The changing face of web search



Prabhakar Raghavan Yahoo! Research



Reasons for you to exit now ...

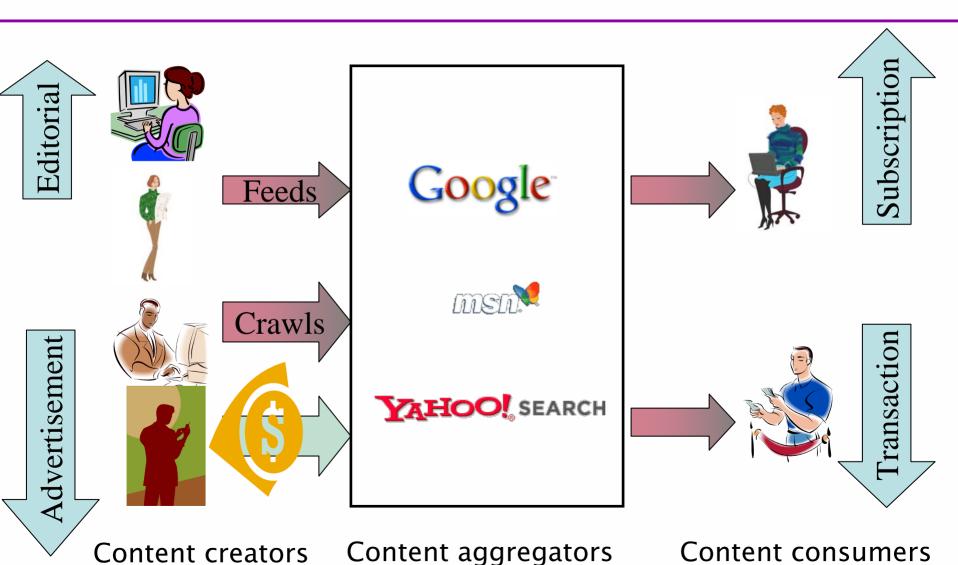
- I gave an early version of this talk at the Stanford InfoLab seminar in Feb
- This talk is essentially identical to the one I gave at STOC 2006 a month ago



- Access to "heterogeneous", distributed information
 - Heterogeneous in creation
 - Heterogeneous in accuracy
 - Heterogeneous in motives
- Multi-billion dollar business
 - Source of new opportunities in marketing
- Strains the boundaries of trademark and intellectual property laws
- A source of unending technical challenges



The coarse-level dynamics



Yahoo! Research

Content consumers



Brief (non-technical) history

- Early keyword-based engines
 - Altavista, Excite, Infoseek, Inktomi, Lycos, ca. 1995-1997
- Paid placement ranking: Goto (morphed into Overture → Yahoo!)
 - Your search ranking depended on how much you paid
 - Auction for keywords: casino was expensive!



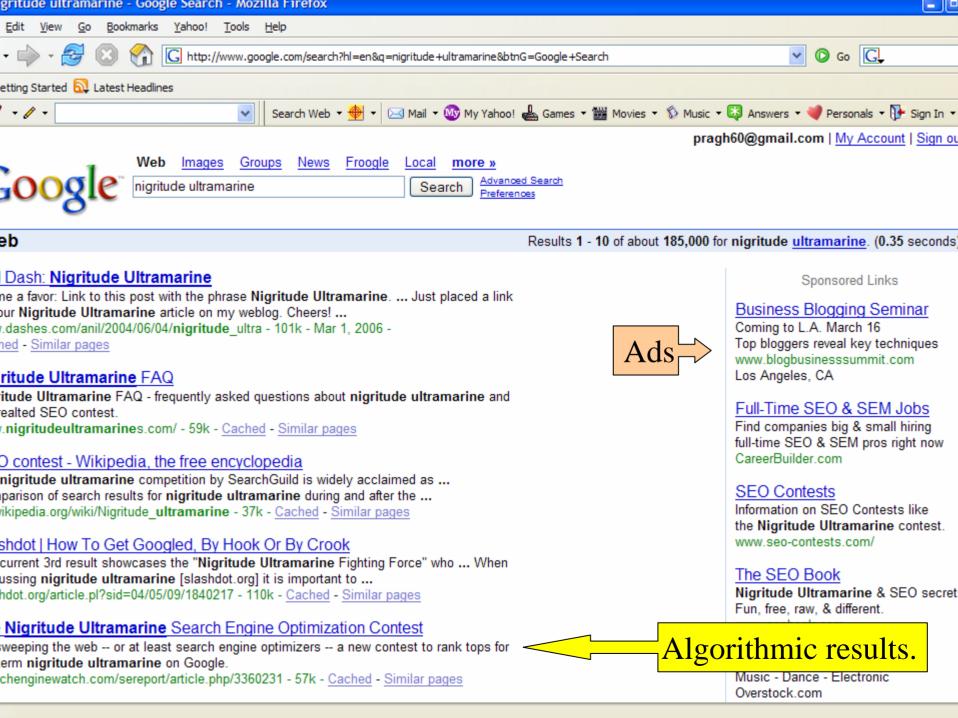
Brief (non-technical) history

- 1998+: Link-based ranking pioneered by Google
 - Blew away all early engines except Inktomi
 - -Great user experience in search of a business model
 - Meanwhile Goto/Overture's annual revenues were nearing \$1 billion



Brief (non-technical) history

- Result: Google added "paid-placement" ads to the side, separate from search results
- 2003: Yahoo follows suit, acquiring Overture (for paid placement) and Inktomi (for search)



"Social" search

Is the Turing test always the right question?

Photos: Explore Flickr • Learn More



Tags / jaguar / clusters

jaguar

SEARCH

(Or, try an advanced search.)











<u>car</u>, <u>cars</u>, <u>auto</u>, <u>etype</u>, <u>automobile</u>, <u>classic</u>, vintage, autoshow, red, show

> See more in this cluster...











zoo, animal, cat, animals, bigcat, seattle, woodlandparkzoo, sleep, edinburgh, caged

> See more in this cluster...











guitar, fender

See more in this cluster...











aircraft, raf

See more in this cluster...

These are the most recent photos tagged with jaguar. See more...















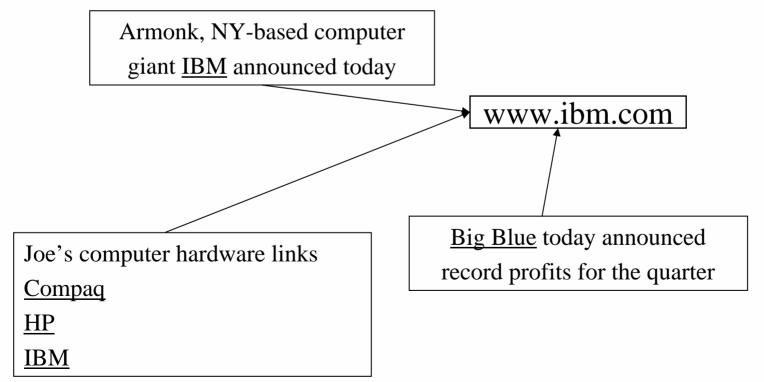
The power of social media

- Flickr community phenomenon
- Millions of users share and tag each others' photographs (why???)
- The wisdom of the crowd can be used to search
- The principle is not new <u>anchor</u> text used in "standard" search
- Don't try to pass the Turing test?



Anchor text

• When indexing a document *D*, include anchor text from links pointing to *D*.





Challenges in social search

- How do we use these tags for better search?
- How do you cope with spam?
- What's the ratings and reputation system?

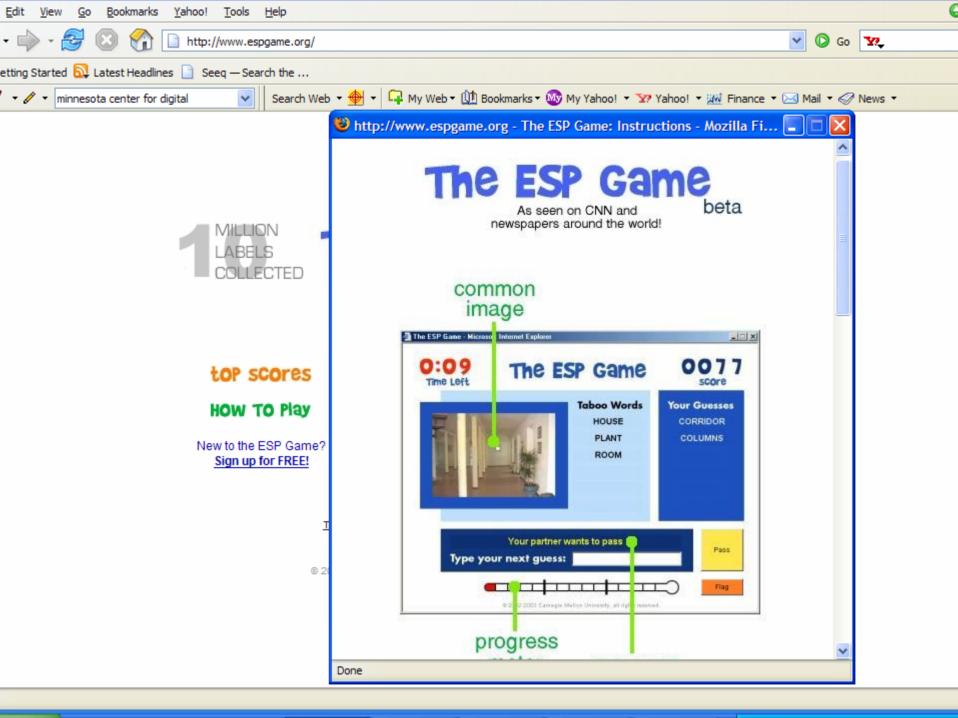
- The bigger challenge: where else can you exploit the power of the people?
- What are the incentive mechanisms?
 - Luis von Ahn (CMU): The ESP Game



Ratings and reputation

- Node reputation: Given a DAG with
 - a subset of nodes called GOOD
 - another subset called BAD
 - Find a measure of goodness for all other nodes.
- Node pair reputation: Given a DAG with a real-valued trust on the edges
 - Predict a real-valued trust for ordered node pairs <u>not</u> joined by an edge



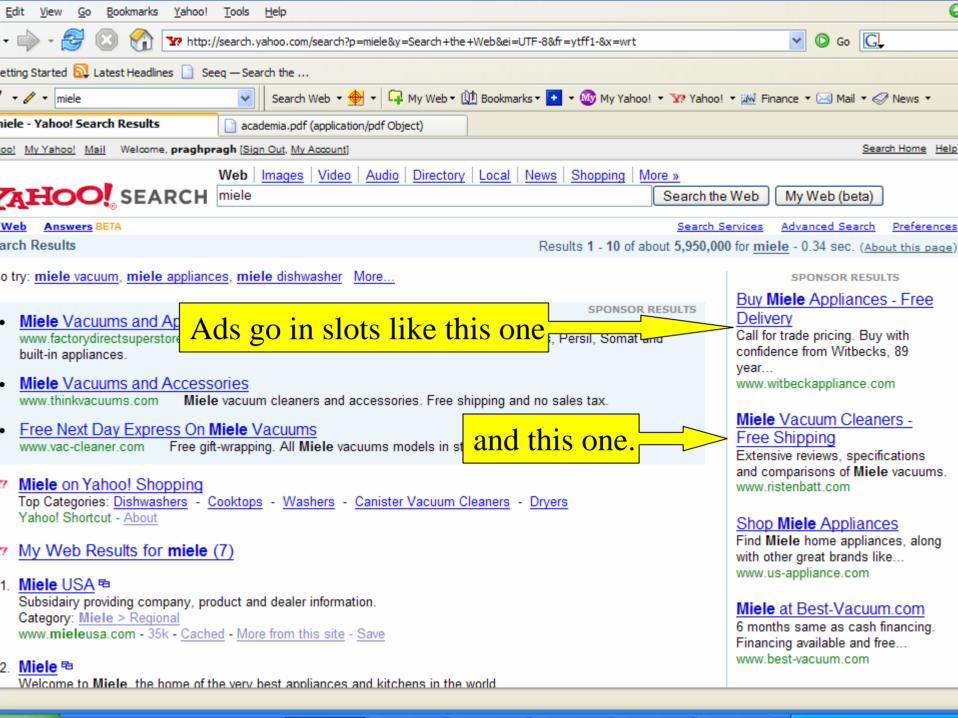


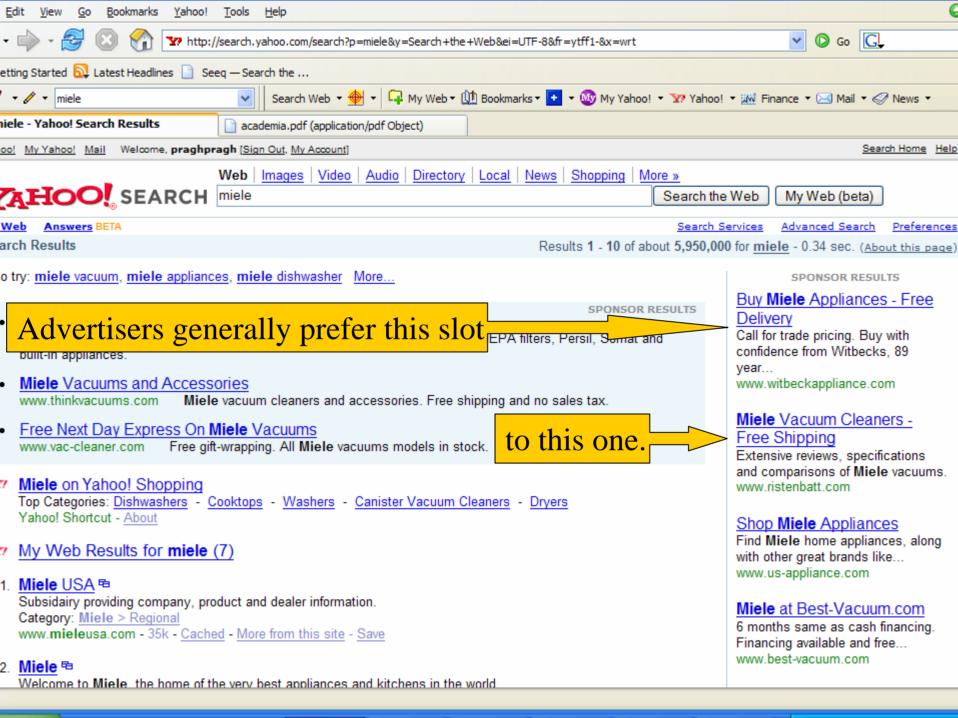
Paid placement

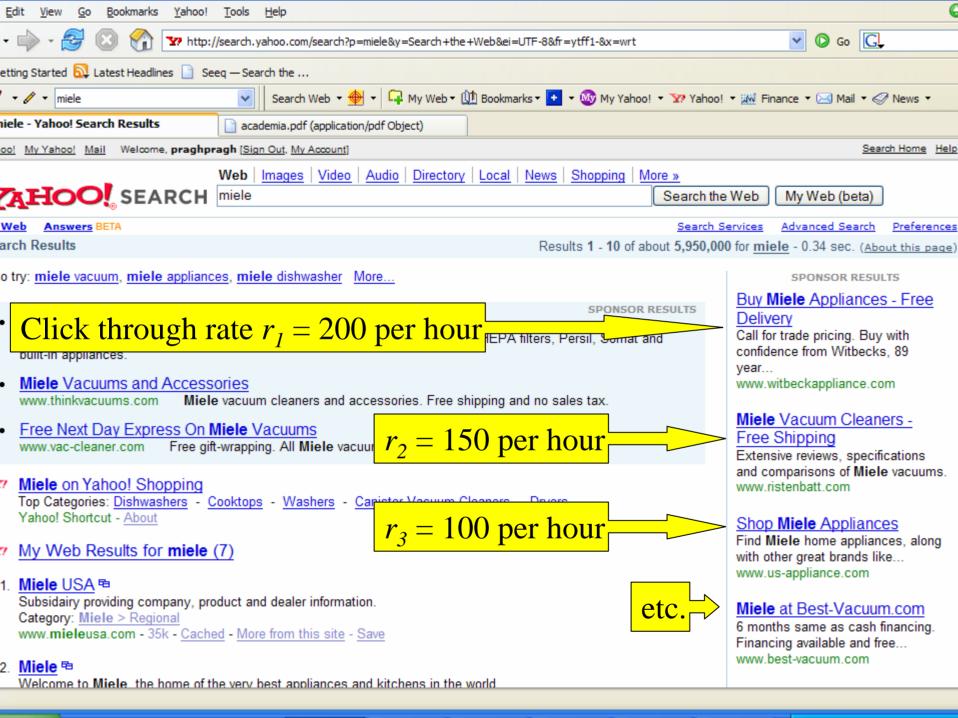
What pays the bills

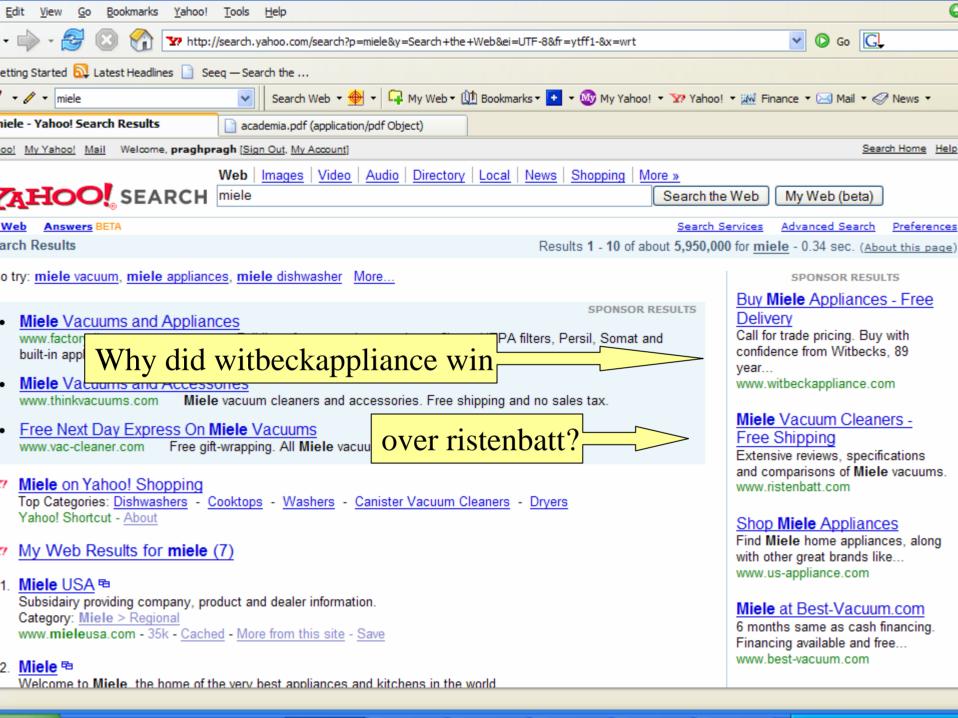


- Of the various advertisers for a keyword, which one(s) get shown?
- What do they pay on a click through?
- The answers turn out to draw on insights from microeconomics







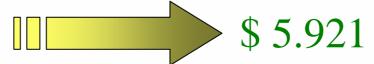


- Click-through rate depends only on the slot, not on the advertisement
- In fact not true; more on this later.



Advertiser's value

- We assume that an advertiser j has a value v_i per click through
 - -Some measure of downstream profit
- Say, click-through followed by
 - 96% of the time, no purchase
 - 0.7% buy Dishwasher, profit \$500
 - 1.2% buy Vacuum Cleaner, profit \$200
 - 2.1% buy Cleaning agents, profit \$1



23



- For the keyword *miele*, say an advertiser has a value of \$10 per click.
- How much should he bid?
- How much should he be charged?

The <u>value</u> of a slot for an advertiser, what he <u>bids</u> and what he is <u>charged</u>, may all be different.

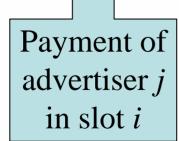


Advertiser's payoff in ad slot i

(Click-through rate) x (Value per click) – (Payment to search engine)

=
$$r_i v_j$$
 – (Payment to Engine)

$$= r_i \ v_j - p_{ij}$$



Function of all other bids.



Two auction pricing mechanisms

Not truthful.

- *First price:* The winner of the auction is the highest bidder, and pays his bid.
- <u>Second price</u>: The winner is the highest bidder, but pays the secondhighest bid.
- Engine decides and announces pricing.
- What should an advertiser bid?



Second-price = Vickrey auction

- Consider first a single advt slot
- Winner pays the second-highest bid
- Vickrey: *Truth-telling is a dominant* strategy for each player (advertiser)
 - -No incentive to "game" or fake bids



Auctions and pricing: multiple slots

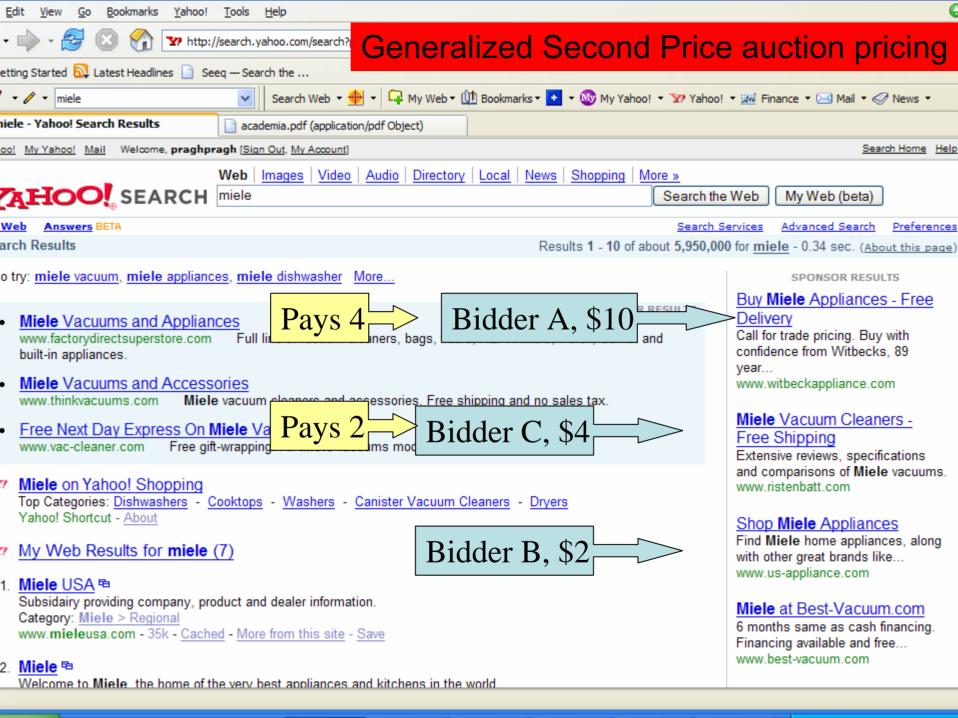
- Overture's (→Yahoo!'s) model:
 - Ads displayed in order of decreasing bid
 - E.g., if advertiser A bids 10, B bids 2, C bids
 4 order ACB
- How do you price slots? Generalized <u>Vickrey</u>?
 - Generalized second-price (GSP)
 - Vickrey-Clark-Groves (<u>VCG</u>): each advertiser pays the *externality* he imposes on others



- Suppose click rates are <u>200</u> in the top slot, <u>100</u> in the second slot
- VCG payment of the second player (C) is 2 x 100 = 200 Externality on third player B.
- For the first player, 4x(200-100) + 200

Externality on C.

Externality on B.





- Truth-telling is a dominant strategy under VCG ...
- Truth-telling not dominant under GSP!

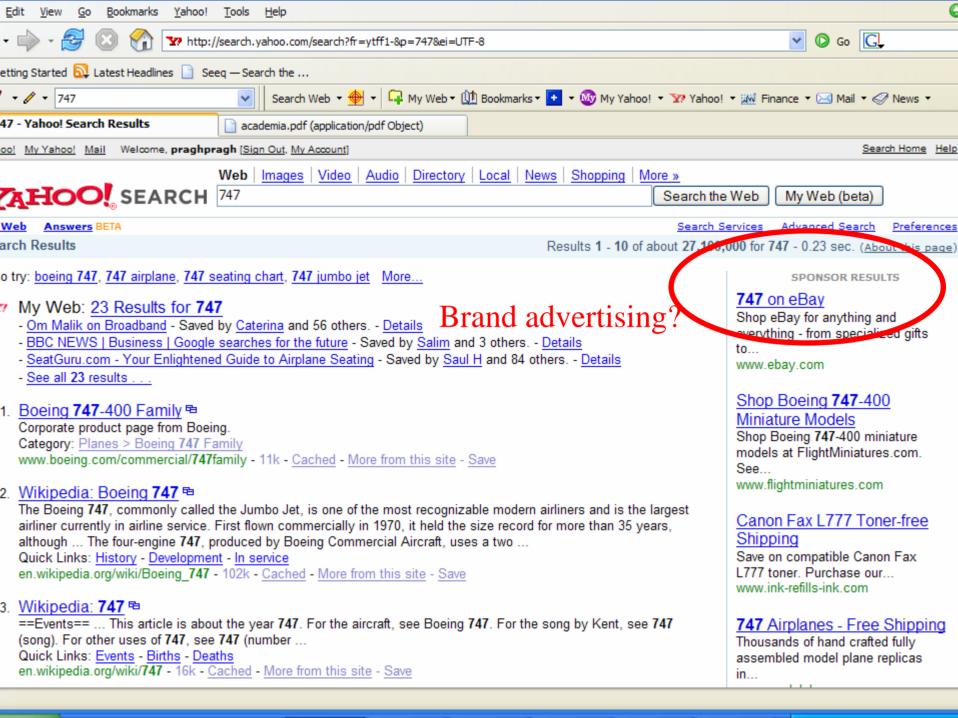
Aggarwal, Goel, Motwani (ACM EC 2006): give a truthful mechanism in a model that precludes VCG.

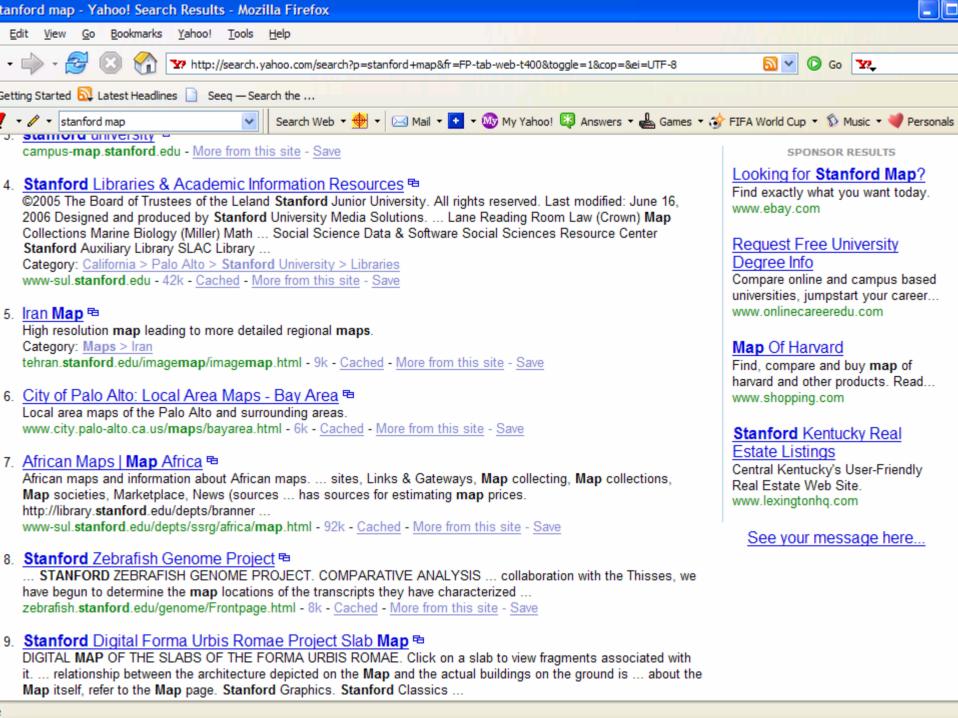


- Static equilibrium of GSP is locally envy-free: no advertiser can improve his payoff by exchanging bids with advertiser in slot above.
- Depending on the mechanism, revenue varies: GSP ≥ VCG.

Locally envy-free mechanisms correspond to *Stable Marriage* solutions.

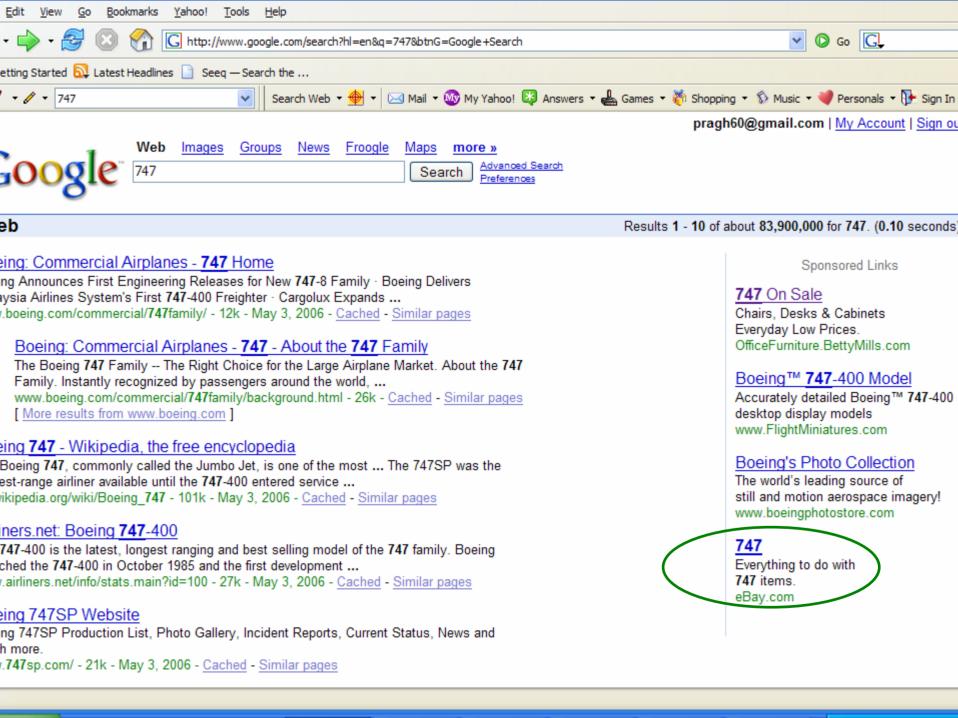
- What's good about bid-ordering and GSP?
 - Advertisers like transparency
- What's wrong with bid-ordering?

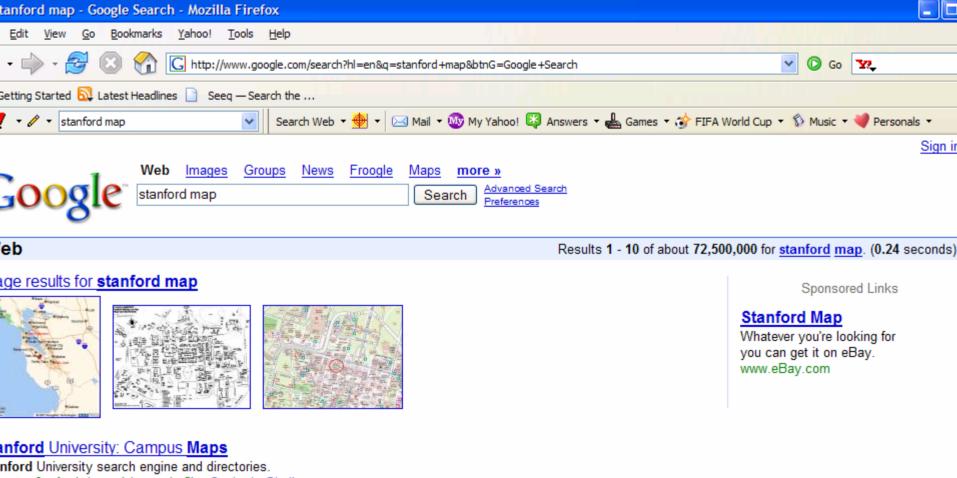




Revenue ordering

- Simplified version of Google's ordering
 - Each ad j has an expected clickthrough denoted CTR_j
 - Advertiser j's bid is denoted b_j
- Then, expected revenue from this advertiser is $R_j = b_{j+1} \times CTR_j$
- Order advertisers by R_i
 - Payment by GSP





v.stanford.edu/search/maps/ - 5k - Cached - Similar pages

Stanford University: Directions to Campus

Stanford University search engine and directories. ... Maps and Directions. Maps | Driving to Campus | Public Transportation | From Local Airports ... www.stanford.edu/home/visitors/directions.html - 10k - Cached - Similar pages

STANFORD UNIVERSITY

Format: PDF/Adobe Acrobat - View as HTML

/Stanford Man Librarios Group

Parking, Visitor, Parking, Produced by, Stanford University, Maps and Records, July

Bus Parking, Visitor, Parking, Visitor Parking, Undergraduate ...

s.stanford.edu/images/03-04 Visitor-Map.pdf - Similar pages



Still primitive understanding

- Advertisers' bids generally placed by robots
 - Currently approved by Engines
 - No room for coalitions
- Granularity of markets to bid on
- Pricing when the number of ad slots is variable 🤙



Burgeoning research area

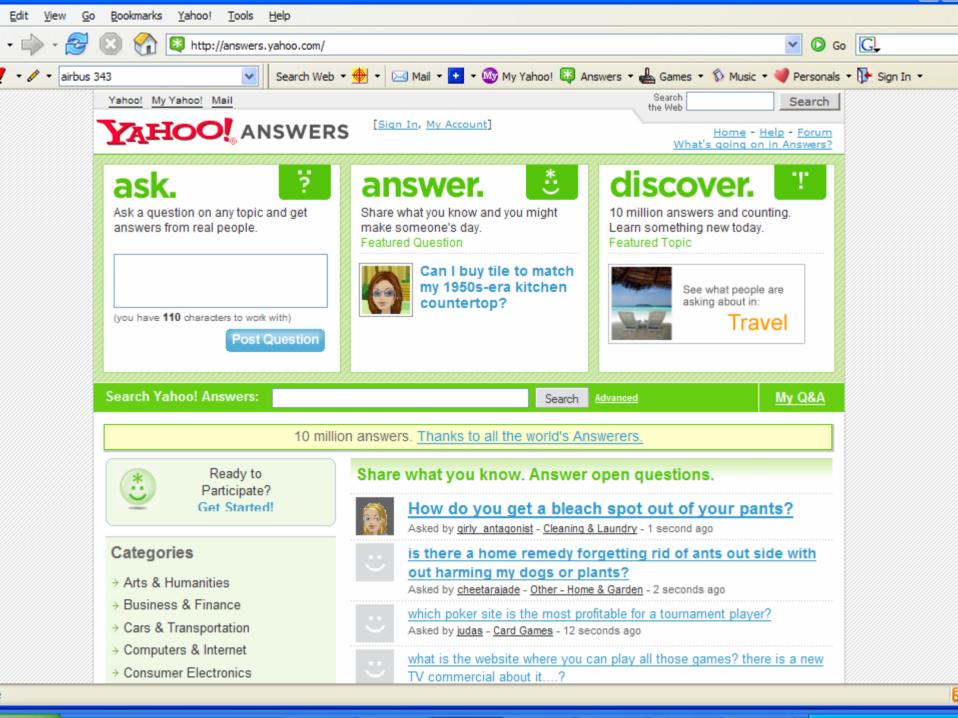
- Marketplace design
 - Multi-billion dollar business, growing fast
 - -Interface of microeconomics and CS
- Many open problems, a few papers, some of them quite realistic





Incentive networks

Joint w/Jon Kleinberg (FOCS 2005)

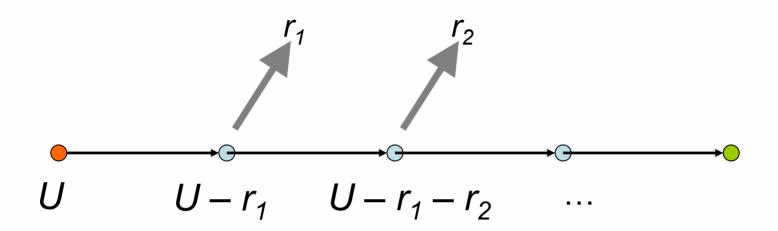




The power of the middleman

- Setting: you have a need
 - -For information, for goods ...
- You initiate a request for it and offer a reward for it, to some person X
 - -Reward = your value U for the answer
- How much should X "skim off" from your offered reward, before propagating the request?



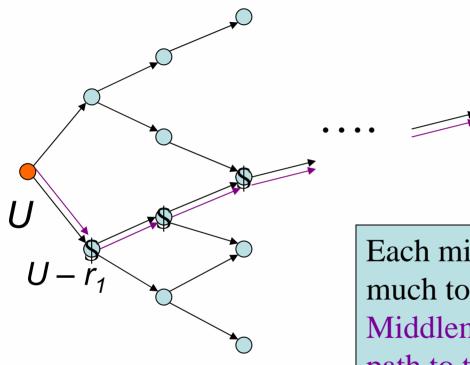


Request propagated repeatedly until it finds an answer. Target not known in advance.

Middlemen get reward only if answer reached.



More generally



Each middleman decides how much to "skim off".

Middleman only gets paid if on the path to the answer.

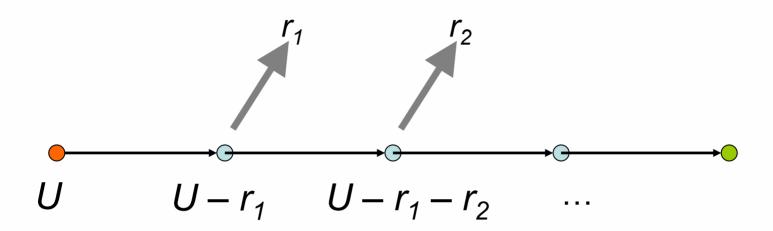


Rewards must be non-trivial

- We will assume that all the $r_i \ge 1$.
- Else, have a form of Zeno's paradox:
 - Source can get away with offering an arbitrarily small reward.
- Equivalently, nodes value their effort in participating.



Back to the line



Under *strategic* behavior by each player, how much should a player skim?

n = answer rarity: probability a node has the answer = 1/n, independently of other nodes.



- For rarity n, it takes about n hops to get to the answer.
- Initial reward must be exponential
 - —A very inefficient network. failure probability.

For a constant

Branching processes

- Branching process: a network where
- Each node has a number of descendants
- Number of descendants is a random variable X
 - drawn from a probability distribution
 - -Expectation[X] = b

Branching processes

 Classical study of population dynamics and random graph evolution.

Basic fact:

- -If b < 1, process dies out
- -If $b \ge 1$, process infinite.



Main results - unique Nash

- For *b*<2, the initial investment <u>must</u> be exponential in the path length from the root to the answer.
- For *b*>2, the initial investment is <u>linear</u> in the path length from the root to the answer.

Criticality at b=2.

Knowing fewer than 2 people is expensive.

Yahoo! Research ⁵¹

- (Sufficient) competition makes incentive networks efficient.
- But ... we haven't fully introduced competition yet.
 - On trees, we have a unique path from the origin to each node.

Yahoo! Research ⁵²



- Full model of competition
 - –When does competition promote efficiency?
- Given a DAG, how does a node compute its strategy?



- Web search is scientifically young
- It is intellectually diverse
 - The human element
 - The social element
- The science must capture economic, legal and sociological reality.

Yahoo! Research ⁵⁴

Thank you.



Questions?

pragh@yahoo-inc.com

http://research.yahoo.com