

CSE 312

# Foundations of Computing II


## Lecture 23: Markov chains and Pagerank



**Aleks Jovicic**

Slide Credit: Based on Stefano Tessaro's slides for 312 19au incorporating ideas from Anna Karlin, Ryan O'Donnell, Alex Tsun, Rachel Lin, Hunter Schafer & myself 😊

# Agenda

- Markov Chains 
- Stationary Distributions
- Example
- Application: PageRank

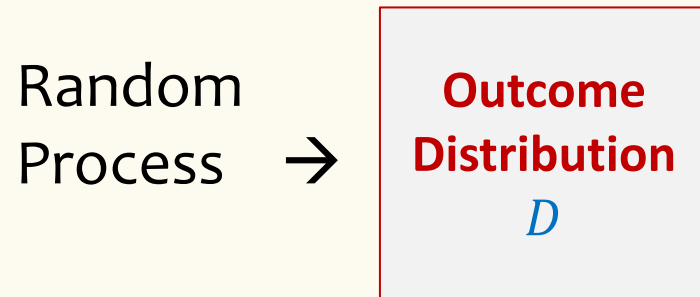
# So far, a single-shot random process

Random  
Process

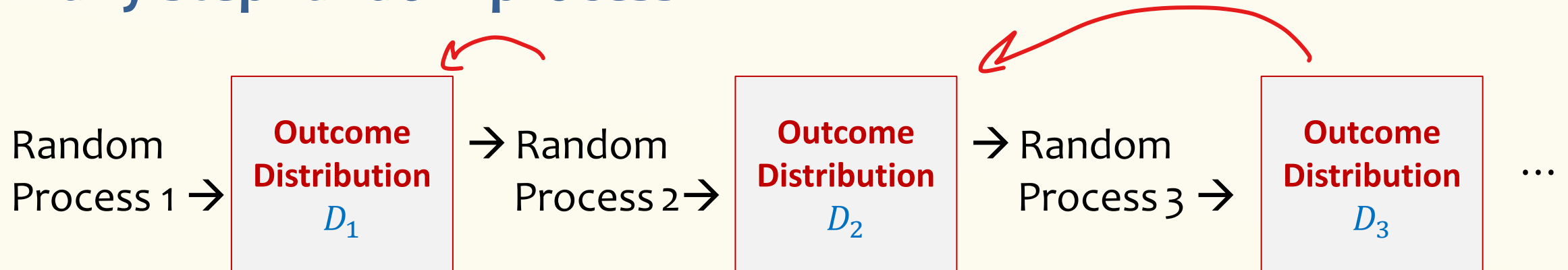


$D$   
-----  
0, 1, 2, 3  
^ ^ ^ ^

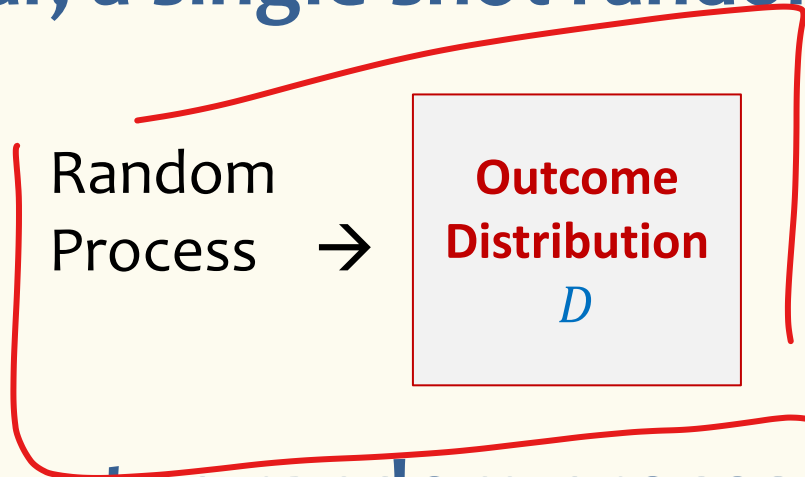
# So far, a single-shot random process



# Many-step random process

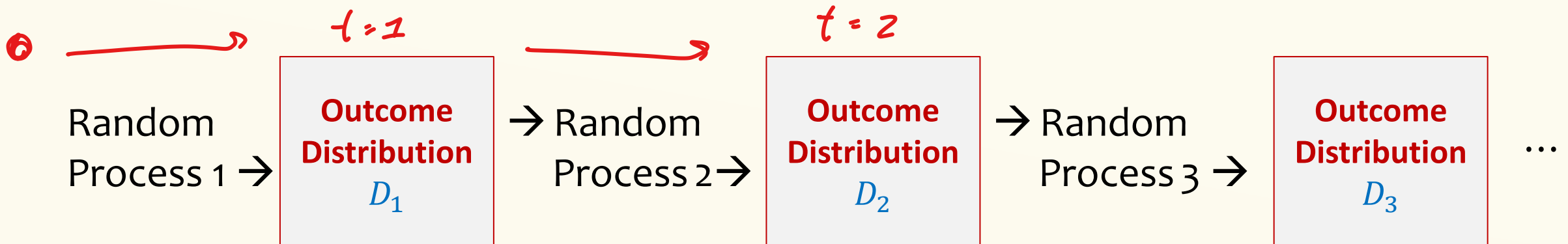


# So far, a single-shot random process



**Today:**  
see a very special type of DTSP  
Called a **Markov Chain**

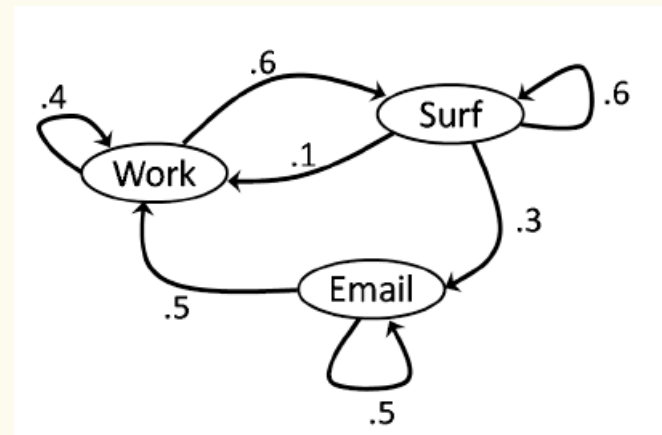
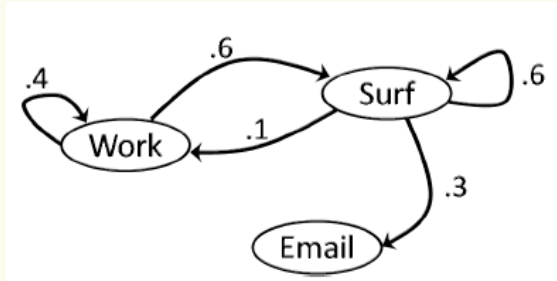
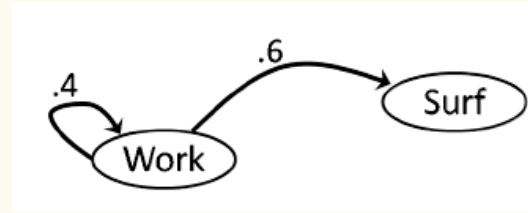
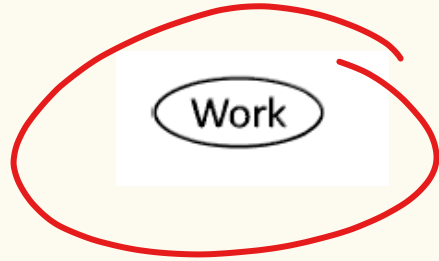
# Many-step random process



**Definition:** A discrete-time stochastic process (DTSP) is a sequence of random variables  $X^{(0)}, X^{(1)}, X^{(2)}, \dots$  where  $X^{(t)}$  is the value at time  $t$ .

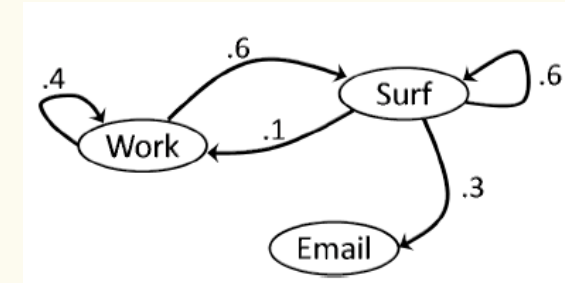
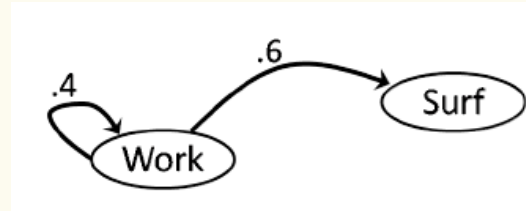
# A day in my life (I wish)

$X_0 = w$



# A day in my life (I wish)

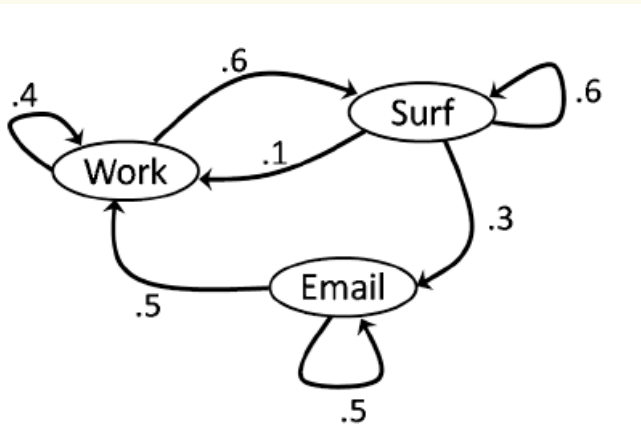
$t = 0$



This type of probabilistic finite automaton is called a **Markov Chain**. The next state depends only on the current state and not on the history.

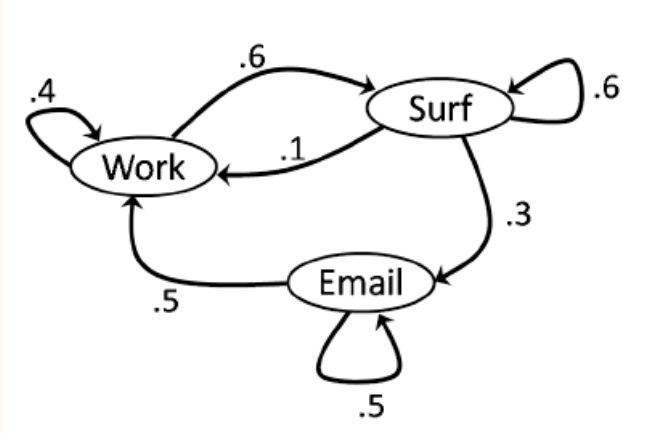
For ANY  $t \geq 0$ , if I was working at time  $t$ , then at  $t+1$  with probability 0.4 I continue working, with probability 0.6, I switch to surfing, and with probability 0, I switch to emailing.

This is called **History Independent** (similar to memoryless).



# A day in my life (I wish)

Many interesting questions.



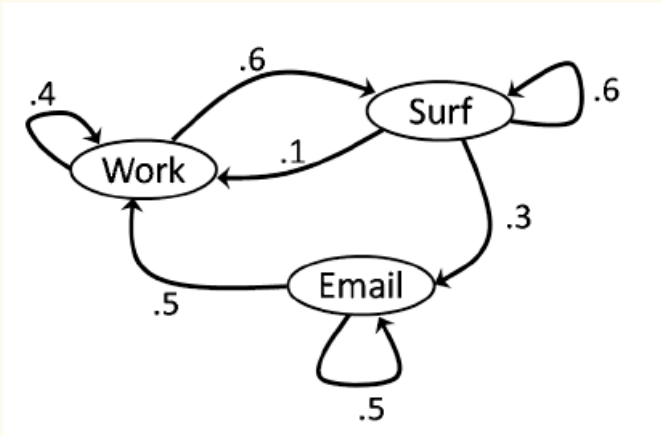
1. What is the probability that I work at time 1?
2. What is the probability that I work at time 2?

$X^{(t)}$  state I'm in at time  $t$  (random variable)

| $t$                                       | 0 | 1 | 2 |
|---|---|---|---|
| $q_w^{(t)} = \Pr(X^{(t)} = \text{work})$  |   |   |   |
| $q_s^{(t)} = \Pr(X^{(t)} = \text{surf})$  |   |   |   |
| $q_e^{(t)} = \Pr(X^{(t)} = \text{email})$ |   |   |   |



# A day in my life (I wish)

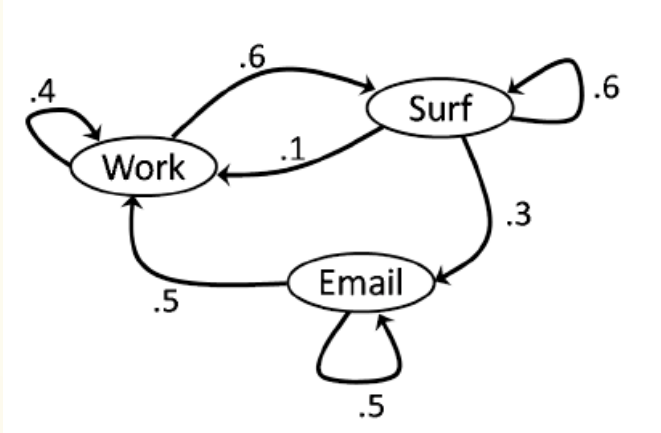


## Many interesting questions

1. What is the probability that I work at time 1?
2. What is the probability that I work at time 2?
3. What is the probability that I work at time  $t=100$ ?
4. What is the probability that I'm working at some random time far in the future?

# A day in my life (I wish)

What is the probability I'm in each state at time  $t$ , as a function of the probability distribution over states at time  $t-1$



$X^{(t)}$  state I'm in at time  $t$  (random variable)

$t-1$

$t$

$$q_W^{(t-1)} = \Pr(X^{(t-1)} = \text{work})$$

$$q_W^{(t)} =$$

$$q_S^{(t-1)} = \Pr(X^{(t-1)} = \text{surf})$$

$$q_S^{(t)} =$$

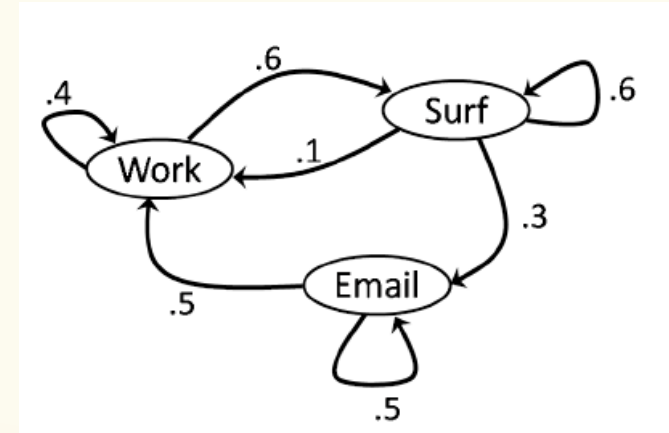
$$q_E^{(t-1)} = \Pr(X^{(t-1)} = \text{email})$$

$$q_E^{(t)} =$$

$$\begin{pmatrix} q_w^{(t)} & q_S^{(t)} & q_E^{(t)} \end{pmatrix} = \begin{pmatrix} q_w^{(t-1)} & q_S^{(t-1)} & q_E^{(t-1)} \end{pmatrix} \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

## Transition Probability Matrix

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$



$$\rightarrow \mathbf{q}^{(t)} = \mathbf{q}^{(t-1)} P$$

$$\mathbf{q}^{(t)} = (q_w^{(t)}, q_S^{(t)}, q_E^{(t)})$$

Apply  $q^{(t)} = q^{(t-1)} P$  inductively.

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$$\rightarrow q^{(t)} = q^{(0)} P^t$$

# The t-step walk $P^t$

Recall  $q^{(t)} = q^{(0)} P^t$

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$$P^2 = \begin{matrix} & W & S & E \\ W & (.22 & .6 & .18) \\ S & (.25 & .42 & .33) \\ E & (.45 & .3 & .25) \end{matrix}$$

$$P^3 = \begin{matrix} & W & S & E \\ W & (.238 & .492 & .270) \\ S & (.307 & .402 & .291) \\ E & (.335 & .450 & .215) \end{matrix}$$


$$P^{10} \approx \begin{matrix} & W & S & E \\ W & (.2940 & .4413 & .2648) \\ S & (.2942 & .4411 & .2648) \\ E & (.2942 & .4413 & .2648) \end{matrix}$$

$$P^{30} \approx \begin{matrix} & W & S & E \\ W & (.29411764705 & .44117647059 & .26470588235) \\ S & (.29411764706 & .44117647058 & .26470588235) \\ E & (.29411764706 & .44117647059 & .26470588235) \end{matrix}$$

$$P^{60} \approx \begin{matrix} & W & S & E \\ W & (.294117647058823 & .441176470588235 & .264705882352941) \\ S & (.294117647068823 & .441176470588235 & .264705882352941) \\ E & (.294117647068823 & .441176470588235 & .264705882352941) \end{matrix}$$

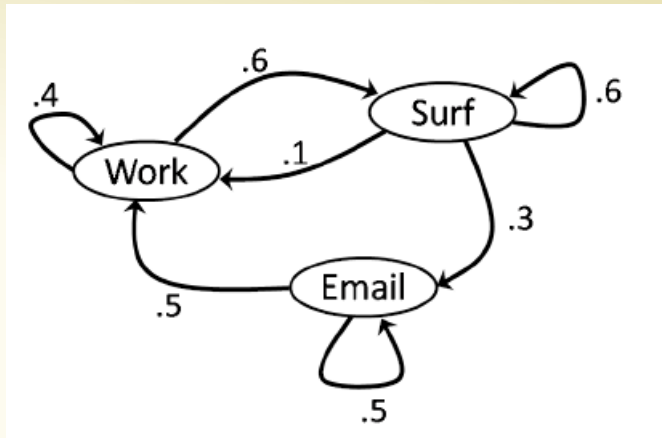
What does this say about  $q^{(t)}$ ?

# Agenda

- Markov Chains
- Stationary Distributions 
- Example
- Application: PageRank

## Observation

If  $q^{(t)} = q^{(t-1)}$  then it will never change again!

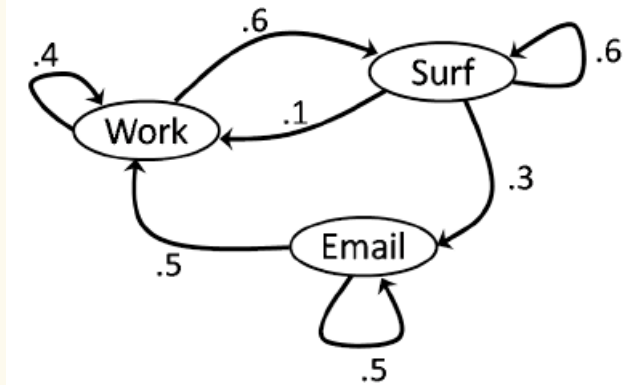


Called a “stationary distribution” and has a special name

$$\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$$

Solution to  $\boldsymbol{\pi} = \boldsymbol{\pi} P$

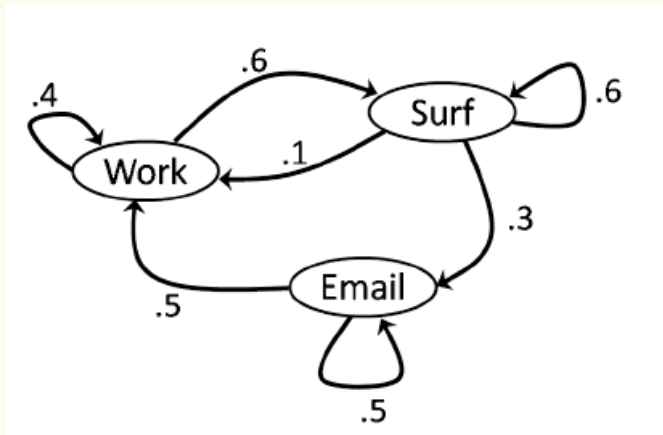
$$\begin{pmatrix} \pi_W & \pi_S & \pi_E \end{pmatrix} = \begin{pmatrix} \pi_W & \pi_S & \pi_E \end{pmatrix} \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$





# Solving for Stationary Distribution

$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$



*Stationary Distribution satisfies*

- $\boldsymbol{\pi} = \boldsymbol{\pi}P$ , where  $\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$
- $\pi_W + \pi_S + \pi_E = 1$

$$\Rightarrow \pi_W = \frac{15}{34}, \pi_S = \frac{10}{34}, \pi_E = \frac{9}{34}$$

$\Rightarrow$  As  $t \rightarrow \infty$ ,  $\mathbf{q}^{(t)} \rightarrow \boldsymbol{\pi}$  !!

# Markov Chains in general

- A set of  $n$  states  $\{1, 2, 3, \dots, n\}$
- The state at time  $t$  is denoted by  $X^{(t)}$
- A transition matrix  $\mathbf{P}$ , dimension  $n \times n$ 
$$P_{ij} = \Pr(X^{(t+1)} = j \mid X^{(t)} = i)$$
- $\mathbf{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \dots, q_n^{(t)})$  where  $q_i^{(t)} = \Pr(X^{(t)} = i)$
- Transition:  $\mathbf{LTP} \rightarrow \mathbf{q}^{(t)} = \mathbf{q}^{(t-1)} \mathbf{P} \implies \mathbf{q}^{(t)} = \mathbf{q}^{(0)} \mathbf{P}^t$
- A stationary distribution  $\pi$  is the solution to:
$$\pi = \pi \mathbf{P}, \text{ normalized so that } \sum_{i \in [n]} \pi_i = 1$$

# The Fundamental Theorem of Markov Chain

If a Markov chain is “irreducible” and “aperiodic”, then it has a unique stationary distribution.

Moreover, as  $t \rightarrow \infty$ , for all  $i, j$ ,  $\mathbf{P}_{ij}^t = \pi_j$

# Stationary Distribution of a Markov Chain

**Definition.** The **stationary distribution of a Markov Chain** with  $n$  states is the  $n$ -dimensional row vector  $\pi$  (which must be a probability distribution – nonnegative and sums to 1) such that

$$\pi P = \pi$$

Intuition: Distribution over states at next step is the same as the distribution over states at the current step

# Stationary Distribution of a Markov Chain


Intuition:  $\mathbf{q}^{(t)}$  is the distribution of being at each state at time  $t$  computed by  $\mathbf{q}^{(t)} = \mathbf{q}^{(0)} \mathbf{P}^t$ . As  $t$  gets large  $\mathbf{q}^{(t)} \approx \mathbf{q}^{(t+1)}$ .

**Theorem.** The **Fundamental Theorem of Markov Chains** says that (under some minor technical conditions), for a Markov Chain with transition probabilities  $\mathbf{P}$  and for any starting distribution  $\mathbf{q}^{(0)}$  over the states

$$\lim_{t \rightarrow \infty} \mathbf{q}^{(0)} \mathbf{P}^t = \boldsymbol{\pi}$$

where  $\boldsymbol{\pi}$  is the stationary distribution of  $\mathbf{P}$  (i.e.,  $\boldsymbol{\pi} \mathbf{P} = \boldsymbol{\pi}$ )

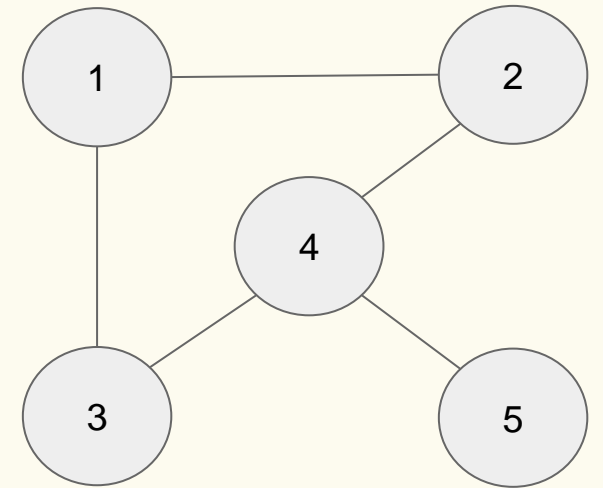
# Agenda

- Markov Chains
- Stationary Distributions
- **Example** 
- Application: PageRank

## Another Example: Random Walks

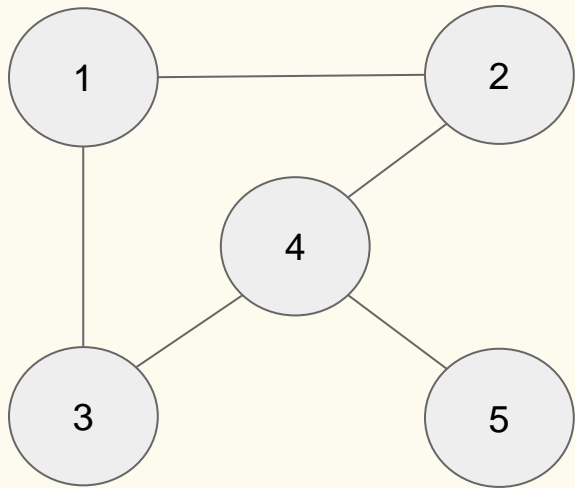
Suppose we start at node 1, and at each step transition to a neighboring node with equal probability.

This is called a "random walk" on this graph.



# Example: Random Walks

Start by defining transition probs.



$$P_{ij} = \Pr(X^{(t+1)} = j \mid X^{(t)} = i)$$

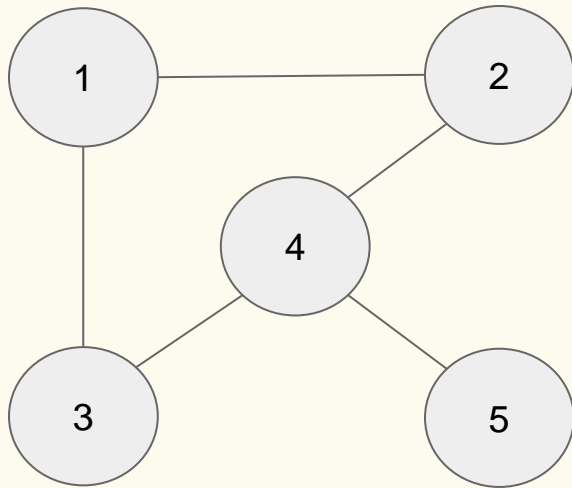
$$q_i^{(t)} = \Pr(X^{(t)} = i) = (q^{(0)} P^t)_i$$

|            | To $s_1$ | To $s_2$ | To $s_3$ | To $s_4$ | To $s_5$ |
|------------|----------|----------|----------|----------|----------|
| From $s_1$ |          |          |          |          |          |
| From $s_2$ |          |          |          |          |          |
| From $s_3$ |          |          |          |          |          |
| From $s_4$ |          |          |          |          |          |
| From $s_5$ |          |          |          |          |          |



# Example: Random Walks

Start by defining transition probs.



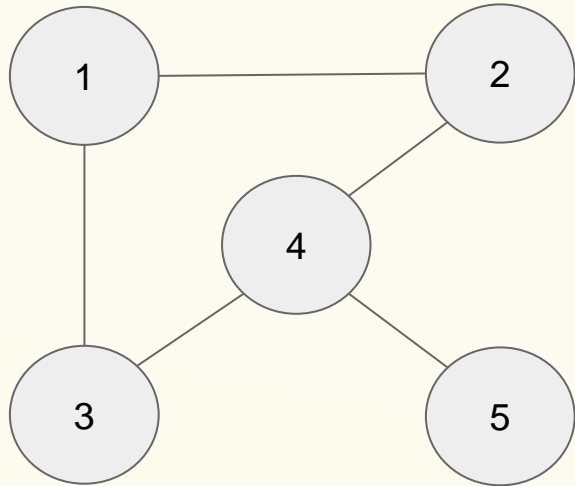
$$P_{ij} = \Pr(X^{(t+1)} = j \mid X^{(t)} = i)$$

$$q_i^{(t)} = \Pr(X^{(t)} = i) = (q^{(0)} P^t)_i$$

|            | To $s_1$ | To $s_2$ | To $s_3$ | To $s_4$ | To $s_5$ |
|------------|----------|----------|----------|----------|----------|
| From $s_1$ | 0        | 1/2      | 1/2      | 0        | 0        |
| From $s_2$ | 1/2      | 0        | 0        | 1/2      | 0        |
| From $s_3$ | 1/2      | 0        | 0        | 1/2      | 0        |
| From $s_4$ | 0        | 1/3      | 1/3      | 0        | 1/3      |
| From $s_5$ | 0        | 0        | 0        | 1        | 0        |


# Example: Random Walks

Suppose we know that  $X^{(0)} = 2$ . What is  $\Pr(X^{(2)} = 3)$ ?



$$\begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

# Agenda

- Markov Chains
- Stationary Distributions
- Example
- **Application: PageRank** 

# PageRank: Some History

The year was 1997

- Bill Clinton in the White House
- Deep Blue beat world chess champion (Kasparov)

The internet was not like it was today. Finding stuff was hard!

- In Nov 1997, only one of the top 4 search engines actually found itself when you searched for it

# The Problem

Search engines worked by matching words in your queries to documents.

Not bad in theory, but in practice there are lots of documents that match a query.

- Search for Bill Clinton, top result is ‘Bill Clinton Joke of the Day’
- Susceptible to spammers and advertisers

# The Fix: Ranking Results

- Start by doing filtering to relevant documents (with decent textual match).
- Then **rank** the results based on some measure of ‘quality’ or ‘authority’.

Key question: How to define ‘quality’ or ‘authority’?

Enter two groups:

- Jon Kleinberg (professor at Cornell)
- Larry Page and Sergey Brin (Ph.D. students at Stanford)

## Both groups had the same brilliant idea

Larry Page and Sergey Brin (Ph.D. students at Stanford)

- Took the idea and founded Google, making billions



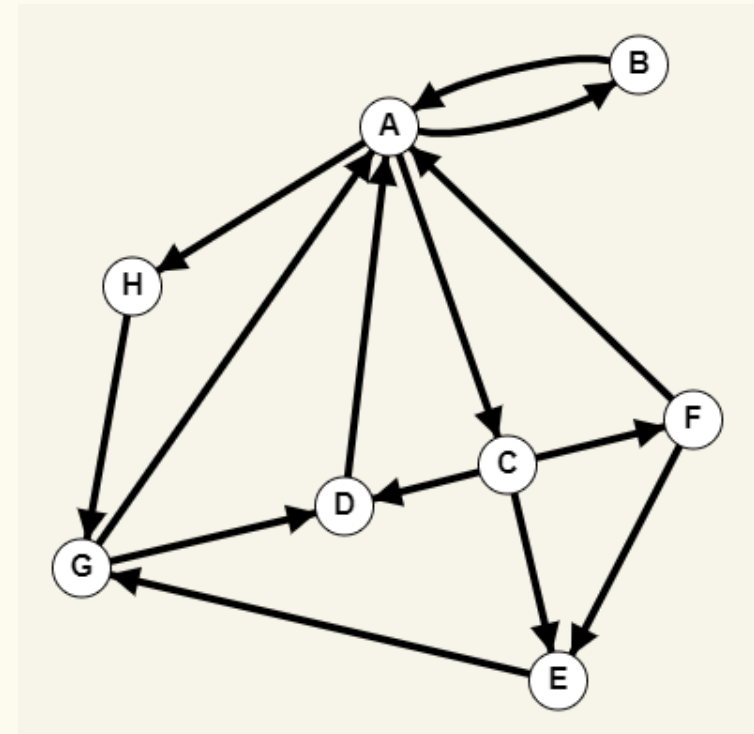
Jon Kleinberg (professor at Cornell)

- MacArthur genius prize, Nevanlinna Prize and many other academic honors



## PageRank - Idea

Take into account directed graph structure of the web. Use **hyperlink analysis** to compute what pages are high quality or have high authority. Trust the internet itself define what is useful via its links.

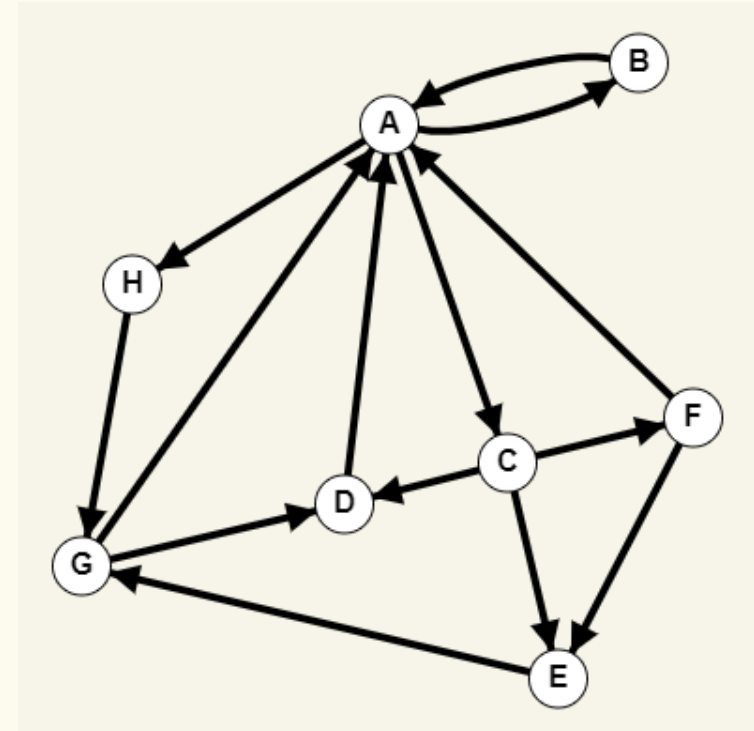




# PageRank - Idea

Idea 1: think of each link as a citation “vote of quality”

Rank pages by in-degree?



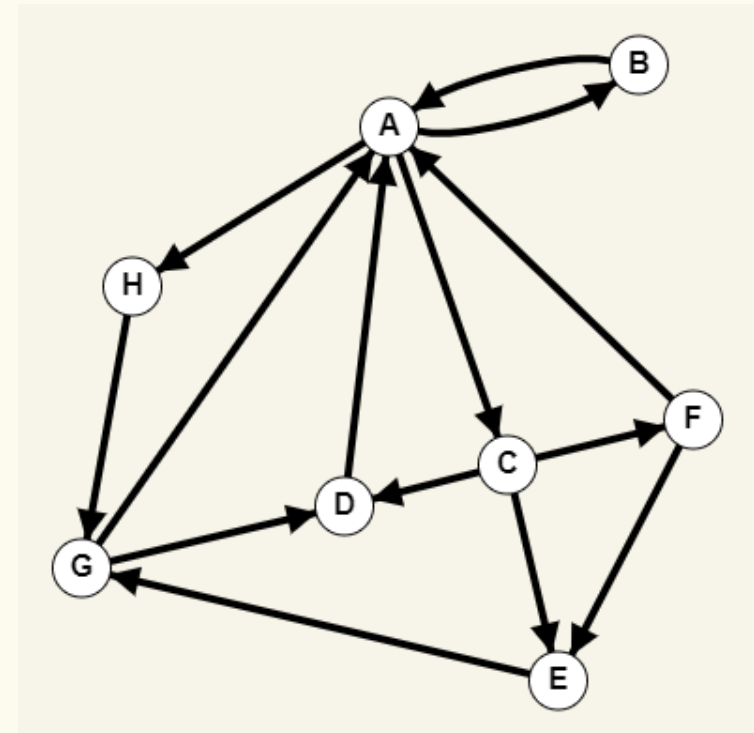
# PageRank - Idea

Idea 1: think of each link as a citation “vote of quality”

Rank pages by in-degree?

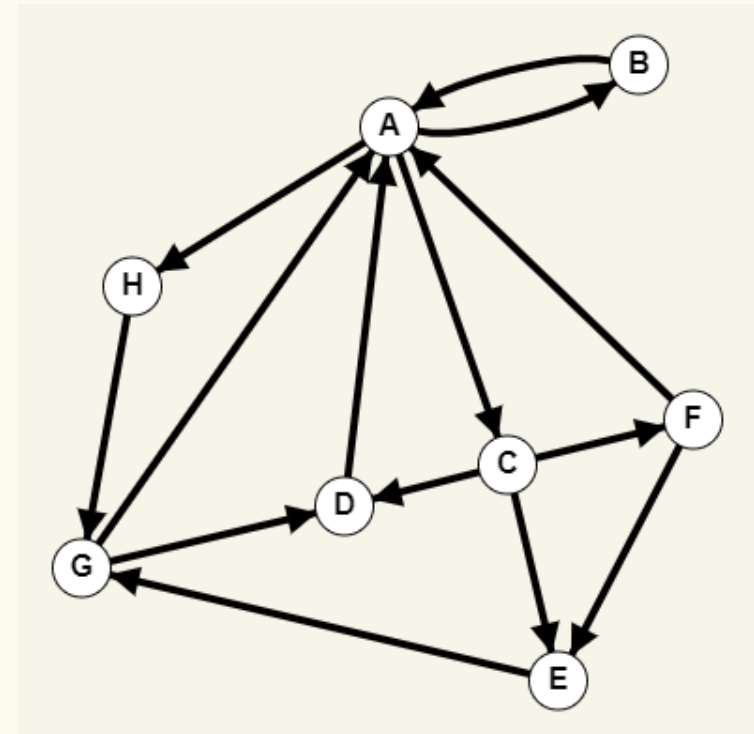
Problems:

- Spamming
- Some linkers not discriminating
- Not all links created equal



# PageRank - Idea

Idea 2: perhaps we should weight the links somehow and then use the weights of the in-links to rank pages



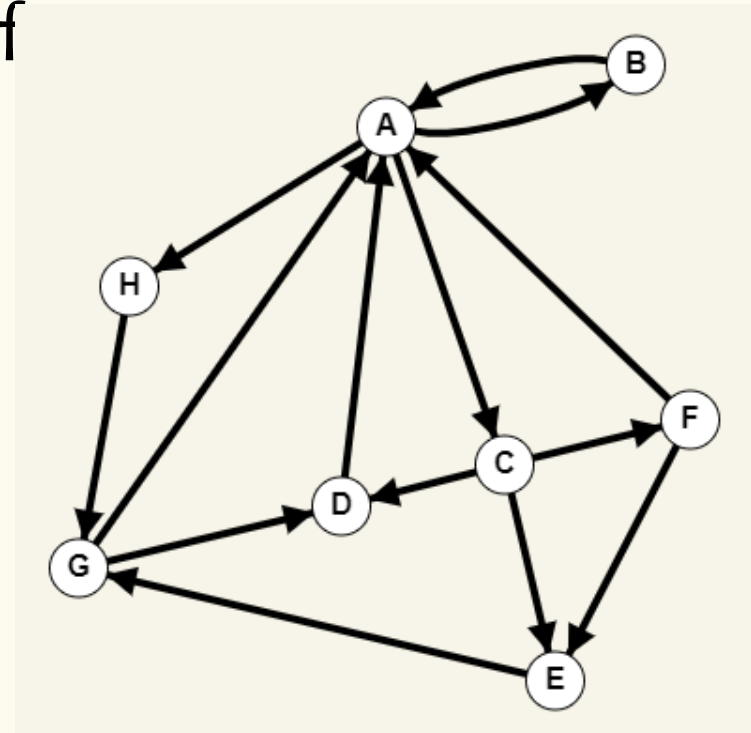
# Inching towards Pagerank



Web page has high quality if it's linked to by lots of high quality pages.

A page is high quality if it links to lots of high quality pages

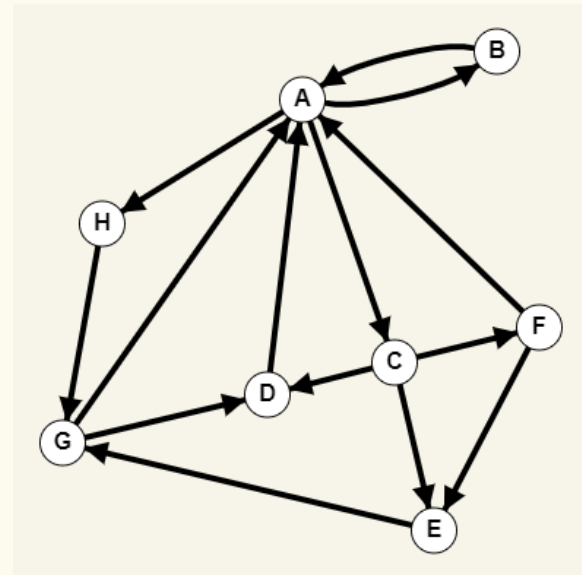
recursive definition!



# Inching towards Pagerank



- If web page  $x$  has  $d$  outgoing links, one of which goes to  $y$ , this contributes  $1/d$  to the importance of  $y$ .
- But we want to take into account the importance of  $x$ .



## Gives the following equations

Idea: Use the transition matrix defined by a random walk on the web  $P$  to compute quality of webpages. Namely, find  $q$  such that

$$qP = q$$



Look familiar?

This is the stationary distribution for the Markov chain defined by a random surfer. Starts at some node (webpage) and randomly follows a link to another.

- Use stationary distribution of her surfing patterns after a long time as notion of quality

## Issues with PageRank

- How to handle dangling nodes (dead ends)?
- How to handle Rank sinks – group of pages that only link to each other?

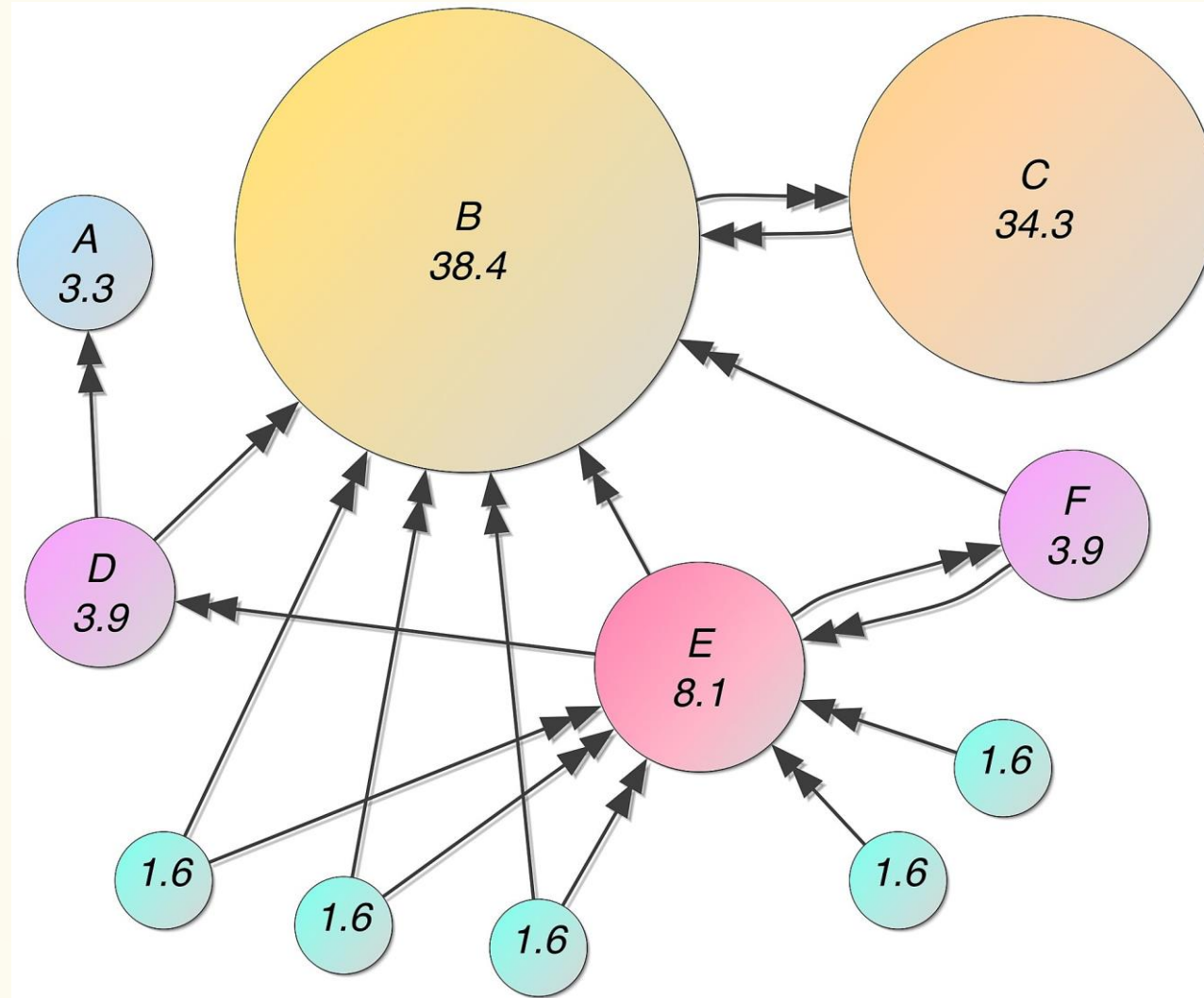
Both solutions can be solved by “teleportation”

# Final PageRank Algorithm

- Make a Markov Chain with one state for each webpage on the internet with the transition probabilities  $P_{ij} = \frac{1}{outdeg(i)}$ .
- Use a modified random walk. At each point in time, if the surfer is at some webpage  $x$ .
  - With probability  $p$ , take a step to one of the neighbors of  $x$  (equally likely)
  - With probability  $1 - p$ , “teleport” to a uniformly random page in the whole internet.
- Compute stationary distribution  $\pi$  of this perturbed Markov chain.
- Define the PageRank of a webpage  $i$  as the stationary probability  $\pi_i$ .
- Find all pages with decent textual match to search and then order those pages by PageRank!



# PageRank - Example



## It Gets More Complicated

While this basic algorithm was the defining idea that launched Google on their path to success, this is far from the end to optimizing search.

Nowadays, Google has a LOT more secret sauce to ranking pages most of which they don't reveal for 1) competitive advantage and 2) avoid gaming their algorithm.