CSE 312

Foundations of Computing II

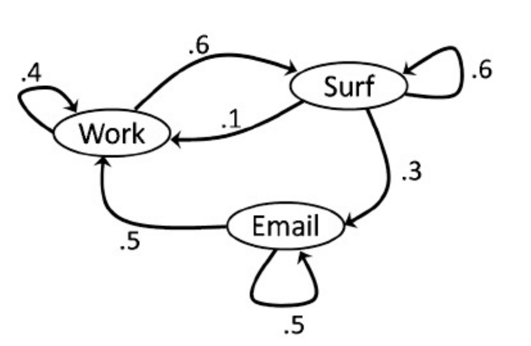
Lecture 25: Markov chains and Pagerank

Agenda

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

Review Markov chain example

How do we interpret this diagram?



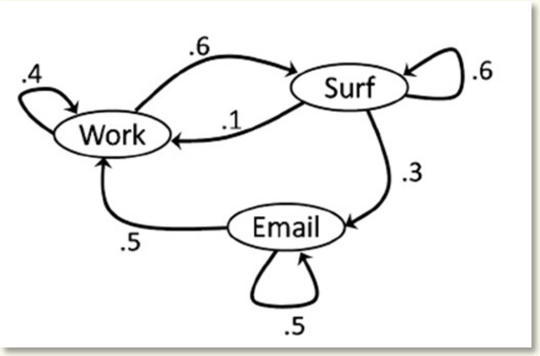
This kind of random process is called a **Markov Chain**

At *each* time step t

- Can be in one of 3 **states**
  - Work, Surf, Email
- If I am in some state s at time t
  - the **labels of out-edges** of s **give the probabilities** of moving to each of the states at time $t+1$ (as well as staying the same)
    - so **labels on out-edges sum to** 1

e.g. If in Email, there is a 50-50 chance it will be in each of Work or Email at the next time step, but it will never be in state Surf in the next step.

Review Transition Probability Matrix and distribution of $X^{(t)}$



$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}} = [q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}]$$

Vector-matrix multiplication

$\boldsymbol{M}$ is the Transition Probability Matrix

Probability vector for state variable $X^{(t)}$ at time t: $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$

For all $t \geq 0$, $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\boldsymbol{M}$

Equivalently, $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t$ for all $t \geq 0$

Review Finite Markov Chains

- A set of n **states** $\{1, 2, 3, \ldots\ n\}$
- The state at time t is denoted by $X^{(t)}$
- A **transition matrix** $\boldsymbol{M}$, dimension $n \times n$

$$\boldsymbol{M}_{ij} = P\left(X^{(t+1)} = j \;\middle|\; X^{(t)} = i\right)$$

- $\boldsymbol{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$ where $q_i^{(t)} = P(X^{(t)} = i)$
- Transition: LTP $\Rightarrow \boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\, \boldsymbol{M}$ so $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\, \boldsymbol{M}^t$
- A **stationary distribution** $\boldsymbol{\pi}$ is the solution to:

$\boldsymbol{\pi} = \boldsymbol{\pi}\, \boldsymbol{M}$, normalized so that $\Sigma_{i\in[n]}\pi_i = 1$

Review Stationary Distribution of a Markov Chain

Definition. The **stationary distribution of a Markov Chain** with n states is the n-dimensional row vector $\boldsymbol{\pi}$ (which must be a probability distribution; that is, it must be nonnegative and sum to 1) such that

$$\boldsymbol{\pi M} = \boldsymbol{\pi}$$

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

Review Stationary Distribution of a Markov Chain

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

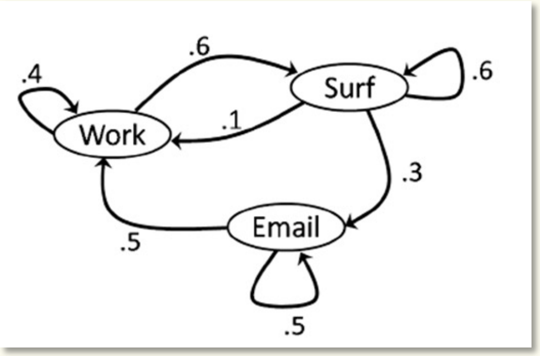
Fundamental Theorem of Markov Chains : For a Markov Chain that is aperiodic* and irreducible*, with transition probabilities $\boldsymbol{M}$ and for any starting distribution $\boldsymbol{q}^{(0)}$ over the states

$$\lim_{t\to\infty} \boldsymbol{q}^{(0)}\boldsymbol{M}^t = \boldsymbol{\pi}$$

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

**These concepts are way beyond us but they turn out to cover a very large class of Markov chains of practical importance.*

Computing the Stationary Distribution



$$[\pi_W, \pi_S, \pi_E] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} = [\pi_W, \pi_S, \pi_E]$$

Stationary Distribution satisfies

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

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

How did we get this?

Solve system of equations:

$$\begin{cases} 0.4 \cdot \pi_W + 0.1 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_W \\ 0.6 \cdot \pi_W + 0.6 \cdot \pi_S = \pi_S \\ 0.3 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_E \end{cases}$$

$$\pi_W + \pi_S + \pi_E = 1$$

Computing the Stationary Distribution

$$[\pi_W, \pi_S, \pi_E] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} = [\pi_W, \pi_S, \pi_E]$$

Solve system of equations:

(1) $0.4 \cdot \pi_W + 0.1 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_W$

$0.6 \cdot \pi_W + 0.6 \cdot \pi_S = \pi_S$ ⇒ $0.6 \cdot \pi_W = 0.4 \cdot \pi_S$ (2)

$0.3 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_E$ ⇒ $0.3 \cdot \pi_S = 0.5 \cdot \pi_E$ (3)

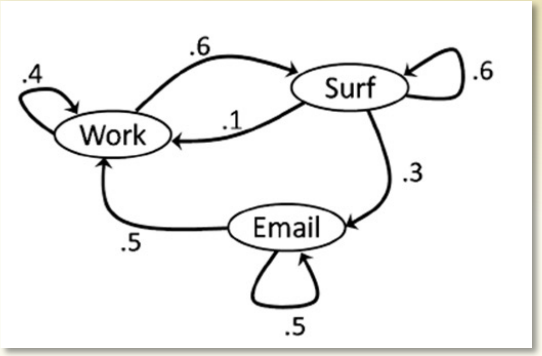
$\pi_W + \pi_S + \pi_E = 1$

But more equations than unknowns ???

One of the equations for $\boldsymbol{\pi} = \boldsymbol{\pi M}$ always depends on the others:

From (3) we get
$0.1 \cdot \pi_S + 0.5 \cdot \pi_E = 0.4 \cdot \pi_S$
so (1) becomes equivalent to (2)

Computing the Stationary Distribution



$$[\pi_W, \pi_S, \pi_E] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} = [\pi_W, \pi_S, \pi_E]$$

Solve system of equations:

Choose the simplest ones to work with

$$0.4 \cdot \pi_W + 0.1 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_W$$

$$0.6 \cdot \pi_W + 0.6 \cdot \pi_S = \pi_S$$

$$0.3 \cdot \pi_S + 0.5 \cdot \pi_E = \pi_E$$

$$\pi_W + \pi_S + \pi_E = 1$$

But more equations than unknowns ???

$\Rightarrow$ $0.6 \cdot \pi_W = 0.4 \cdot \pi_S$ $\Rightarrow$ $\pi_W = (2/3)\, \pi_S$

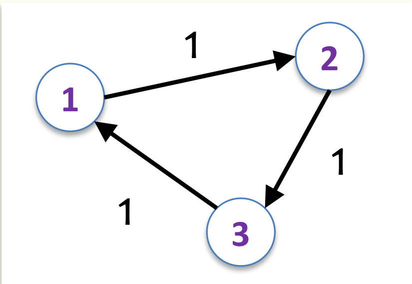
$\Rightarrow$ $0.3 \cdot \pi_S = 0.5 \cdot \pi_E$ $\Rightarrow$ $(3/5) \cdot \pi_S = \pi_E$

$\Rightarrow$ $(2/3)\, \pi_S + \pi_S + (3/5) \cdot \pi_S = 1$

$\Rightarrow$ $((10 + 15 + 9)/15) \cdot \pi_S = 1$

$\Rightarrow$ $\pi_S = 15/34$

Another stationary distribution example

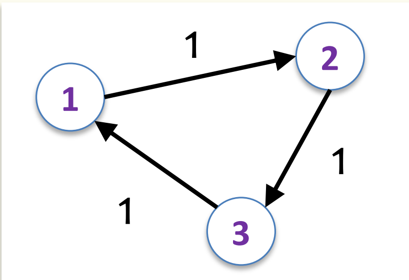


$$\boldsymbol{M} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Poll: pollev.com/rachel312
The stationary distribution for this Markov Chain …

a. is [1/3,1/6,1/2]
b. is [1/2,1/2,1/2]
c. is [1/2,1/4,1/4]
d. is [1/3,1/3,1/3]
e. doesn't exist

Another stationary distribution example



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

This is an example of a Markov Chain that is periodic and does not converge

Poll: pollev.com/rachel312

The stationary distribution for this Markov Chain …

a. is [1/3,1/6,1/2]
b. is [1/2,1/2,1/2]
c. is [1/2,1/4,1/4]
d. is [1/3,1/3,1/3]
e. doesn't exist

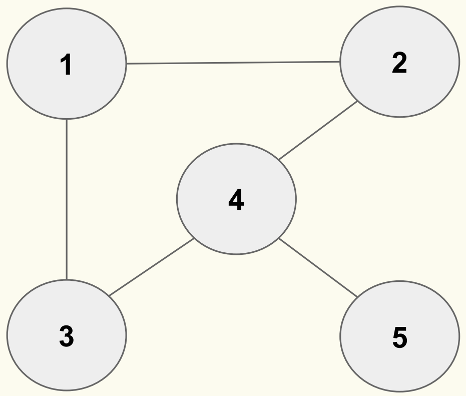
If $X^{(0)} = 1$ then the sequence of states $X^{(0)}, X^{(1)}, X^{(2)}, \dots$ is always

1,2,3,1,2,3,1,2,3,1,2,3,1,2,3,1,2,3, …

Another Example: Random Walks

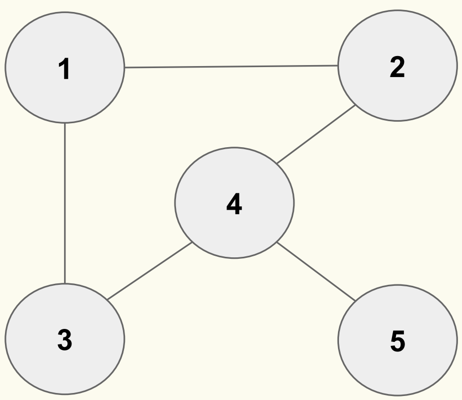
Suppose we start at node 1, and at each step transition to a neighboring node with equal probability.

This is called a “random walk” on this graph.



Example: Random Walks on an Undirected Graph

Start by defining transition probs.



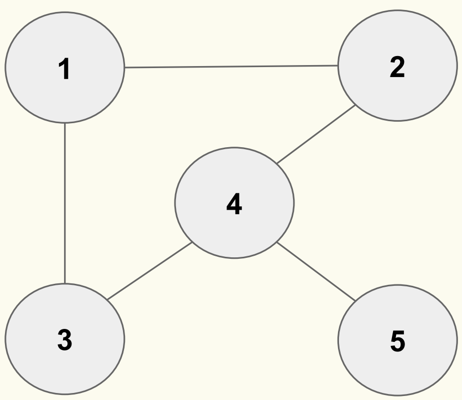
	To 1	To 2	To 3	To 4	To 5
From **1**					
From **2**					
From **3**					
From **4**					
From **5**					

$$\boldsymbol{M}_{ij} = P\left(X^{(t+1)} = j \mid X^{(t)} = i\right)$$

$$\boldsymbol{q}_i^{(t)} = P\left(\mathrm{X}^{(t)} = i\right) = \left(\boldsymbol{q}^{(0)}\boldsymbol{M}^t\right)_i$$

Example: Random Walks on an Undirected Graph

Start by defining transition probs.



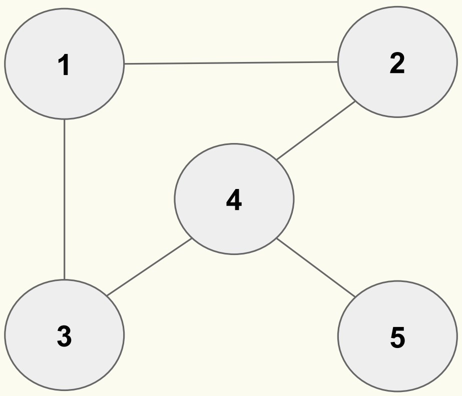
	To 1	To 2	To 3	To 4	To 5
From **1**	0	1/2	1/2	0	0
From **2**	1/2	0	0	1/2	0
From **3**	1/2	0	0	1/2	0
From **4**	0	1/3	1/3	0	1/3
From **5**	0	0	0	1	0

$$\boldsymbol{M}_{ij} = P\left(X^{(t+1)} = j \mid X^{(t)} = i\right)$$

$$\boldsymbol{q}_i^{(t)} = P\left(\mathrm{X}^{(t)} = i\right) = \left(\boldsymbol{q}^{(0)}\boldsymbol{M}^t\right)_i$$

Example: Random Walks on an Undirected Graph

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



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

Brain Break

Agenda

- Recap: Markov Chains
- 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 (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)

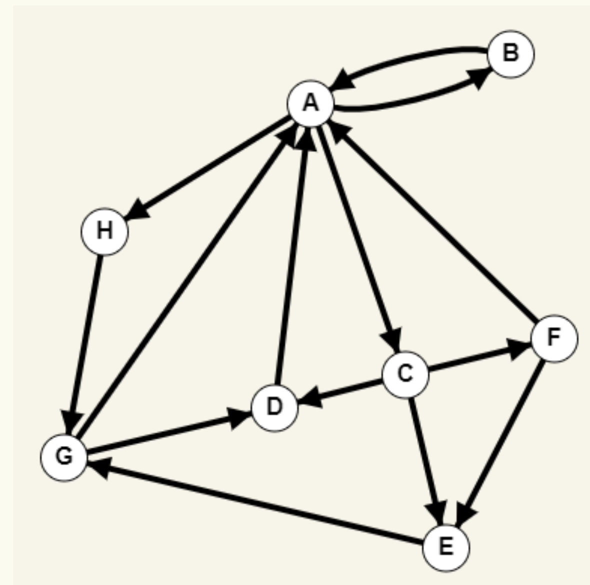
- MacArthur genius prize, Nevanlinna Prize and many other academic honors

PageRank - Idea

Take into account the directed graph structure of the web.

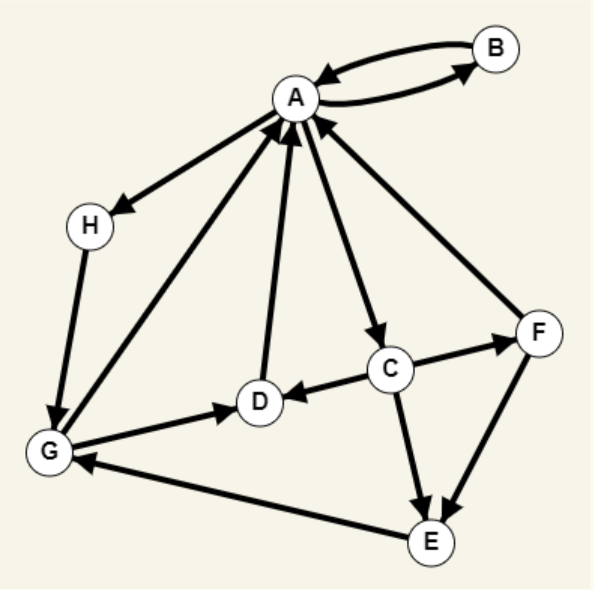
Use **hyperlink analysis** to compute what pages are high quality or have high authority.

Trust the Internet itself define what is useful via its links.



PageRank - Idea

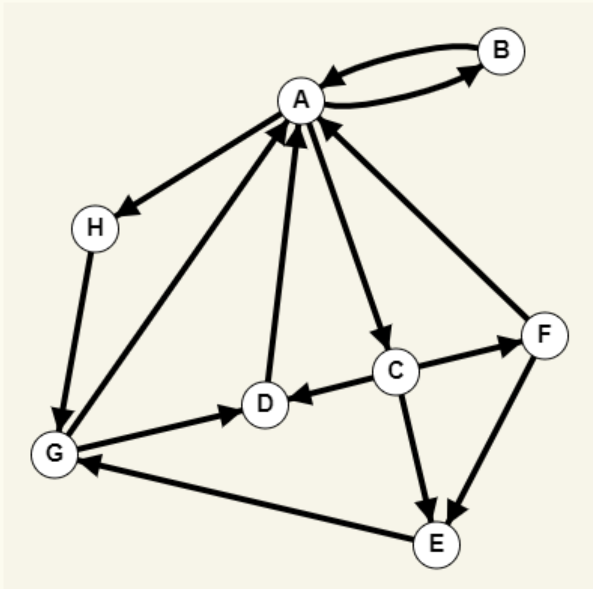
Idea 1 : Think of each link as a citation "vote of quality"



Rank pages by in-degree?

PageRank - Idea

Idea 1 : Think of each link as a citation "vote of quality"



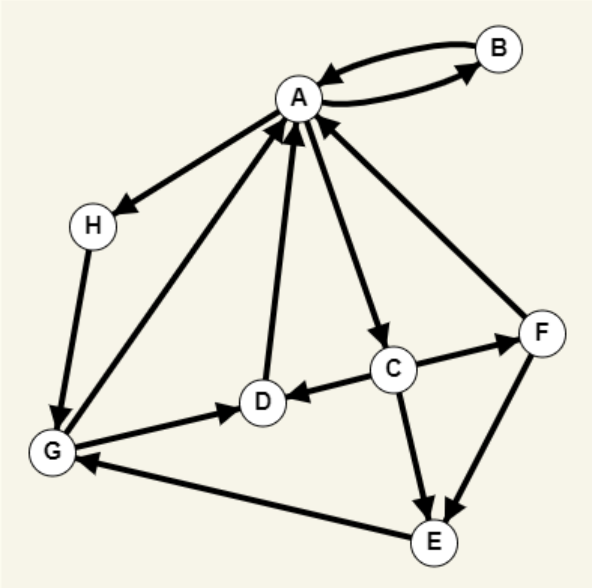
Rank pages by in-degree?

Problems:

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

PageRank - Idea

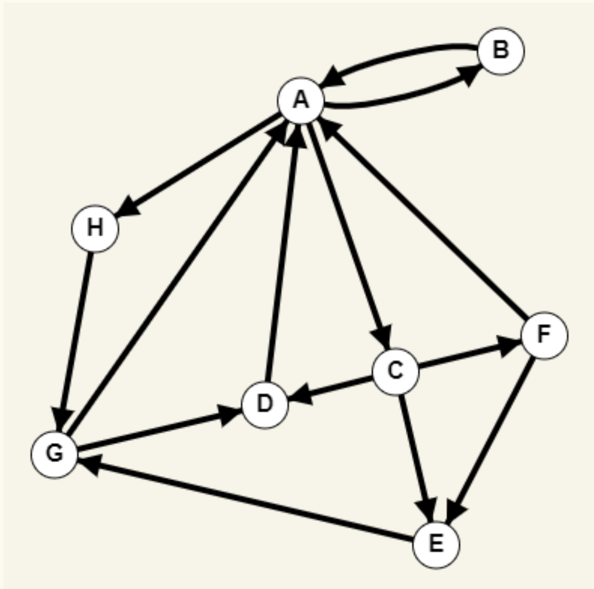
Idea 2 : Perhaps we should weight the links somehow and then use the weights of the in-links to rank pages



Inching towards PageRank

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

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



That's a recursive definition!

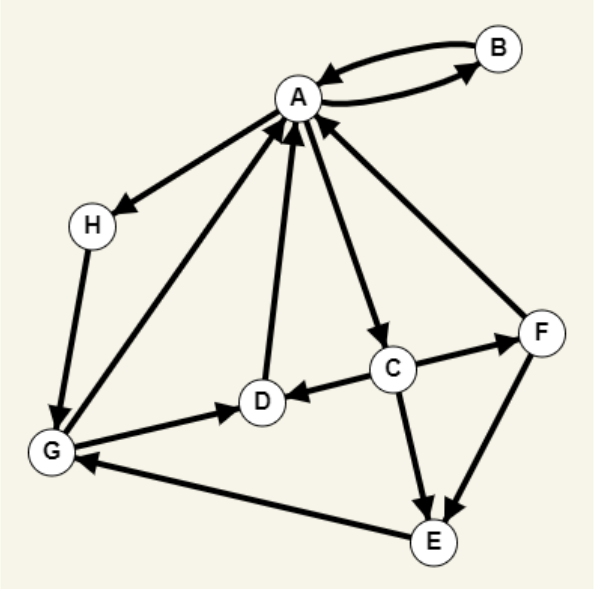
Inching towards PageRank

- If web page x has d outgoing links, one of which goes to y, this contributes $1/d$ to the importance of y

- But $1/d$ of what?

  We want to take into account the importance of x too…

  …so it actually contributes $1/d$ of the importance of x



This gives the following equations

Idea: Use the transition matrix $\boldsymbol{M}$ defined by a *random walk* on the web to compute quality of webpages.

Namely: Find $\boldsymbol{q}$ such that $\boldsymbol{qM} = \boldsymbol{q}$ **Seem familiar?**

This is the stationary distribution for the Markov chain defined by a random web 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

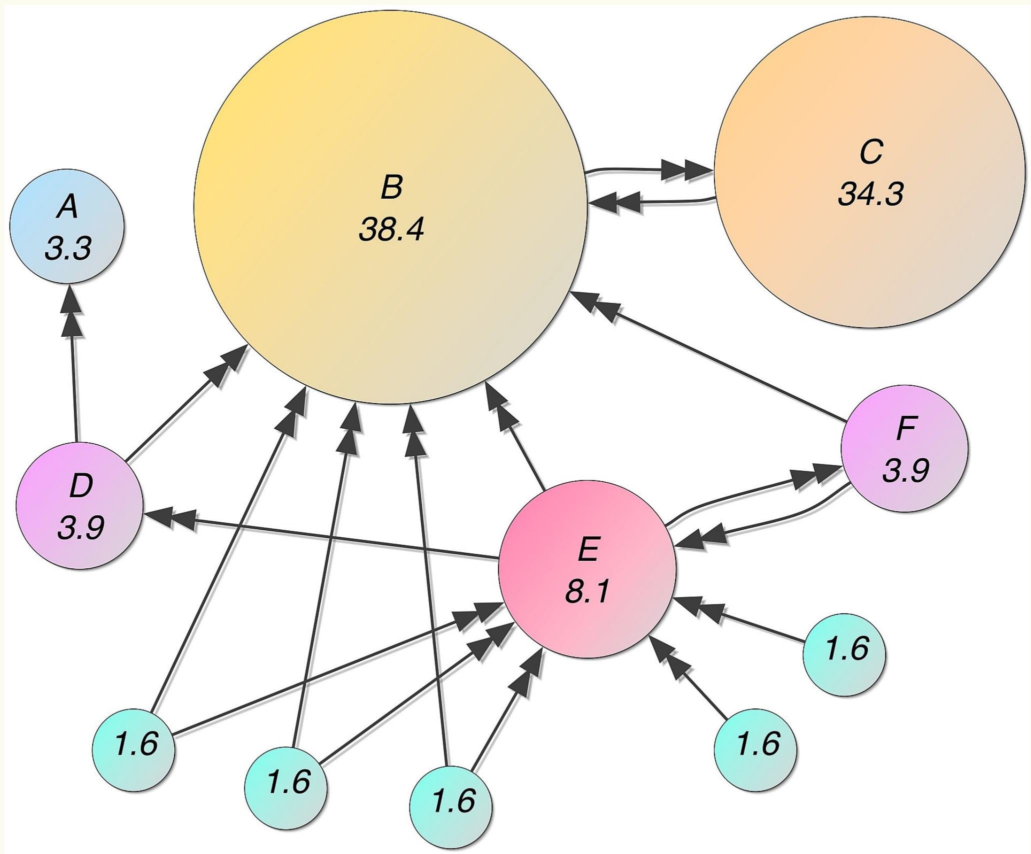
- How to handle dangling nodes (dead ends that don't link to anything) ?
- How to handle Rank sinks – group of pages that only link to each other ?

Both solutions can be solved by "teleportation"

Final PageRank Algorithm

1. Make a Markov Chain with one state for each webpage on the Internet with the transition probabilities $\boldsymbol{M}_{ij} = \frac{1}{outdeg(i)}$.
2. Use a modified random walk. At each point in time if the surfer is at some webpage i:
   - If i has outlinks:
     - With probability p, take a step to one of the neighbors of i (equally likely)
     - With probability $1 - p$, "teleport" to a uniformly random page in the whole Internet.
   - Otherwise, always "teleport"
3. Compute stationary distribution $\boldsymbol{\pi}$ of this perturbed Markov chain.
4. Define the PageRank of a webpage i as the stationary probability π_i.
5. 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 and other web search engines have a LOT more secret sauce to rank pages, most of which they don't reveal 1) for competitive advantage and 2) to avoid gaming of their algorithms.