Welcome! Ask Qs or say hi in chat before/during/after class!

CSE 312

Foundations of Computing II

Lecture 27: Markov Chains & PageRank

PAUL G. ALLEN SCHOOL OF COMPUTER SCIENCE & ENGINEERING

Rachel Lin, Hunter Schafer

Slide Credit: Based on Stefano Tessaro's slides for 312 19au incorporating ideas from Alex Tsun's and Anna Karlin's slides for 312 20su and 20au

Music: Kishi Bashi

Agenda

- Recap: Markov Chains
  - Intuition
  - Computing Probabilities
  - Matrix Notation
- Stationary Distributions
- PageRank

So far, a single-shot random process

Random Process → **Outcome Distribution** D

Last time / Today :
See a very special type of DTSP called **Markov Chains**

Many-step random process

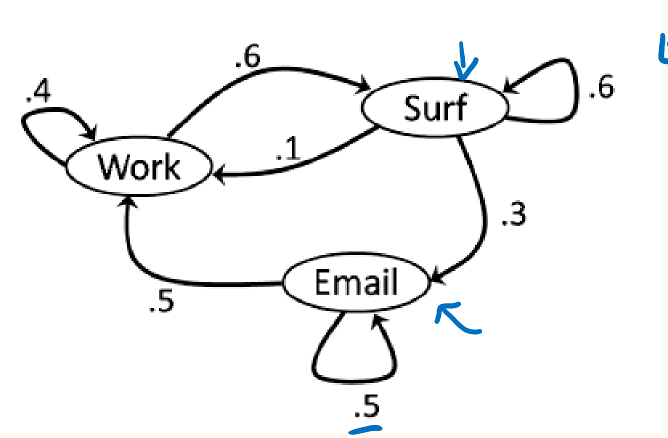
Random Process 1 → **Outcome Distribution** D_1 → Random Process 2 → **Outcome Distribution** D_2 → Random Process 3 → **Outcome Distribution** D_3 ⋯

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

Formalizing Markov Chain

Markov Property

$Pr(X^{(t+1)} = j \mid X^{(t)} = i, \text{history}) = Pr(X^{(t+1)} = j \mid X^{(t)} = i)$

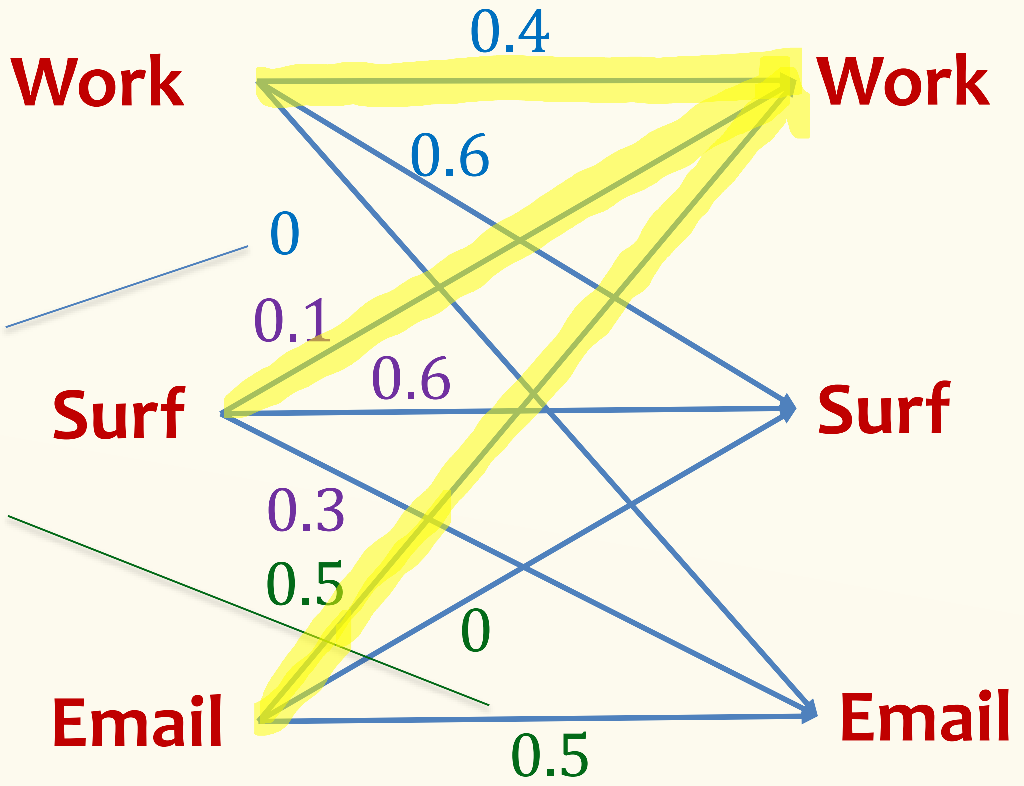


$\boldsymbol{X^{(t)}}$ my state at t $\boldsymbol{X^{(t+1)}}$ my state at t+1

$p_{WE} = \mathrm{P}(X^{(t+1)} = E \mid X^{(t)} = W) = 0$

$p_{EE} = \mathrm{P}(X^{(t+1)} = E \mid X^{(t)} = E) = 0.5$

From To



3. What is the prob that I work at t+1 = 100?

By LTP: $p_W^{(t+1)} = P(X^{(t+1)} = W) = \Sigma_{U \in \{W,S,E\}} \mathrm{P}(X^{(t+1)} = W \mid X^{(t)} = U)\, \mathrm{P}(X^{(t)} = U)$

Transition Matrix

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

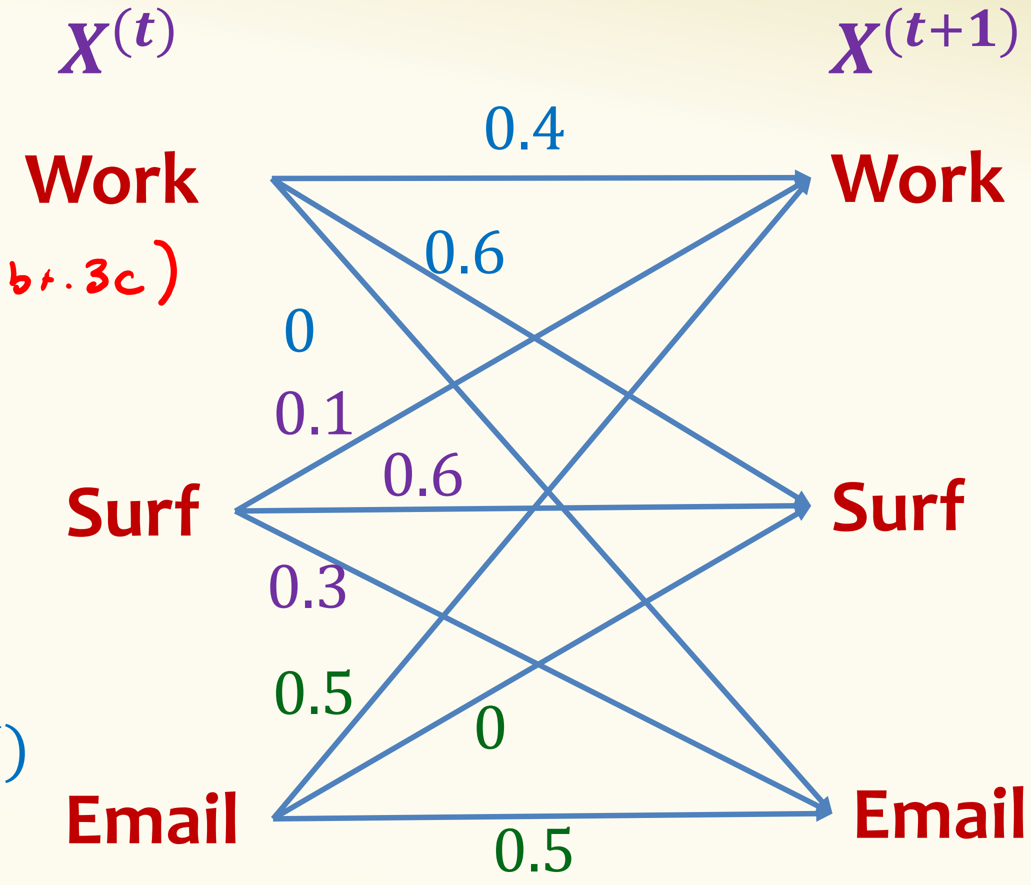
$$P = \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

$X^{(t)} = (a, b, c)$

$X^{(t+1)} = (a, b, c)\begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$

$= (.4a + .1b + .5c,\ .6a + .6b,\ .3b + .3c)$

$$X^{(t)} = (p_W^{(t)} \quad p_S^{(t)} \quad p_E^{(t)})$$



LTP: $p_W^{(t+1)} = P\left(X^{(t+1)} = W\right)$

$= \Sigma_{U \in \{W,S,E\}} \mathrm{P}\left(X^{(t+1)} = W \middle| \mathrm{X}^{(t)} = U\right) \mathrm{P}(\mathrm{X}^{(t)} = \mathrm{U})$

➔ $X^{(t+1)} = X^{(t)}\, P$

➔ $X^{(t)} = X^{(0)}\, P^t$

3. What is the prob that I work at t = 100?

Closed formula: $p_W^{(t)} = X^{(t)}[1] = (X^{(0)}\, P^t)\,[1]$

Agenda

- Recap: Markov Chains
  - Intuition
  - Computing Probabilities
  - Matrix Notation
- Stationary Distributions
- PageRank

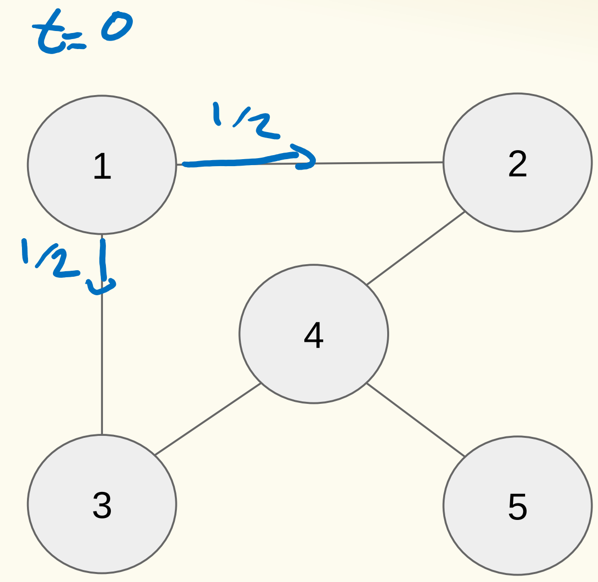
Example: Random Walks

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

How does the probability of me being at each node look as we let this process goes on?

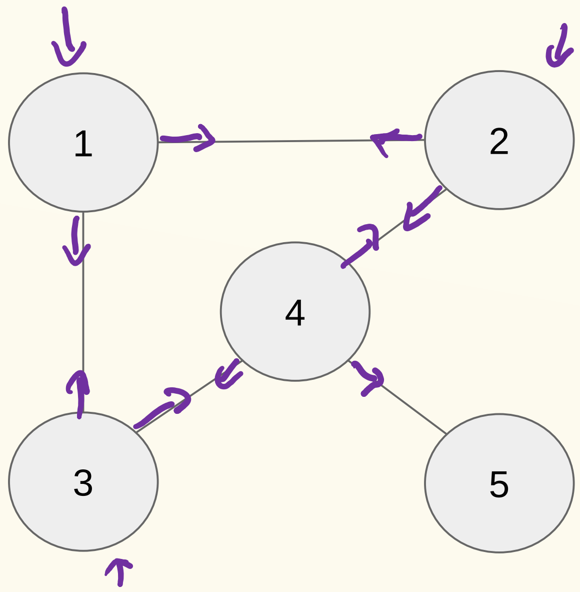
$$\Pr(X^{(t+1)}=2 \mid X^{(t)}=1)=1/2$$

$$\Pr(X^{(t+1)}=3 \mid X^{(t)}=1)=1/2$$



Example: Random Walks

Start by defining transition probs.



	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

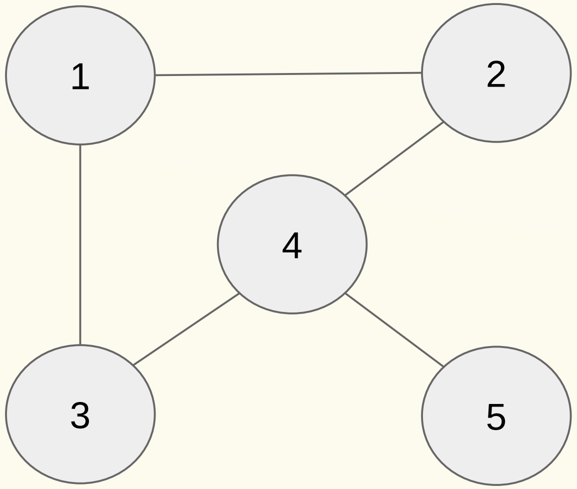
$$P_{ij} = P[i,j] = \Pr\left(X^{(t+1)} = j \mid X^{(t)} = i\right)$$

$$p_i^{(t)} = \Pr\left(X^{(t)} = i\right) = \text{[illegible]}$$

$$= X^{(t)}[i] = \left(X^{(0)} P^t\right)[i]$$

Example: Random Walks

Start by defining transition probs.



	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

$$P_{ij} = P[i,j] = \Pr\left(X^{(t+1)} = j \mid X^{(t)} = i\right)$$

$$p_i^{(t)} = \Pr\left(X^{(t)} = i\right) =$$

$$= X^{(t)}[i] = \left(X^{(0)} P^t\right)[i]$$

Example: Random Walks

Compute $\Pr\left(X^{(2)} = 3 \mid X^{(0)} = 2\right) =$ $\sum_{i=1}^{5} \Pr(X^{(2)}=3 \mid X^{(0)}=2, X^{(1)}=i) \Pr(X^{(1)}=i)$ [LTP]

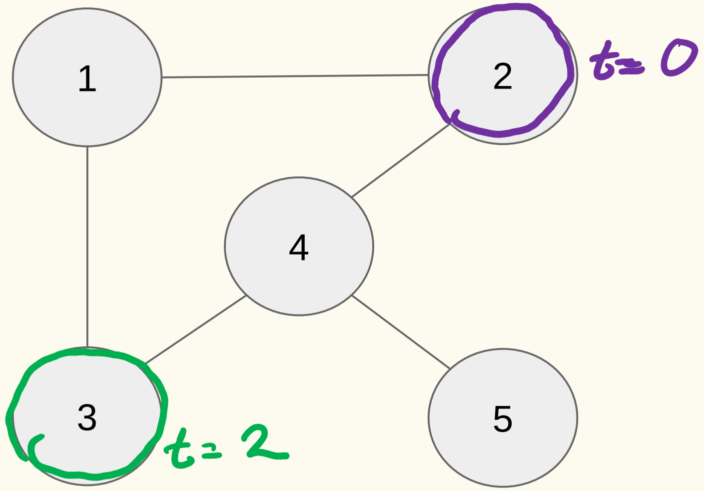
$= \sum_{i=1}^{5} \Pr(X^{(2)}=3 \mid X^{(1)}=i) \Pr(X^{(1)}=i)$ [Markov prop.]

$= \Pr(X^{(2)}=3 \mid X^{(1)}=1) \Pr(X^{(1)}=1) + \Pr(X^{(2)}=3 \mid X^{(1)}=4) \Pr(X^{(1)}=4)$

$= P_{13} \Pr(X^{(1)}=1) + P_{43} P(X^{(1)}=4)$

$= P_{13} P_{21} + P_{43} P_{24}$

$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{12}$



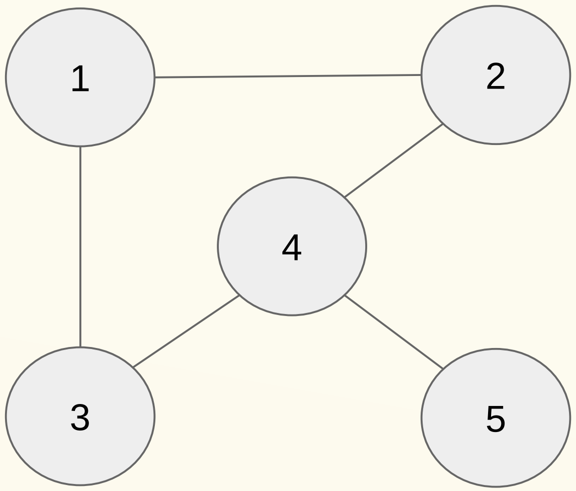
$$\begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

. . . . –

$P(X^{(1)}=1) = \underbrace{P(X^{(1)}=1 \mid X^{(0)}=2)}_{P_{21}} \underbrace{\Pr(X^{(0)}=2)}_{1}$

Example: Random Walks

Compute $\Pr(X^{(2)} = 3 \mid X^{(0)} = 2)$



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

$X^{(0)} = (0,1,0,0,0)$

$$X^{(1)} = \underbrace{(0, 1, 0, 0, 0)}_{X^{(0)}} \underbrace{\begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}}_{P}$$

$= (\frac{1}{2}, 0, 0, \frac{1}{2}, 0)$ [using matrix multiply]

$X^{(2)} = X^{(1)}P = (1/2, 0, 0, 1/2, 0) \cdot (\ldots P \ldots)$

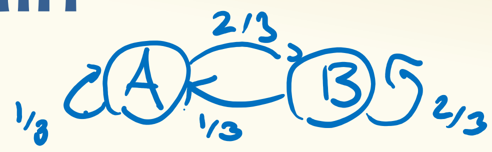
$= (0, \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{3}, \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{3}, 0, \frac{1}{2}\cdot\frac{1}{3})$

$= (0, 5/12, 5/12, 0, 1/6)$

$\Pr(X^{(2)} = 3 \mid X^{(0)} = 2)$

Stationary Distribution of a Markov Chain

Example:



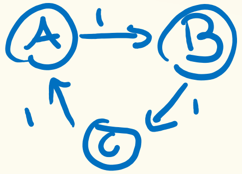
$\Rightarrow \pi = [\frac{1}{3}, \frac{2}{3}]$ (A, B)

Definition. The **stationary distribution of a Markov Chain** with n states (which doesn't always exist), is the n-dimensional row vector π (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

Example:



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

$\pi P = \pi$ if $\pi = [\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$

can verify

Stationary Distribution of a Markov Chain

Example of triangle graph from last slide does not meet def for convergence!

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

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

$$\lim_{t \to \infty} X^{(0)} P^t = \pi$$

where π is the stationary distribution of P (i.e., $\pi P = \pi$)

Brain Break

Agenda

- Recap: Markov Chains
  - Intuition
  - Computing Probabilities
  - Matrix Notation
- Stationary Distributions
- 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 (that part is easier). Then **rank** the results based on some measure of ‘quality’ or ‘authority’.

Key question: Who defines ‘quality’ or ‘authority’?

Enter two groups:

- Jon Kleinberg (professor at Cornell, MacArthur Genius Prize)
- Larry Page and Sergey Brin (Ph.D. students at Stanford, founded Google)

PageRank - Idea

~~Idea: Rank by in-degree~~

Use **hyperlink analysis** to compute what pages are high quality or have high authority. Trust the internet itself define what is useful.

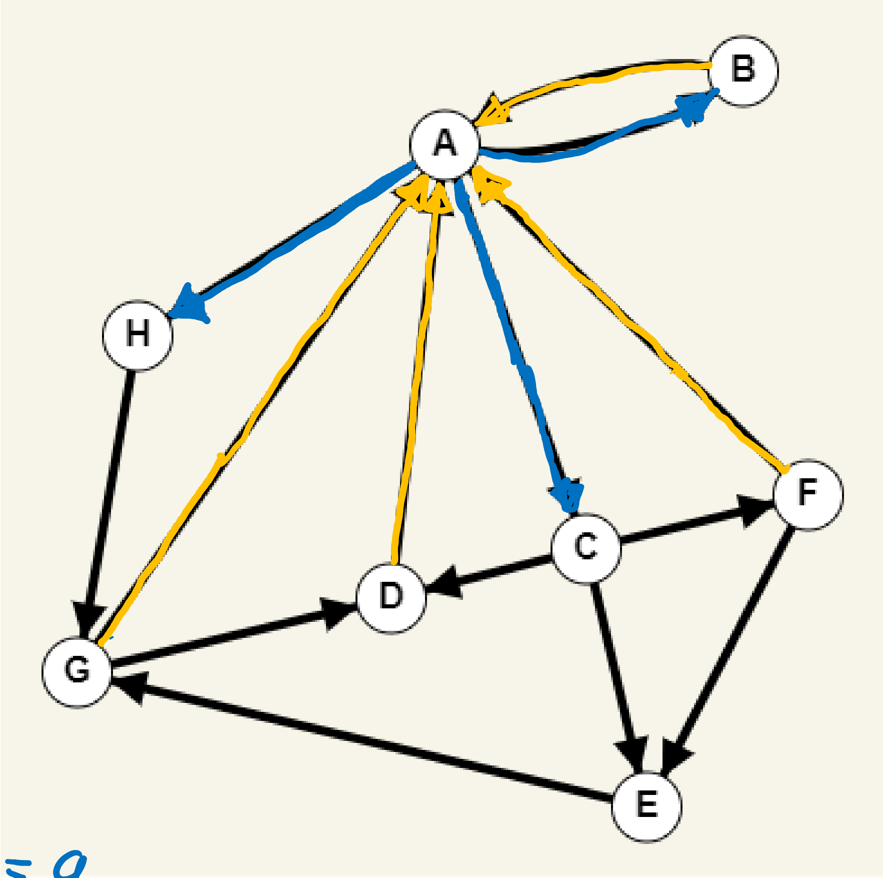
Define q to be quality vector s.t. q_A = quality of A

$q_A = \frac{1}{2} q_G + 1 q_D + \frac{1}{2} q_F + 1 q_B$

$q_H = \frac{1}{3} q_A$

Define $P_{ij} = \begin{cases} 1/\text{outdeg}(i) & \text{if } i \to j \\ 0 & \text{otherwise} \end{cases}$

$\Rightarrow$ Find q s.t. $qP = q$



PageRank

Idea: Use this transition matrix P to compute quality of webpages. Namely, find q such that

$$qP = q$$

Seems like trying to find the stationary distribution of a Markov chain? Where is the Markov chain here? 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

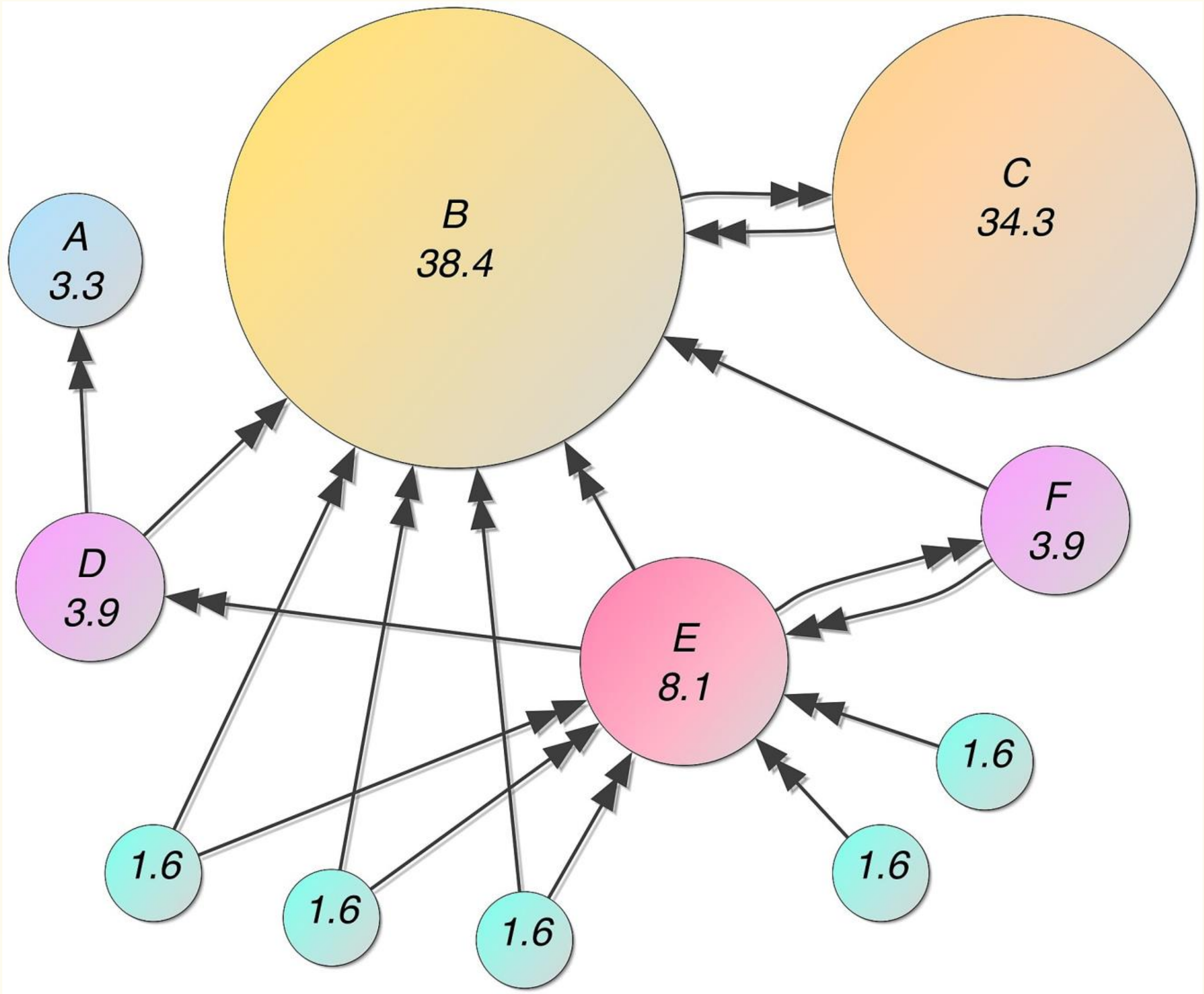
- How to handle dangling nodes (dead ends)?
- How to handle Rank sinks – group of pages that only link to each other?

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, 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 π of this perturbed Markov chain.
- Define the PageRank of a webpage x as the stationary probability π_x.
- Order pages by PageRank

PageRank - Example



It Gets More Complicated

While this basic algorithm was the defining thing that launched Google on their path to success, this is not 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.