



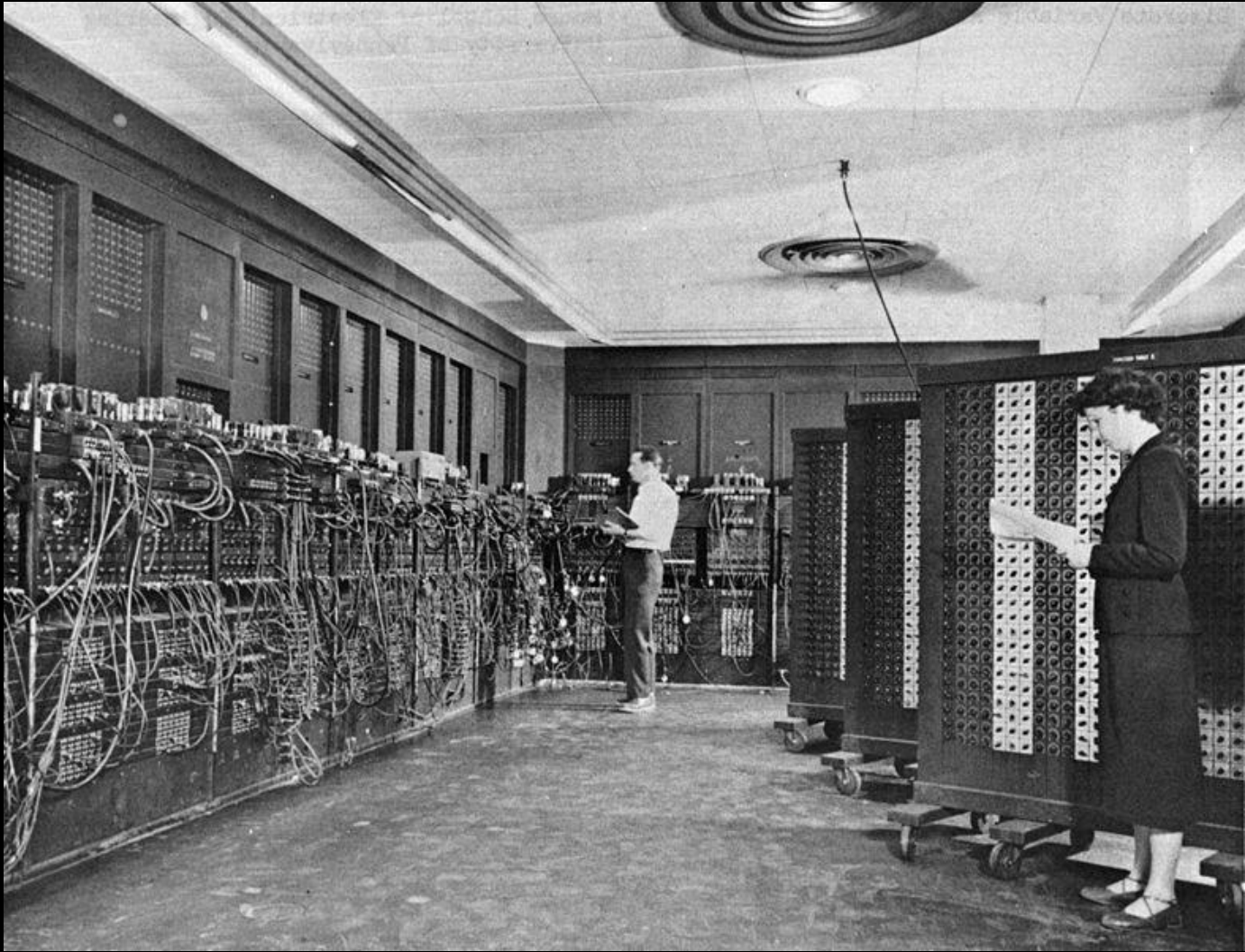
The quantitative study of user behavior online

user behavior online

The quantitative study of

Prabhakar Raghavan
Yahoo! Labs

Over 50 years of computers



ENIAC: Less computing power than an iphone



Today – Yahoo! data center

Thousands of computers, each million times faster than ENIAC



Some trends

- Moore's Law: Computers double in speed every 18 months
 - Every 15 years, computers get 1000 times faster!!
- All information in the world about everything, is available everywhere, forever
 - The millions of computers at Google, Yahoo, Facebook, Amazon ... are available to you all the time



The most important trend

- Most computer cycles are used
 - Not for computing ...
 - ... but for communications
 - Not by scientists or specialists
 - ... but by ordinary consumers
- Ergo: the big question is not
 - What can be computed ...
 - ... but what users will do with computing

The scientific shift

- Computer scientists study computation
- Social scientists study people
- Now – we must study the combination!

In this lecture

- Some vignettes of such studies
- **Why this is hard**
 - But important
- **The new opportunity: the Web**
 - Biggest observatory of social behavior

Learning about users on the Web

Depth

- **Lab/field studies**
 - Eye-tracking, interviews
 - Pursue defined tasks
- **Instrumented panels**
 - Toolbars, clients
 - Log what users do on their own
- **Click logs**
 - Can experiment with interfaces
 - No clue why people click what they click

Scale

Grimes et al. 2007

- “the field study, the instrumented user panel, and the raw query log, provide complementary sources of data”
- “logs can only measure the how and the what, rather than the why. For example, if we have a sequence of queries, we only know the sequence of queries, but we have no evidence of why the user is typing in that particular sequence.”



Vignettes

General theme

- What questions can we answer by these methods?
- Quantitative sociology meets data mining

All behavior has social meaning
Alfred Adler



Long Tail, Heavy Tail?

Observation:

Many products, each purchased very few times,
together make up a large fraction of all purchases

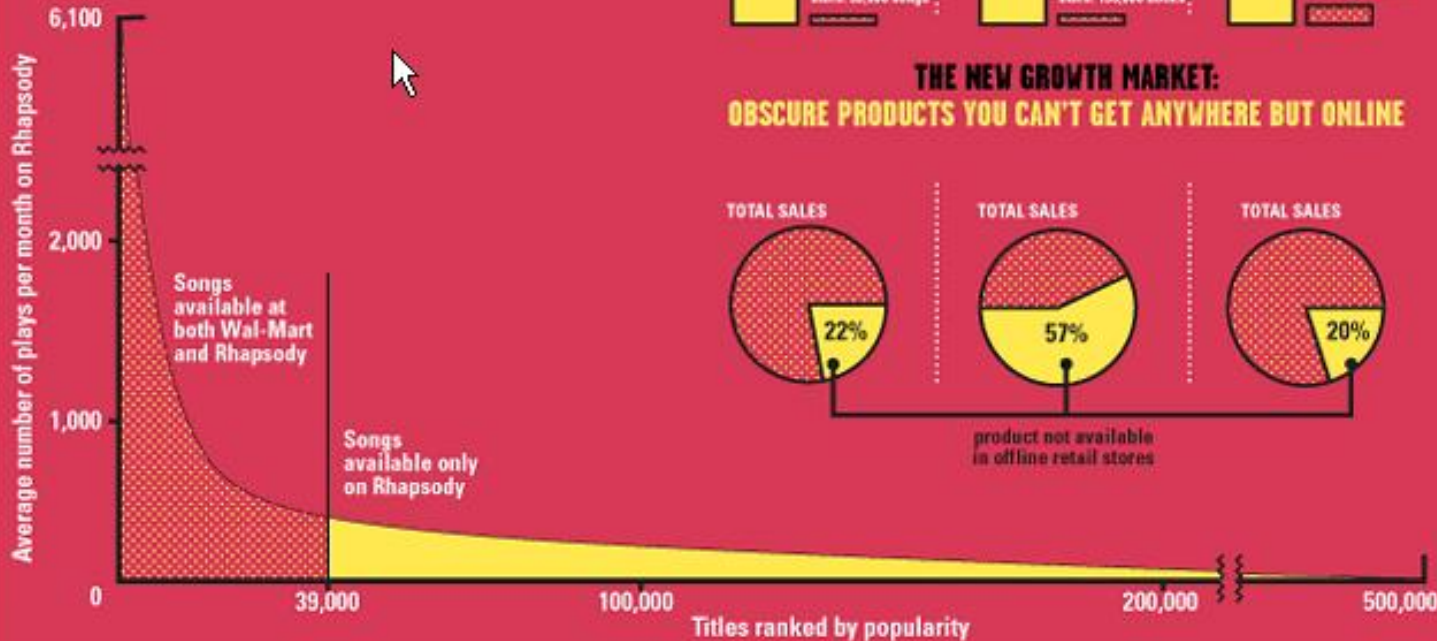
Movies ... Songs ... Queries to a search engine

....

“The Long tail”

ANATOMY OF THE LONG TAIL

Online services carry far more inventory than traditional retailers. Rhapsody, for example, offers 19 times as many songs as Wal-Mart's stock of 39,000 tunes. The appetite for Rhapsody's more obscure tunes (charted below in yellow) makes up the so-called Long Tail. Meanwhile, even as consumers flock to mainstream books, music, and films (right), there is real demand for niche fare found only online.



How Endless Choice Is Creating Unlimited Demand

The Long Tail



Why the Future of Business
Is Selling Less of More

CHRIS ANDERSON

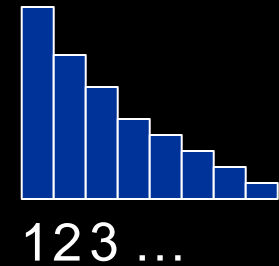
"Anderson's insights influence Google's strategic thinking in a profound way.
READ THIS BRILLIANT AND TIMELY BOOK."
—ERIC SCHMIDT, CEO, GOOGLE



Chris Anderson, Wired, Oct 2004

Heavy tails

- Decreasing histograms over numbers $[1, n]$
 - E.g., movie popularities

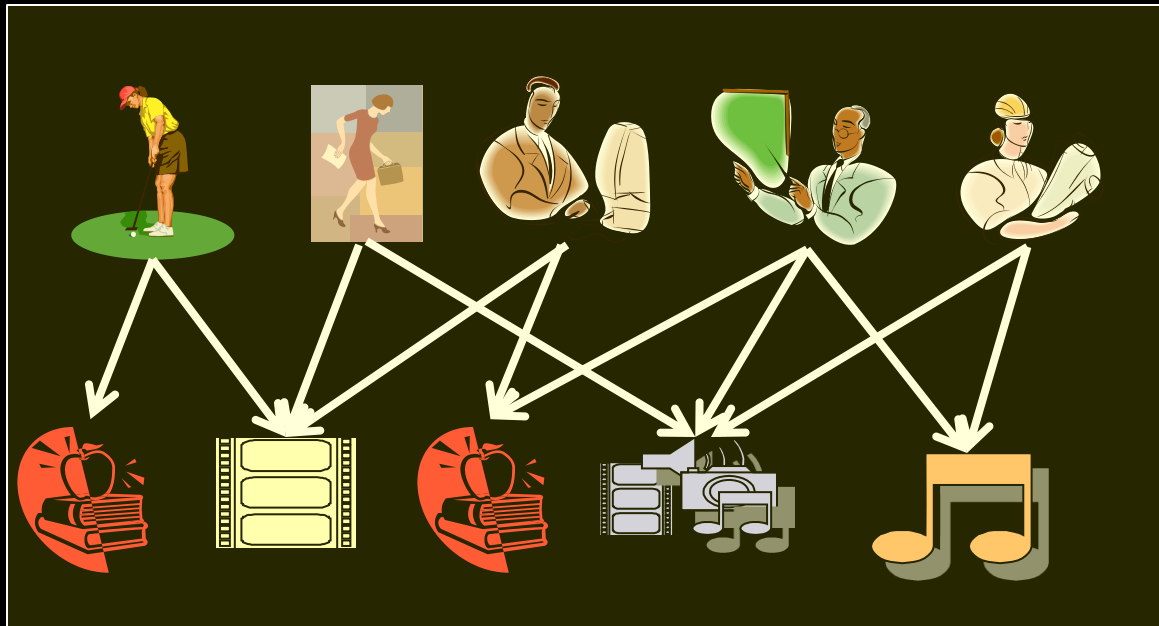


- For any fixed k (say, 10000), the fraction in all buckets $>k$ is quite big
- Arise in observed statistics arising from human behavior
 - Number of friends, search queries, popular songs, books ...

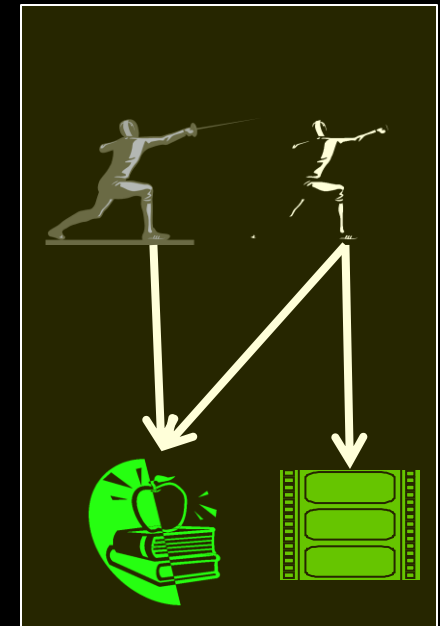
Heavy tail of user interests

- Many queries, each asked very few times, make up a large fraction of all queries
 - Movies watched, blogs read, words used ...

One explanation



Normal people

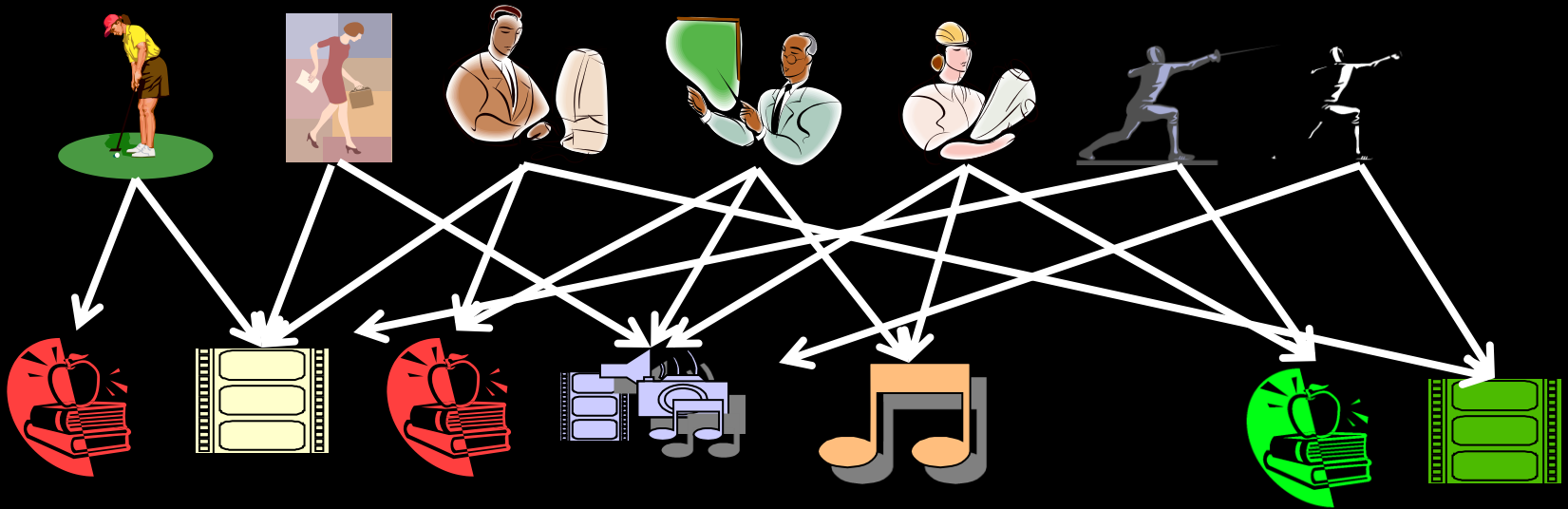


Weirdos

Heavy tail of user interests

- Many queries, each asked very few times, make up a large fraction of all queries
 - Movies watched, blogs read, words used ...

The reality



We are all partially weird!

Broder et al 2009

Why the heavy tail matters



If you're building a market, you chase the tail ...

Not because the worst-sellers make a lot of money

But because they matter to a lot of people

The change for Computer Science

- **Social phenomena repeatedly create such heavy tails**
 - We can observe them on the Web at scale
 - Traditional analysis is inadequate
- **We need a new style of analysis**
 - New experimental tools
 - New mathematical tools



Social influence

Six degrees of separation ...

Social attention and influence

- In 1940's Paul Lazarsfeld :
 - “Who talks to whom about what, and with what effect?”
- Difficult to answer
 - Measuring “who talks to whom” hard at scale
 - Measure “who influences whom” even harder
- Web 2.0 brings the answer within reach
 - Fascinating implications for web companies, users ... and for science

Is it a small world?

ORIGIN OF 6 DEGREES

- 1960's: Milgram and Travers "small world" experiment
- Subjects given letter for target individual
- Could only send to a friend
- Protocol generated 300 "letter chains" - 64 reached target
- Led to the famous "six degrees" phrase

6 DEGREES – WEB EDITION

- 2001-02: recreate w/email
- Milgram: one target, 300 chains
- Now: 18 targets around world, 24,163 chains, 61,168 hands, 166 countries
- 400 reached targets

Dodds *et al.* 2001

From Connections → Influence

Cultural Markets (songs ...)

- “Hits” are many times more successful than average
 - Success seems obvious in retrospect, but hard to predict
- Can inequality and unpredictability be explained by social influence?
- Problem: Experiment would require 1,000s of participants
 - Each “market” requires hundreds of participants
 - Need to compare many markets

From Connections → Influence

- Experiment with subjects asked to rate new songs
- **Control group:** songs presented in random order
- **Treatment group:** songs presented in order of popularity for other users



Salganik et al. 2008, 2009

Influence in cultural markets

- **Individuals are influenced by the choices of others**
 - The stronger the social signal, the more they are influenced
- **Collective decisions are also influenced**
 - Popular songs are more popular (and unpopular songs are less popular)
 - However, which songs become popular becomes harder to predict
- **The paradox of social influence:**
 - Individuals have more information on which to base choices
 - But collective choice (i.e. what becomes popular) reveals less and less about individual preferences
- **Manipulating social influence not so easy**
 - Can create self-fulfilling prophecies at level of individual songs, but not for entire market

Influence and twitter

- Twitter is ideally suited to answer questions about influencers
 - Fully-observable network of individuals who explicitly opt-in to follow each other
 - Twitterers are expressly motivated to be influential
- **Relatively easy to track diffusion**
 - **Popularity of URL shorteners means can track tweets over hops**
- Objective is to predict influence as function of
 - # Followers, # Friends, # Reciprocated Ties
 - # Tweets, Time of joining
 - Past influence score



The Kardashian question

- **Large cascades are rare, hence:**
 - “Social epidemics” are extremely rare
 - Probably impossible to predict them or how they will start
 - Better to trigger many small cascades
- **\$10,000 per tweet isn't good value**
- **But “Ordinary Influencers” are promising**
 - Only influence one other person on average
 - Average influence is close to zero (0.28)

Understanding the observations

- Can create model, prove theorems
- Can address “which influencers should we target?”
 - Can run experiments at scale
- But ... no sociological understanding yet of why/how all this happens
 - No experimental loop



How users see
search results

Image and product search



martin logan ma...10 .jpg
700 x 700 | 32k
hifi4me.de



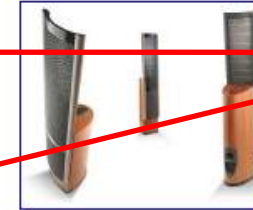
Martin Logan .JPG
294 x 639 | 20k
aheadstereo.com



aerius martin logan .jpg
580 x 438 | 46k
audiocostruzioni.com



martin logan cls .jpg
580 x 420 | 38k
audiocostruzioni.com



martin logan clarity .jpg
600 x 509 | 41k
ixbt.com



giovannits mart...an .jpg
1024 x 768 | 74k
milossaluciano.com



martin logan .jpg
450 x 358 | 83k
audiojunkies.com



Purity Dark Che... 2 .jpg
1024 x 628 | 104k
limar.hu



Fresco i Purity...le .jpg
993 x 768 | 146k
limar.hu



Matinee Purity...le .jpg
884 x 768 | 106k
limar.hu



Helos 10 Lifestyle .jpg
768 x 768 | 87k
limar.hu



Vista Dark Cher...le .jpg
1024 x 661 | 118k
limar.hu



Ordered by decreasing score in row-major order

Users don't see row-major

- Variety of evidence that users' eye scans don't go row-major
 - Eye tracking at search engines
- Visual cues not well understood
- Diversity

More general 2-d layout

- The problem goes beyond image/product results
 - Search engines doing general two-dimensional results presentation
 - Heterogeneous objects being laid out in the results page

Richer use of 2-d real estate

The image shows a screenshot of a Yahoo! search results page for the query "madonna". The page layout is as follows:

- Top Left:** The Yahoo! logo and a search bar containing the text "madonna".
- Left Rail:** A vertical sidebar on the left side of the page. It contains several sections:
 - "View Notes (2)" with a pencil icon.
 - "SearchScan - On" with a green checkmark icon.
 - A red text label "Left rail" pointing to the sidebar.
 - "272,000,000 results for madonna:"
 - "Show All" with a globe icon.
 - Links to "Wikipedia", "YouTube", "MTV", "MySpace", and "IMDb", each with its respective logo.
 - A red text label "News image" pointing to the "MySpace" link.
 - "Video Sites" with a video camera icon.
- Main Content Area:**
 - A large image of Madonna in a black top with a white fur collar. Below it is a "Watch Music Videos" section with three small video thumbnails.
 - A link to "Madonna - Official Site" with the URL "www.madonna.com" and sub-links for "Albums", "Lyrics", "Photos", and "Videos".
 - A "Videos" section with the heading "Play Full-Length Tracks" and a list of songs: "4 Minutes", "Crazy For You", "Papa Don't Preach", and "Live To Tell". A link for "More Madonna songs" is at the bottom.
 - A "Madonna - News Results" section with a small thumbnail image of Madonna. It contains three news items:
 - "Madonna - Dolce Gabbana Feared Hiring Madonna For Ad Campaign" by ContactMusic - 4 hours ago.
 - "Angelina Jolie and Brad Pitt 'To Do a Madonna'" by Showbiz Spy - 7 hours ago.
 - "Madonna - Haiti Telethon Success" by ContactMusic - Jan 23 06:57am.A red text label "News" points to the second news item.
 - A link to "Madonna (Entertainer) - Wikipedia" with a short bio: "Exhaustive bio and discography of Madonna's early life, career, 'Sex' controversy, electronic club mix phase, and more." and the URL "en.wikipedia.org/wiki/Madonna (entertainer) - 467k - Cached".

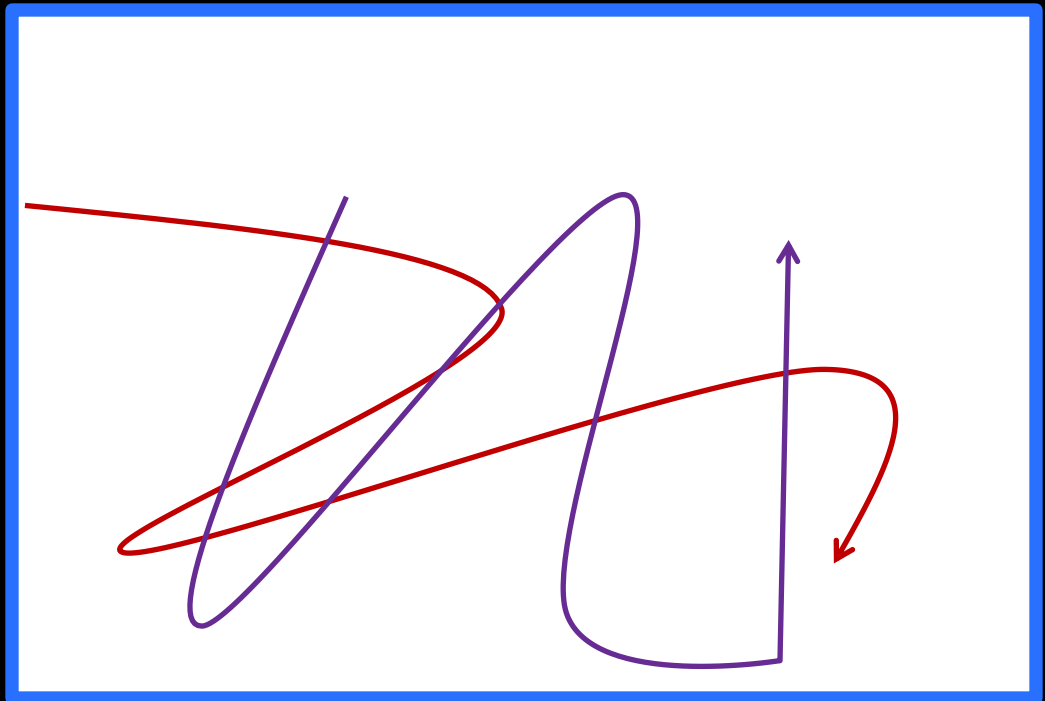


How do we lay out all these links on the screen?

New approach +Chierichetti, Kumar

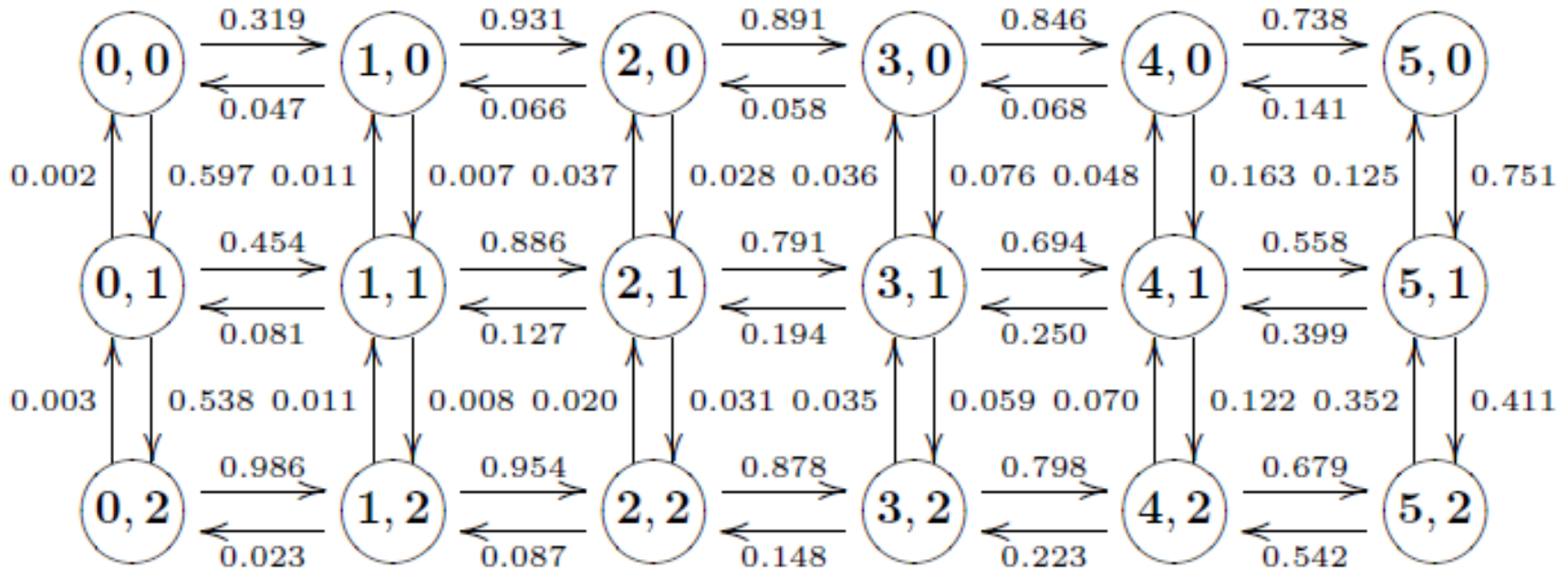
- **Analyze click logs**

- Where do users click on the screen, and in what order



Analyze 100's of millions of users

Calculate a model of users



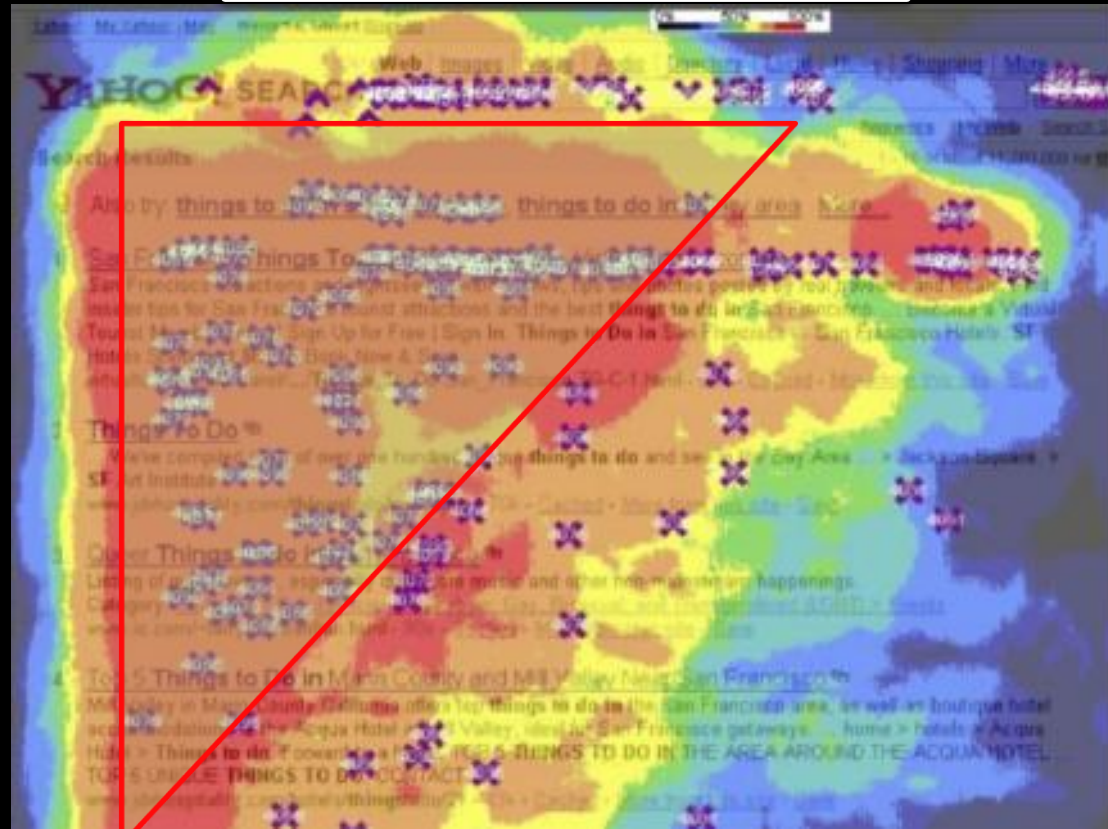
Model tells us where users' eyes go on the screen

What does this model tell us?

- Large scale validation of eye-tracking

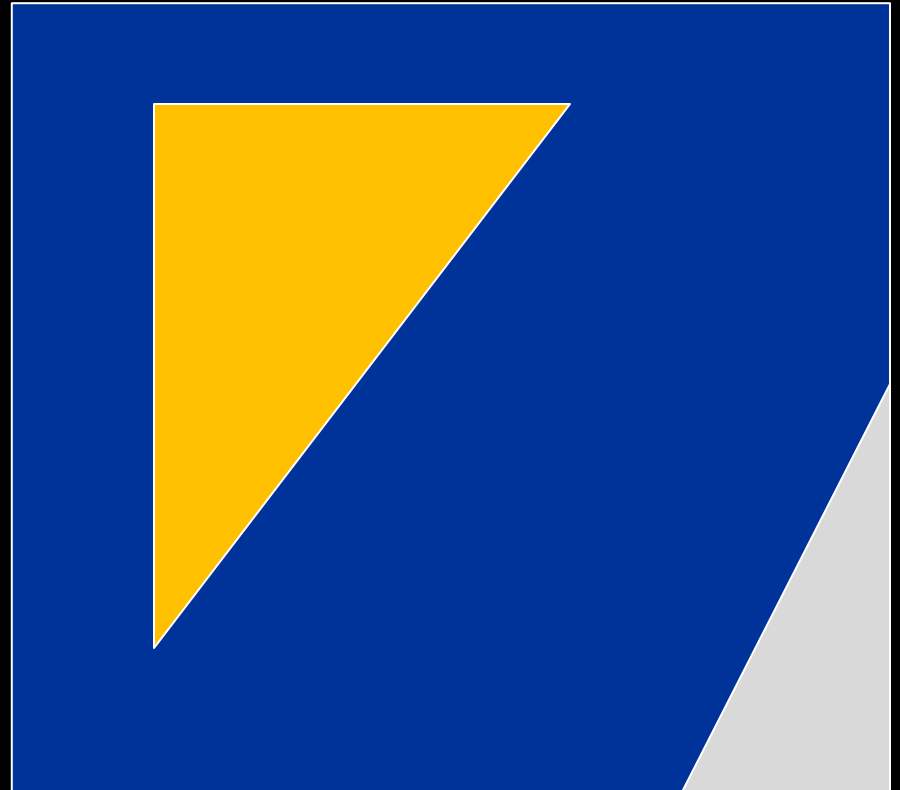
- First results page different from rest
- *Golden triangle*

Eye tracking heat map

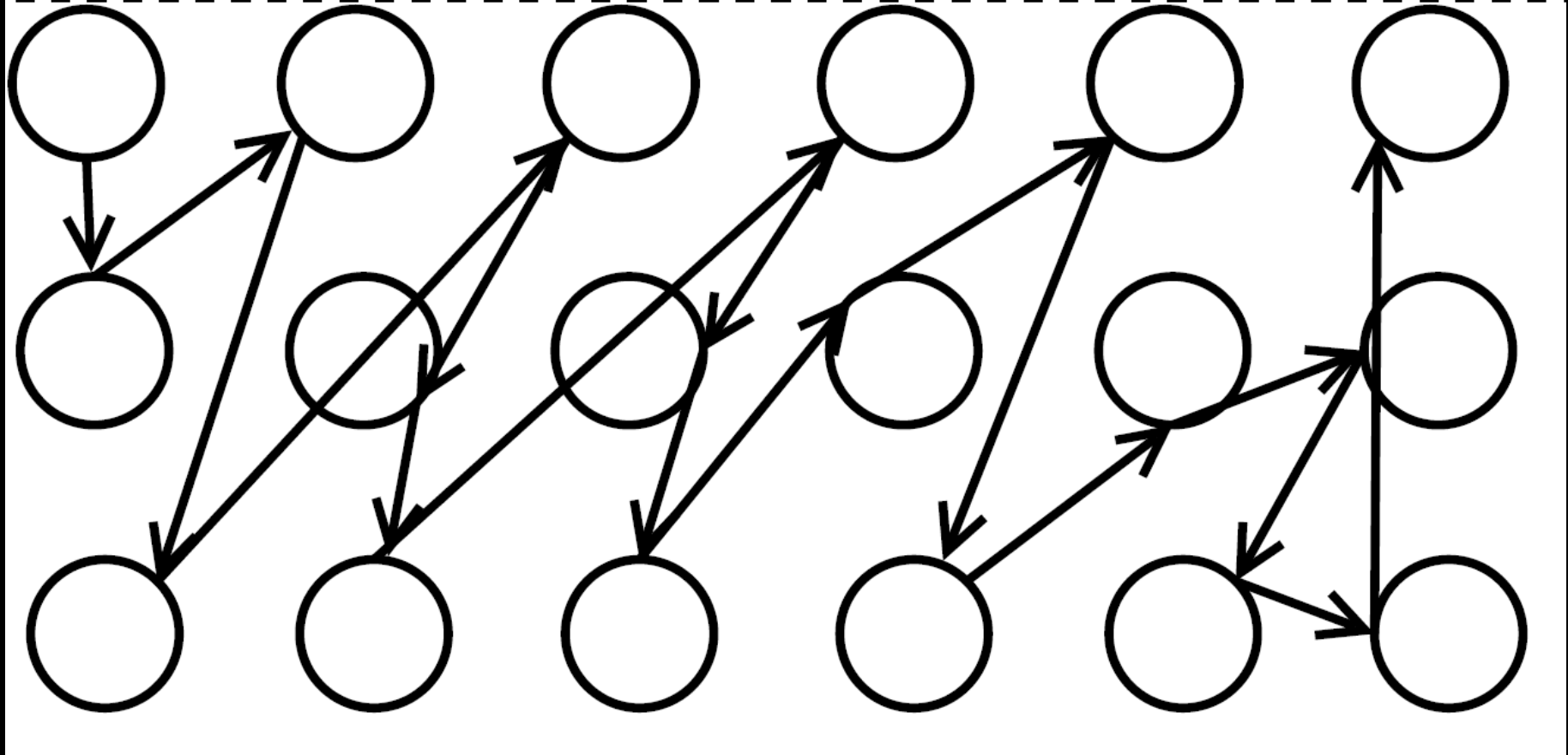


What's new

- Silver triangle
- Now can place images where users look



Where to put the images



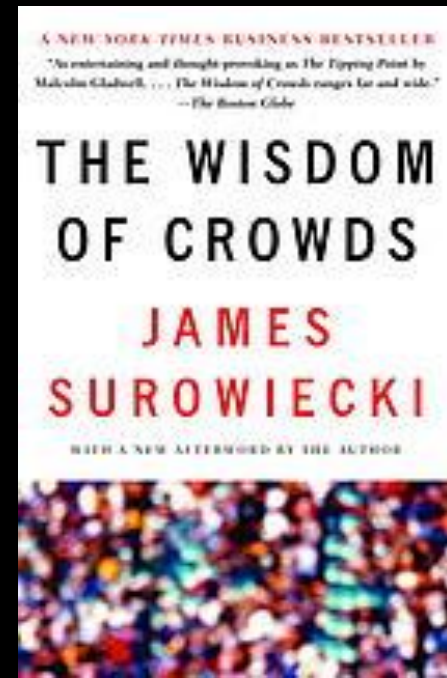


Prediction Markets

“The Wisdom of the Crowds”
„The Wisdom of the Crowds“

Prediction markets

- Idea: a group of people can collectively make a prediction better than an expert
 - *The Wisdom of the Crowds*
- E.g., employees in a project predict when the project will finish
 - Generally, a better prediction than the official project timeline!



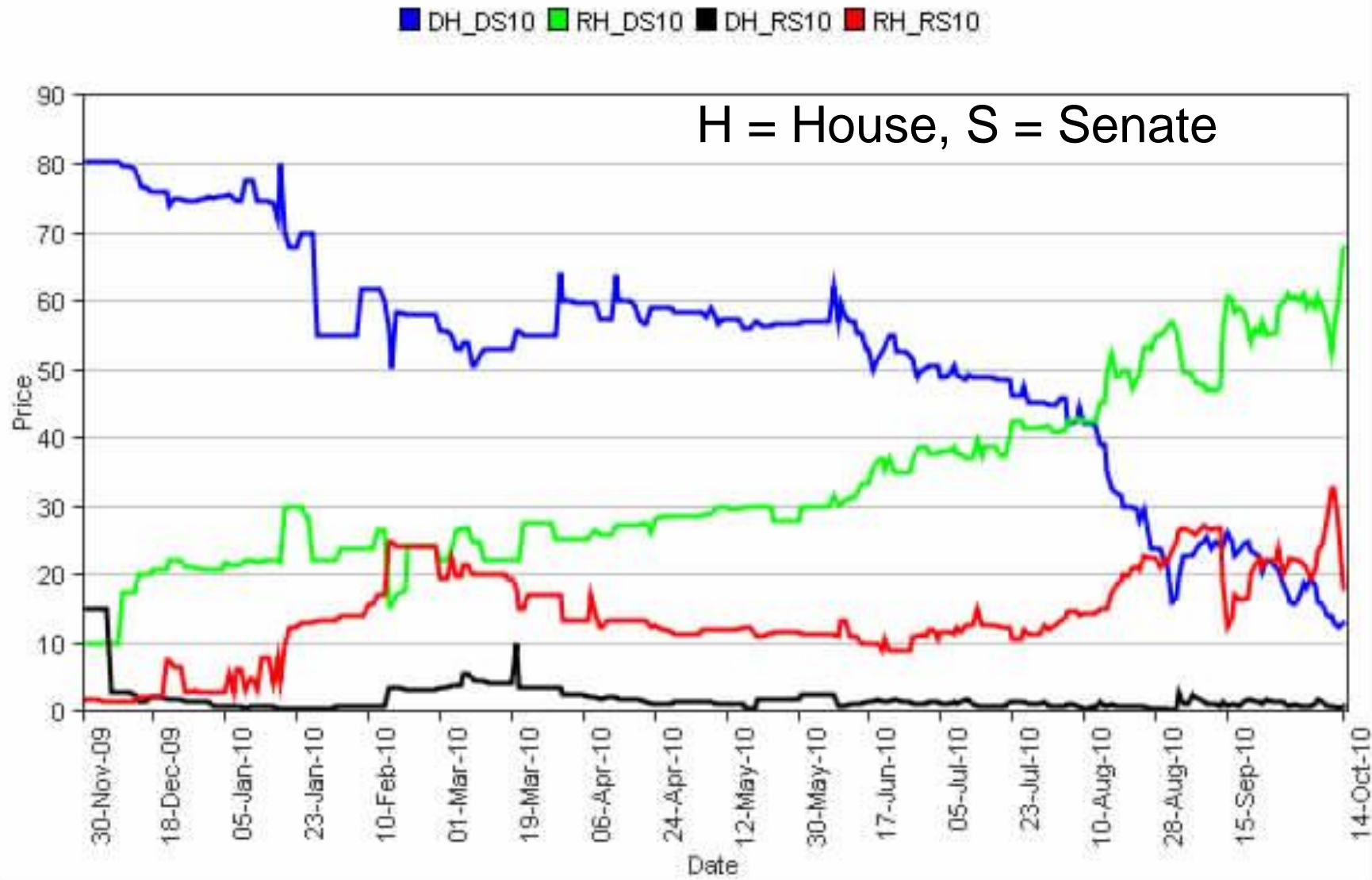
Prediction markets

- **Markets where you buy/sell predictions**
 - *E.g.*, “stock price” that the next AS Roma game will be a draw is 92 cents
 - Buy and sell stocks as crowd predicts
 - Over time, these prices are a good predictor of actual outcome
- **Not an opinion poll!**

Example: Iowa Electronic Markets

- Operated by the Business School at the University of Iowa
 - Used for academic studies
- Can buy/sell outcomes of US elections
- <http://tippie.uiowa.edu/iem/index.cfm>

IEM prices for 2010 US elections



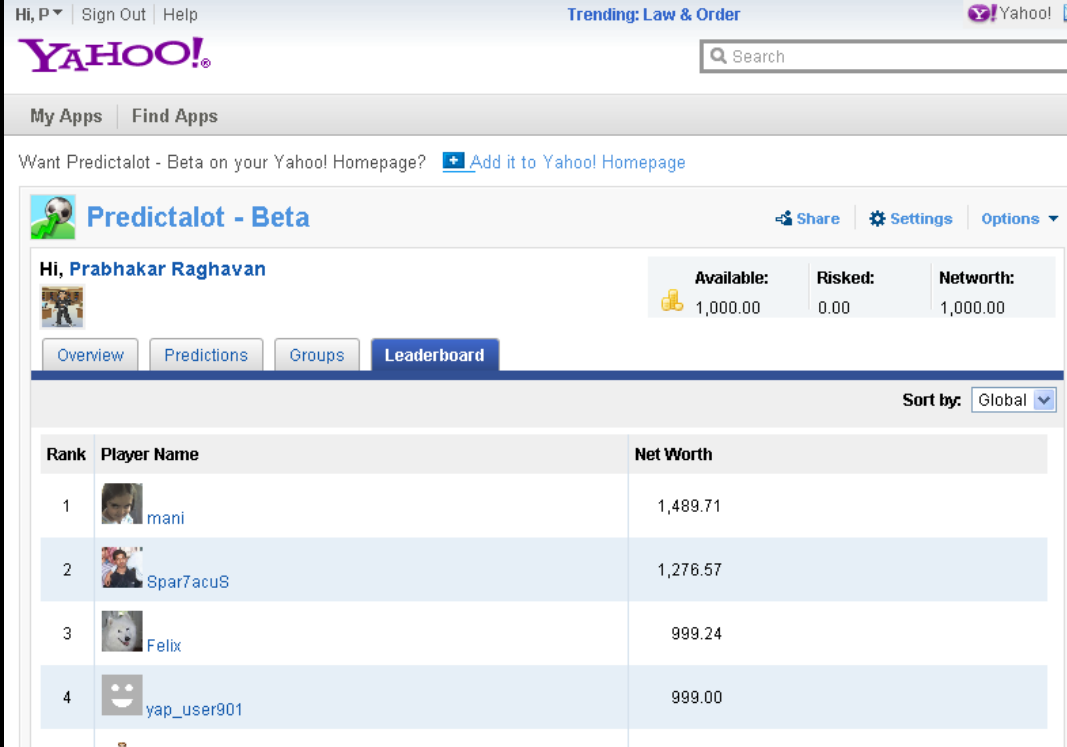
Are prediction markets truly powerful?

- Goel *et al.* 2010
 - Studied thousands of football games, music, ...
- Prediction markets can be better than other techniques such as machine learning and polls
 - But the advantage is very slight
- Thus, the Wisdom of the Crowds may not be as powerful as believed!

Combinatorial prediction markets

Predictalot and variants

- Can buy combinations: e.g., all four semi-final teams will be from Europe, etc.
- Predictions of various participants will interact



The screenshot shows the Yahoo! Predictalot - Beta app interface. At the top, it says "Hi, P" with links for "Sign Out" and "Help". The trending topic is "Law & Order". The user's name is "Prabhakar Raghavan". The app shows the user's current status: Available: 1,000.00, Risked: 0.00, and Networth: 1,000.00. Below this is a "Leaderboard" tab with a table of participants.

Rank	Player Name	Net Worth
1	mani	1,489.71
2	Spar7acu8	1,276.57
3	Felix	999.24
4	yap_user901	999.00



Looking ahead ...

What have we seen so far

- Many vignettes of human behavior
 - Studied at varying scales
- The Web affords us an observatory into social behavior

We have the Observatory

- **What are some big questions?**
 - Qualitative studies of user engagement
 - **What do people want to do?**
 - Quantifying user engagement
 - Inventing new genres of online experience
 - Optimizing user experiences
 - Whose happiness are we optimizing?
 - **Social choice questions**
 - **Privacy**

What are the academic challenges?

- Social scientists – study audiences at small scale, but deep understanding
- **Computer scientists – large scale, but little understanding**
- How can we combine these academic disciplines?

A new convergence

- **Social sciences meet massive computing**
 - Already happening in computational microeconomics

Grazie!

– <http://labs.yahoo.com>

