CSE 312

Foundations of Computing II

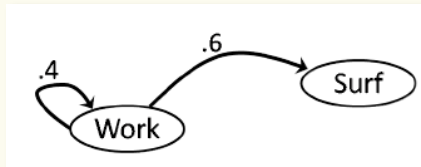
Lecture 24: Application of Markov Chains + glimpse of auction theory

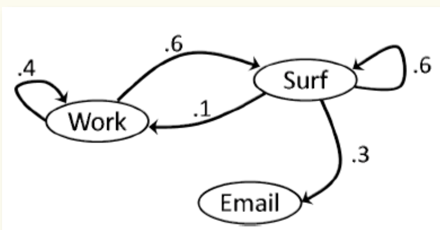
Agenda

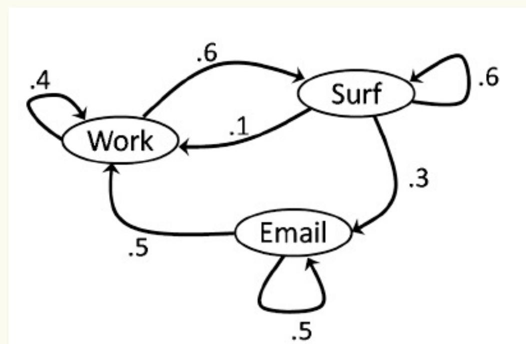
- Recap Markov Chains
- Application: PageRank
- A bit of auction theory

A typical day in my life….

time $t = 0$

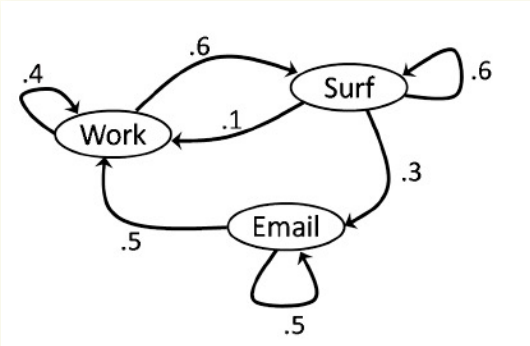






A typical day in my life

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

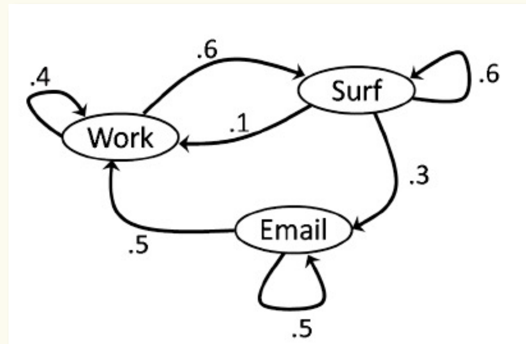
– If I am in some state s at time t

- 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.

This diagram looks vaguely familiar if you took CSE 311 …

Markov chains are a special kind of *probabilistic (finite) automaton*

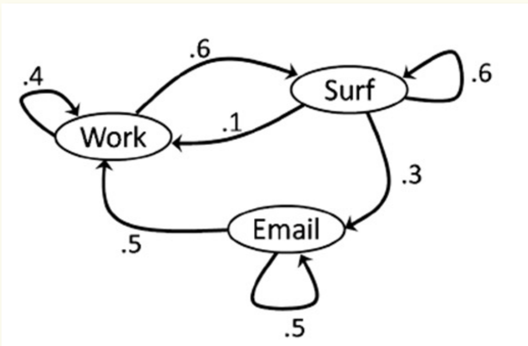


The diagrams look a bit like those of Deterministic Finite Automata (DFAs) you saw in 311 except that…

- There are no input symbols on the edges
  - Think of there being only one kind of input symbol "another tick of the clock" so no need to mark it on the edge
- They have multiple out-edges like an NFA, except that they come with probabilities

But just like DFAs, the only thing they remember about the past is the state they are currently in.

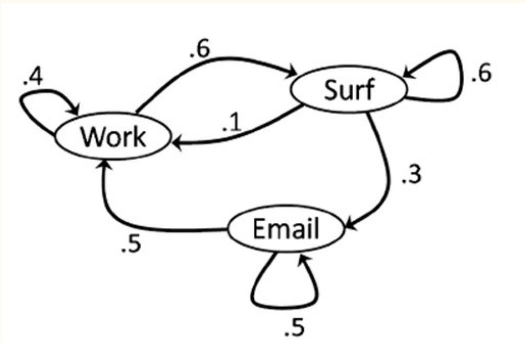
Many interesting questions about Markov Chains



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?**

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



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

Vector-matrix multiplication

Write $\boldsymbol{q}^{(t)} = [q_W^{(t)}, q_S^{(t)}, q_E^{(t)}]$

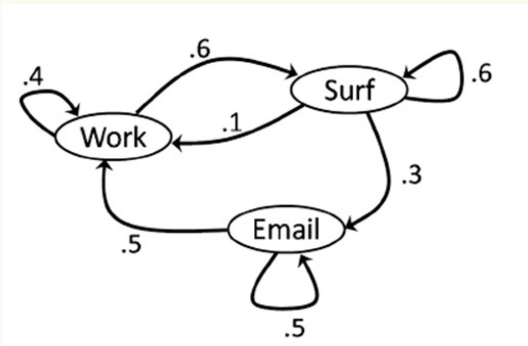
Then for all $t \geq 0$, $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}\boldsymbol{M}$

$q_W^{(t)} = P(X^{(t)} = \text{Work})$

$q_S^{(t)} = P(X^{(t)} = \text{Surf})$

$q_E^{(t)} = P(X^{(t)} = \text{Email})$

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?**

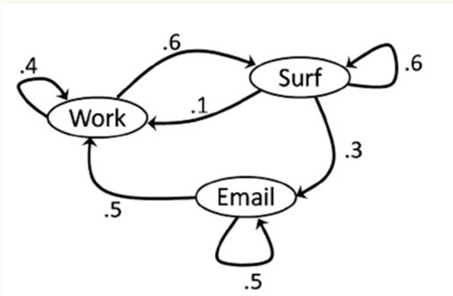
$$\boxed{\boldsymbol{q}^{(t)} = \boldsymbol{q}^{(0)} \boldsymbol{M}^t \text{ for all } t \geq 0}$$

What does $\boldsymbol{q}^{(t)}$ look like for really big t ?

$M^t_{SE} = P(E \text{ at time } t \mid \text{surfing at time } 0)$

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

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



$\boldsymbol{M}$

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

$\boldsymbol{M}^2$

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

$\boldsymbol{M}^3$

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

$\boldsymbol{M}^{10}$

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

$\boldsymbol{M}^{30}$

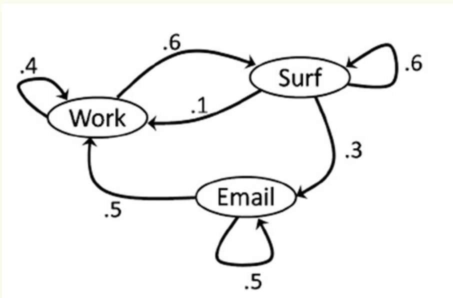
$$\begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .29411764705 & .44117647059 & .26470588235 \\ .29411764706 & .44117647058 & .26470588235 \\ .29411764706 & .44117647059 & .26470588235 \end{pmatrix} \end{array}$$

$\boldsymbol{M}^{60}$

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

What does th
say about $\boldsymbol{q}^{(t}$

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



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

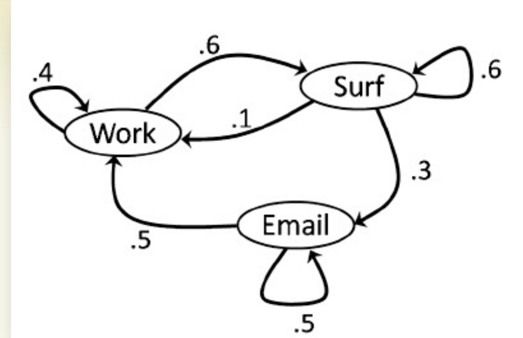
$$[q_W^{(0)}, q_S^{(0)}, q_E^{(0)}] \cdot \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix}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

- In the long run, the starting state doesn't really matter!!
- So at any point in the (slightly distant) future, the change I'm surfing the web is about 44%

- Suppose that we believe that the distribution on states converges to some fixed probability vector $\boldsymbol{\pi} = (\pi_W, \pi_S, \pi_E)$

- Can we figure out what $\boldsymbol{\pi}$ vis just by looking at M?

Observation



If $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}$ 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 $\boldsymbol{\pi} = \boldsymbol{\pi}\, \boldsymbol{M}$

Solving for Stationary Distribution

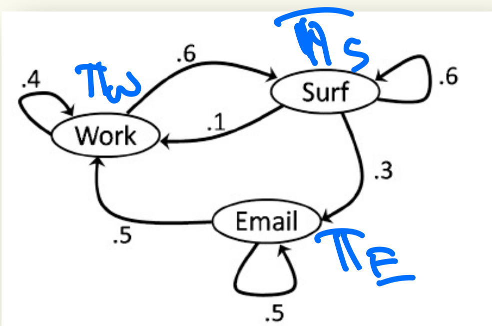
$$(\pi_W, \pi_S, \pi_E) \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{pmatrix} = (\pi_W, \pi_S, \pi_E)$$

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

$$\pi_W \cdot 0.6 + \pi_S \cdot 0.6 + \pi_E \cdot 0 = \pi_S$$

$$\pi_W \cdot 0 + \pi_S \cdot 0.3 + \pi_E \cdot 0.5 = \pi_E$$

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



Solving for Stationary Distribution

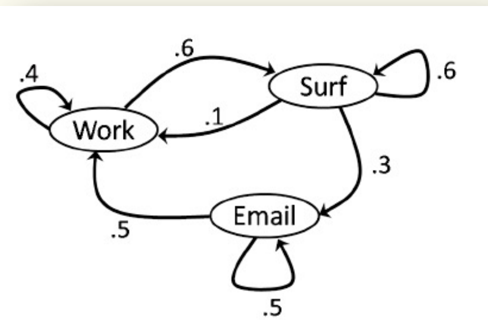
$$(\pi_W, \pi_S, \pi_E) \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{pmatrix} = (\pi_W, \pi_S, \pi_E)$$

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

$$\pi_W \cdot 0.6 + \pi_S \cdot 0.6 + \pi_E \cdot 0 = \pi_S$$

$$\pi_W \cdot 0 + \pi_S \cdot 0.3 + \pi_E \cdot 0.5 = \pi_E$$

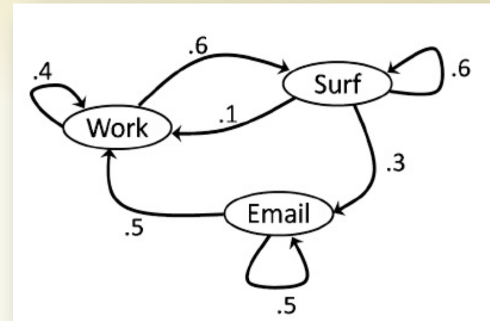
$$\pi_W + \pi_S + \pi_E = 1$$



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

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

Markov Chains recap



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

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

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

- $\boldsymbol{M}^{t}{}_{ij}$ = Pr (in state j after t steps| start in state i).
- Nice Markov chains are not sensitive to initial distribution of states. $\boldsymbol{M}^t \rightarrow \boldsymbol{W}$, where all rows in $\boldsymbol{W}$ are the same probability vector $\boldsymbol{\pi}$
- A **stationary distribution** $\boldsymbol{\pi}$ is the solution to:

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

$$\boldsymbol{M}^{60} \quad \begin{array}{c} \\ W \\ S \\ E \end{array} \begin{array}{c} \begin{array}{ccc} W & S & E \end{array} \\ \begin{pmatrix} .294117647058823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \\ .294117647068823 & .441176470588235 & .264705882352941 \end{pmatrix} \end{array}$$

The Fundamental Theorem of Markov Chains

Theorem. Any nice* Markov chain
has a <u>unique</u> stationary distribution $\boldsymbol{\pi}$.
Moreover, as $t \to \infty$, for all i, j, $\lim_{t\to\infty} \boldsymbol{M}_{ij}^t = \pi_j$

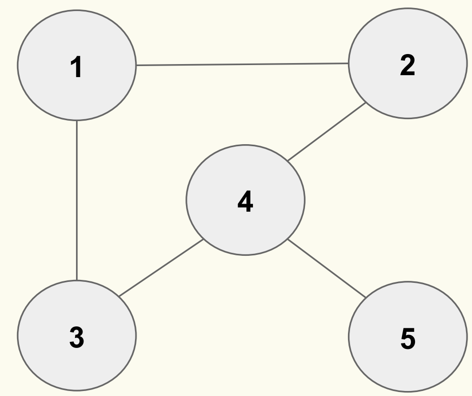
connected

aperiodic and irreducible: 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.



Agenda

- Recap Markov Chains
- Application: PageRank
- A bit of auction theory

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)

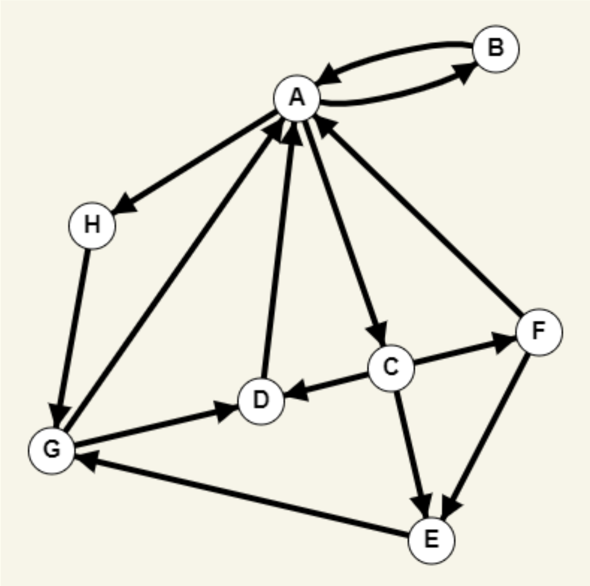
– MacArthur genius prize, Nevanlinna Prize and many other academic honors

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.

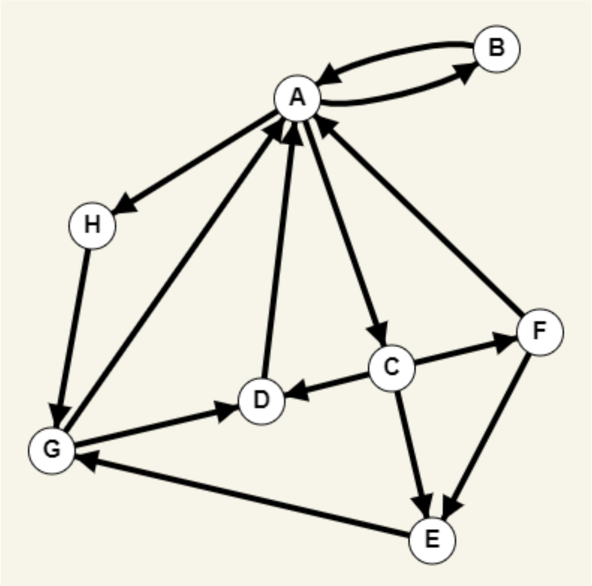
Trust the Internet itself to define what is useful via its links.



PageRank - Idea

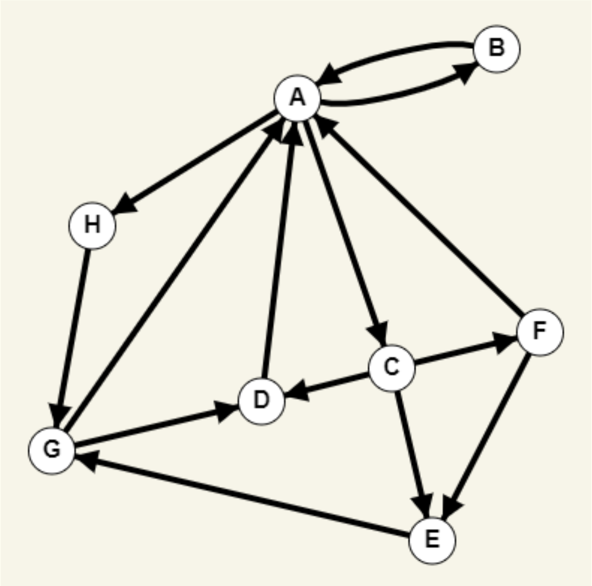
Idea 1 : Think of each link as a citation "vote of quality"

Rank pages by in-degree?



PageRank - Idea

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



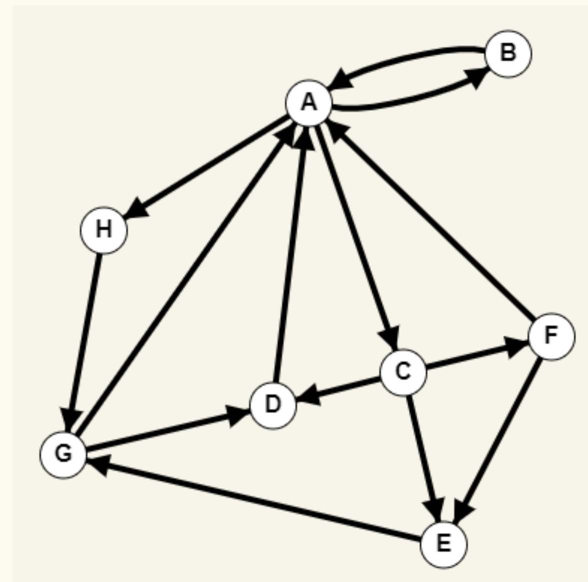
Rank pages by in-degree?

Problems:

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

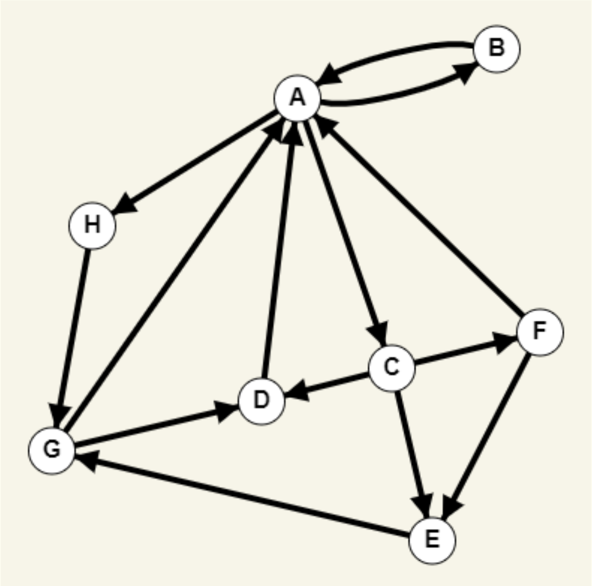
PageRank - Idea

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

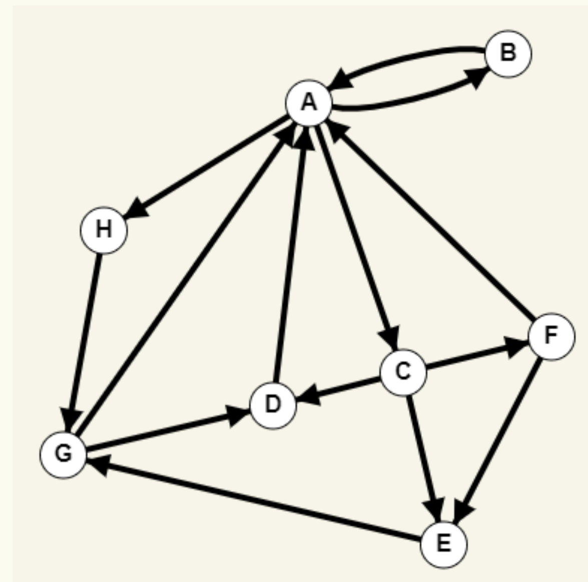


2. A page is high quality if it links to lots of high quality pages

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?

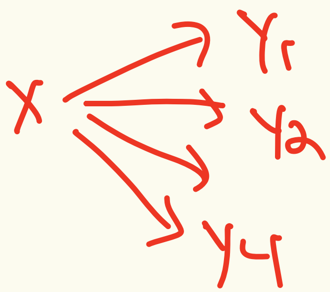
  We want to take into account the importance of x too…

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

q_x: quality of page x

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

$$\overbrace{\left(q_A, q_B, q_C \dots\right)}^{\vec{q}}$$



$$\begin{array}{c c} & \begin{array}{cccc} Y_1 & Y_2 & Y_3 & Y_4 \end{array} \\ X & \begin{pmatrix} 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0\frac{1}{4} & \frac{1}{4} \\ \alpha & \frac{1}{n} & \frac{1}{n} & & & \frac{1}{n} \end{pmatrix} \end{array}$$

$$M$$

$$\vec{q} = \vec{q} M$$

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?**

stationary prob vector

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"

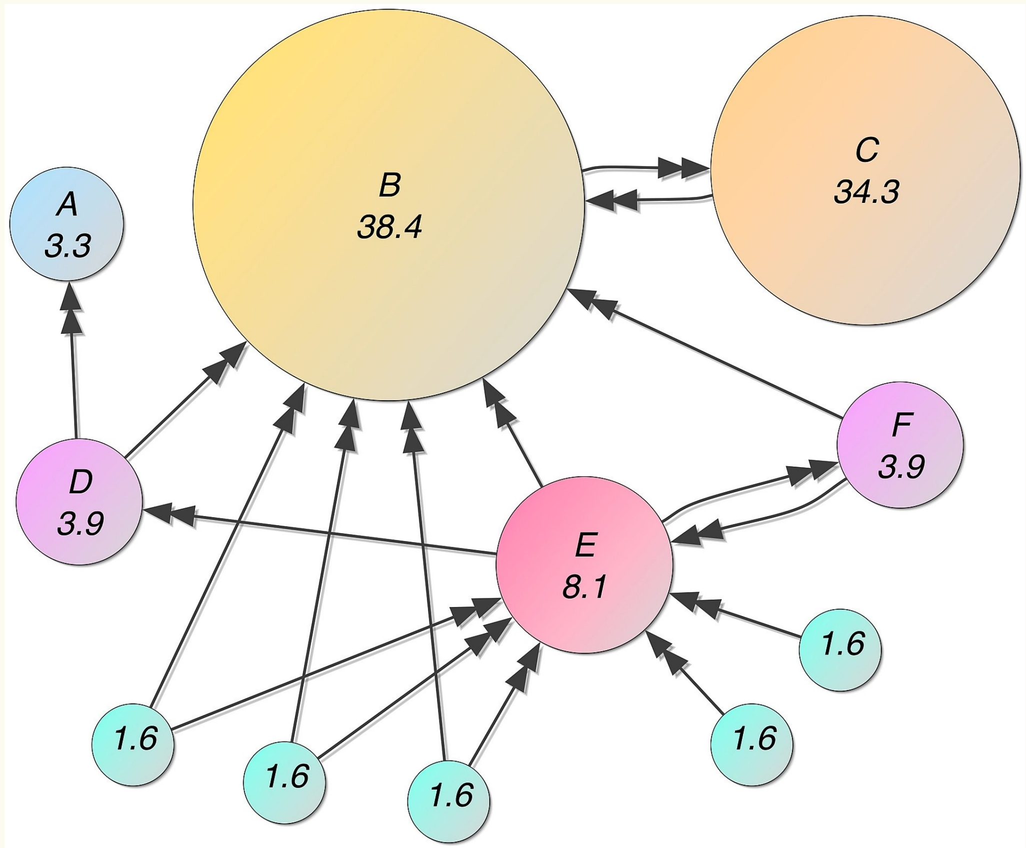
Final PageRank Algorithm

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!

$\pi = (\pi_1, \pi_2 \ldots \pi_n)$

$\pi_i : \cdot$

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

- Recap Markov Chains
- Application: PageRank
- A bit of auction theory

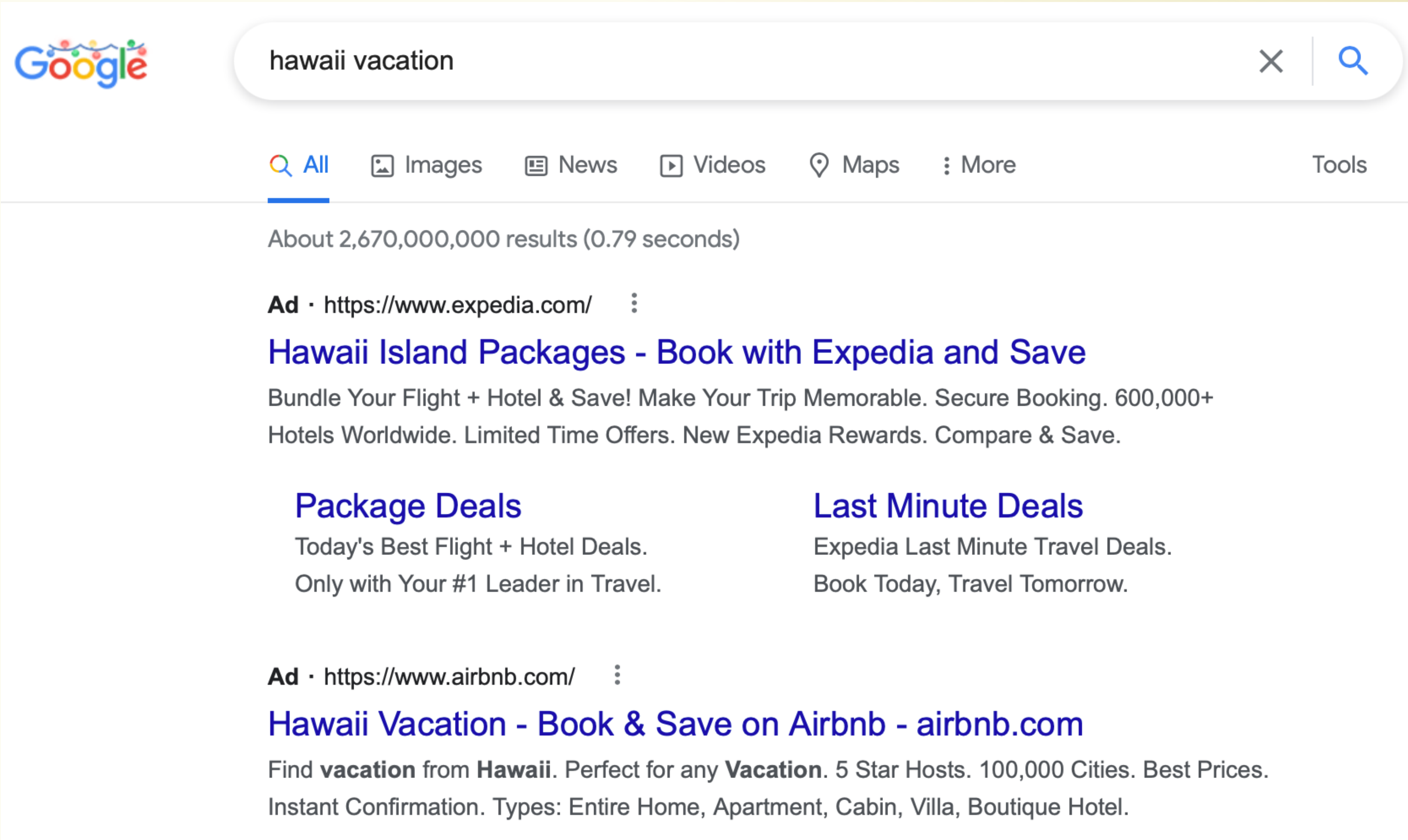
Auctions

- Some goods on eBay and amazon are sold via auction.
- Companies like Google and Meta 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.

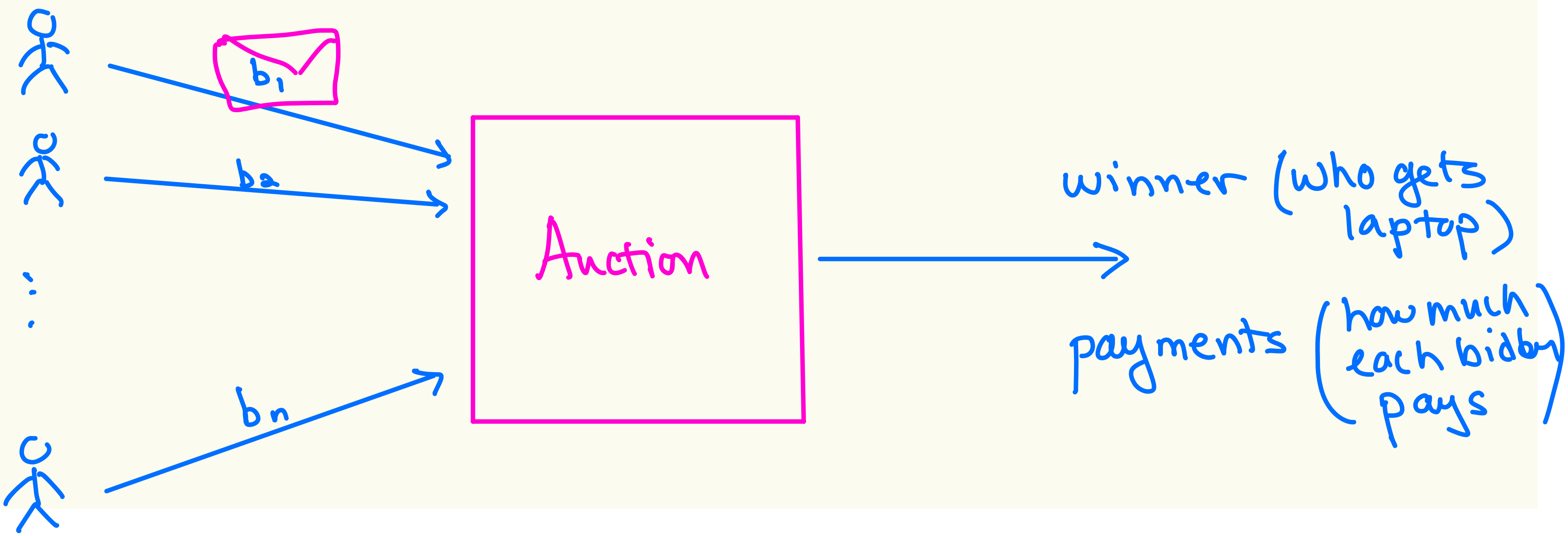

Google
hawaii vacation
All
Images
News
Videos
Maps
More
Tools
About 2,670,000,000 results (0.79 seconds)
Ad · https://www.expedia.com/
Hawaii Island Packages - Book with Expedia and Save
Bundle Your Flight + Hotel & Save! Make Your Trip Memorable. Secure Booking. 600,000+ Hotels Worldwide. Limited Time 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?



mapping from bids $\longrightarrow$ winner payment

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	1	2	3	4
	Bids	100	81	35	24
Payments	1st price	100	0	0	0
	2nd price	81	0	0	0
	All pay	100	81	35	24

Bidder model

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

Bidder is trying to maximize their $\boxed{\text{“utility”}}$ –
the value of the item they get – price they pay.

100 – 86 = \$14

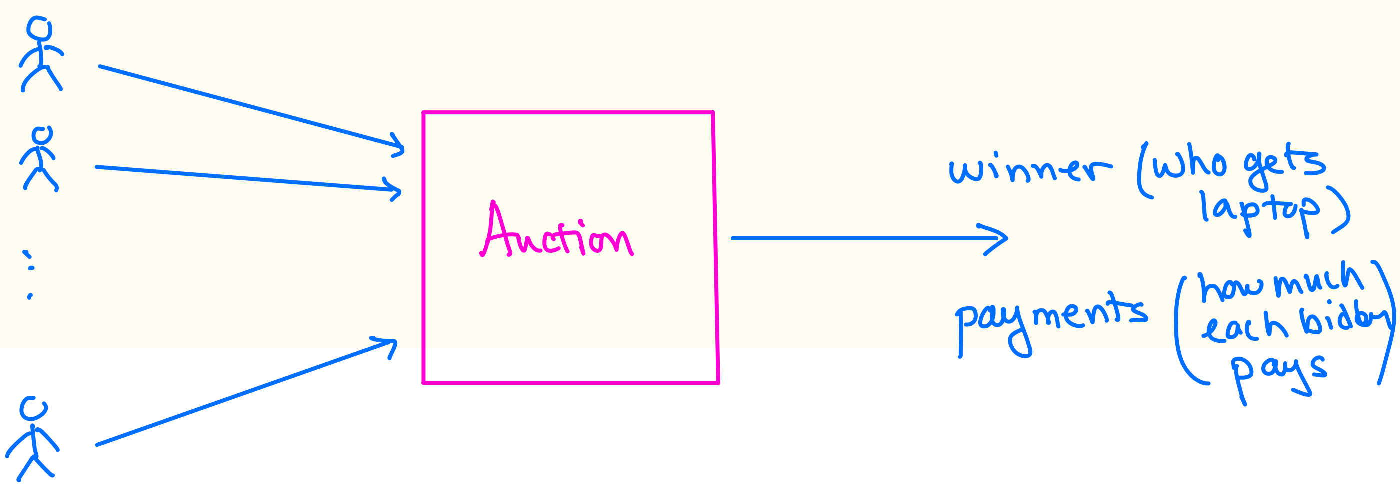
100 – 102 = \$–2

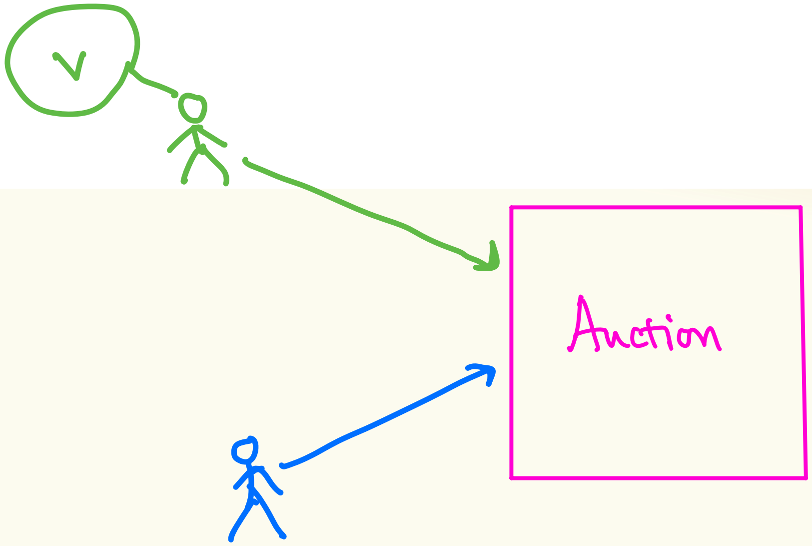
Theorem

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




v
Auction

	2nd price	1st price	all-pay auction
v			
Expected auctioneer reven			

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!