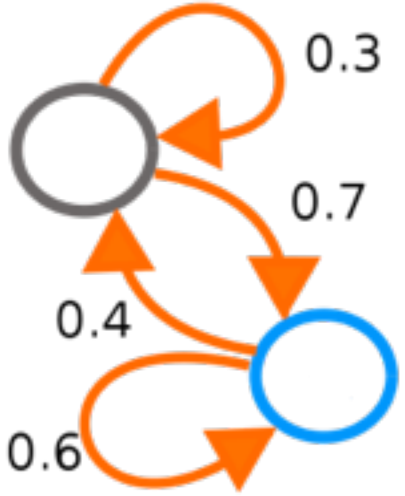
15-251

Great Theoretical Ideas in Computer Science

Markov Chains



April 9, 2015

A day in the life of me

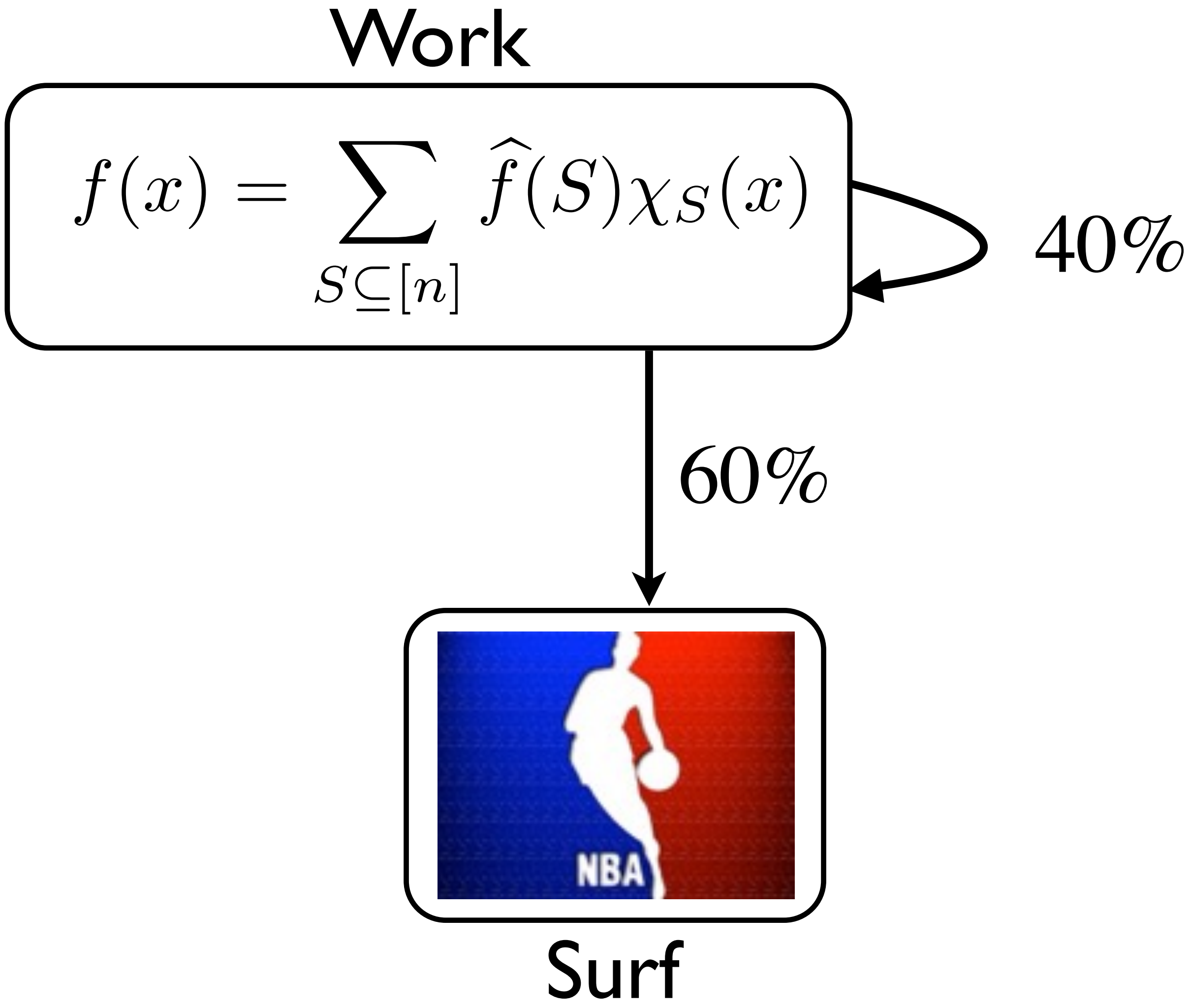
9:00am

Work

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$$

A day in the life of me

9:01am



Work

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$$

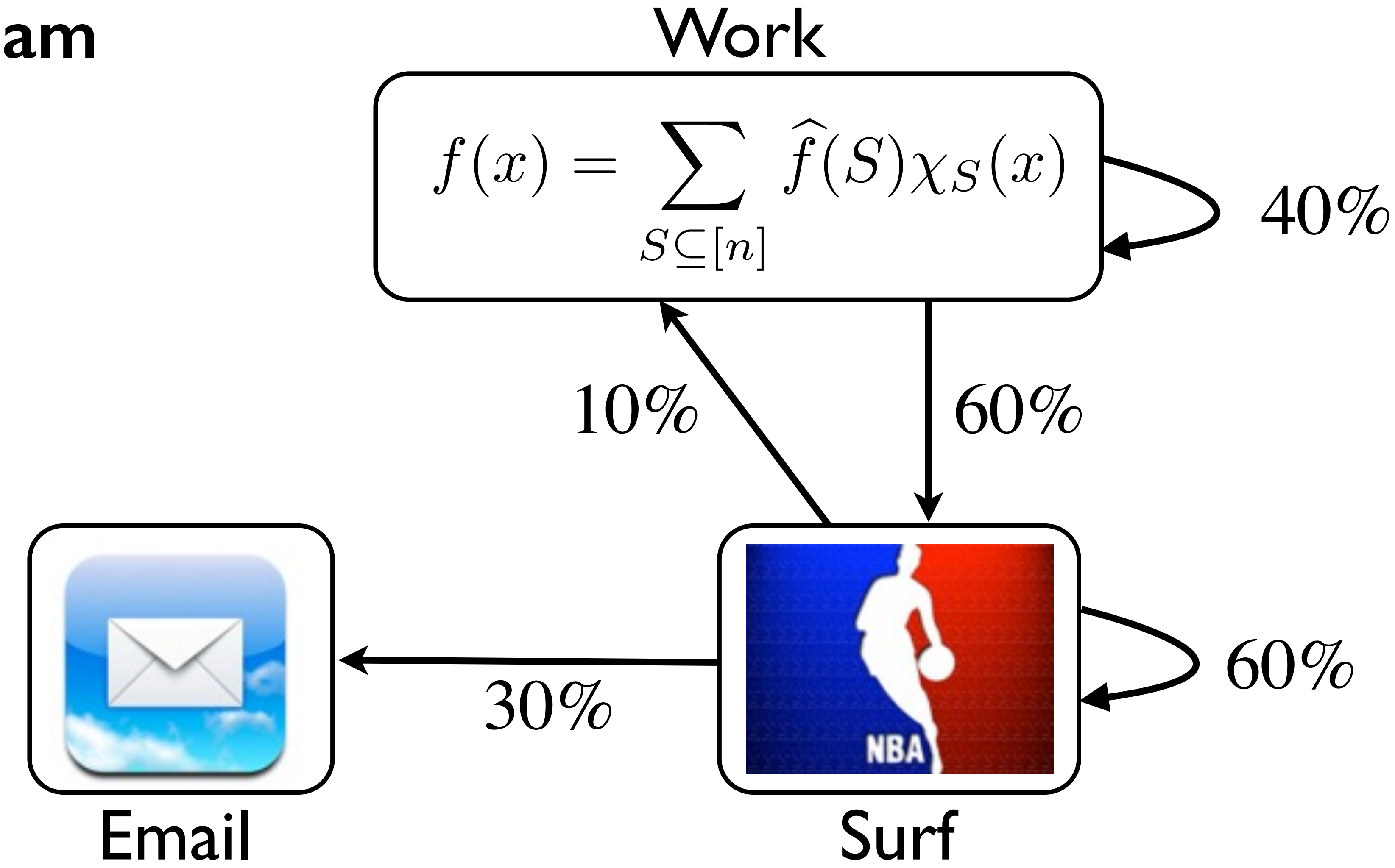
40%

60%

Surf

A day in the life of me

9:02am



Work

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x)$$

40%

10%

60%

30%

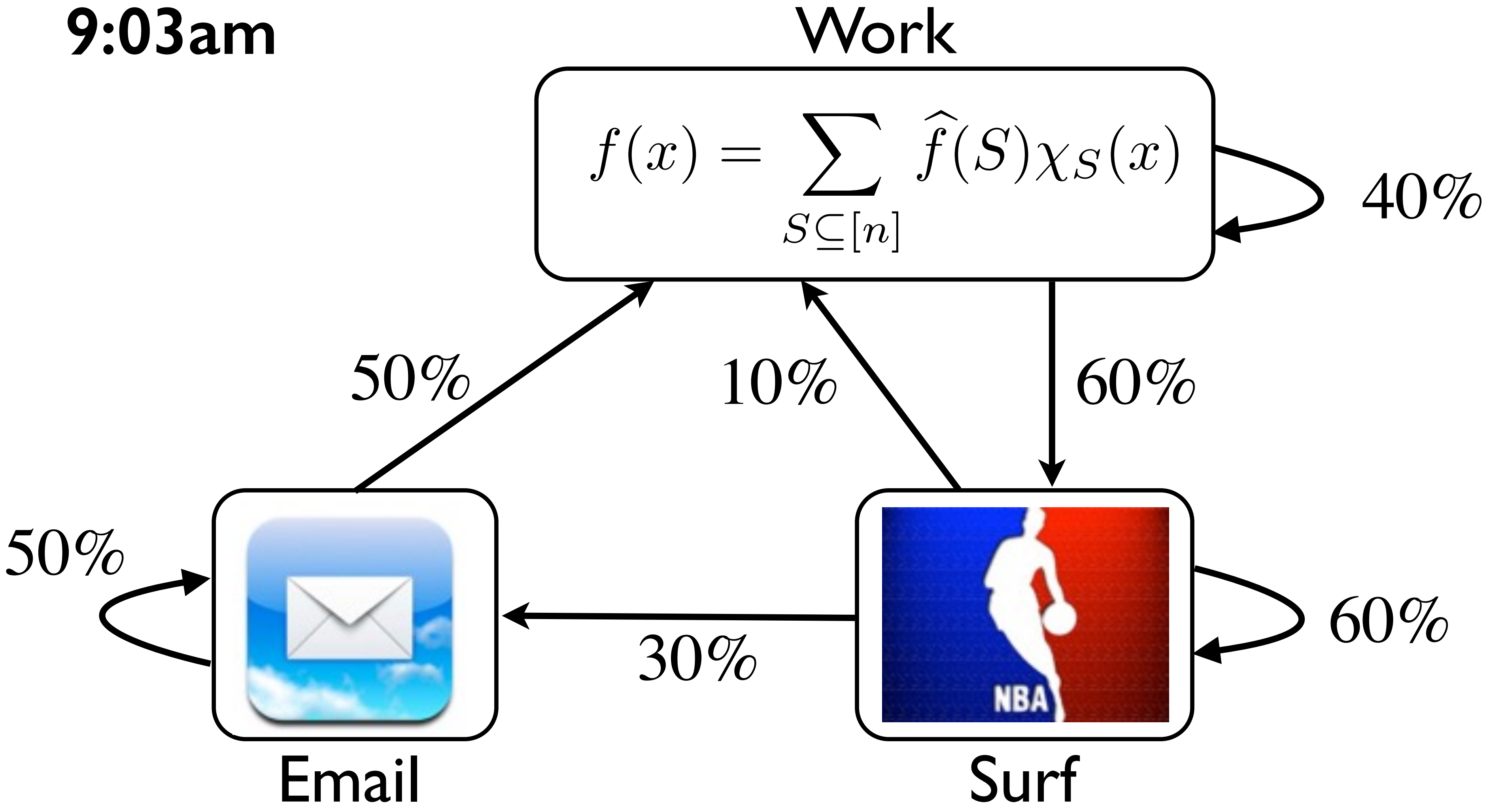
60%

Email

Surf

A day in the life of me

9:03am



Work

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$$

40%

50%

10%

60%

50%

30%

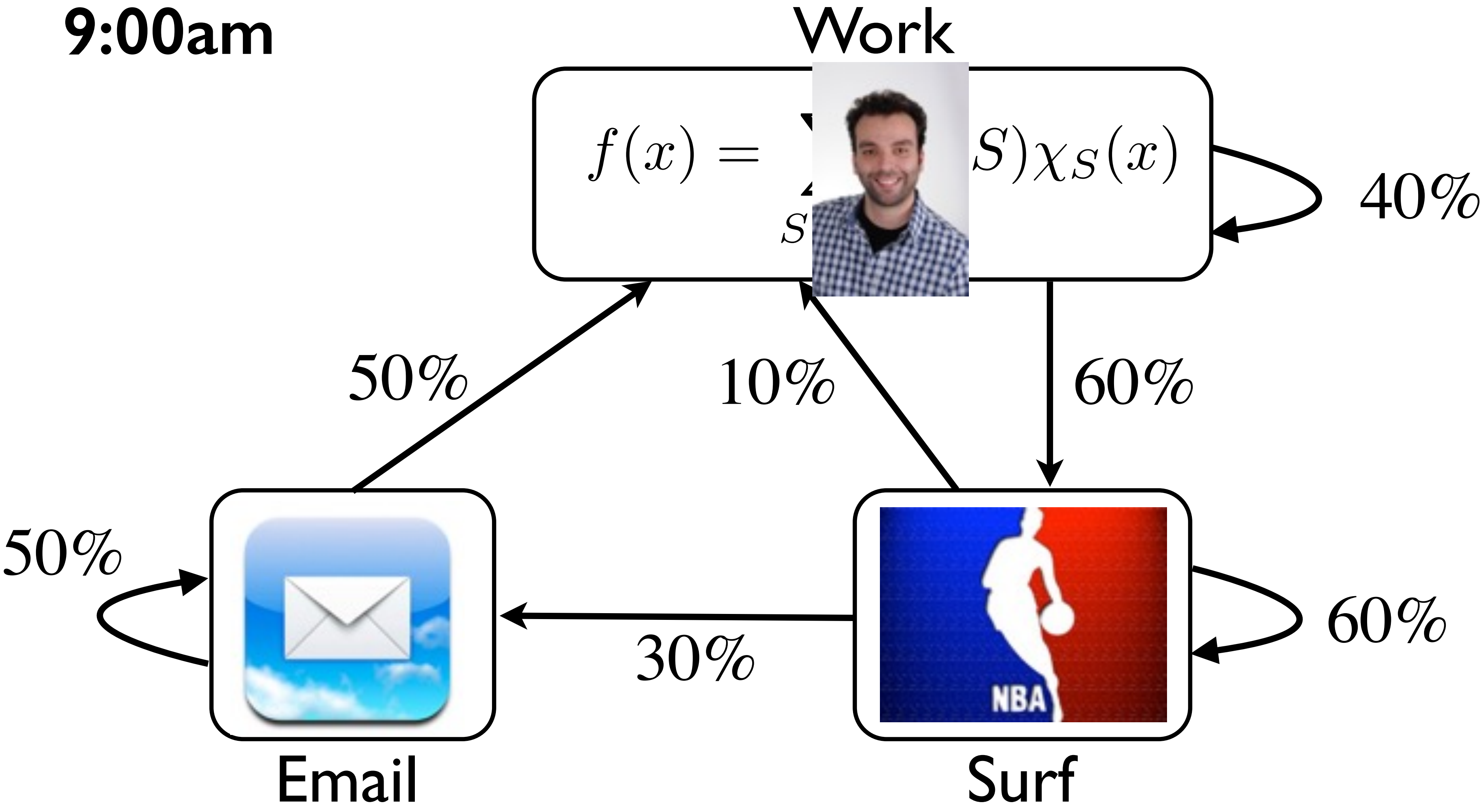
60%

Email

Surf

A day in the life of me

9:00am



Work

$f(x) = \sum_{S}$ $S)\chi_S(x)$

40%

50%

10%

60%

50%

30%

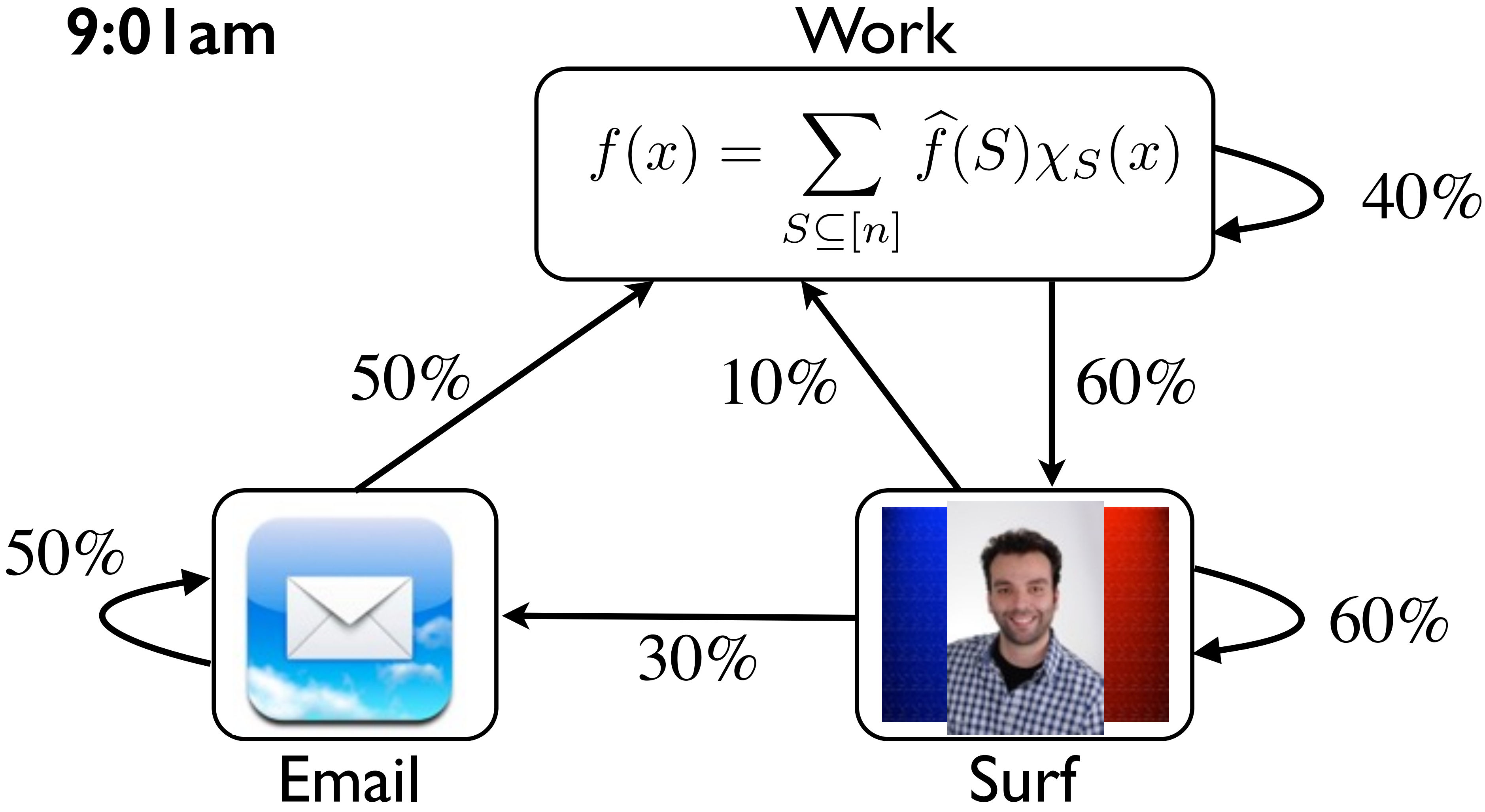
60%

Email

Surf

A day in the life of me

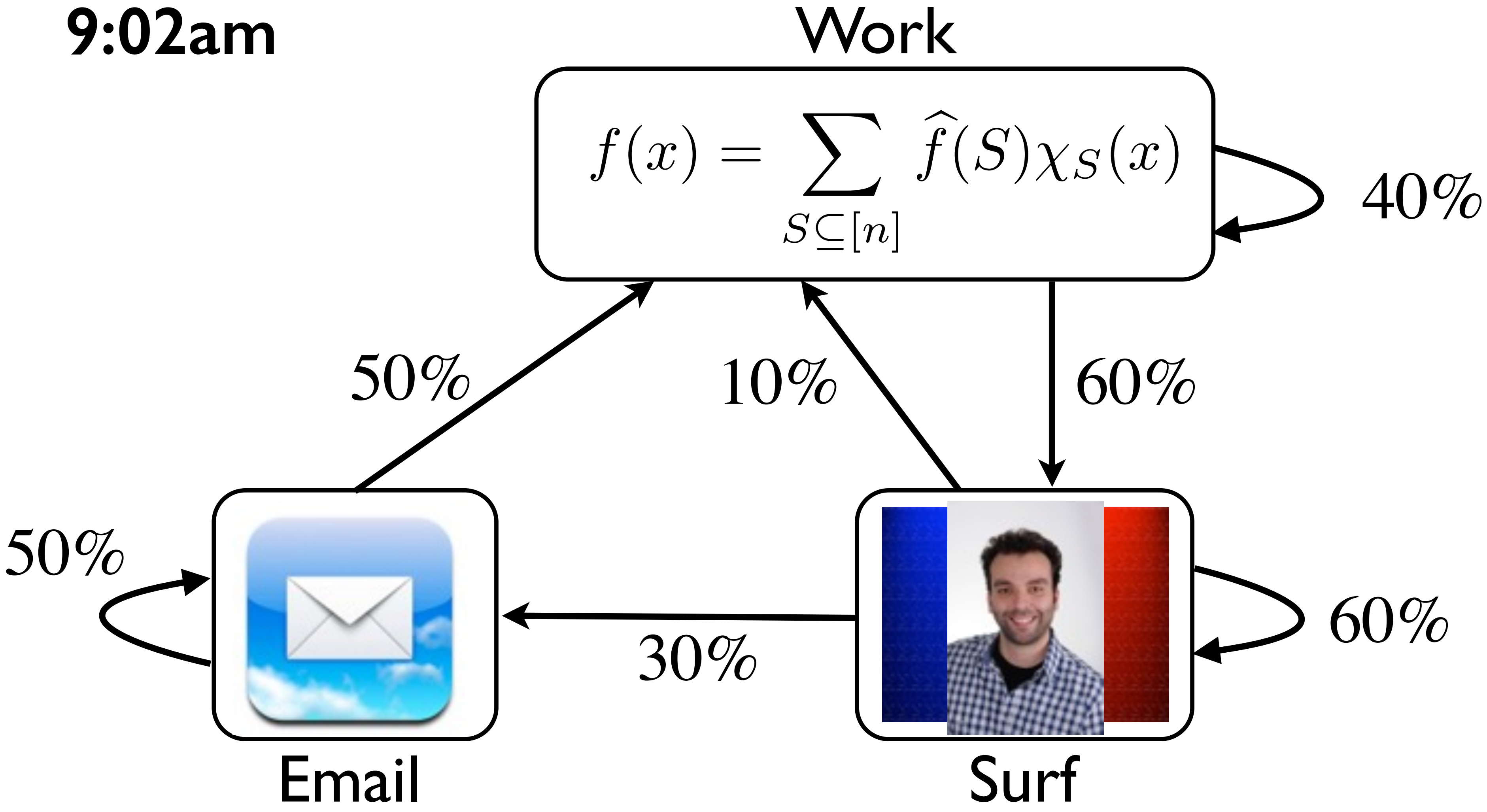
9:01am



$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$$

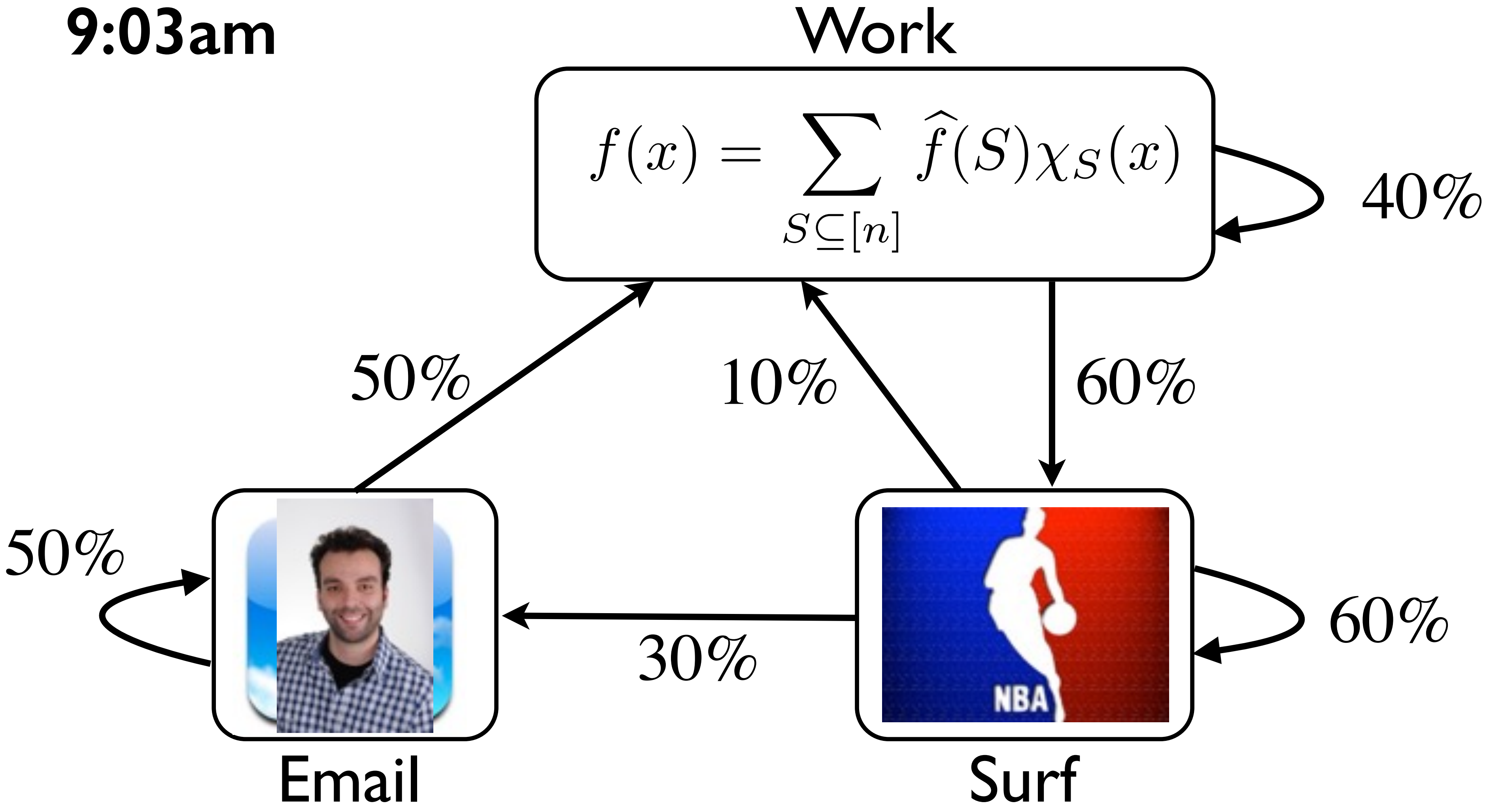
A day in the life of me

9:02am



A day in the life of me

9:03am



Work

$$f(x) = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S(x)$$

40%

50%

10%

60%

50%

Email

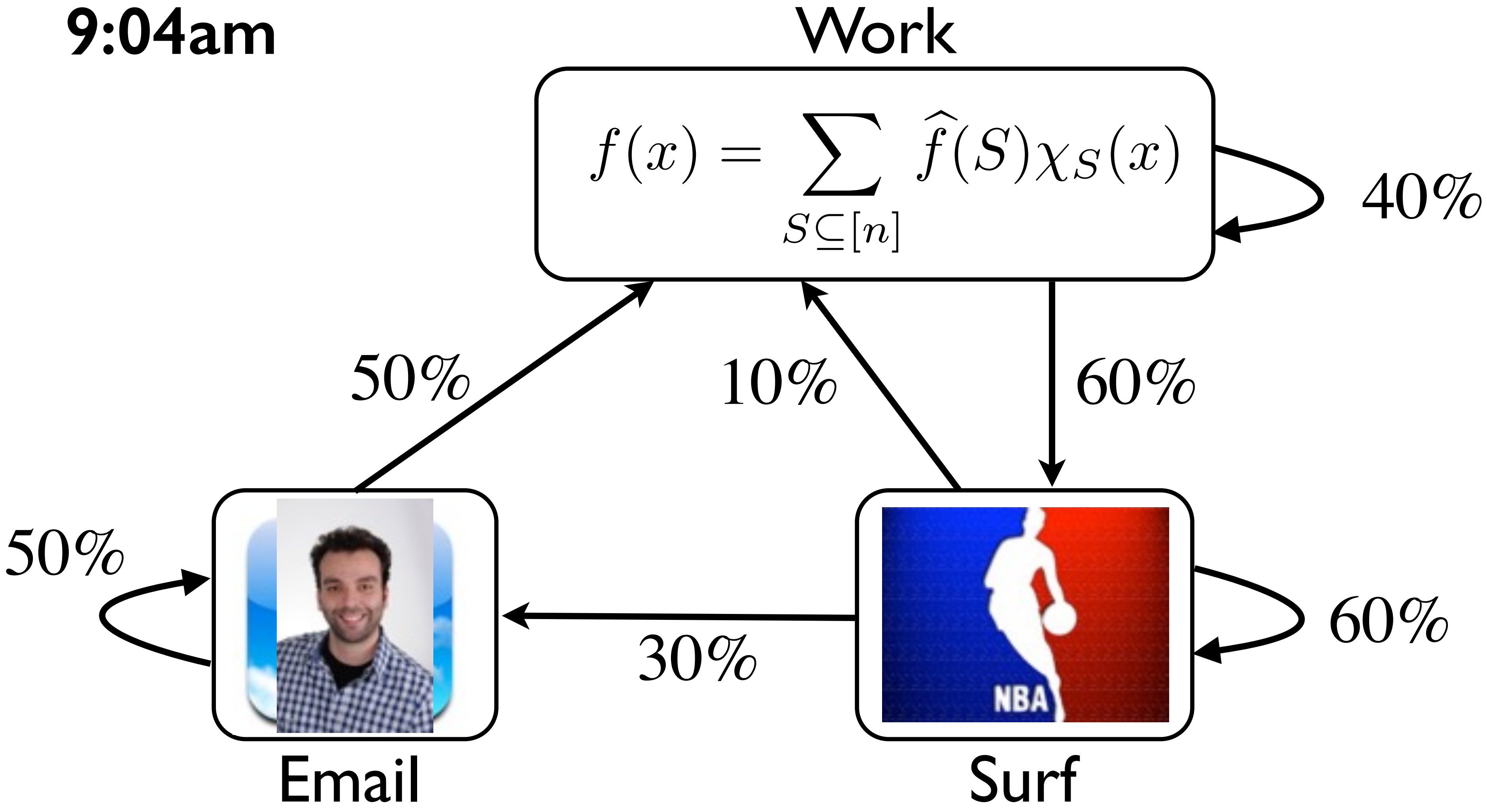
30%

Surf

60%

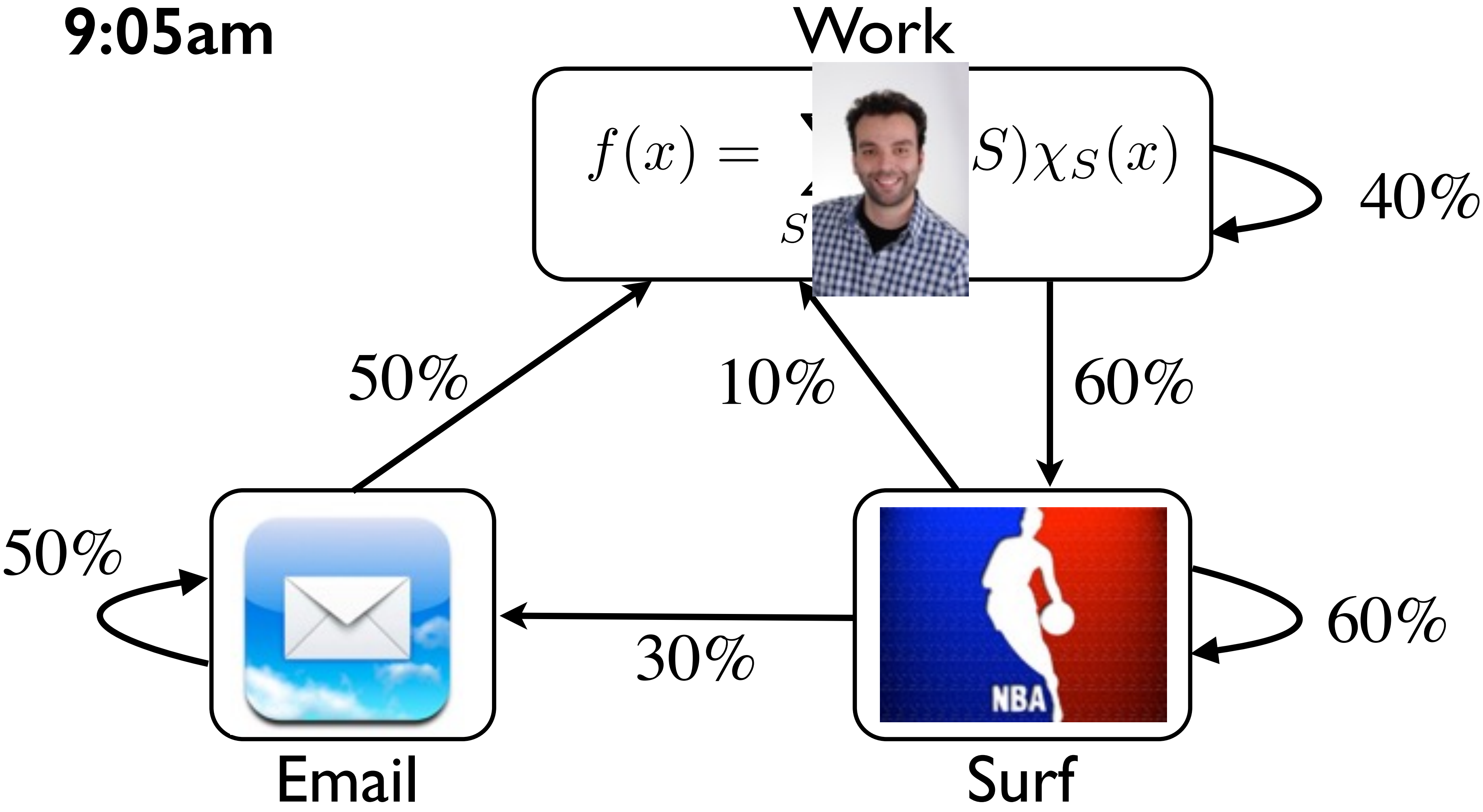
A day in the life of me

9:04am



A day in the life of me

9:05am



Markov Model

Markov Model

Andrey Markov (1856 - 1922)

- Russian mathematician.
- Famous for his work on random processes.

Markov Model

Andrey Markov (1856 - 1922)

- Russian mathematician.
- Famous for his work on random processes.

A model for the evolution of a random system.

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

Cool Things About 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.

Outline

Motivating examples and applications

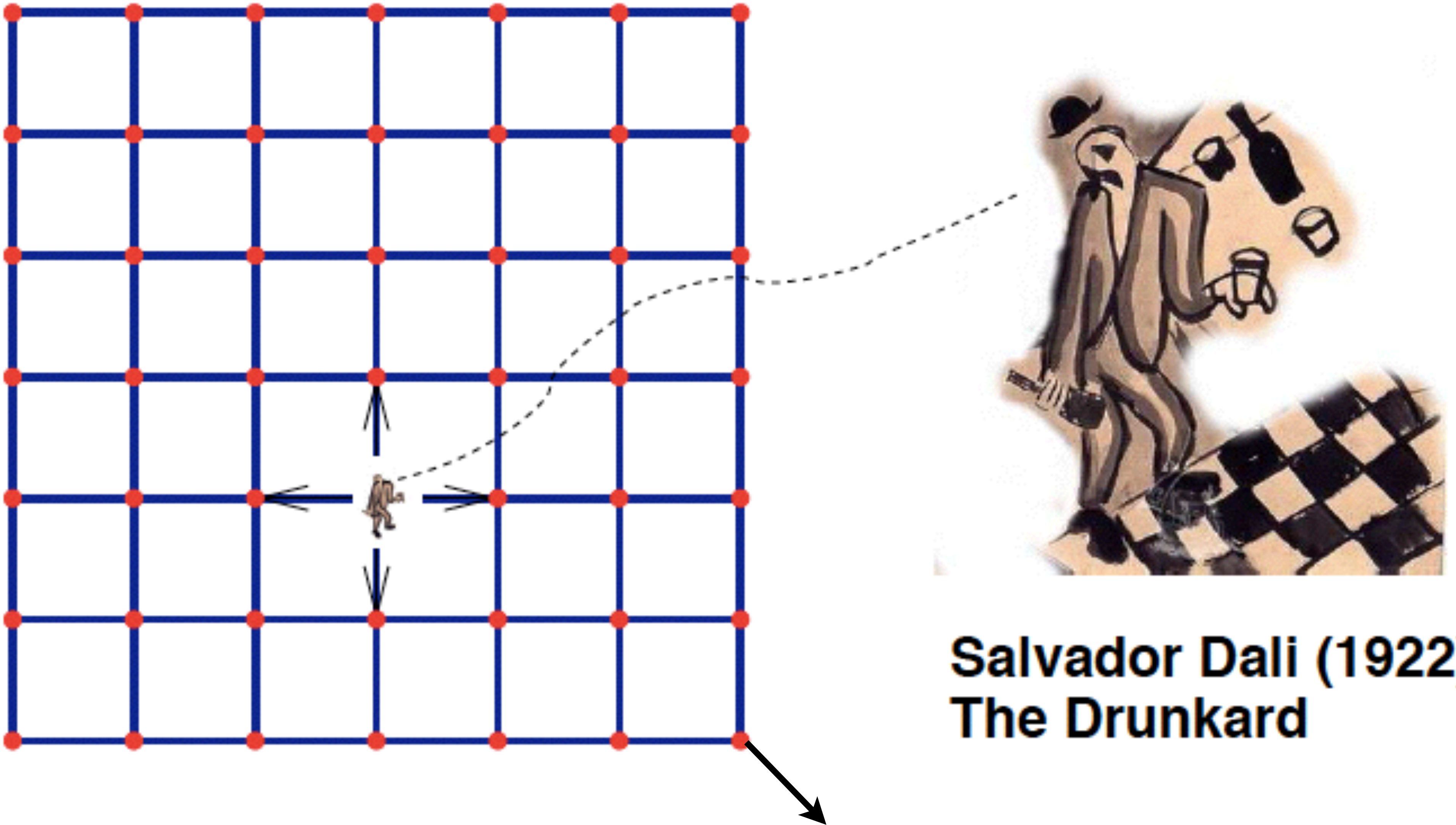
Basic mathematical representation and properties

Applications

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

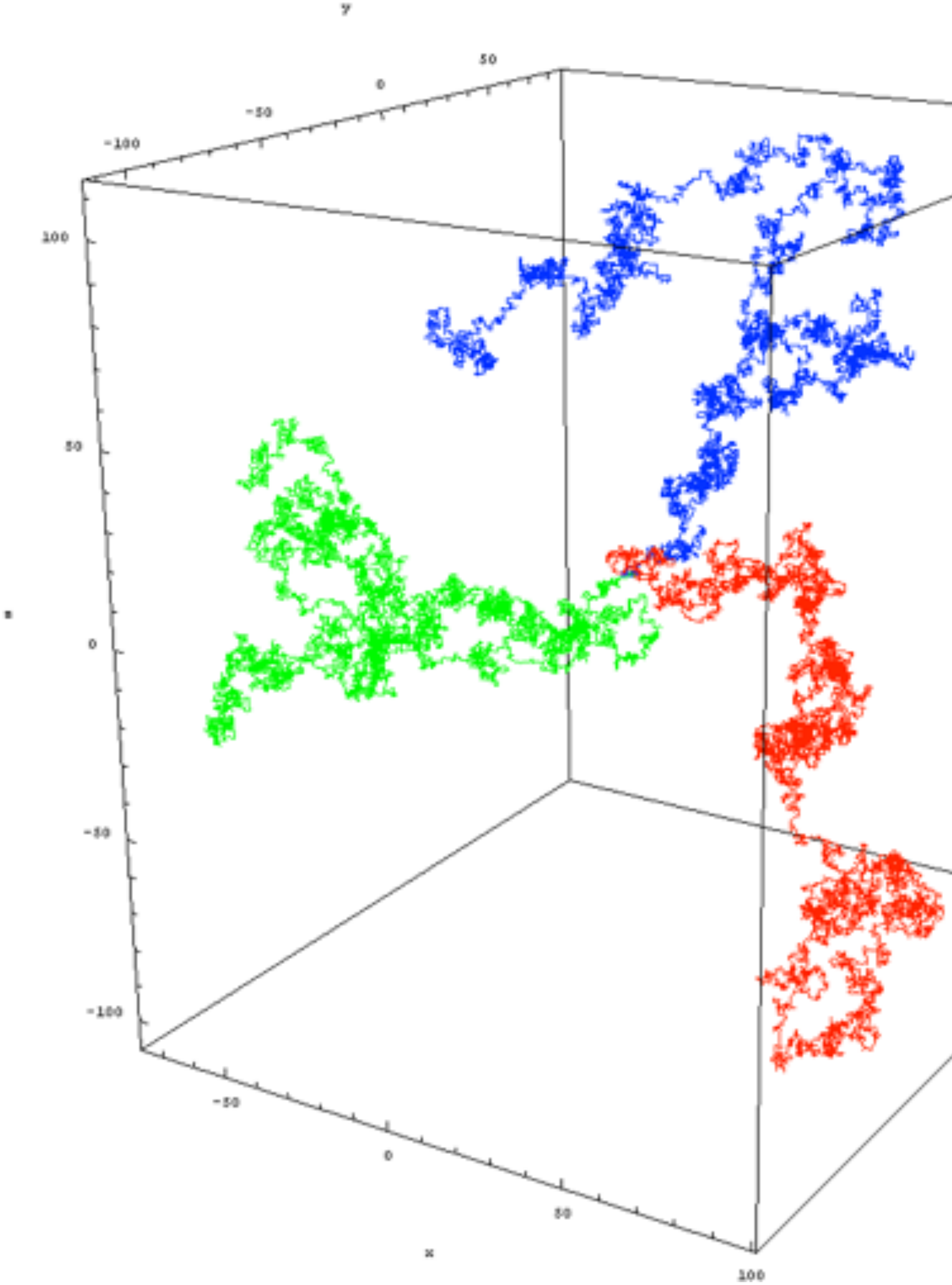
Some Examples of Markov Models

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:

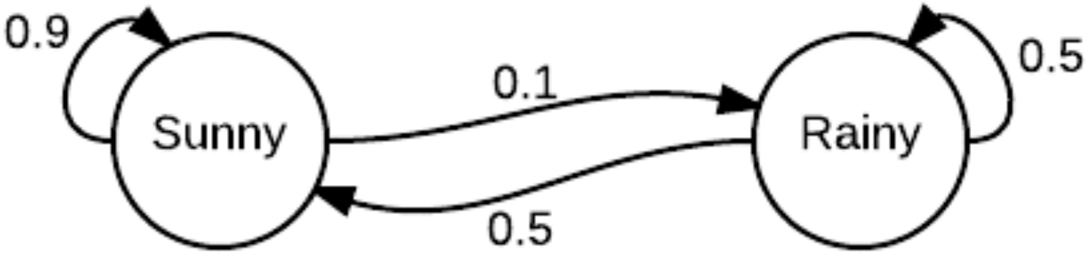
Pr(sunny to rainy) = 0.1

Pr(sunny to sunny) = 0.9

Pr(rainy to rainy) = 0.5

Pr(rainy to sunny) = 0.5

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



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

Example: Life Insurance

Goal of 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

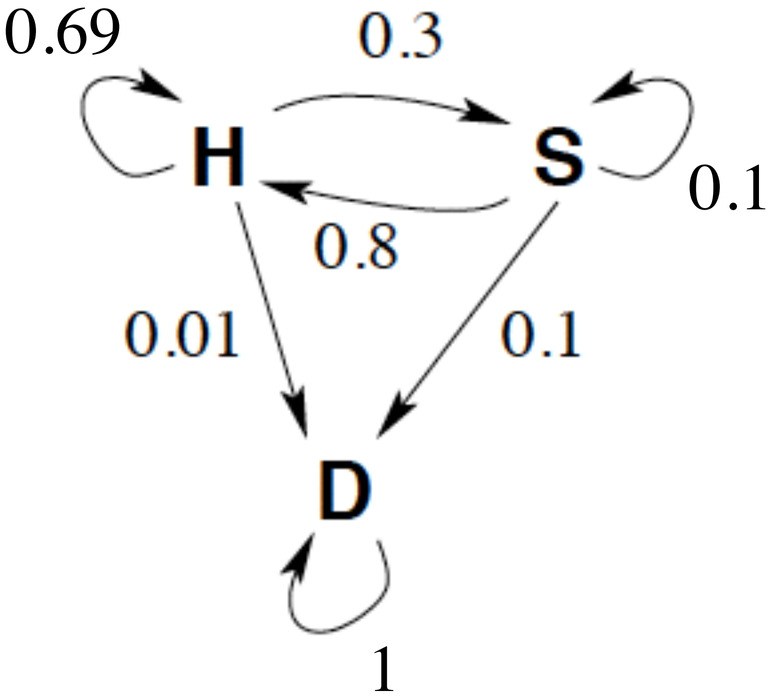
Example: Life Insurance

Goal of 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:



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

Some Applications of Markov Models

Application: Algorithmic Music Composition

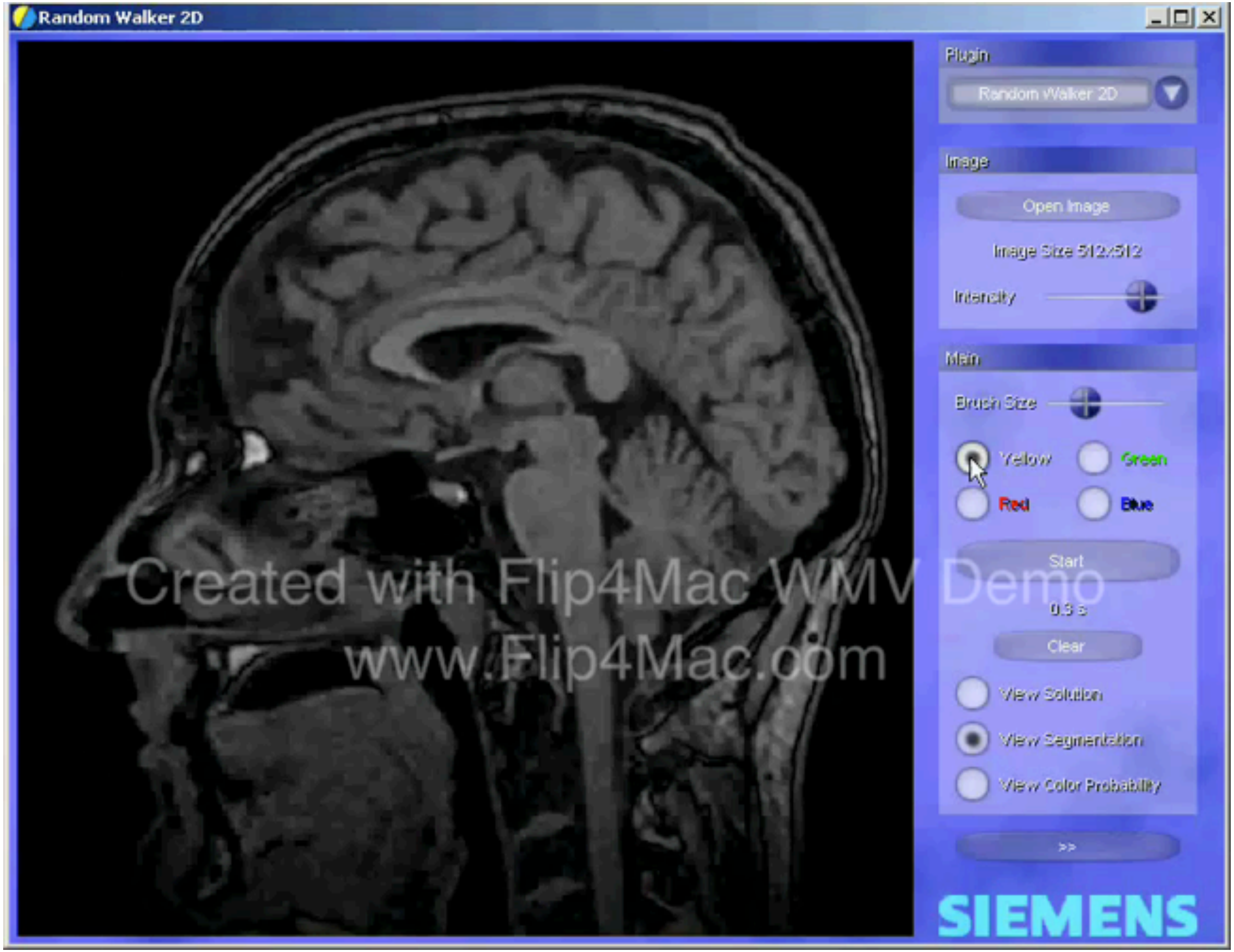
Nicholas Vasallo

Megalithic Copier #2: Markov Chains

(2011)

written in Pure Data

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

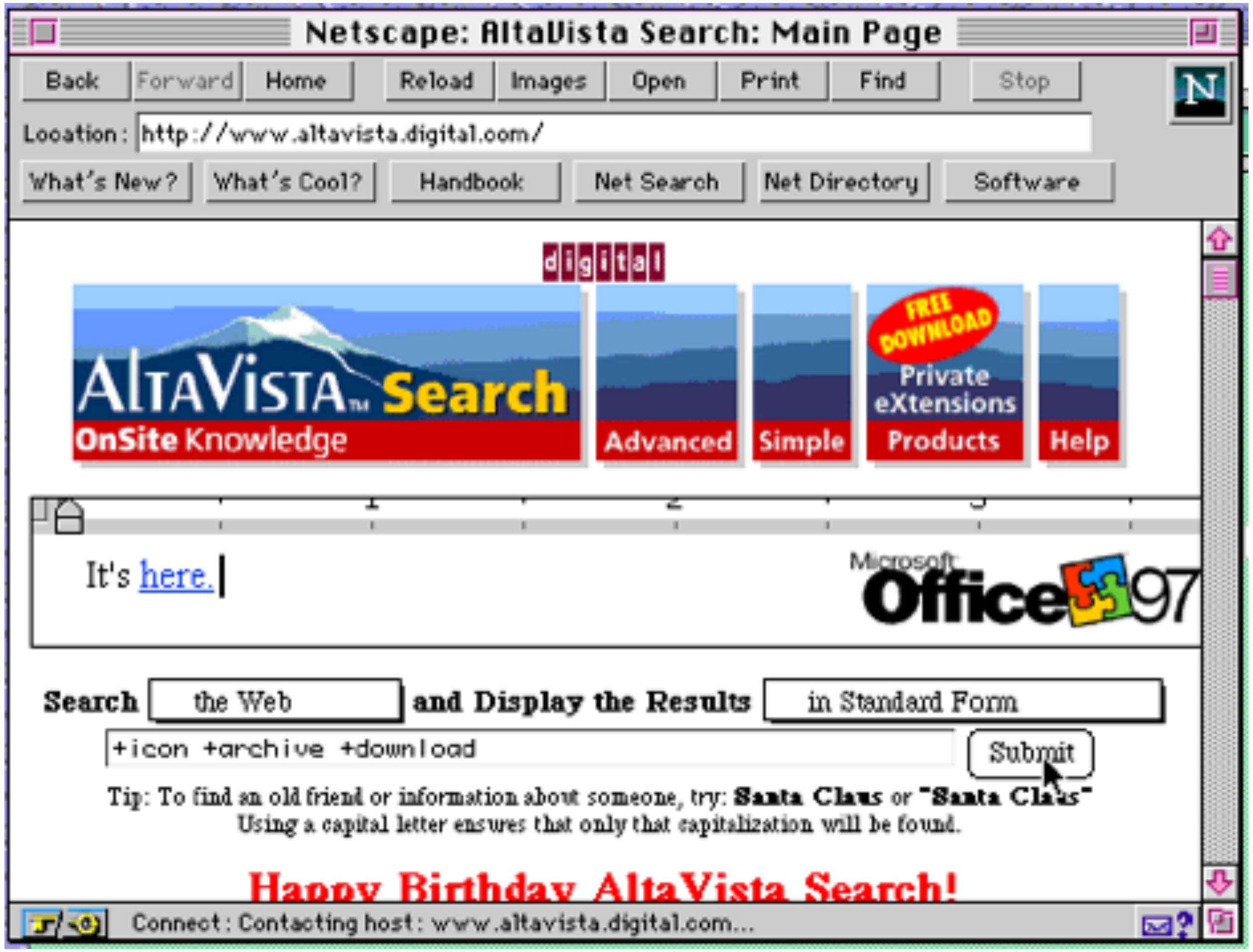
Application: Speech Recognition

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

test

Application: Google PageRank

1997: Web search was horrible



Location: http://www.altavista.digital.com/

It's here.

Search the Web and Display the Results in Standard Form

+icon +archive +download

Tip: To find an old friend or information about someone, try: **Santa Claus** or **"Santa Claus"**
Using a capital letter ensures that only that capitalization will be found.

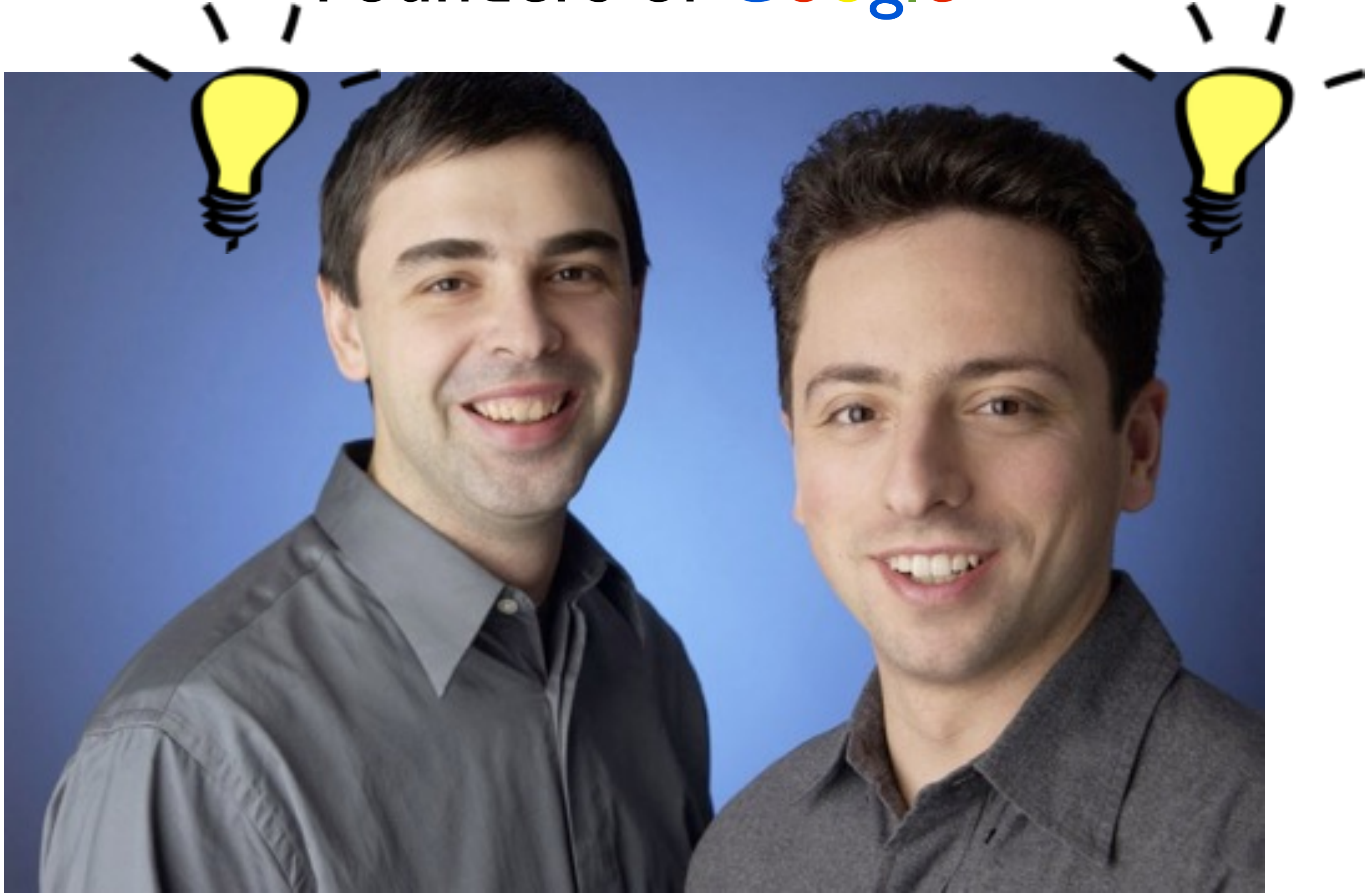
Happy Birthday AltaVista Search!

Connect: Contacting host: www.altavista.digital.com...

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?

2 important factors:

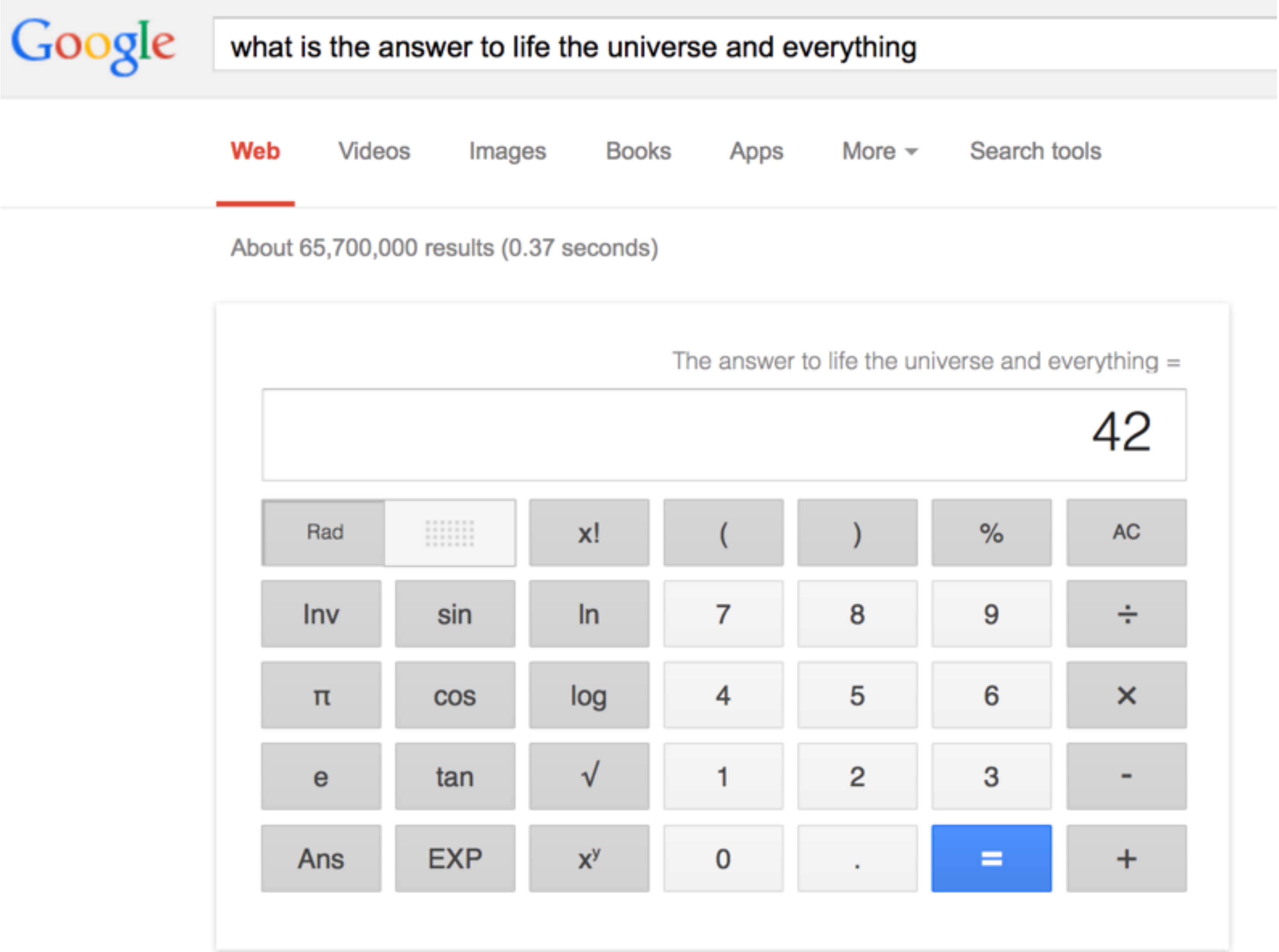
Relevance of the page.

Reputation of the page.

The number and reputation of links pointing to you.

Reputation is measured using PageRank.

PageRank is calculated using a Markov chain.



what is the answer to life the universe and everything

About 65,700,000 results (0.37 seconds)

The answer to life the universe and everything =

42

Outline

Motivating examples and applications

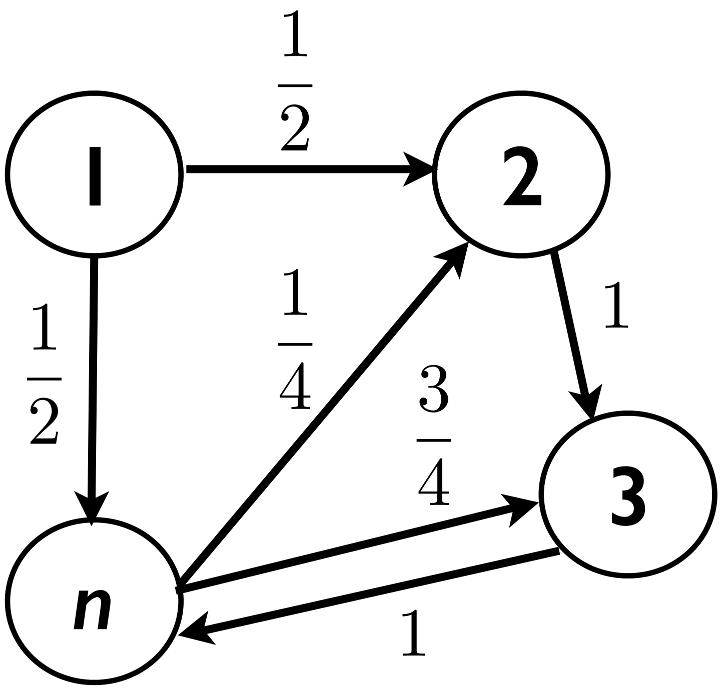
Basic mathematical representation and properties

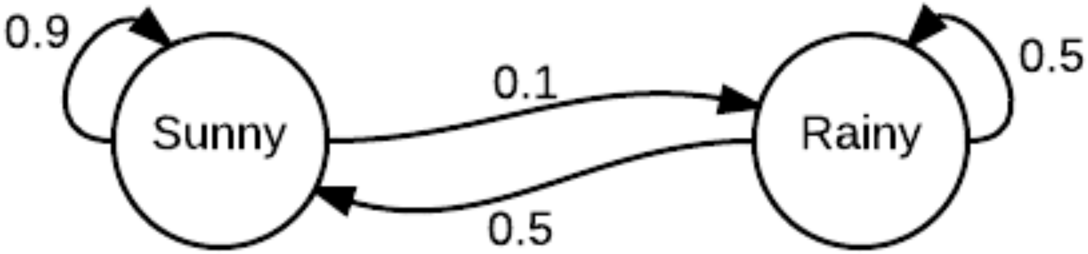
Applications

The Setting

There is a system with *n* possible states/values.

At each time step, the state changes probabilistically.

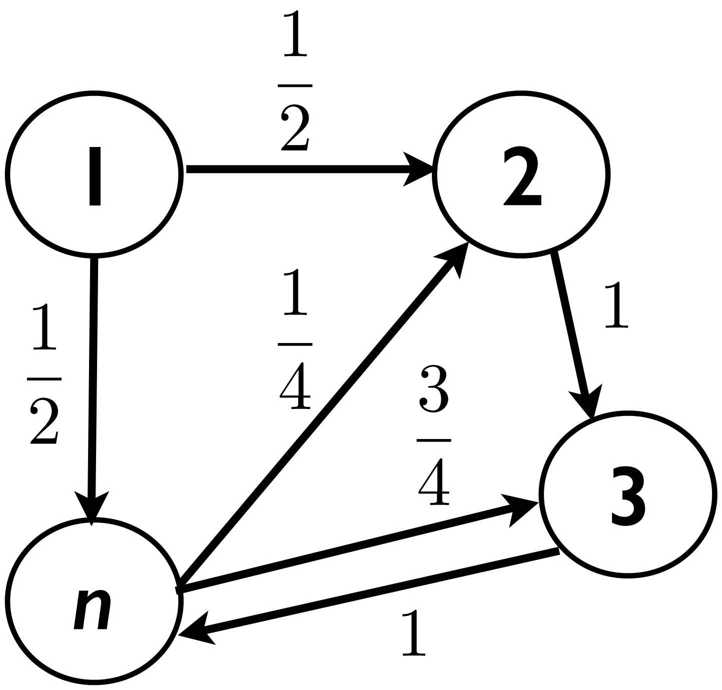


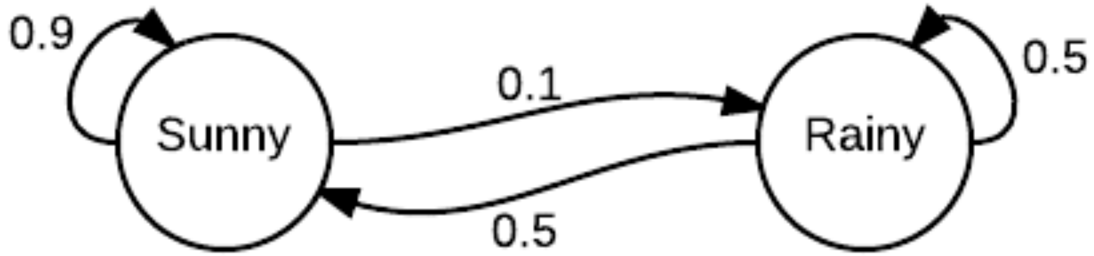


The Setting

There is a system with *n* possible states/values.

At each time step, the state changes probabilistically.





Memoryless

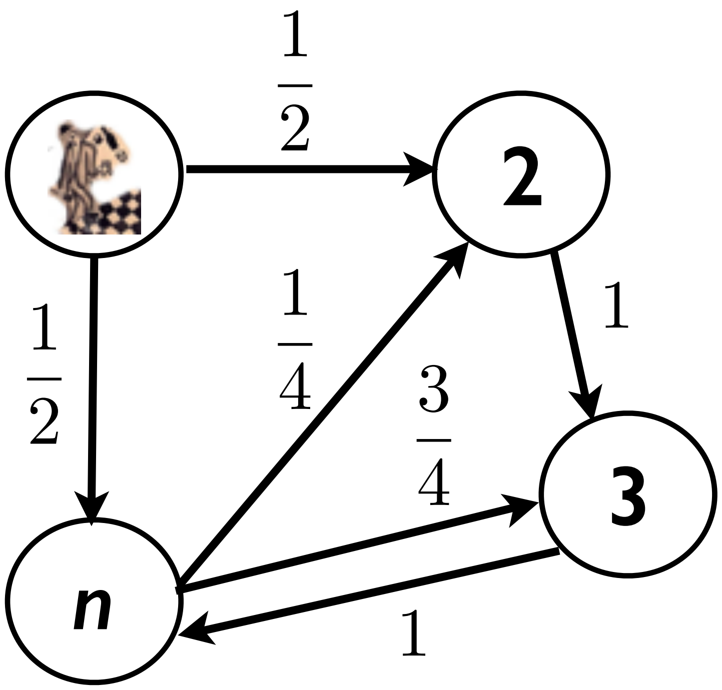
The next state only depends on the current state.

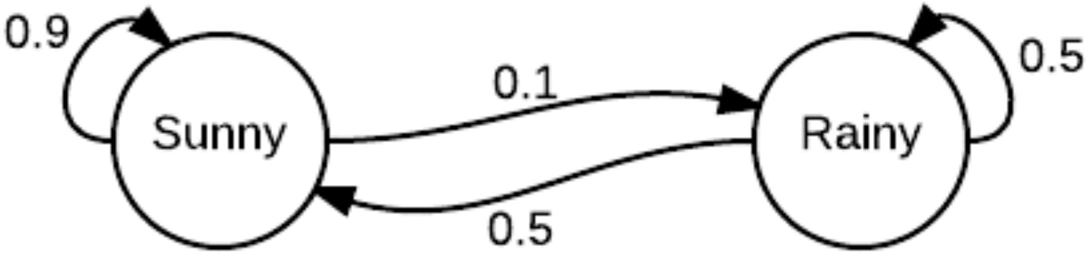
Evolution of the system: random walk on the graph.

The Setting

There is a system with *n* possible states/values.

At each time step, the state changes probabilistically.



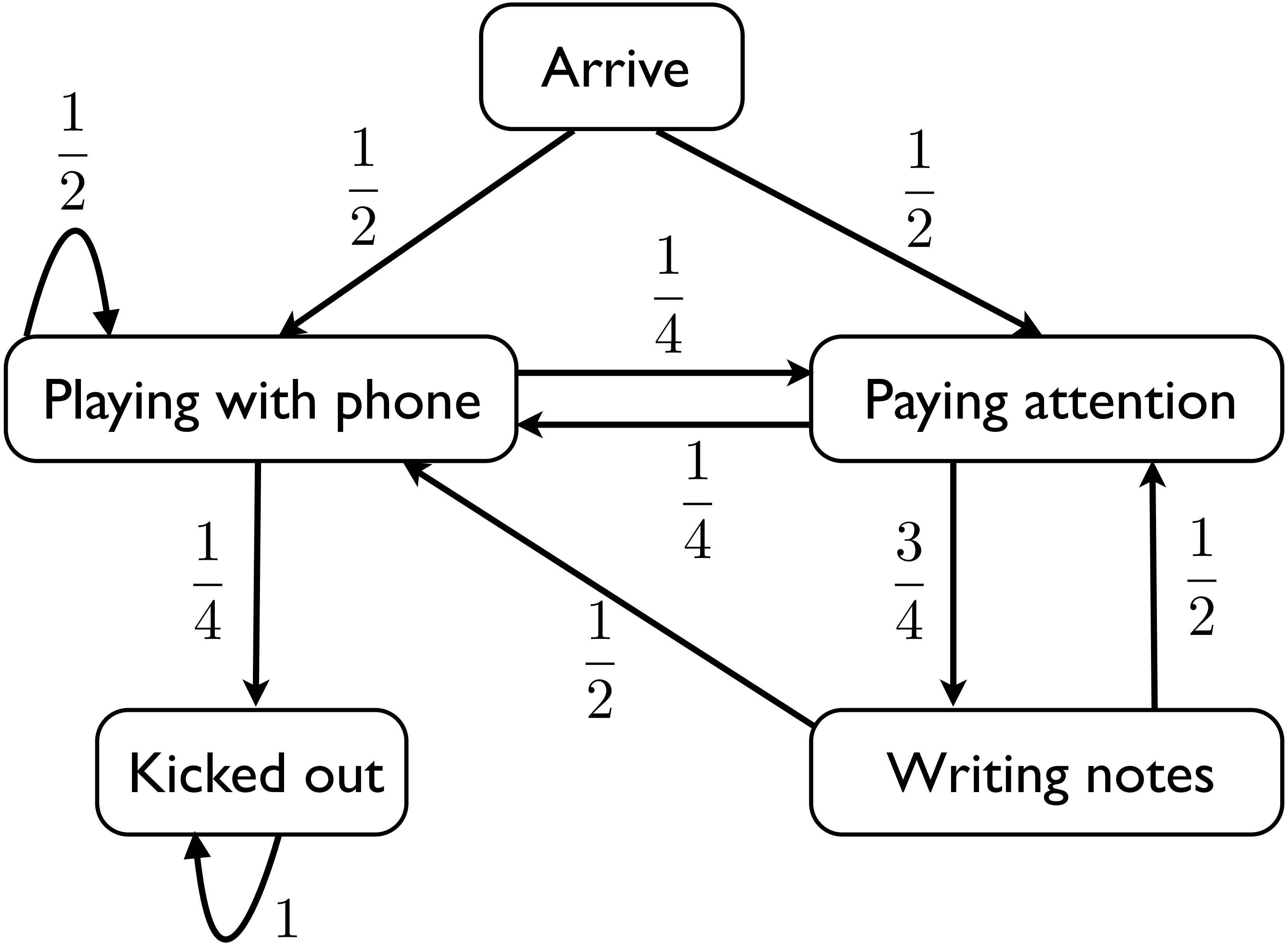


Memoryless

The next state only depends on the current state.

Evolution of the system: random walk on the graph.

Example: Markov Model for a Lecture



The Setting

There is a system with n possible states/values.

At each time step, the state changes probabilistically.

Let $X_t =$ the state of the system at time *t*

Evolution of the system: $X_0, X_1, X_2, \ldots, X_t, \ldots$

X_0 is the initial state.

Memoryless:

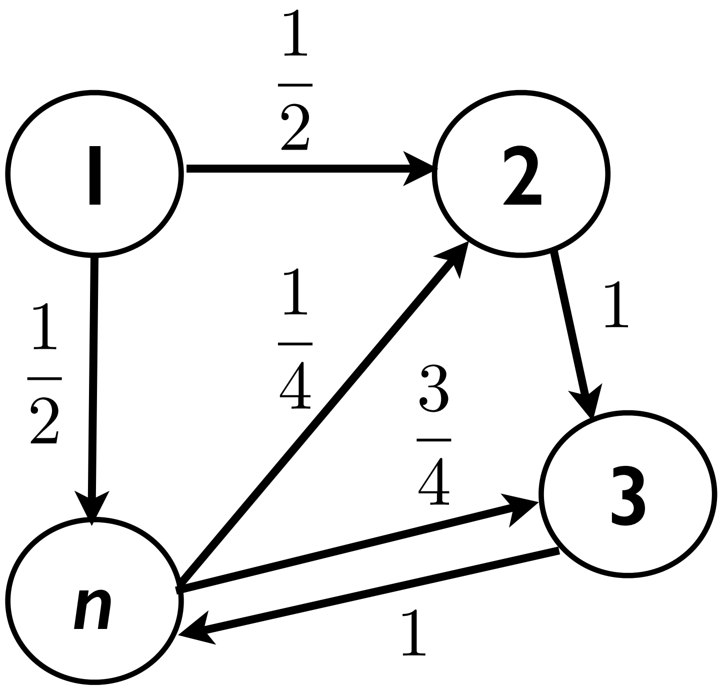
The probability that X_t is in a certain state is determined by the state of X_{t-1}:

$$\Pr[X_t = x | X_0 = x_0, X_1 = x_1, \ldots X_{t-1} = x_{t-1}]$$

$$= \Pr[X_t = x | X_{t-1} = x_{t-1}]$$

The Setting

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



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

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

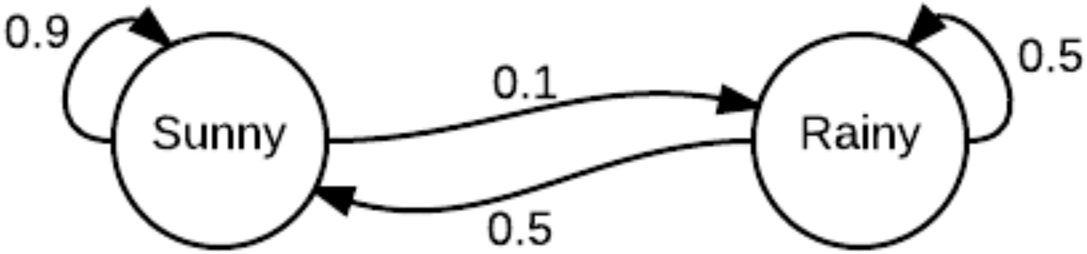
$$\Pr[X_1 = n | X_0 = 1] = \frac{1}{2}$$

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

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

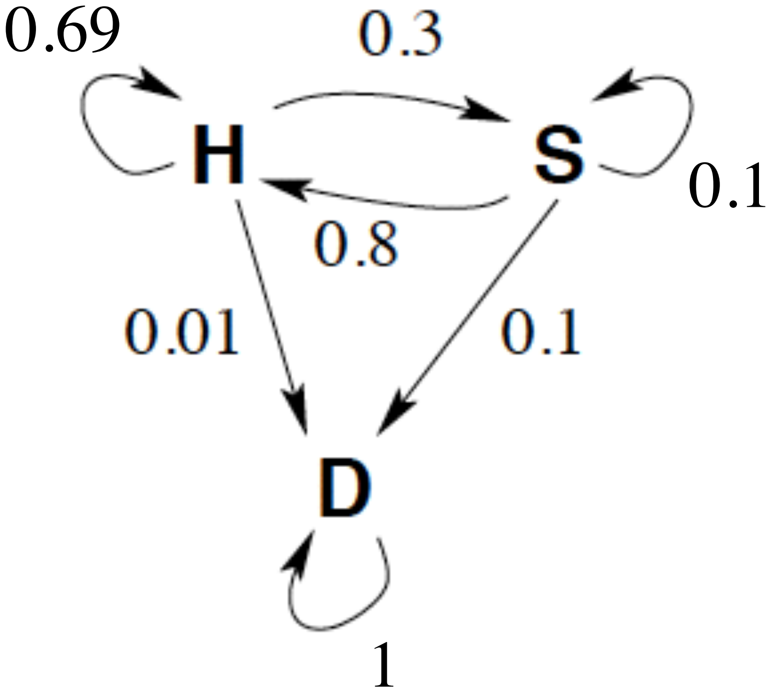
$$\Pr[X_t = 3 | X_{t-1} = 2] = 1$$

The Setting: Equivalent representations



$$\begin{array}{c|cc} & \mathbf{S} & \mathbf{R} \\ \hline \mathbf{S} & 0.9 & 0.1 \\ \mathbf{R} & 0.5 & 0.5 \end{array}$$

Transition Matrix



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

Transition Matrix

Simplifying assumptions for 251

Finite number of states.

The underlying graph is strongly connected.

Some Fundamental and Natural Questions

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

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

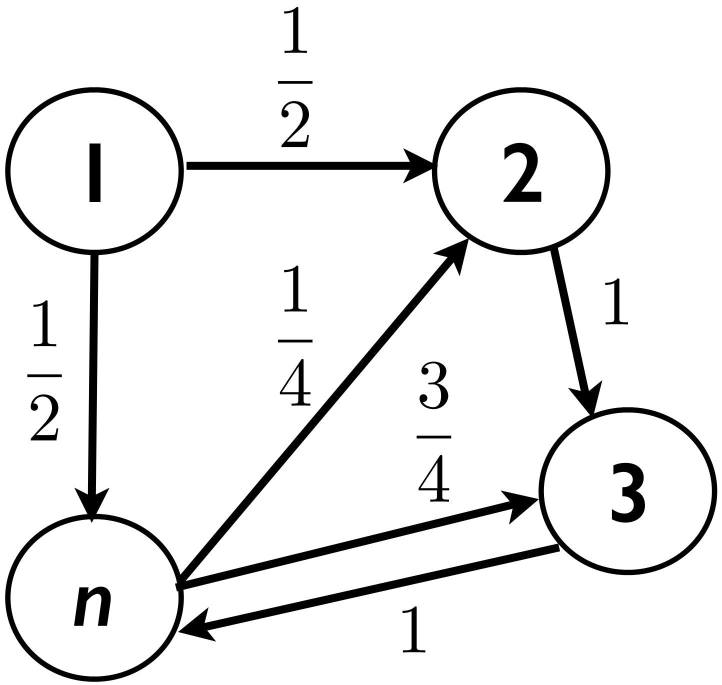
What is the expected time of having visited every state (given some initial state)?

...

Mathematical representation of the evolution

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

How can we mathematically represent the evolution?



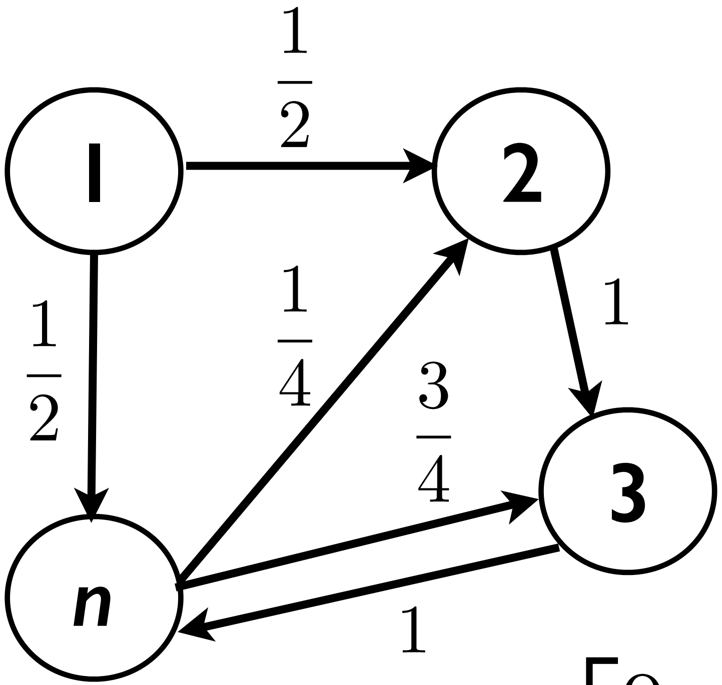
$$\begin{array}{cc} & \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \end{array} \\ \begin{array}{c} \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{4} \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4}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

Transition Matrix

Mathematical representation of the evolution

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

How can we mathematically represent the evolution?



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

$$\underset{\pi_0}{\begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \left[1\right. & 0 & 0 & \left.0\right] \end{array}} \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{bmatrix} = \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} \\ \left[0\right. & \frac{1}{2} & 0 & \left.\frac{1}{2}\right] \end{array}$$

Mathematical representation of the evolution

The probability of states after 1 step:

the new state (probabilistic)

$$\underset{\pi_0}{\begin{array}{c}\begin{array}{cccc}\mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4}\end{array}\\ \begin{bmatrix}1 & 0 & 0 & 0\end{bmatrix}\end{array}} \begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & \frac{1}{4} & \frac{3}{4} & 0\end{bmatrix} = \underset{\pi_1}{\begin{array}{c}\begin{array}{cccc}\mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4}\end{array}\\ \begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix}\end{array}}$$

The probability of states after 2 steps:

the new state (probabilistic)

$$\underset{\pi_1}{\begin{array}{c}\begin{array}{cccc}\mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4}\end{array}\\ \begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\end{bmatrix}\end{array}} \begin{bmatrix}0 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & \frac{1}{4} & \frac{3}{4} & 0\end{bmatrix} = \underset{\pi_2}{\begin{array}{c}\begin{array}{cccc}\mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4}\end{array}\\ \begin{bmatrix}0 & \frac{1}{8} & \frac{7}{8} & 0\end{bmatrix}\end{array}}$$

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:

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

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

Mathematical representation of the evolution

In general:

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

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

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

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

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

Homework:

Prove this.

Remarkable Property 1

What happens in the long run?

Suppose the Markov chain is "aperiodic".

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

As $t \to \infty$

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

Remarkable Property I

As $t \to \infty$

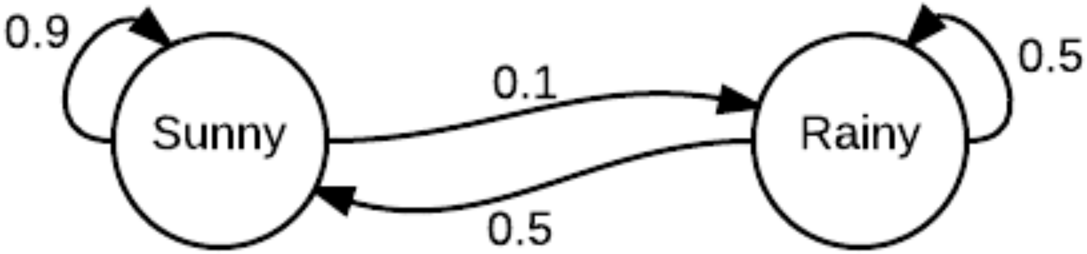
$$\begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} \text{Transition} \\ \text{Matrix} \\ \\ \end{bmatrix}^t \longrightarrow \pi$$

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

stationary/invariant distribution

This π is unique.

Remarkable Property 1



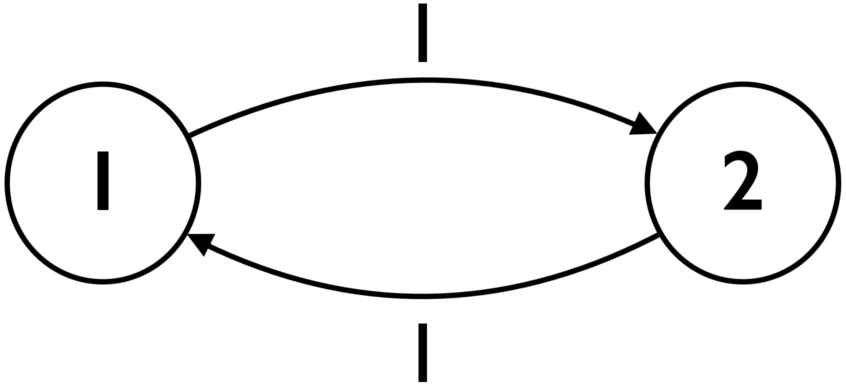
Stationary distribution is

$$\begin{bmatrix} \frac{5}{6} & \frac{1}{6} \end{bmatrix}$$

In the long run, it is sunny 5/6 of the time,
it is rainy 1/6 of the time.

Remarkable Property 1

What is a "periodic" Markov chain?



$$\pi_0 = \begin{bmatrix}1 & 0\end{bmatrix}$$
$$\pi_1 = \begin{bmatrix}0 & 1\end{bmatrix}$$
$$\pi_2 = \begin{bmatrix}1 & 0\end{bmatrix}$$
$$\pi_3 = \begin{bmatrix}0 & 1\end{bmatrix}$$
$$\pi_4 = \begin{bmatrix}1 & 0\end{bmatrix}$$
$$\dots$$

There is still a stationary distribution.

$$\pi = \begin{bmatrix}1/2 & 1/2\end{bmatrix}$$

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

But it is not a limiting distribution.

Remarkable Property 2

Let T_{ij} = time of reaching state *j*
when you start at state *i*

Then $$\mathbb{E}[T_{ii}] = \frac{1}{\pi[i]} \, .$$

Known as the Mean Recurrence Theorem.

Remarkable Property 2

Let $T_{ij} =$ time of reaching state *j*
when you start at state *i*

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

Intuition:

If we walk for *N* steps,
you would expect to be in state *i* about $\pi[i]N$ times.

$$\cdots \boxed{X_t} \; X_{t+1} \; X_{t+2} \; X_{t+3} \; \boxed{X_{t+4}} \; X_{t+5} \; X_{t+6} \; X_{t+7} \; \cdots \; X_N$$

Average time between successive visits to *i* : $1/\pi[i]$

Summary so far

Markov chains can be characterized by the transition matrix K.

$$K[i,j] = \Pr[X_1 = j | X_0 = i]$$

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

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

There is a unique invariant distribution π: $\pi = \pi K$

For aperiodic Markov Chains: $\pi_t \to \pi$

$$\mathbb{E}[T_{ii}] = \mathbb{E}[\#\text{ steps to go from } i \text{ to } i] = 1/\pi[i]$$

Outline

Motivating examples and applications

Basic mathematical representation and properties

Applications

Outline

Applications

- Theoretical
- Practical

A Theoretical Application:
Connectivity problem

The connectivity problem

Input: An undirected graph $G = (V, E)$, and $s, t \in V$.

Output: Yes if *s* and *t* are connected. No otherwise.

Easy to do in polynomial time with BFS or DFS.

How about using only O(log n) space?

Doesn't seem possible…

Would randomness help?

Not clear.

The connectivity problem

Input: An undirected graph $G = (V, E)$, and $s, t \in V$.

Output: Yes if *s* and *t* are connected. No otherwise.

```
v := s
for k = 1,2,...,N:
    v := random-neighbor(v)
    if v = t, return YES
return NO
```

For N = poly(n), this uses O(log n) space.

But what is the success probability???

If s and t are disconnected, we give correct answer.

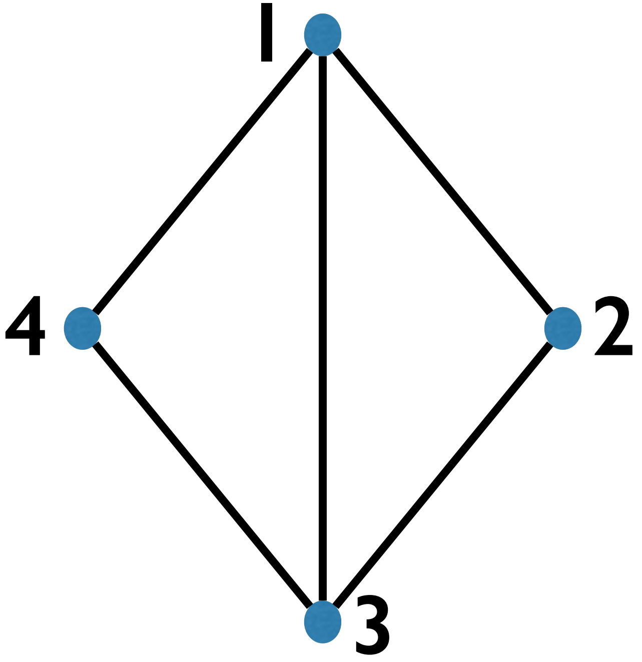
What if s and t are connected?

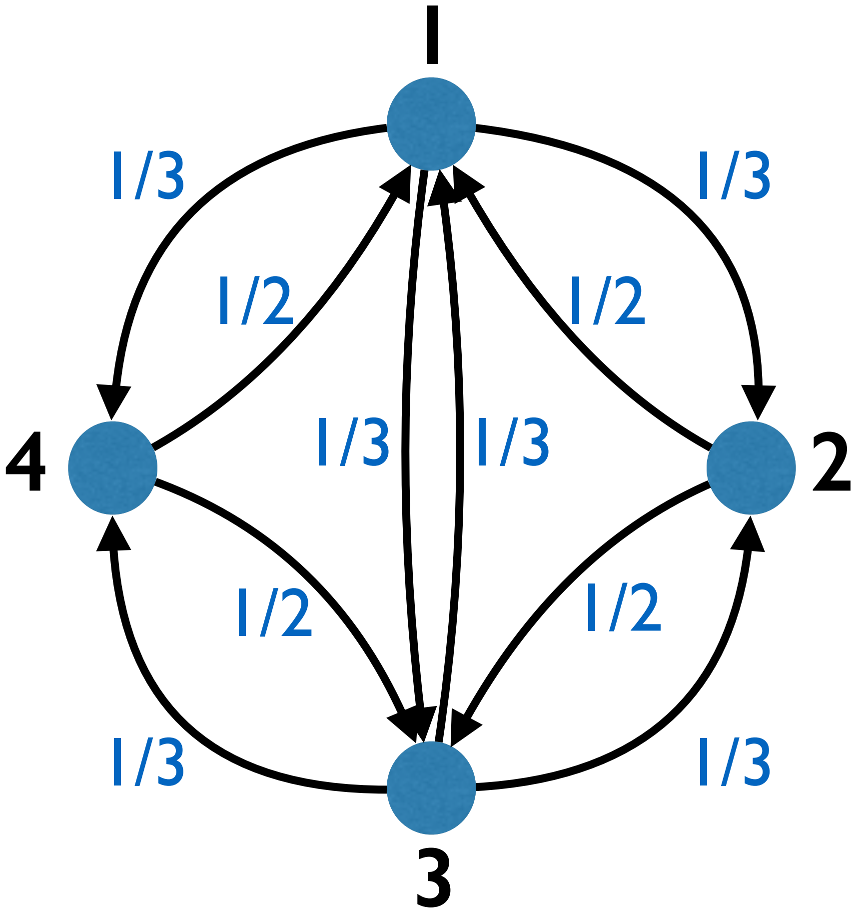
Random walk on undirected graphs

Given an undirected graph with n nodes, m edges.

Start at some vertex.

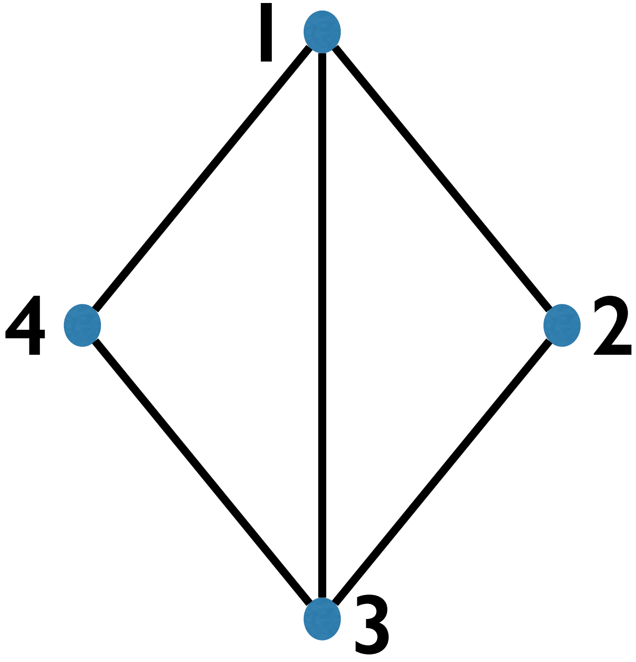
At each step, go to a random neighbor.





Random walk on undirected graphs

How does the transition matrix look like?

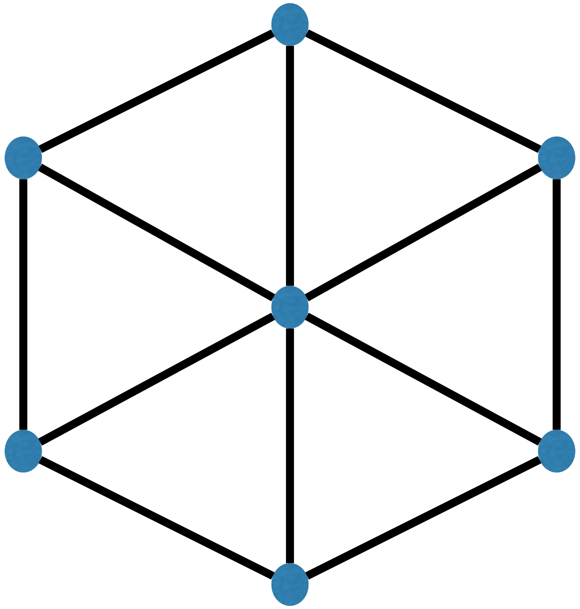


$$A = \begin{array}{c|cccc|l} & 1 & 2 & 3 & 4 & \\ 1 & 0 & 1 & 1 & 1 & \div \deg(1) \\ 2 & 1 & 0 & 1 & 0 & \div \deg(2) \\ 3 & 1 & 1 & 0 & 1 & \div \deg(3) \\ 4 & 1 & 0 & 1 & 0 & \div \deg(4) \end{array}$$

$$K = \begin{array}{c|cccc|} & 1 & 2 & 3 & 4 \\ 1 & 0 & 1/3 & 1/3 & 1/3 \\ 2 & 1/2 & 0 & 1/2 & 0 \\ 3 & 1/3 & 1/3 & 0 & 1/3 \\ 4 & 1/2 & 0 & 1/2 & 0 \end{array}$$

Random walk on undirected graphs

How does the stationary distribution look like?



Seems higher degree should imply higher limiting prob.

Is $\pi[i]$ proportional to $\deg(i)$?

Yes! $\pi = \left[\frac{\deg(1)}{2m}, \frac{\deg(2)}{2m}, \frac{\deg(3)}{2m}, \cdots, \frac{\deg(n)}{2m} \right]$

So: $\mathbb{E}[T_{ii}] = \frac{2m}{\deg(i)}$

Random walk on undirected graphs

How about $\mathbb{E}[T_{ij}]$? (when *i* and *j* are connected)

Pick a path from *i* to *j* : $i = i_1, i_2, i_3, \ldots, i_r = j \quad (r \leq n)$

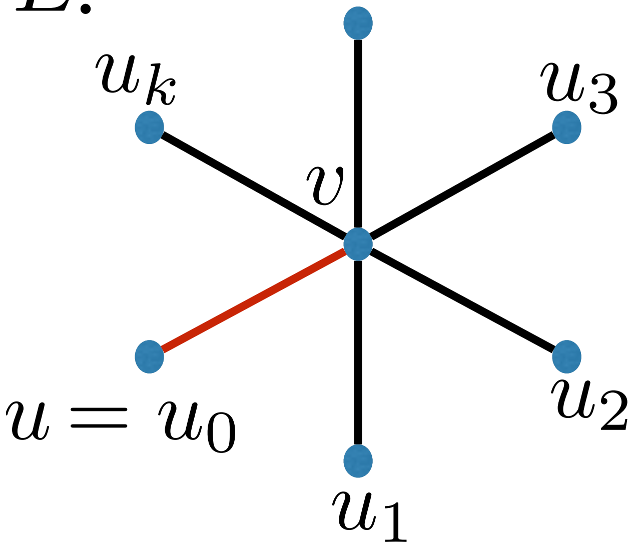
$$\begin{aligned} \mathbb{E}[T_{ij}] &\leq \mathbb{E}[T_{i_1 i_2} + T_{i_2 i_3} + \cdots + T_{i_{r-1} i_r}] \\ &= \mathbb{E}[T_{i_1 i_2}] + \mathbb{E}[T_{i_2 i_3}] + \cdots + \mathbb{E}[T_{i_{r-1} i_r}] \\ &\leq 2m + 2m + \cdots + 2m = 2mn \leq n^3 \end{aligned}$$

because $\mathbb{E}[T_{uv}] \leq 2m$ when $(u, v) \in E$

$$\boxed{\mathbb{E}[T_{ij}] \leq n^3}$$

Random walk on undirected graphs

$\mathbb{E}[T_{uv}] \leq 2m$ **when** $(u, v) \in E$:



$$\frac{2m}{\deg(v)} = \mathbb{E}[T_{vv}]$$

$$= \sum_{i=0}^{k} \Pr[\text{first step } v \text{ to } u_i] \cdot \mathbb{E}[T_{vv} | \text{first step } v \text{ to } u_i]$$

$$= \sum_{i=0}^{k} \frac{1}{\deg(v)} \cdot (1 + \mathbb{E}[T_{u_i v}])$$

$$\geq \frac{1}{\deg(v)} \cdot (1 + \mathbb{E}[T_{u_0 v}]) \quad \implies \quad 2m \geq 1 + \mathbb{E}[T_{uv}]$$

The connectivity problem

Coming back to the algorithm:

```
v := s
for k = 1,2,...,1000n^3:
    v := random-neighbor(v)
    if v = t, return YES
return NO
```

If s and t are disconnected, we give correct answer.

What if s and t are connected?

$$\mathbb{E}[T_{st}] \leq n^3 \implies \Pr[\text{error}] = \Pr[T_{st} > 1000n^3] \leq \frac{1}{1000}$$

Markov's inequality: $\Pr[X > c\mathbb{E}[X]] \leq \frac{1}{c}$

The connectivity problem

For a long time was one of the canonical problems that:

- had a space efficient randomized alg.
- didn't know if it had a space efficient deterministic alg.

Until:

2004:
"Undirected connectivity in log-space"

Omer Reingold

Some Practical 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

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.

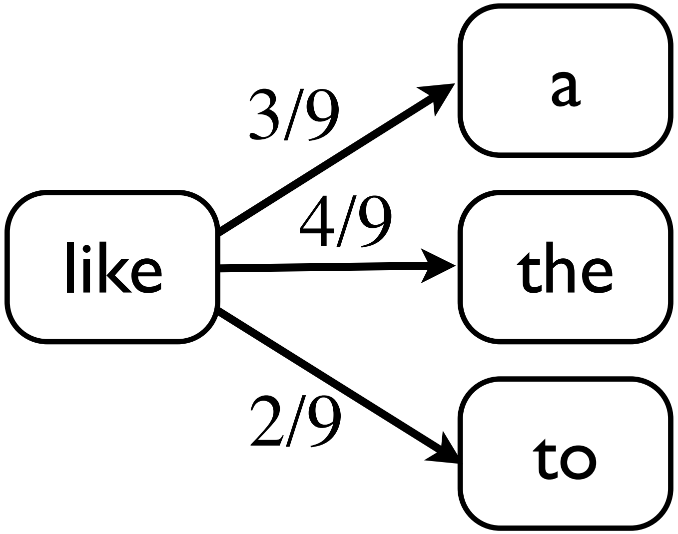
Automatic Text Generation

From the sample document, create a Markov chain.

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

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

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)

Image Segmentation

Simple version

Given an image of an object, figure out:
which pixels correspond to the object,
which pixels correspond to the background

i.e., label each pixel "object" or "background"

User labels a small number of pixels with known labels

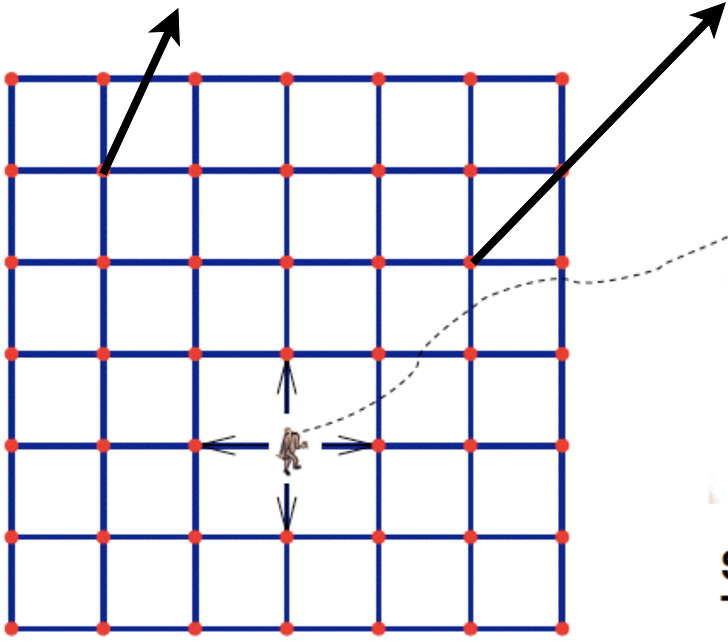
Image Segmentation

Underlying Markov Model

Each pixel is a node/state.

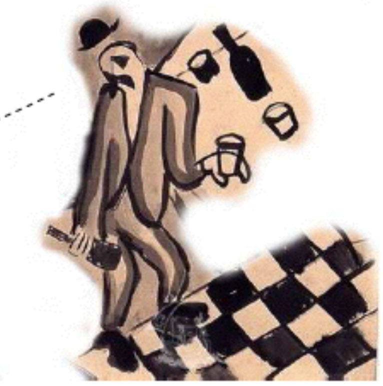
There is an edge between adjacent pixels.

Edge probabilities reflect similarity between pixels.



“background”

“object”



Salvador Dali (1922)
The Drunkard

Which one is more likely:

random walker first visits
“background”
or
“object” ?

Google PageRank

PageRank is a measure of reputation:

The number and reputation of links pointing to you.

The Markov Chain

Every webpage is a node/state. (In total n webpages)

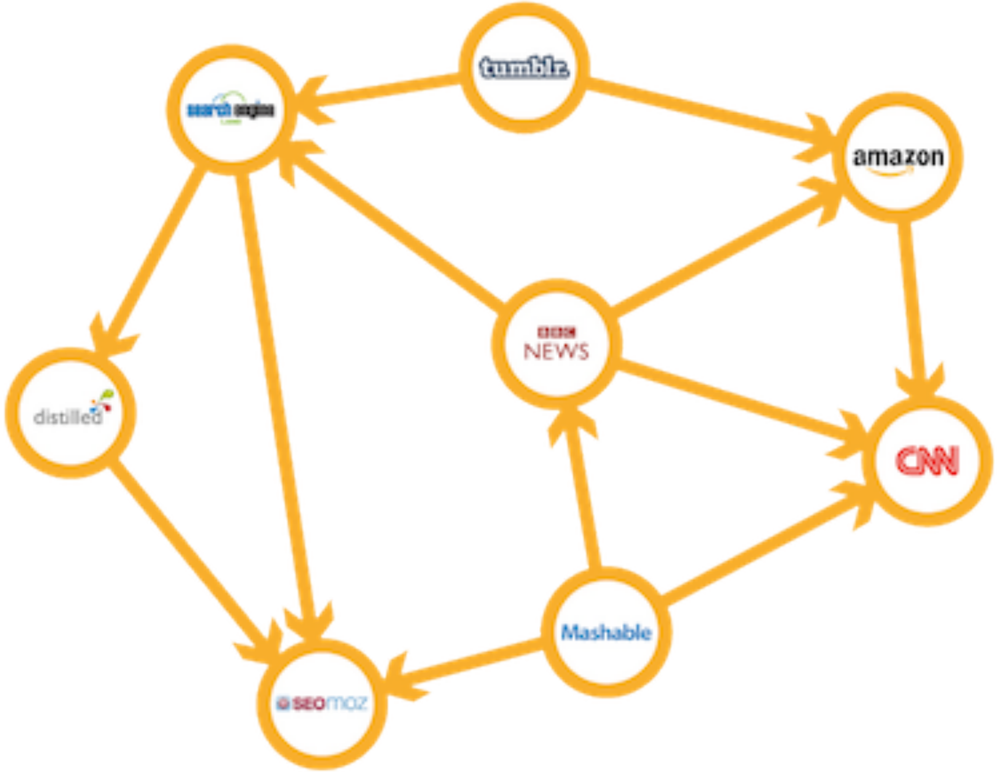
Each hyperlink is an edge.

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

If A has m outgoing edges, each gets label $1/m$

If A has no outgoing edges, put an edge A ---> B for all B (jump to a random page)

Google PageRank



<u>A little tweak:</u>

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

Google PageRank

Stationary distribution:

probability of being in state i in the long run

PageRank of a webpage

=

The stationary probability corresponding to the webpage

Google:

"*PageRank continues to be the heart of our software*"

How are Markov models applied ?

2 common types of applications

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

   Use the Markov model to simulate the process.

   e.g. text generation, music composition.

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

   e.g. Google PageRank, image segmentation

Outline

Motivating examples and applications

Basic mathematical representation and properties

Applications