state ***i*** when starting at state ***j*** ?

What is the expected time of having visited every state (given some initial state)?

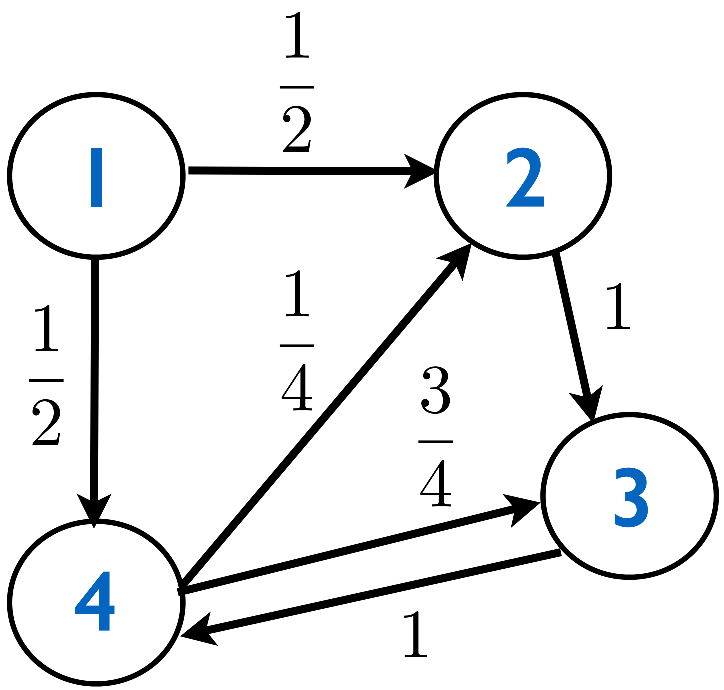
$\vdots$

How do you answer such questions?

Mathematical representation of the evolution

Suppose we start at state 1 and let the system evolve.

How can we mathematically represent the evolution?

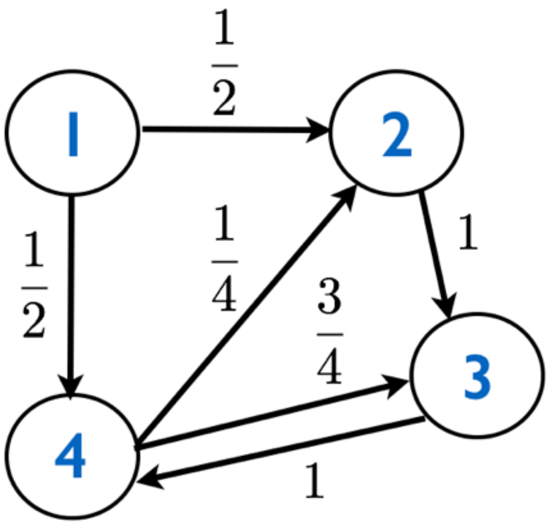


$$\begin{array}{c c} & \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix} \end{array}$$

$$\begin{array}{cc} & \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \pi_0 = & \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \end{array}$$

What is π_1? By inspection, $\pi_1 = \begin{array}{c} \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array}$.

Poll



$$\begin{array}{c c} & \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix} \end{array}$$

Given $\pi_1 = \begin{array}{c} \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array}$, what is π_2 ?

$\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$

$\begin{bmatrix} 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix}$

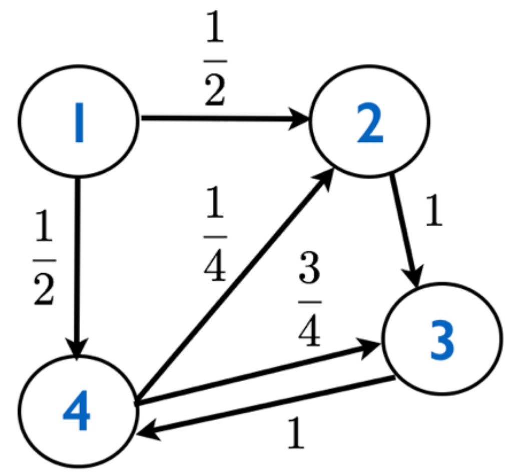
$\begin{bmatrix} 0 & \frac{5}{8} & \frac{3}{8} & 0 \end{bmatrix}$

$\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$

$\begin{bmatrix} 0 & \frac{1}{8} & \frac{7}{8} & 0 \end{bmatrix}$

$\begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}$

Mathematical representation of the evolution



$$\begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 2 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 1 \\ 4 & 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{array}$$

Given π_0 .

What is π_1? $\quad \pi_1[j] = \Pr[X_1 = j]$

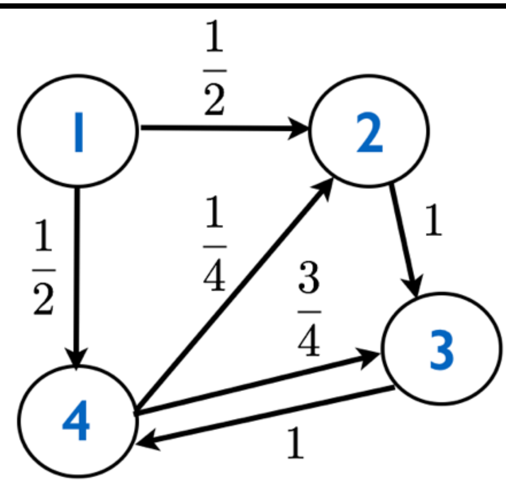
(law of total probability) $\quad = \displaystyle\sum_{i=1}^{4} \Pr[X_1 = j \mid X_0 = i] \Pr[X_0 = i]$

matrix mult.

$$= \sum_{i=1}^{4} K[i,j] \cdot \pi_0[i] = (\pi_0 \cdot K)[j]$$

This is true for any j .

Mathematical representation of the evolution



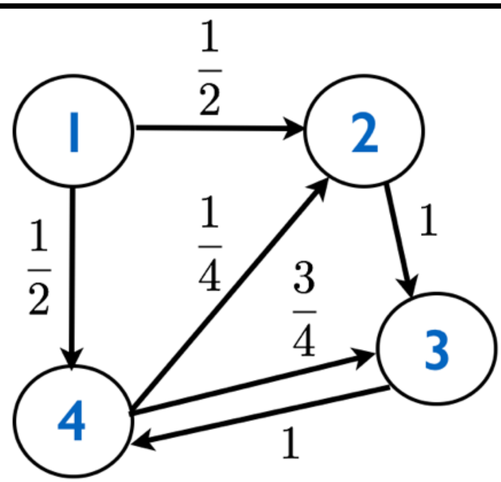
$$
\begin{array}{c|cccc}
 & 1 & 2 & 3 & 4 \\
\hline
1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
2 & 0 & 0 & 1 & 0 \\
3 & 0 & 0 & 0 & 1 \\
4 & 0 & \frac{1}{4} & \frac{3}{4} & 0
\end{array}
$$

The probability of states after 1 step:

$$
\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}}_{\pi_0}
\underbrace{\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix}}_{K}
= \underbrace{\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}}_{\pi_1}
$$

the new state (probabilistic)

Mathematical representation of the evolution



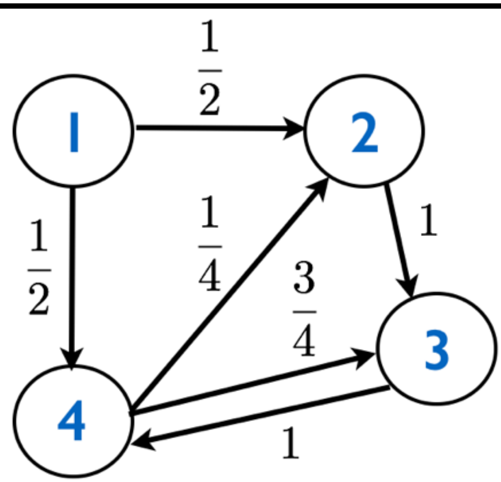
$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix} \end{array}$$

The probability of states after 2 steps:

$$\underbrace{\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}}_{\pi_1} \underbrace{\begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix}}_{K} = \underbrace{\begin{bmatrix} 0 & \frac{1}{8} & \frac{7}{8} & 0 \end{bmatrix}}_{\pi_2}$$

the new state (probabilistic)

Mathematical representation of the evolution



$$\begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix} \end{array}$$

$$\pi_1 = \pi_0 \cdot K$$

$$\pi_2 = \pi_1 \cdot K$$

So $$\pi_2 = (\pi_0 \cdot K) \cdot K$$

$$= \pi_0 \cdot K^2$$

Mathematical representation of the evolution

In general:

If the initial probabilistic state is $\begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} = \pi_0$

$p_i =$ probability of being in state *i*,

$p_1 + p_2 + \cdots + p_n = 1$,

after ***t*** steps, the probabilistic state is:

$$\begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} \text{Transition} \\ \text{Matrix} \end{bmatrix}^{t} = \pi_t$$

Remarkable Property of Markov Chains

What happens in the long run?

i.e., can we say anything about π_t for large t ?

Suppose the Markov chain is "aperiodic".

Then, as the system evolves, the probabilistic state ***converges*** to a limiting probabilistic state.

As $t \to \infty$, for any $\pi_0 = [p_1 \quad p_2 \quad \cdots \quad p_n]$:

$$[p_1 \quad p_2 \quad \cdots \quad p_n] \begin{bmatrix} \text{Transition} \\ \text{Matrix} \end{bmatrix}^t \longrightarrow \pi$$

Remarkable Property of Markov Chains

In other words:

$$\pi_t \to \pi \quad \text{as} \quad t \to \infty\,.$$

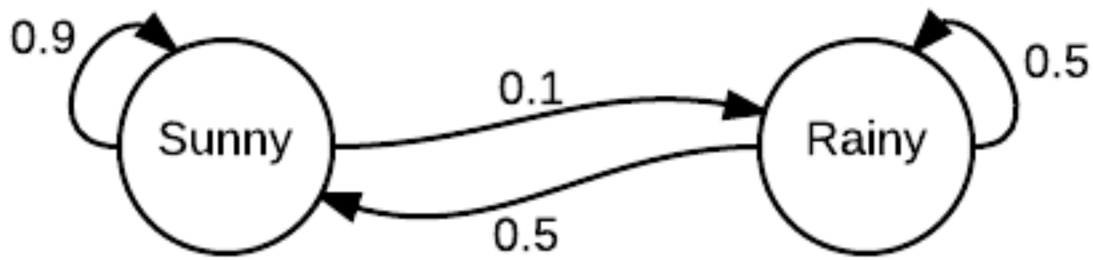
Note:

$$\pi \begin{bmatrix} \text{Transition} \\ \text{Matrix} \\ \\ \end{bmatrix} = \pi$$

stationary/invariant distribution

This π is unique.

Remarkable Property of Markov Chains



Stationary distribution is $\begin{bmatrix} \frac{5}{6} & \frac{1}{6} \end{bmatrix}$.

$$\begin{bmatrix} \frac{5}{6} & \frac{1}{6} \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \end{bmatrix}$$

*In the long run, it is Sunny **5/6** of the time,*
*it is Rainy **1/6** of the time.*

Remarkable Property of Markov Chains

How did I find the stationary distribution?

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}^2 = \begin{bmatrix} 0.86 & 0.14 \\ 0.7 & 0.3 \end{bmatrix}$$

$$\begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}^4 = \begin{bmatrix} 0.8376 & 0.1624 \\ 0.812 & 0.188 \end{bmatrix}$$

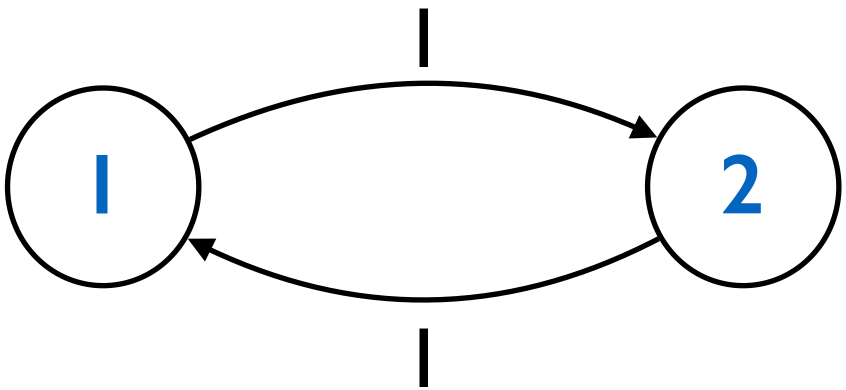
$$\begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}^8 = \begin{bmatrix} 0.833443 & 0.166557 \\ 0.832787 & 0.167213 \end{bmatrix}$$

Exercise: Why do the rows converge to π ?

Remarkable Property of Markov Chains

We needed the Markov chain to be "aperiodic".

What is a "periodic" Markov chain?



$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\pi_0 = [1 \quad 0]$$
$$\pi_1 = [0 \quad 1]$$
$$\pi_2 = [1 \quad 0]$$
$$\pi_3 = [0 \quad 1]$$
$$\vdots$$

There is still a *stationary* distribution.

$$\pi = [1/2 \quad 1/2]$$

$$[1/2 \quad 1/2] \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = [1/2 \quad 1/2]$$

But it is not a *limiting* distribution.

Things to remember

Markov Chains can be characterized by the transition matrix K.

$$K[i,j] = \Pr[X_t = j \mid X_{t-1} = i]$$
$$= \Pr[i \to j \text{ in one step}]$$

What is the probability of being in state *i* after *t* steps?

$$\pi_t = \pi_0 \cdot K^t \qquad\qquad \pi_t[i] = (\pi_0 \cdot K^t)[i]$$

Things to remember

Theorem (Fundamental Theorem of Markov Chains):

Consider a Markov chain that is strongly connected and aperiodic.

- There is a unique invariant/stationary distriution π such that

$$\pi = \pi K.$$

- For any initial distribution π_0 ,

$$\lim_{t \to \infty} \pi_0 K^t = \pi$$

- Let T_{ij} be the number of steps it takes to reach state j provided we start at state i. Then,

$$\mathbf{E}[T_{ii}] = \frac{1}{\pi[i]}.$$

The plan

Motivating examples and applications

Basic mathematical representation and properties

A bit more on applications

How are Markov Chains applied ?

2 common types of applications:

1. Build a Markov chain as a statistical model of a real-world process.

   Use the Markov chain to simulate the process.

   e.g. text generation, music composition.

2. Use a measure associated with a Markov chain to approximate a quantity of interest.

   e.g. Google PageRank, image segmentation

Automatic Text Generation

Generate a superficially real-looking text given a sample document.

Idea:

From the sample document, create a Markov chain.

Use a random walk on the Markov chain to generate text.

Example:

Collect speeches of Obama, create a Markov chain.

Use a random walk to generate new speeches.

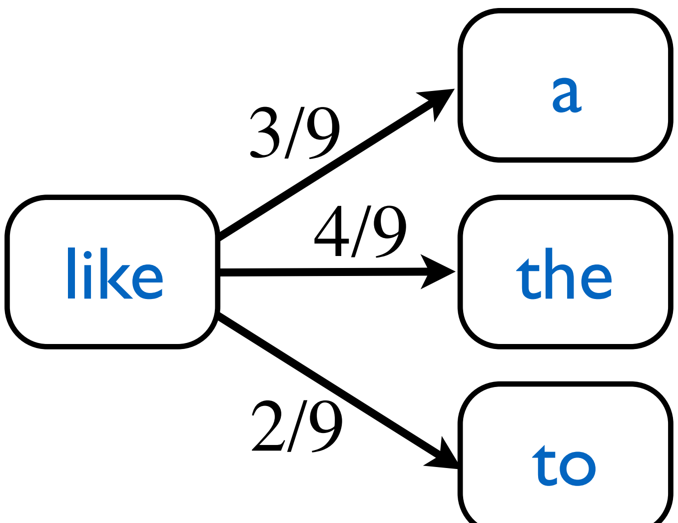
Automatic Text Generation

The Markov Chain:

1. For each word in the document, create a node/state.

2. Put an edge word1 ---> word2
if there is a sentence in which word2 comes after word1.

3. Edge probabilities reflect frequency of the pair of words.



like a 3 times

like the 4 times

like to 2 times

Automatic Text Generation

"I jumped up. I don't know what's going on so I am coming down with a road to opportunity. I believe we can agree on or do about the major challenges facing our country."

Automatic Text Generation

Another use:

Build a Markov chain based on speeches of Obama.

Build a Markov chain based on speeches of Bush.

Given a new quote, can predict if it is by Obama or Bush.

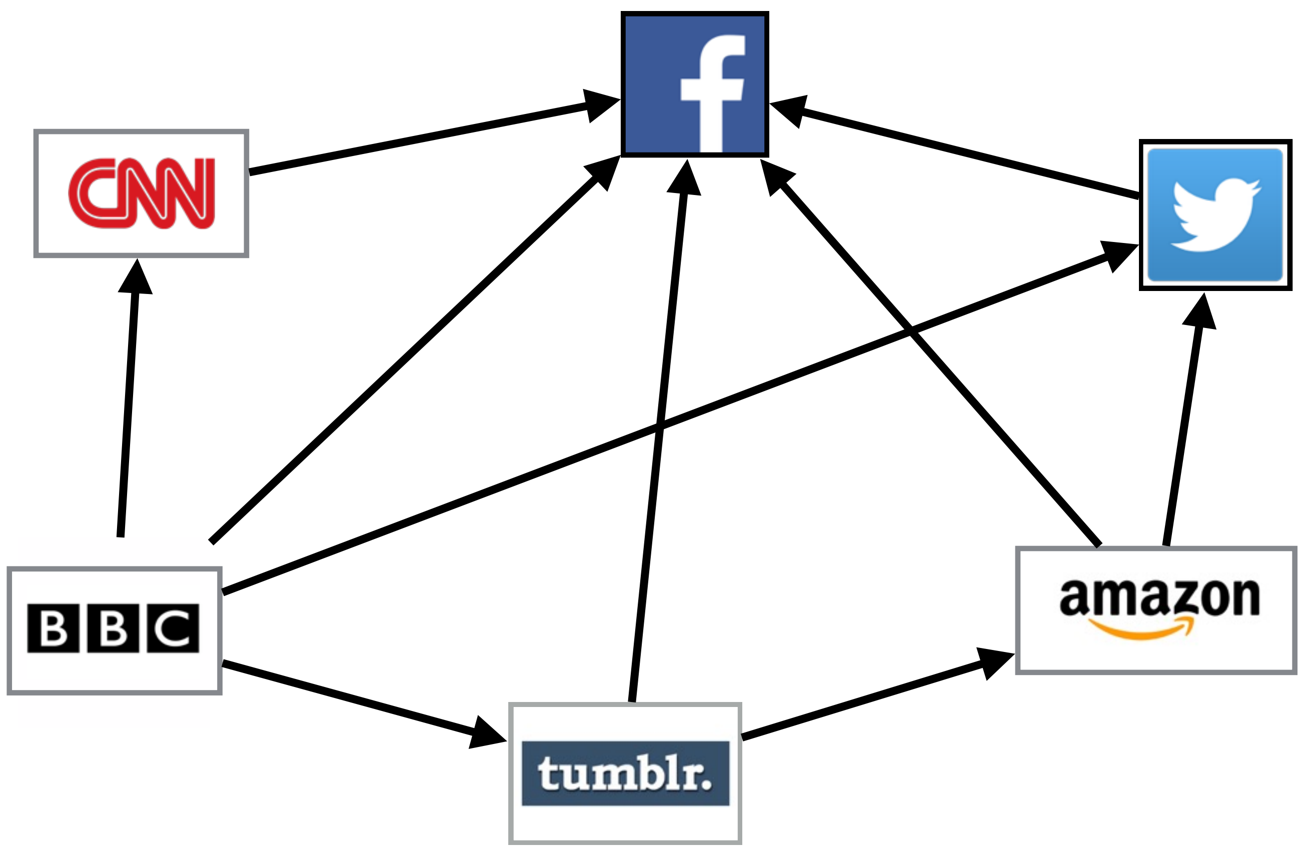
(by testing which Markov model the quote fits best)

Google PageRank

PageRank is a measure of reputation:

The number and reputation of links pointing to you.

The Markov Chain:



Google PageRank

PageRank is a measure of reputation:

The number and reputation of links pointing to you.

The Markov Chain:

1. Every webpage is a node/state.

2. Each hyperlink is an edge:

   if webpage **A** has a link to webpage **B**, **A** ---> **B**

3a. If **A** has *m* outgoing edges, each gets label $1/m$.

3b. If **A** has no outgoing edges, put edge **A** ---> **B** $\forall$ **B**
(jump to a random page)

Google PageRank

A little tweak:

Random surfer jumps to a random page with 15% prob.

Stationary distribution:

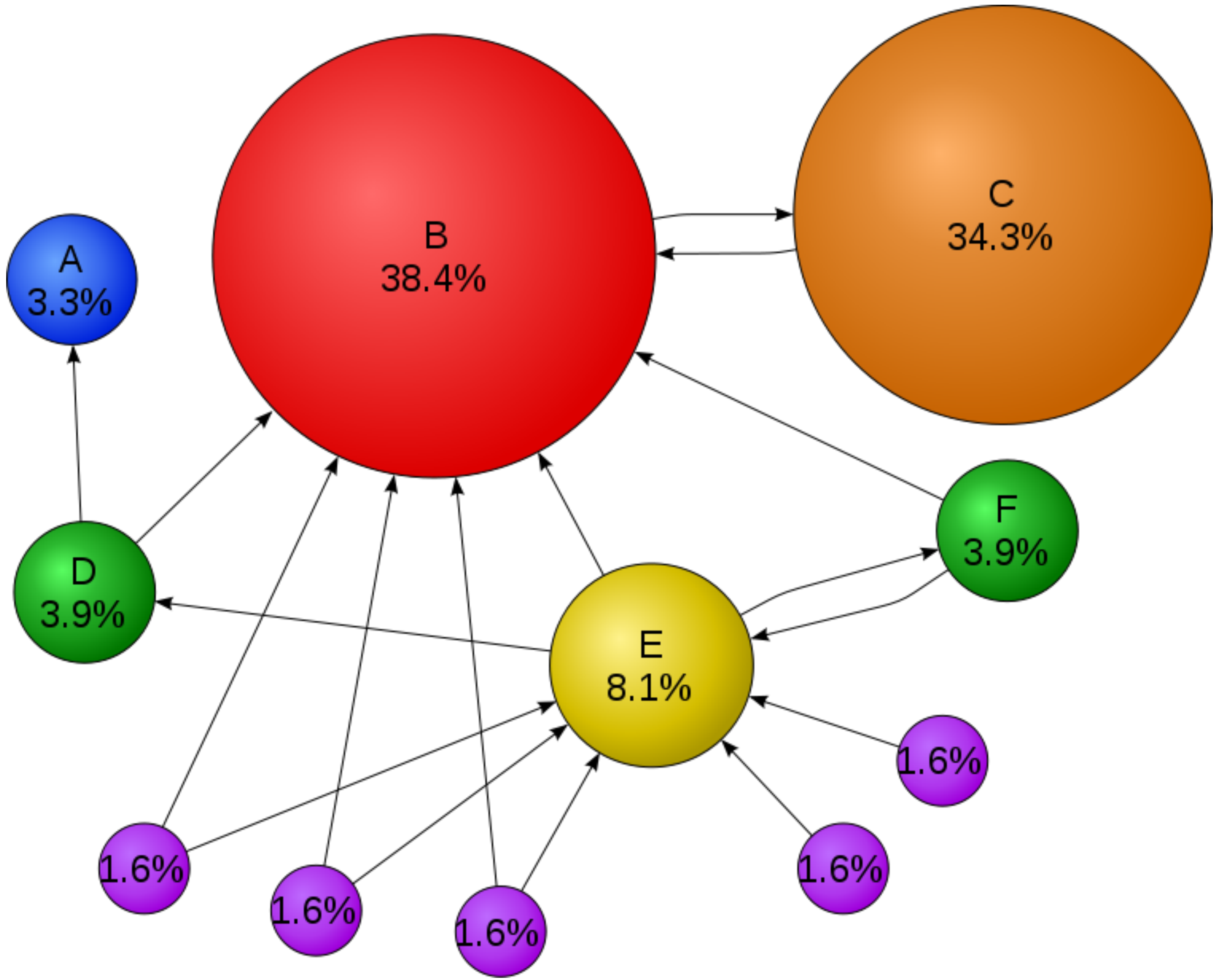
probability of being at webpage **A** in the long run

PageRank of webpage **A**

=

The stationary probability of **A**

Google PageRank



Google PageRank

Google:

“PageRank continues to be the heart of our software.”

How are Markov Chains applied ?

2 common types of applications:

1. Build a Markov chain as a statistical model of a real-world process.

   Use the Markov chain to simulate the process.

   e.g. text generation, music composition.

2. Use a measure associated with a Markov chain to approximate a quantity of interest.

   e.g. Google PageRank, image segmentation

The plan

Motivating examples and applications

Basic mathematical representation and properties

A bit more on applications