CSE 312

Foundations of Computing II

Lecture 24: Finish Markov Chains, applications

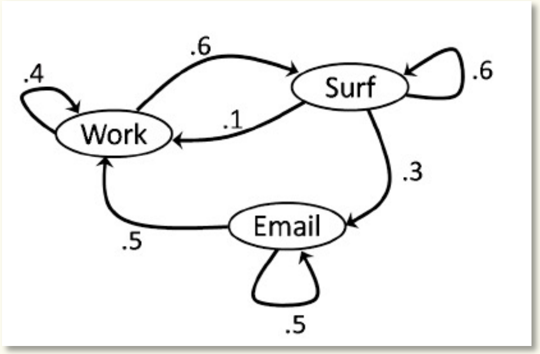
www.slido.com/2167587

Agenda

- Markov Chains
- Two applications:
  - Markov Chain Monte Carlo
  - Pagerank
- A glimpse of auction theory

A typical day in my life

How do we interpret this diagram?



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

At *each* time step t

– I can be in one of 3 **states**

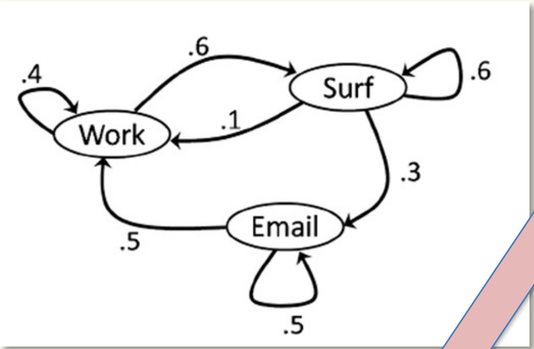
- Work, Surf, Email

– If I am in some state s at time t

- the **labels of out-edges** of s **give the probabilities** of my 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 I am in Email, there is a 50-50 chance I will be in each of Work or Email at the next time step, but I will never be in state Surf in the next step.

An organized way to understand the distribution of $X^{(t)}$

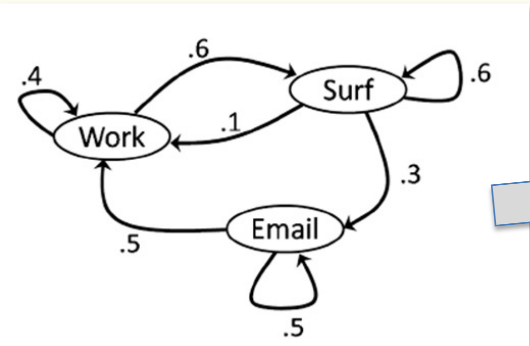


Write as a tuple $(q_W^{(t)}, q_S^{(t)}, q_E^{(t)})$ a.k.a. a row vector:

$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$$

t	0	1	2
$q_W^{(t)} = P(X^{(t)} = \text{Work})$	$\mathbf{1}$	$\mathbf{0.4}$	$= \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$
$q_S^{(t)} = P(X^{(t)} = \text{Surf})$	$\mathbf{0}$	$\mathbf{0.6}$	$= \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$
$q_E^{(t)} = P(X^{(t)} = \text{Email})$	$\mathbf{0}$	$\mathbf{0}$	$= \mathbf{0.4} \cdot 0 + \mathbf{0.6} \cdot 0.3 = 0 + 0.18 = \mathbf{0.18}$

An organized way to understand the distribution of $X^{(t)}$



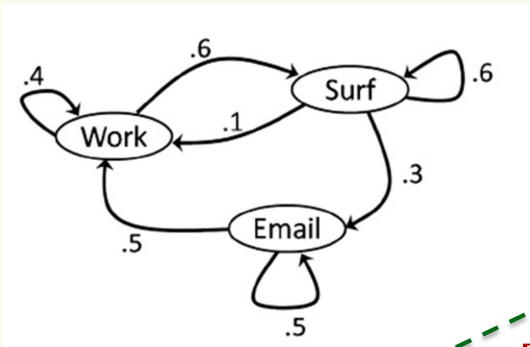
$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \qquad \boldsymbol{M} = \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

Write as a "transition probability matrix" $\boldsymbol{M}$

- one row/col per state. Row=now, Col=next
- each row sums to 1

t	0	1	2
$q_W^{(t)} = P(X^{(t)} = \text{Work})$	$\mathbf{1}$	$\mathbf{0.4}$	$= \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22}$
$q_S^{(t)} = P(X^{(t)} = \text{Surf})$	$\mathbf{0}$	$\mathbf{0.6}$	$= \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60}$
$q_E^{(t)} = P(X^{(t)} = \text{Email})$	$\mathbf{0}$	$\mathbf{0}$	$= \mathbf{0.4} \cdot 0 + \mathbf{0.6} \cdot 0.3 = 0 + 0.18 = \mathbf{0.18}$

An organized way to understand the distribution of $X^{(t)}$



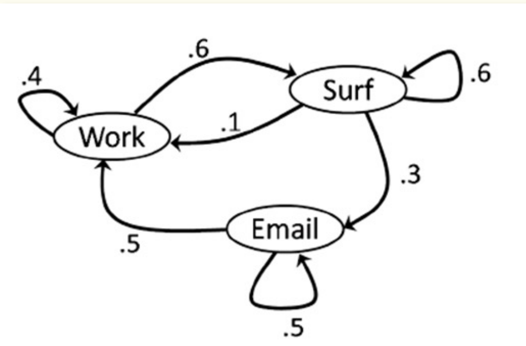
$$[q_W^{(t)}, q_S^{(t)}, q_E^{(t)}] \overset{\boldsymbol{M}}{\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}} = [q_W^{(t+1)}, q_S^{(t+1)}, q_E^{(t+1)}]$$

Vector-matrix multiplication

$$[0.4, 0.6, 0] \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} = [0.22, 0.60, 0.18\,]$$

$$\begin{aligned} q_W^{(1)} &= \mathbf{0.4} & q_W^{(2)} &= \mathbf{0.4} \cdot 0.4 + \mathbf{0.6} \cdot 0.1 = 0.16 + 0.06 = \mathbf{0.22} \\ q_S^{(1)} &= \mathbf{0.6} & q_S^{(2)} &= \mathbf{0.4} \cdot 0.6 + \mathbf{0.6} \cdot 0.6 = 0.24 + 0.36 = \mathbf{0.60} \\ q_E^{(1)} &= \mathbf{0} & q_E^{(2)} &= \mathbf{0.4} \cdot 0 \quad + \mathbf{0.6} \cdot 0.3 = 0 \quad + 0.18 = \mathbf{0.18} \end{aligned}$$

An organized way to understand the 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

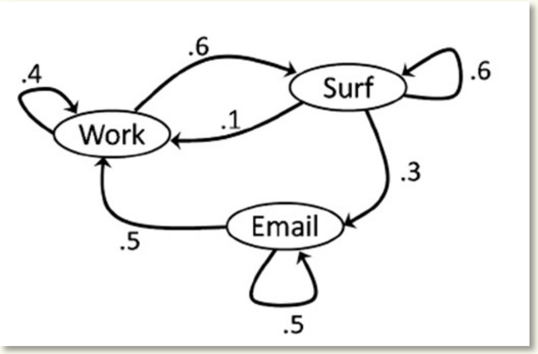
$$q_W^{(t)} \cdot 0.4 + q_S^{(t)} \cdot 0.1 + q_E^{(t)} \cdot 0.5 = q_W^{(t+1)}$$

$$q_W^{(t)} \cdot 0.6 + q_S^{(t)} \cdot 0.6 + q_E^{(t)} \cdot 0 = q_S^{(t+1)}$$

$$q_W^{(t)} \cdot 0 + q_S^{(t)} \cdot 0.3 + q_E^{(t)} \cdot 0.5 = q_E^{(t+1)}$$

Write $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$ Then for all $t \geq 0$, $\boldsymbol{q}^{(t)}\boldsymbol{M} = \boldsymbol{q}^{(t+1)}$

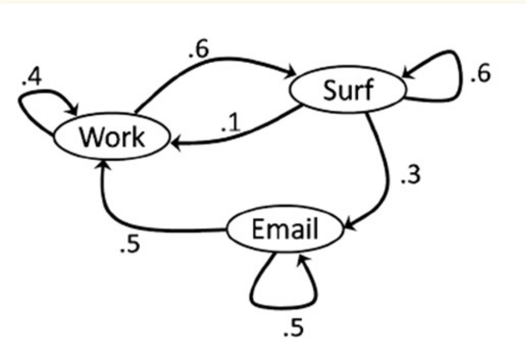
By induction … we can derive



$$\begin{array}{c} \boldsymbol{M} \\ \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix} \end{array}$$

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

Many interesting questions about Markov Chains



Given: In state Work at time $t = 0$

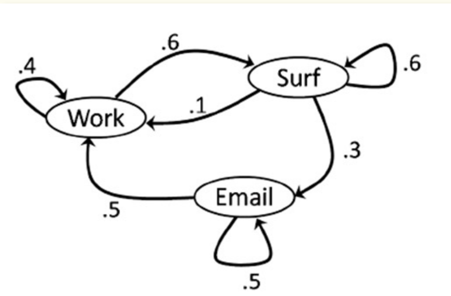
1. **What is the probability that I am in state s at time 1?**
2. **What is the probability that I am in state s at time 2?**
3. **What is the probability that I am in state s at some time t far in the future?**

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

What does $\boldsymbol{M}^t$ look like for really big t ?

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

$\boldsymbol{M}^t$ as t grows



$\boldsymbol{M}$

$$\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

$\boldsymbol{M}^2$

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

$\boldsymbol{M}^3$

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

$\boldsymbol{M}^{10}$

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

$\boldsymbol{M}^{30}$

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

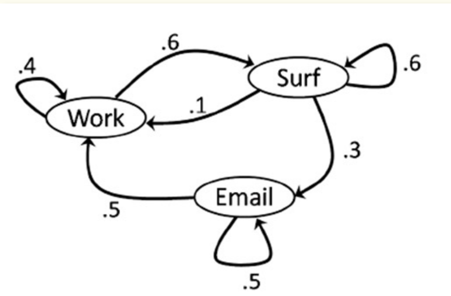
$\boldsymbol{M}^{60}$

$$\begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

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

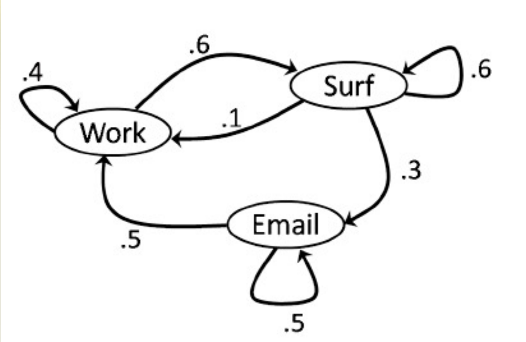
$\boldsymbol{M}^t$ as t grows

$$\boxed{\boldsymbol{q}^{(60)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^{60}}$$



$$[q_W^{(0)}, q_S^{(0)}, q_E^{(0)}] \cdot \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array} = [q_W^{(60)}, q_S^{(60)}, q_E^{(60)}]$$

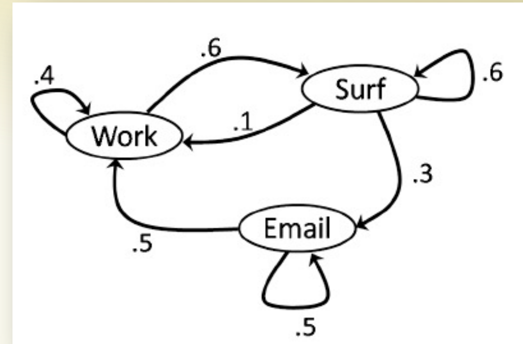
Observation



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

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

Observation



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

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

Called a **stationary distribution** and has a special name

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

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

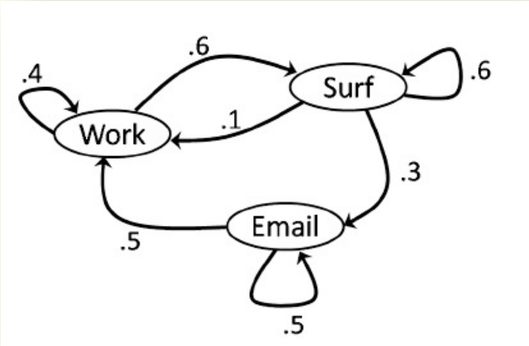
Solving for Stationary Distribution

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

Stationary Distribution satisfies

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

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



As $t \to \infty$, $\boldsymbol{q}^{(t)} \to \boldsymbol{\pi}$ no matter what distribution $\boldsymbol{q}^{(0)}$ is !!

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** $\boldsymbol{M}$, dimension $n \times n$

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

- $\boldsymbol{q}^{(t)} = (q_1^{(t)}, q_2^{(t)}, \dots, q_n^{(t)})$ where $q_i^{(t)} = P(X^{(t)} = i)$
- Transition: LTP $\Rightarrow$ $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\ \boldsymbol{M}$ so $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\ \boldsymbol{M}^t$

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

Fundamental Theorem of Markov Chains

Recall $\boldsymbol{q}^{(t)}$ is the distribution of being at each state at time t computed by $\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)}\boldsymbol{M}^t$. 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

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

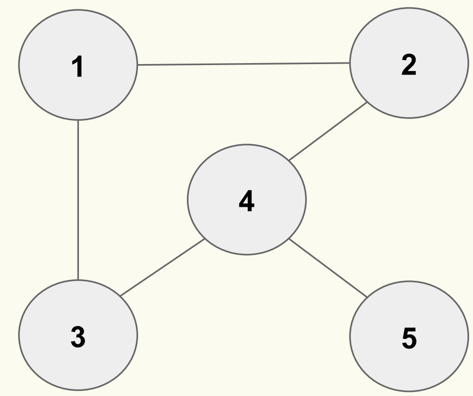
where $\boldsymbol{\pi}$ is the stationary distribution of $\boldsymbol{M}$ (i.e., $\boldsymbol{\pi}\boldsymbol{M} = \boldsymbol{\pi}$)

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

Another Example: Random Walks

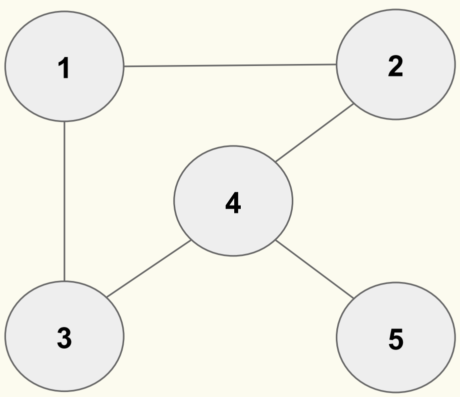
Suppose we start at node 1, and at each step transition to a neighboring node with equal probability.

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



Example: Random Walks on an Undirected Graph

Start by defining transition probs.



$$
\begin{array}{c|ccccc}
 & \text{To } \mathbf{1} & \text{To } \mathbf{2} & \text{To } \mathbf{3} & \text{To } \mathbf{4} & \text{To } \mathbf{5} \\
\hline
\text{From } \mathbf{1} & & & & & \\
\text{From } \mathbf{2} & & & & & \\
\text{From } \mathbf{3} & & & & & \\
\text{From } \mathbf{4} & & & & & \\
\text{From } \mathbf{5} & & & & &
\end{array}
$$

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

Agenda

- Markov Chains
- Two applications:
  - Markov Chain Monte Carlo
  - Pagerank
- A glimpse of auction theory

Markov Chain Monte Carlo

- Technique for sampling from high dimensional distributions
- Computational method for studying large, very complex sets.
- In some cases, a technique for getting good approximate solutions to complex optimization problems
- Used in every field of science and engineering
- "*To someone working in my part of the world, asking about applications of MCMC is like asking about applications of the quadratic formula. The results are really used in every aspect of scientific inquiry*" --- Persi Diaconis, Stanford

MCMC

- Idea: simulate a random walk that moves among possible configurations of a system and will converge to a useful distribution over these configurations.
- Example: Knapsack Problem (NP-complete)
  - Input: collection of n items, for each item
    - Value
    - Weight
  - Goal: output subset S of items of maximum total value, that has total weight < W.

MCMC for knapsack

- Define a Markov chain with states being possible solutions and transition probabilities that have higher probabilities on "good solutions"
- Simulate the Markov chain for many iterations until reach a "good" state.

MCMC for Knapsack Problem

Algorithm 1 MCMC for 0-1 Knapsack Problem

subset $\leftarrow$ vector of n zeros (indexed by 0 to $n-1$), where subset is always a binary vector in $\{0,1\}^n$ that represents whether or not we have each item. (This means that we initially start with an empty knapsack).
best_subset $\leftarrow$ subset
for $t = 1, \ldots,$ NUM_ITER **do**
    $k \leftarrow$ a uniformly random integer in $\{0, 1, \ldots, n-1\}$.
    new_subset $\leftarrow$ subset but with subset$[k]$ flipped ($0 \to 1$ or $1 \to 0$).
    $\Delta \leftarrow$ value(new_subset) $-$ value(subset)
    if new_subset satisfies weight constraint (total weight $\leqslant W$) **then**
        if $\Delta > 0$ OR ($T > 0$ AND Unif$(0,1) < e^{\Delta/T}$) **then**
            subset $\leftarrow$ new_subset
    if value(subset) $>$ value(best_subset) **then**
        best_subset $\leftarrow$ subset

Brain Break

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)

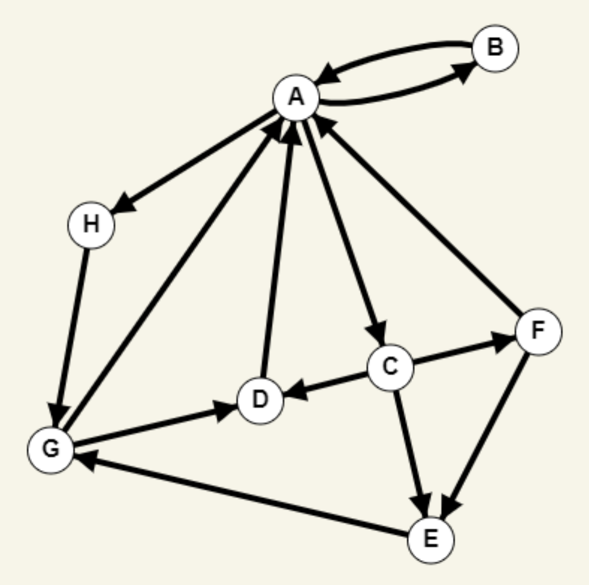
– MacArthur genius prize, Nevanlinna Prize and many other academic honors

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.

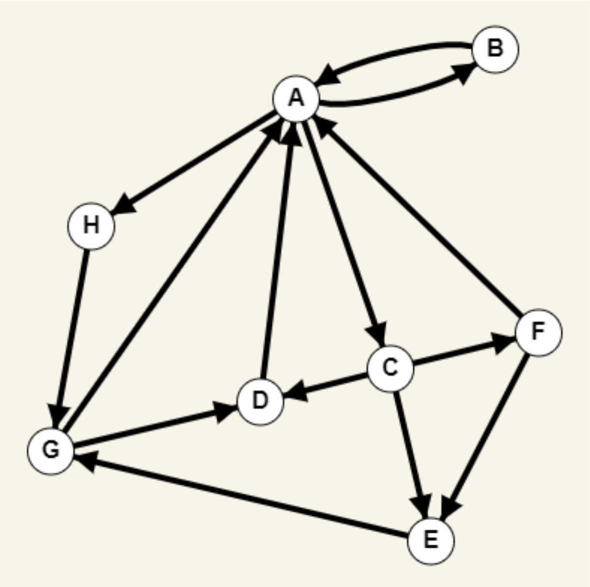
Trust the Internet itself define what is useful via its links.



PageRank - Idea

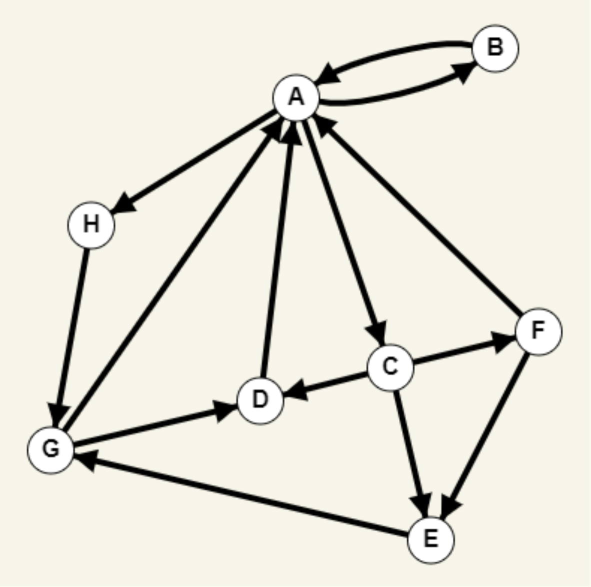
Idea 1 : Think of each link as a citation "vote of quality"

Rank pages by in-degree?



PageRank - Idea

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



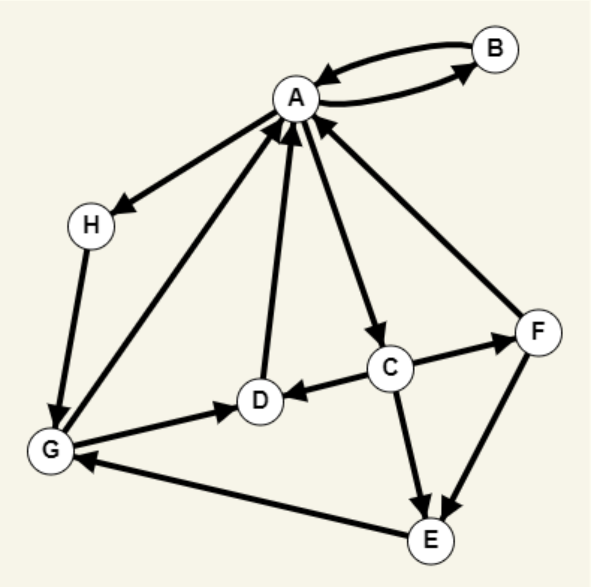
Rank pages by in-degree?

Problems:

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

PageRank - Idea

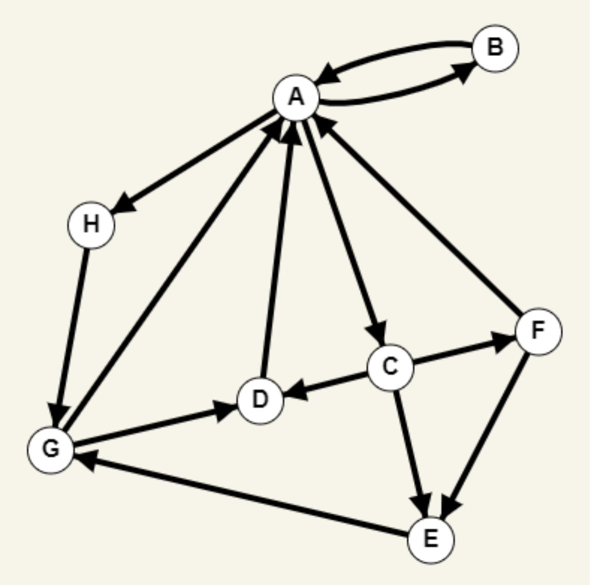
Idea 2 : Perhaps we should weight the links somehow and then use the weights of the in-links to rank pages



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!

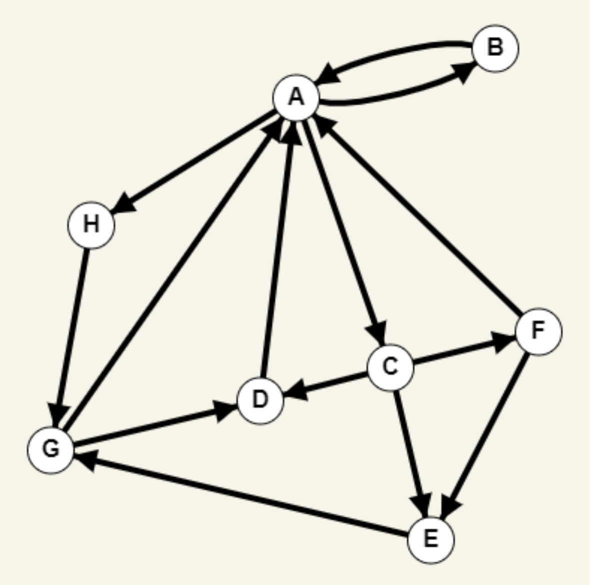
Inching towards PageRank

- 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 M defined by a *random walk* on the web to compute quality of webpages.

Namely: Find q such that $qM = q$ **Seem familiar?**

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 $\boldsymbol{qM = q}$ **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

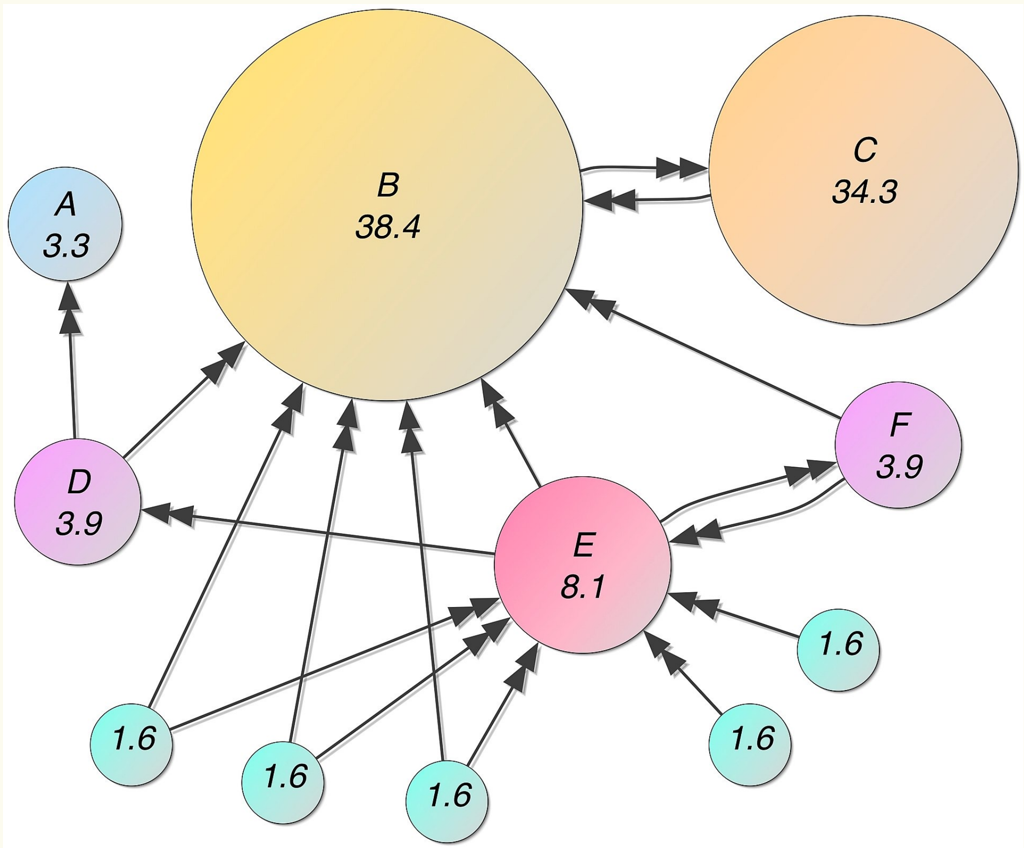
- How to handle dangling nodes (dead ends that don't link to anything) ?
- How to handle Rank sinks – group of pages that only link to each other ?

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.

Agenda

- Markov Chains
- Two applications:
  - Markov Chain Monte Carlo
  - Pagerank
- A glimpse of auction theory

Auctions

- Some goods on eBay and amazon are sold via auction.
- Companies like Google and Facebook make most of their money by selling ads.
- The ads are sold via auction.
    - Advertisers submit bids for certain “keywords”

Facebook Ads bidding... 🤔 Is this an auction?

Yes! That’s the first thing you need to understand to master bidding management of Facebook Ads. **When you’re creating a new campaign, you’re joining a huge, worldwide auction.**

You’ll be competing with hundreds of thousands of advertisers to buy what Facebook is selling: Real estate on the News Feed, Messenger, Audience Network, and mobile apps to display your ads to the users.

hawaii vacation

All | Images | News | Videos | Maps | More | Tools

About 2,670,000,000 results (0.79 seconds)

Ad · https://www.expedia.com/

Hawaii Island Packages - Book with Expedia and Save

Bundle Your Flight + Hotel & Save! Make Your Trip Memorable. Secure Booking. 600,000+ Hotels Worldwide. Limited Time Offers. New Expedia Rewards. Compare & Save.

Package Deals

Today's Best Flight + Hotel Deals.
Only with Your #1 Leader in Travel.

Last Minute Deals

Expedia Last Minute Travel Deals.
Book Today, Travel Tomorrow.

Ad · https://www.airbnb.com/

Hawaii Vacation - Book & Save on Airbnb - airbnb.com

Find **vacation** from **Hawaii**. Perfect for any **Vacation**. 5 Star Hosts. 100,000 Cities. Best Prices. Instant Confirmation. Types: Entire Home, Apartment, Cabin, Villa, Boutique Hotel.

An auction is a …

- Game
  - Players: advertisers
  - Strategy choices for each player: possible bids
  - Rules of the game – made up by Google/Facebook/whoever is running the auction
- What do we expect to happen? How do we analyze mathematically?

Special case: Sealed bid single item auction

- Say I decide to run an auction to sell my laptop and I let you be the bidders.
- If I want to make as much money as possible – what should I choose as the rules of the auction?

Special case: Sealed bid single item auction

- Say I decide to run an auction to sell my laptop and I let you be the bidders.
- If I want to make as much money as possible – what should the rules of the auction be?

Some possibilities:

- **First price auction:** highest bidder wins; pays what they bid.
- **Second price auction:** highest bidder wins; pays second highest bid.
- **All pay auction:** highest bidder wins: all bidders pay what they bid.

Which of these will make me the most money?

Special case: Sealed Bid single item auction

Some possibilities:

- **First price auction:** highest bidder wins; pays what they bid.
- **Second price auction:** highest bidder wins; pays second highest bid.
- **All pay auction:** highest bidder wins: all bidders pay what they bid.

Bidder model

Each bidder has a value, say v_i for bidder i.

Bidder is trying to maximize their “utility” –
 the value of the item they get – price they pay.

Theorem

A second price auction is **truthful**. In other words, it is always in each bidder's best interest to bid their true value.

Bayes-Nash equilibrium

Suppose that $V_1 \sim F_1, V_2 \sim F_2, \ldots, V_n \sim F_n$.
A bidding strategy $\beta_i(\cdot)$ is a **Bayes-Nash equilibrium** if $\beta_i(v_i)$ is a **best response in expectation** to $\beta_j(V_j)\ \forall j \neq i$.

Revenue Equivalence Theorem

In equilibrium, no matter what distribution the bids are drawn from, the expected auctioneer revenue is the same in all three auctions!