CSE 312

Foundations of Computing II

Lecture 25: Markov chains and Pagerank

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Anna R. Karlin

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

Agenda

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

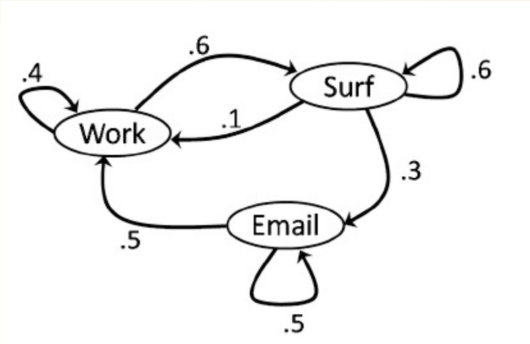
Many-step random process

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.

Transition Probability Matrix

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

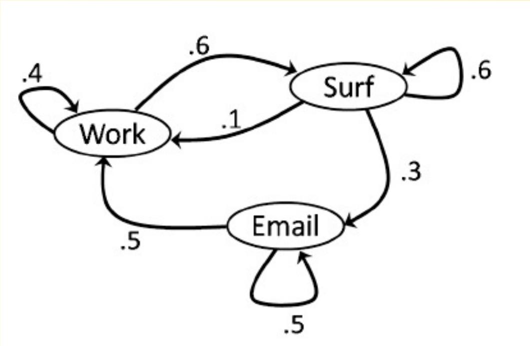


$\boldsymbol{X^{(t)}}$ state at time *t* (random variable)

$$\left(q_W^{(t)}, \quad q_S^{(t)}, \quad q_E^{(t)}\right) = \left(\Pr\left(X^{(t)} = \text{ work}\right), \Pr\left(X^{(t)} = \text{surf}\right), \Pr\left(X^{(t)} = \text{email}\right)\right)$$

Transition Probability Matrix

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



$\left(q_W^{(t)},\quad q_S^{(t)},\quad q_E^{(t)}\right) = \left(\Pr\left(X^{(t)} = \text{ work}\right), \Pr\left(X^{(t)} = \text{surf}\right), \Pr\left(X^{(t)} = \text{email}\right)\right)$

$$\left(q_W^{(t)},\quad q_S^{(t)},\quad q_E^{(t)}\right) = \left(q_W^{(t-1)},\quad q_S^{(t-1)},\quad q_E^{(t-1)}\right) \begin{pmatrix} .4 & .6 & 0 \\ .1 & .6 & .3 \\ .5 & 0 & .5 \end{pmatrix}$$

➔ $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(t-1)}\ \boldsymbol{P}$ $\qquad$ $\boldsymbol{q}^{(t)} = \left(q_W^{(t)},\quad q_S^{(t)},\quad q_E^{(t)}\right)$

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

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

➔ $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\,\boldsymbol{P}^{t}$

The t-step walk $\boldsymbol{P}^t$

$$\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(t-1)}\,\boldsymbol{P}$$
$$\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\,\boldsymbol{P}^t$$

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

$$P^2 = \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .22 & .6 & .18 \\ .25 & .42 & .33 \\ .45 & .3 & .25 \end{pmatrix} \end{array}$$

$$P^3 = \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .238 & .492 & .270 \\ .307 & .402 & .291 \\ .335 & .450 & .215 \end{pmatrix} \end{array}$$

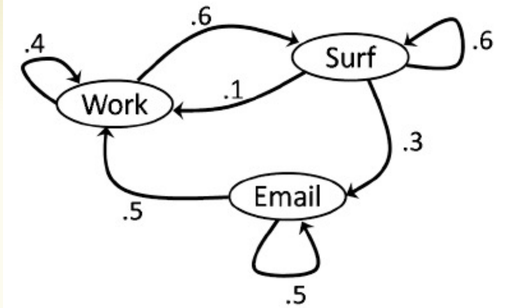
$$P^{10} \approx \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .2940 & .4413 & .2648 \\ .2942 & .4411 & .2648 \\ .2942 & .4413 & .2648 \end{pmatrix} \end{array}$$

$$P^{30} \approx \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .29411764705 & .44117647059 & .26470588235 \\ .29411764706 & .44117647058 & .26470588235 \\ .29411764706 & .44117647059 & .26470588235 \end{pmatrix} \end{array}$$

$$P^{60} \approx \begin{array}{cc} & \begin{array}{ccc} W & S & E \end{array} \\ \begin{array}{c} W \\ S \\ E \end{array} & \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

What do these numbers tell us about $\boldsymbol{q}^{(t)}$?

Observation



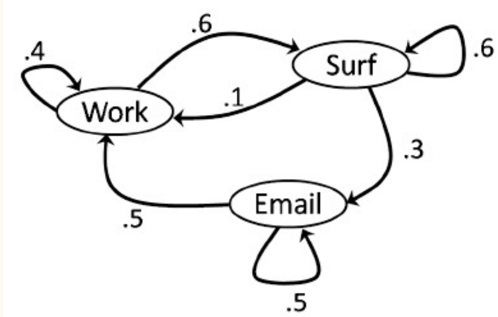
If $q^{(t)}$ = $q^{(t-1)}$ then it will never change again!

Called a "stationary distribution" and has a special name

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

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

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



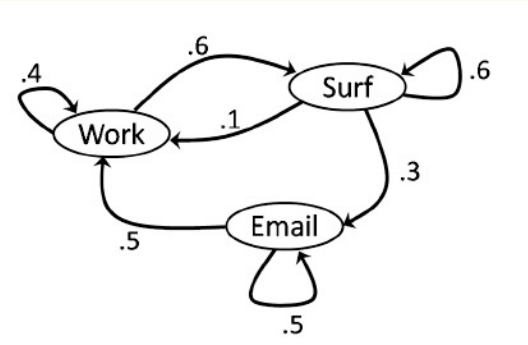
Solving for Stationary Distribution

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

Stationary Distribution satisfies

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

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



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

The Fundamental Theorem of Markov Chains

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

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

Finite Markov Chains

- Defined by a set of states and a transition probability matrix:
  - A set of n states S={1, 2, 3, … n}
  - The state at time t is denoted by $X^{(t)}$
  - A transition matrix ***P***, dimension $n \times n$

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

  - Has **Markov property**: State at time t depends only on state at time $t - 1$
  - This does not mean that state at time t is independent of state at times 0, $\ldots, t-2$! Just that all of the dependency is captured by $X^{(t-1)}$

State at time t and matrix powers

- $\Pr\left(X^{(1)} = j \middle| \mathrm{X}^{(0)} = i\right) = \boldsymbol{P}_{ij}$
- $\Pr\left(X^{(2)} = j \middle| \mathrm{X}^{(0)} = i\right) =$

State at time t and matrix powers

- $\Pr\left(X^{(1)} = j \,\middle|\, \mathrm{X}^{(0)} = i\right) = \boldsymbol{P}_{ij}$
- $\Pr\left(X^{(2)} = j \,\middle|\, \mathrm{X}^{(0)} = i\right) = (\boldsymbol{P^2})_{ij}$
- $\Pr\left(X^{(3)} = j \,\middle|\, \mathrm{X}^{(0)} = i\right) =$

Finite Markov Chains

- Defined by a set of states and a transition probability matrix:
  - A set of n states $\{1, 2, 3, \ldots\ \mathrm{n}\}$
  - The state at time t is denoted by $X^{(t)}$
  - A transition matrix $\boldsymbol{P}$, dimension $n \times n$

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

  - More generally, $\Pr\left(X^{(t)} = j \middle| \mathrm{X}^{(0)} = i\right) = \boldsymbol{P^t}_{ij}$
  - Similarly, $\Pr\left(X^{(t+s)} = j \middle| \mathrm{X}^{(s)} = i\right) = \boldsymbol{P^t}_{ij}$

$\boldsymbol{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$ where $q_i^{(t)} = \Pr(X^{(t)} = i)$

Finite Markov Chains

- Defined by a set of states and a transition probability matrix:
  - A set of n states $\{1, 2, 3, \ldots \text{ n}\}$
  - The state at time t is denoted by $X^{(t)}$
  - A transition matrix $\boldsymbol{P}$, dimension $n \times n$

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

  - More generally, $\Pr\left(X^{(t)} = j \middle| \mathrm{X}^{(0)} = i\right) = \boldsymbol{P^t}_{ij}$
  - Similarly, $\Pr\left(X^{(t+s)} = j \middle| \mathrm{X}^{(s)} = i\right) = \boldsymbol{P^t}_{ij}$
  - $\boldsymbol{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$ where $q_i^{(t)} = \Pr(\mathrm{X}^{(t)} = i)$
  - $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(t-1)}\, \boldsymbol{P} \implies \boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\, \boldsymbol{P}^t$

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 – nonnegative and sums to 1) such that

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

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

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 $\boldsymbol{q}^{(0)}$ over the states

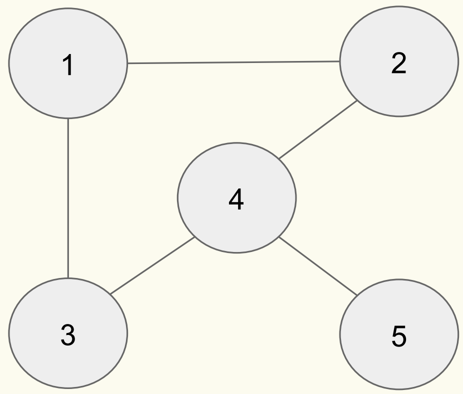
$$\lim_{t\to\infty} \boldsymbol{q}^{(0)} \boldsymbol{P}^{t} = \boldsymbol{\pi}$$

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

Another Example: Random Walks

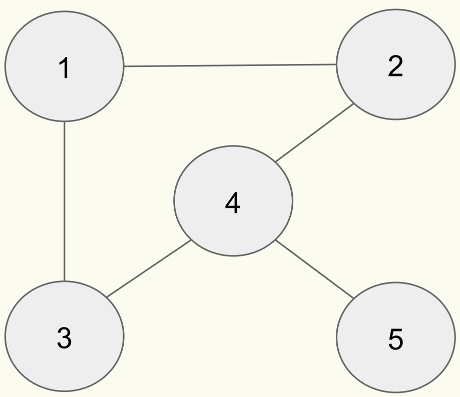
Suppose we start at node 1, and at each step transition to a neighboring node with equal probability.

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



Example: Random Walks

Start by defining transition probs.



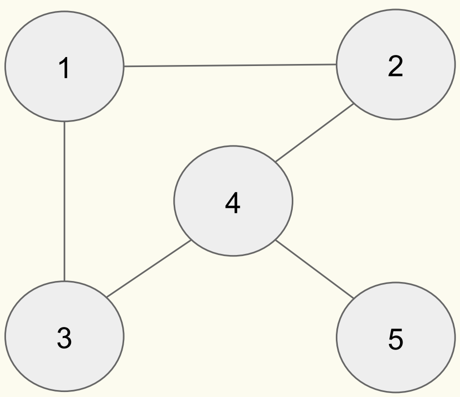
$$
\begin{array}{c|ccccc}
 & \text{To } \mathbf{s_1} & \text{To } \mathbf{s_2} & \text{To } \mathbf{s_3} & \text{To } \mathbf{s_4} & \text{To } \mathbf{s_5} \\
\hline
\text{From } \mathbf{s_1} & & & & & \\
\text{From } \mathbf{s_2} & & & & & \\
\text{From } \mathbf{s_3} & & & & & \\
\text{From } \mathbf{s_4} & & & & & \\
\text{From } \mathbf{s_5} & & & & &
\end{array}
$$

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

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

Example: Random Walks

Start by defining transition probs.



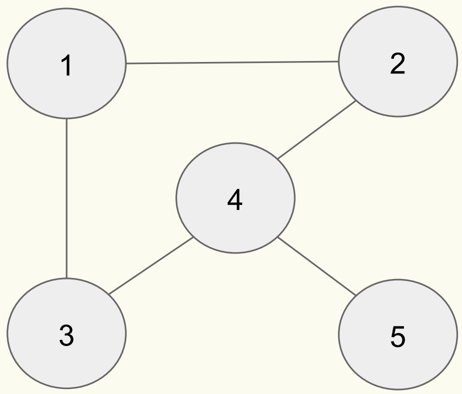
$$
\begin{array}{c|ccccc}
 & \text{To } \mathbf{s_1} & \text{To } \mathbf{s_2} & \text{To } \mathbf{s_3} & \text{To } \mathbf{s_4} & \text{To } \mathbf{s_5} \\
\hline
\text{From } \mathbf{s_1} & 0 & 1/2 & 1/2 & 0 & 0 \\
\text{From } \mathbf{s_2} & 1/2 & 0 & 0 & 1/2 & 0 \\
\text{From } \mathbf{s_3} & 1/2 & 0 & 0 & 1/2 & 0 \\
\text{From } \mathbf{s_4} & 0 & 1/3 & 1/3 & 0 & 1/3 \\
\text{From } \mathbf{s_5} & 0 & 0 & 0 & 1 & 0
\end{array}
$$

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

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

Example: Random Walks

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



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

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)

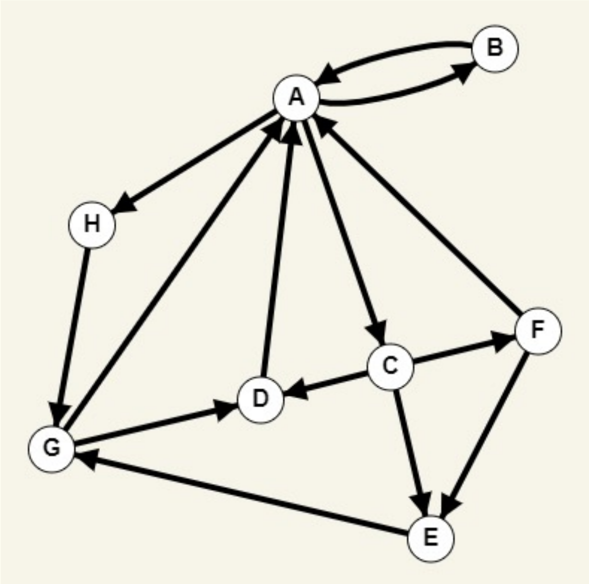
- Took the idea and founded Google, making billions

Jon Kleinberg (professor at Cornell)

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

PageRank - Idea

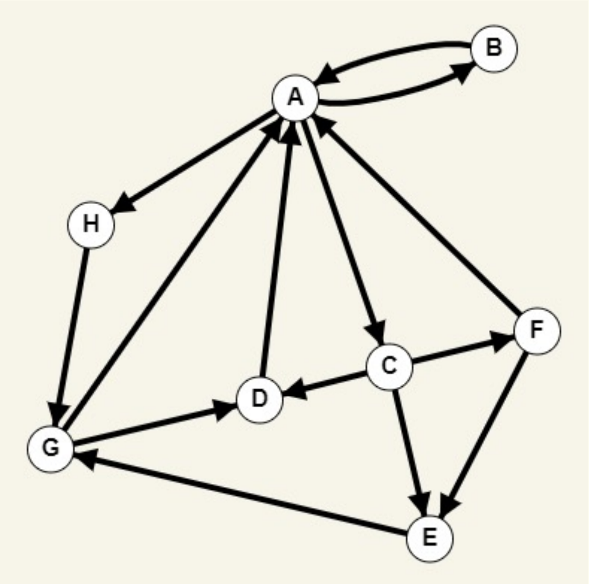
Take into account directed graph structure of the web. Use **hyperlink analysis** to compute what pages are high quality or have high authority. Trust the internet itself define what is useful via its links.



PageRank - Idea

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

Rank pages by in-degree?



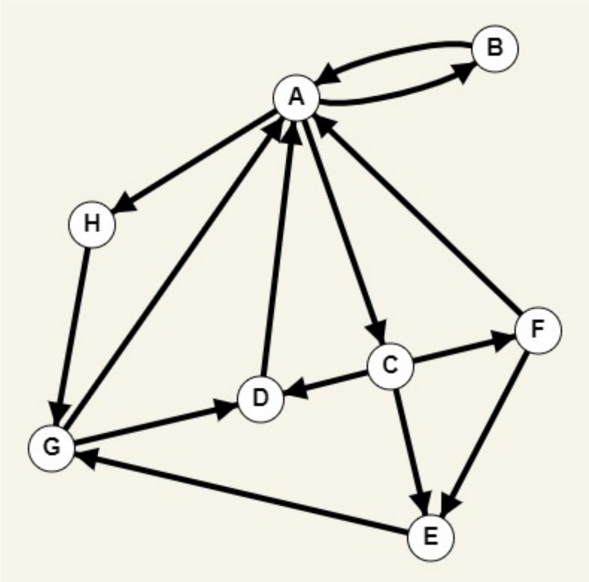
PageRank - Idea

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

Rank pages by in-degree?

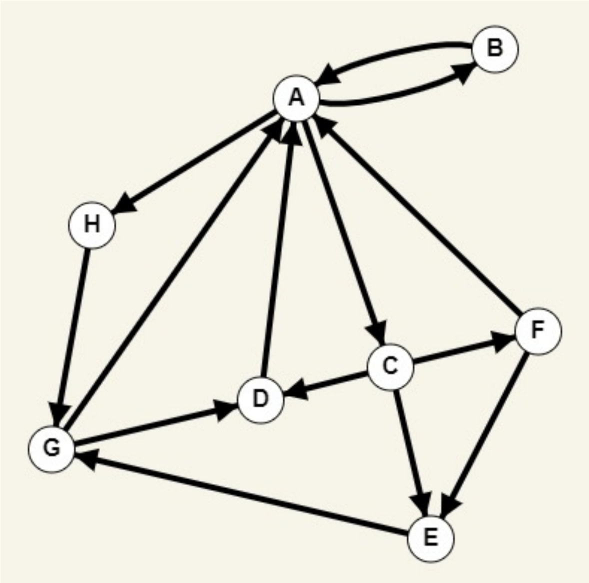
Problems:

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



PageRank - Idea

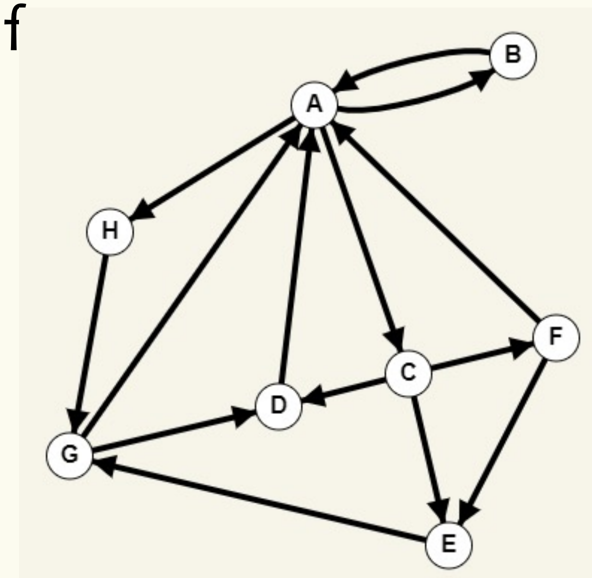
Idea 2: perhaps we should weight the links somehow and then use the weights of the in-links to rank pages



Inching towards Pagerank

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

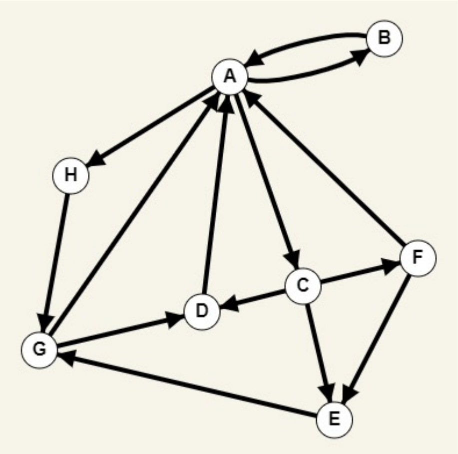
A page is high quality if it links to lots of high quality pages



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 $\boldsymbol{P}$ to compute quality of webpages. Namely, find $\boldsymbol{q}$ such that

$$\boldsymbol{qP} = \boldsymbol{q}$$

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

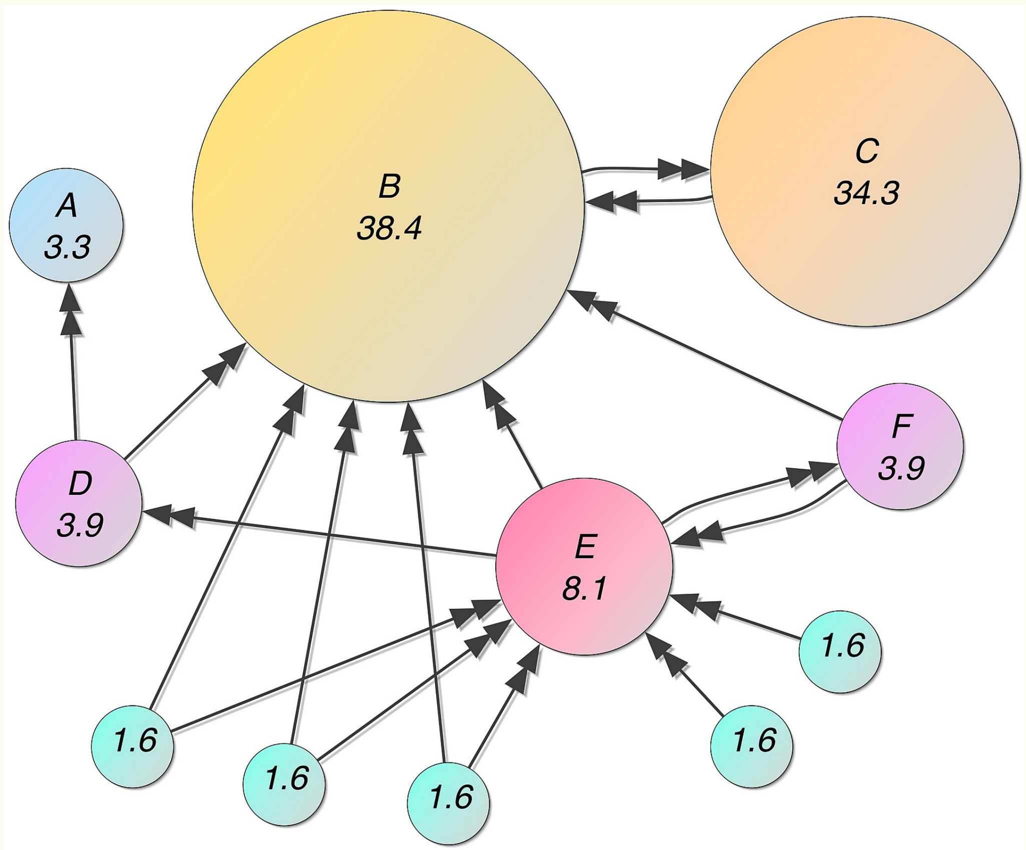
- How to handle dangling nodes (dead ends)?
- How to handle Rank sinks – group of pages that only link to each other?

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 $\boldsymbol{P}_{ij} = \frac{1}{outdeg(i)}$.
- 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 $\boldsymbol{\pi}$ of this perturbed Markov chain.
- Define the PageRank of a webpage i as the stationary probability π_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.