



USC Viterbi
School of Engineering

Stochastic Models of Social Dynamics on the Web

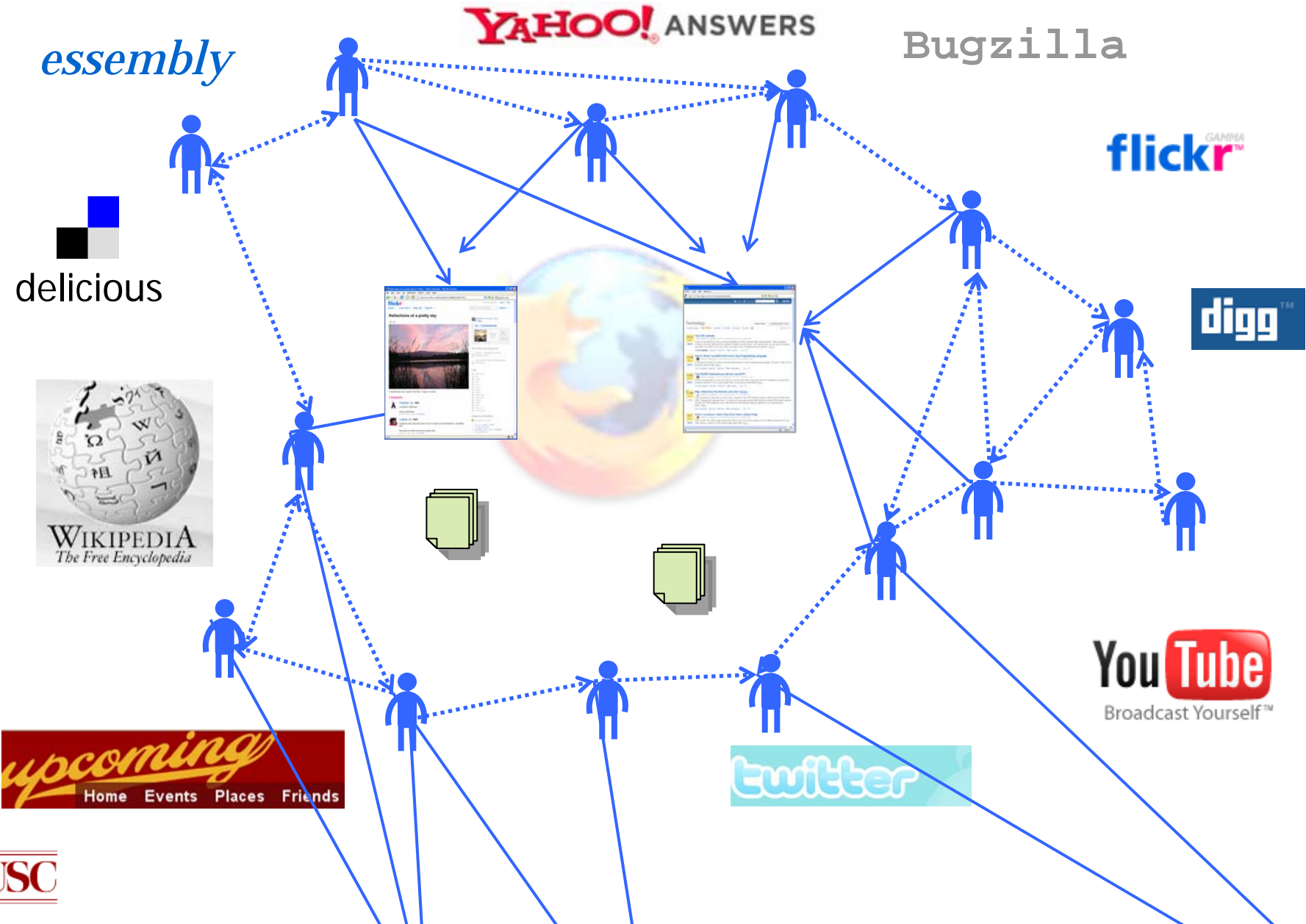
Kristina Lerman

USC Information Sciences Institute
lerman@isi.edu

Thanks: Tad Hogg (HP Labs)



The Social Web



Aggregating Social Knowledge

Image creator

private albums

- Mackay May 2008 (Set)
- Birds (Set)
- Birds (Pool)
- Canberra (Pool)
- Field Guide: Birds of the World (Pool)

public groups

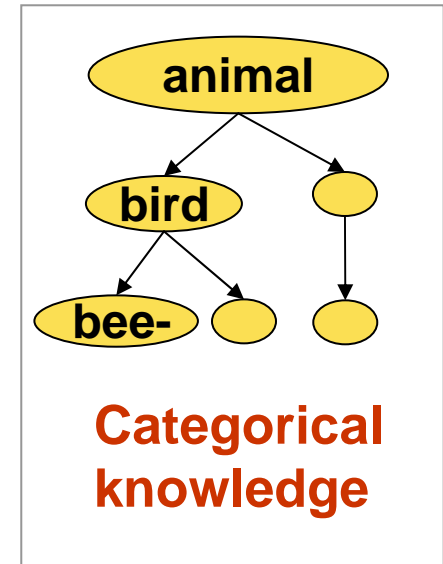
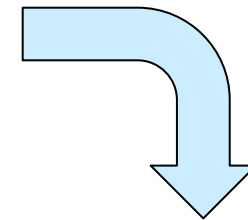
- Birds, Birds, Birds (Pool)
- BIRDPIX (3/day) (Pool)
- Australian Birds (Pool)
- Birds – Kingfishers, Pittas, and Bee-eaters (Pool)
- Birds of Queensland (Pool)

tags

- Rainbow bee-eater
- Merops ornatus
- Australia
- Queensland
- Mackay Gardens

Comments

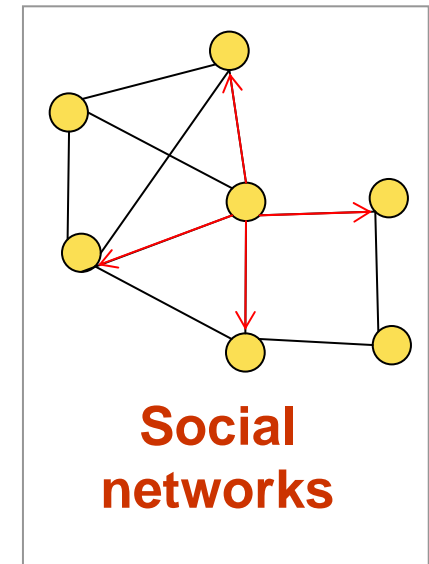
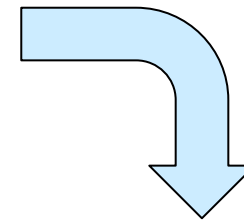
- laRuth pro** says: Gorgeous colours on that bird! Posted 10 months ago.
- aaardvaark pro** says: It the bee-eaters won't come to you, you go to the bee-eaters! It's interesting to be where your local birds migrate to. Posted 10 months ago.
- frangipenny IS HERE!! pro** says: Hi, I'm an admin for a group called [Birds - Kingfishers, Pitta's, Rollers & Bee-eaters](#), and we'd love to have this added to the group! Posted 10 months ago.

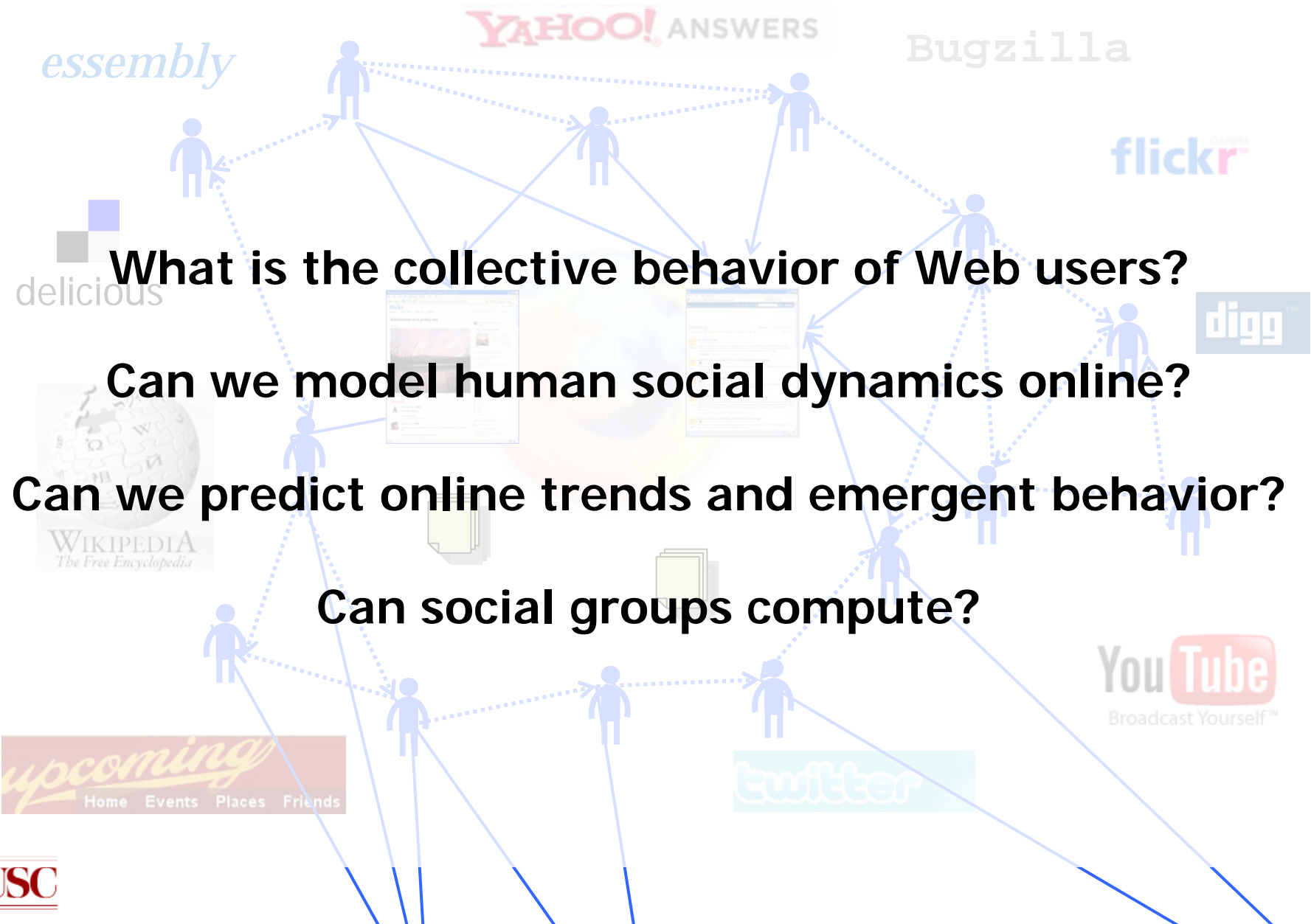


Information Spread on Social Networks



The screenshot shows a web browser window displaying the Twitter profile of 'StopAhmadi'. The browser tabs include 'Twitter / Search - #iranele...' and 'Raymond Jahan (StopAhm...'. The address bar shows 'http://twitter.com/StopAhmadi'. The profile header includes the Twitter logo, navigation links (Home, Profile, Find People, Settings, Help, Sign out), a profile picture of a man with glasses, the name 'StopAhmadi', and a 'Follow' button. The bio reads: 'Name Raymond Jahan', 'Web http://www.flickr...', and 'Bio Dedicated Twitter account for Mousavi supporters'. Statistics show '53 following' and '9,761 followers'. The 'Updates' section shows '1,248' updates. A tweet is visible: '“Educate & inform the whole mass of people. They R the only sure reliance 4 preservation of our liberty”. over 'n out.' followed by 'about 2 hours ago from Seismic Desktop'. Other tweets discuss Iranian government actions and a DDoS script.





What is the collective behavior of Web users?

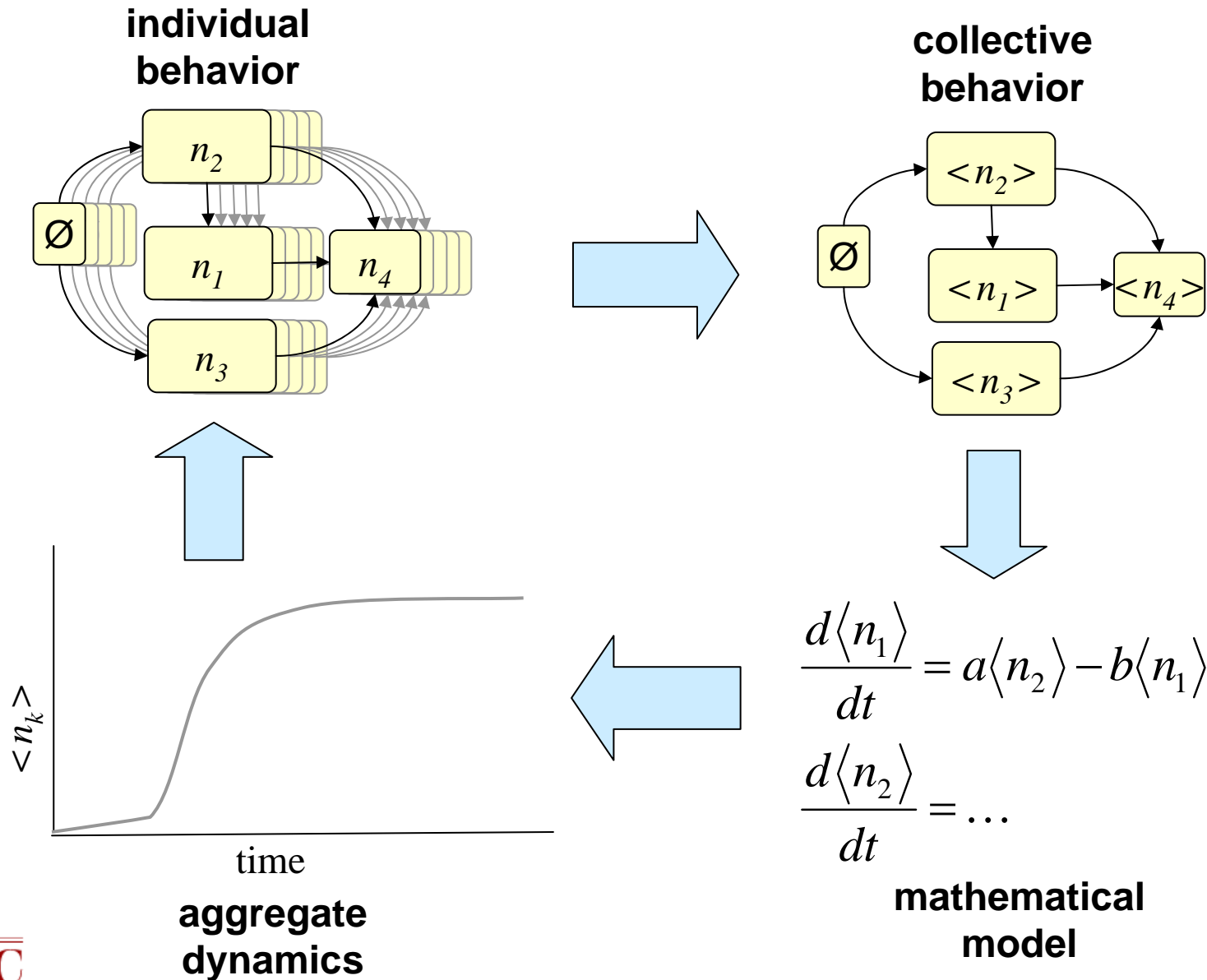
Can we model human social dynamics online?

Can we predict online trends and emergent behavior?

Can social groups compute?

- **Developed a mathematical framework to model the social dynamics of Web users**
- **Modeling enables**
 - Predicting trends or behaviors
 - *E.g., which newly contributed content will become popular*
 - Designing web sites
 - *E.g., how to display information to users*
 - Altering user incentives
 - *E.g., improve content quality, participation or collaboration*
- **Applied the general framework to study Digg**

Stochastic Modeling summary



Social news aggregator Digg allows people to

- submit news stories
- vote on them
- collectively select best stories for the front page

Digg's front page emerges from the opinions of its many users



Life cycle of a story on Digg

Upcoming stories queue (16,000+)

Technology / Upcoming

Profile Friends' Activity Submit New

All Technology World & Business Science Gaming Lifestyle Entertainment

Popular Upcoming News Videos Images Customize

Technology Recommendations (0) All (15737)

sort by: Most Recent Most Diggs Most Comments

1 digg [Video] Becoming an Organized and Productive Blogger: Google
LisaMorosky.com — This is the second video in a series about how bloggers can get organized in order to increase productivity. This video is all about Google Calendar and how you can use it to create a weekly work schedule, keep tabs on your tasks, share your availability with others, and much more.
0 Comments Share Bury lisamorosky submitted 1 min ago

1 digg Free Download Notepad++ 5.5.0 / FileHorse.com
filehorse.com — Notepad++ is an open source text editor, which is suitable for programmers or advanced computer users, seeking better alternative to the Windows ...
0 Comments Share Bury mozarelli submitted 1 min ago

1 digg How to build an Ajax slideshow?
developertutorials.com — How to build a ajax slideshow as http://workawesome.com/ with two buttons to change contents?
0 Comments Share Bury Ducani submitted 1 min ago

1 digg Importing data into a Rails application | John Beynon
john.beynon.org.uk — There will come a time when you need to import data into your Rails application. Our time was now.
0 Comments Share Bury neilmiddleton0 submitted 1 min ago



Front page (~100 stories)

All News, Videos, & Images

Profile Friends' Activity Submit New

Technology World & Business Science Gaming Lifestyle Entertainment

Popular Upcoming News Videos Images Customize

News, Images, Videos Most Recent Top in 24 Hr 7 Days 30 Days 365 Days

319 diggs FDA to be aggressive in tobacco regulation
courier-journal.com — If there is any doubt about how aggressive the federal Food and Drug Administration intends to be in regulating tobacco, take a look at a letter the agency sent out last week.
173 Comments Share Bury badqat made popular 14hr 35min ago

408 diggs Can a Daily Pill Really Boost Your Brain Power?
guardian.co.uk — In America, university students are taking illegally obtained prescription drugs to make them more intelligent. Here, an investigation into the brave new world of neuro enhancement...
83 Comments Share Bury openthink made popular 14hr 45min ago

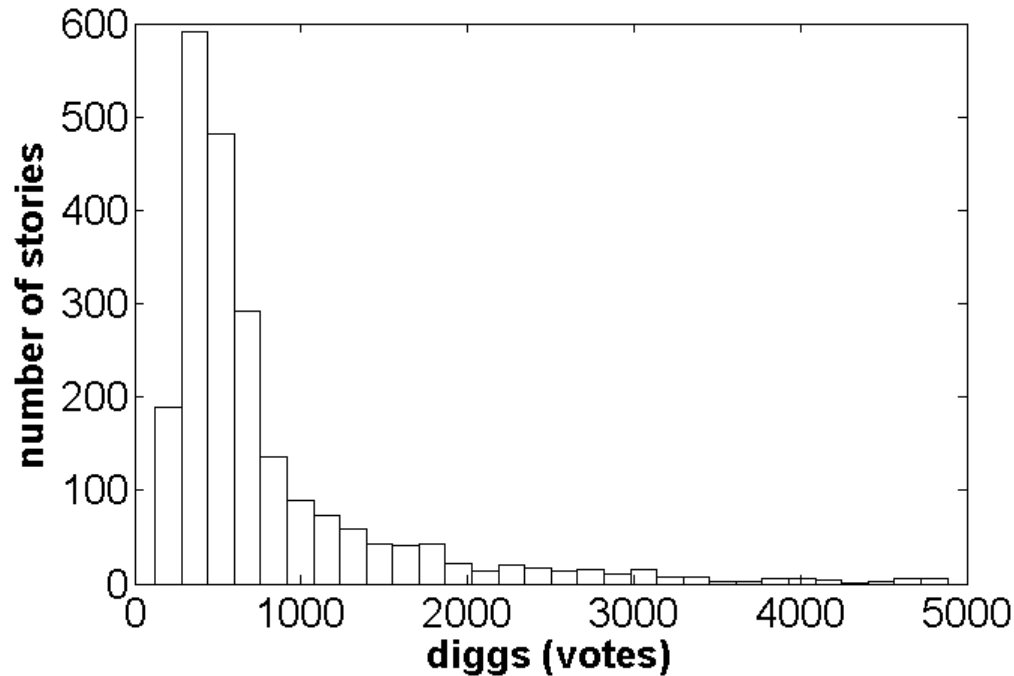
Sponsored by Digg

2215 diggs What happens when your Mom cancels your WoW account...
revision3.com - Parents just don't understand. We don't either - what WAS he trying to do with that remote control?
Share Bury

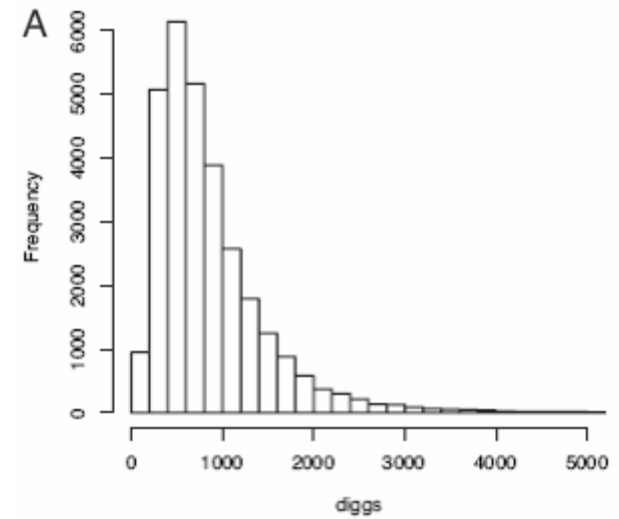
168 diggs Three suspects arrested in U.S. terrorism probe
reuters.com — A Colorado man, his father and an accused accomplice in New York were arrested on Saturday and charged with lying to federal agents about a plot to blow up unspecified targets in the United States, the U.S. Department of Justice said.
60 Comments Share Bury 11cuisinart made popular 14hr 54min ago



Distribution of votes



~2000 front page stories submitted in
June, 2009




Wu & Huberman, 2007

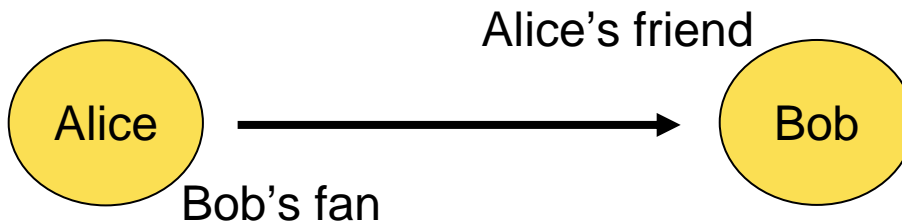
~30,000 front page stories
submitted in 2006

- How do stories accumulate votes?
- Why do some stories become very popular, but most don't?
- Can we explain the distribution of votes?
- Can we predict which stories will be promoted?

Stochastic modeling can help answer these questions

- **Statistical models of aggregate behavior**
 - behavior for Digg, Wikipedia, YouTube, ...
e.g., [Wu & Huberman 2007; Wilkinson 2008]
- **Predicting popularity**
 - Inherently unpredictable
 - *Even though some content (books, music, ...) becomes much more successful than other, even experts can't predict which*
[Salganik, Dodds & Watts, 2006]
 - Predicting popularity of social media content
 - *Number of votes on Digg, views on Youtubs, etc.*
[Kaltenbrunner, Gomez & Lopez, 2007; Szabo & Huberman, 2008]
- **Predicting quality of online content**
 - Dynamics of Youtube views as feature of quality
[Crane & Sornette 2008]

- Construct stochastic model of dynamics of social voting on Digg
- Phenomenology of 
 - Users submit and vote on news stories
 - Digg promotes popular stories to front page
 - Digg allows social networking
 - *Users can designate Friends*
 - *and view their friends' activity on Digg*
 - *Directed social network*
 - *Friends of user **A** are everyone **A** is watching*
 - *Fans of **A** are all users who are watching **A***



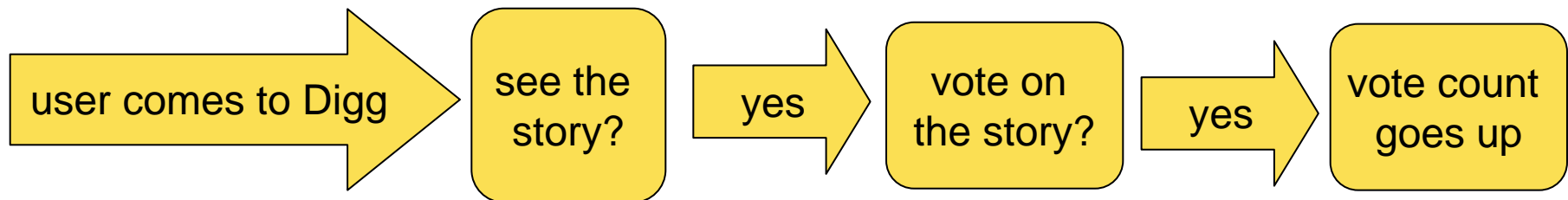
A Look at the Friends Interface

The screenshot shows a web browser window with the URL <http://digg.com/users/telechic/friends/>. The page has tabs for Profile, Friends' Activity, History, and Settings. The main content area is titled "All Recent Activity" and lists various items with their digg counts and submission times. A "Filter" sidebar on the right allows users to narrow down the activity. Two red ovals highlight the "diggs" and "submissions" filter options. Arrows point from these ovals to the main activity list, with text labels "Stories friends voted on" and "Stories friends submitted".

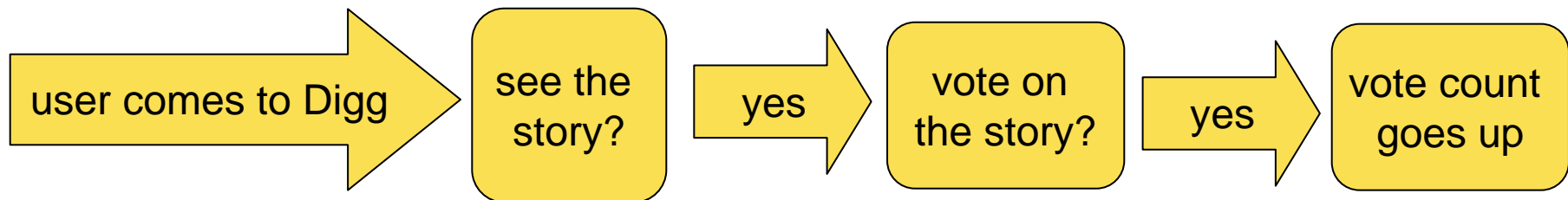
Filter	Count
all recent activity	285 recent
diggs	268 recent
diggs in upcoming	139 recent
submissions	13 recent
comments	
favorites	3 recent
profile activity	1 recent

Activity	Count	Time
German Youth Would Vote Pirate Party	498	2 hr 52 min ago
Even Skeletor Needs A Pet	83	2 hr 54 min ago
The New Retirees [Pic]	115	2 hr 54 min ago
Ambulance Called To David Hasselhoff's House	62	3 hr ago
Mad Men & 30 Rock top Emmy Awards	90	3 hr 2 min ago
7 More Amazing Engineering Wonders of Today & Tomorrow	226	3 hr 3 min ago
'Mad Men,' '30 Rock' take top series Emmys	85	3 hr 6 min ago
MUSE Refusing To Lip Sync On Talk Show (Video)	86	3 hr 8 min ago
Whoops! Students 'Going Google'Get to Read Each Others email	42	3 hr 17 min ago

- **visibility and interestingness** → votes
 - visibility: does user **see** the story?
 - interest: does user **like** the story?



- **visibility and interestingness** → votes
 - visibility: does user **see** the story?
 - *user interface*
 - *browse upcoming stories*
 - *browse front page*
 - *recommended by friends*
 - *search*
 - interest: does user **like** the story?
 - *appeal*
 - *novelty, ...*



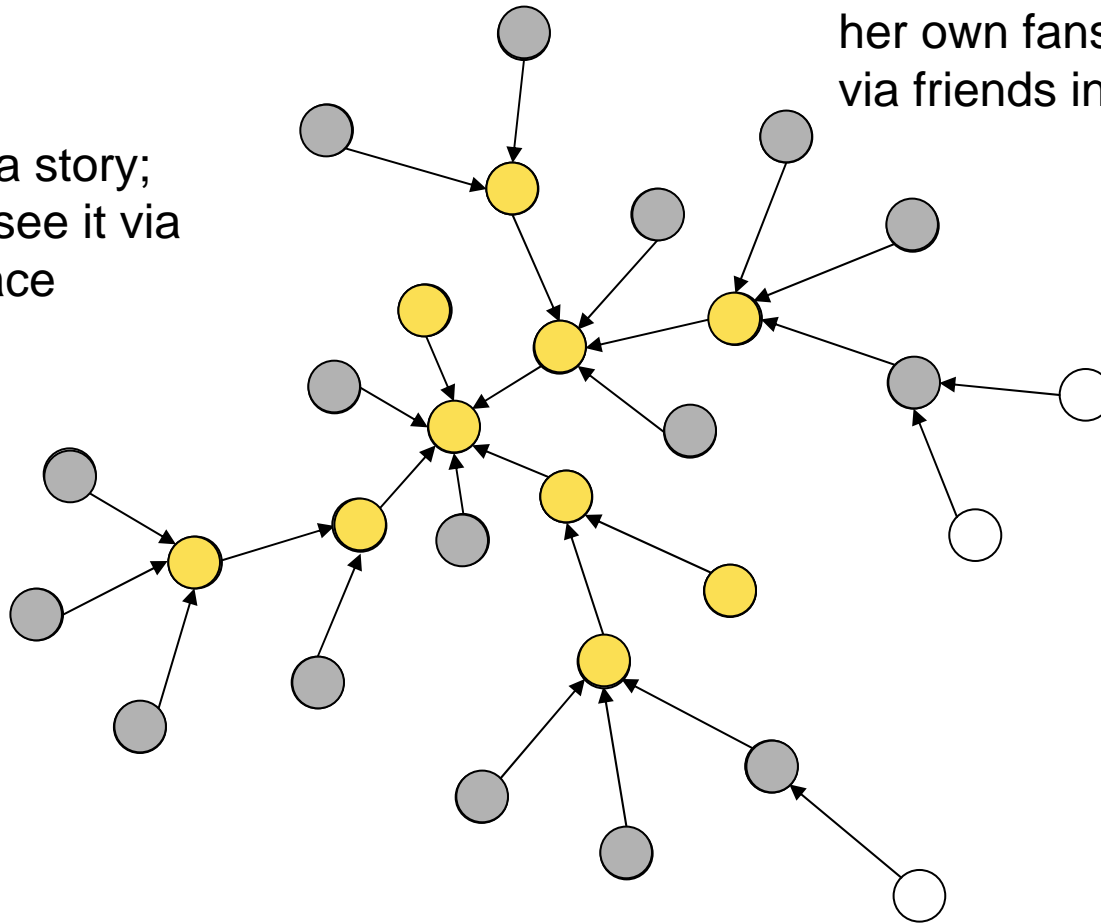
- **Digg shows upcoming stories as a list**
 - Most recently submitted story first
 - 15 stories per page
- **visibility decreases with distance from top of list**
- **A given story**
 - moves down the list as new stories are added
 - eventually moves to later pages
 - *User must click to view subsequent pages*
 - After **promotion to the front page**, story is no longer visible on the upcoming stories pages

- **Digg shows popular (front page) stories as a list**
 - Most recently promoted story first
 - 15 stories per page
- **visibility decreases with distance from top of list**
- **A given story**
 - moves down the list as new stories are promoted
 - eventually moves to later pages
 - *User must click to view subsequent pages*
 - Not visible before **promotion to the front page**

Visibility through Friends Interface

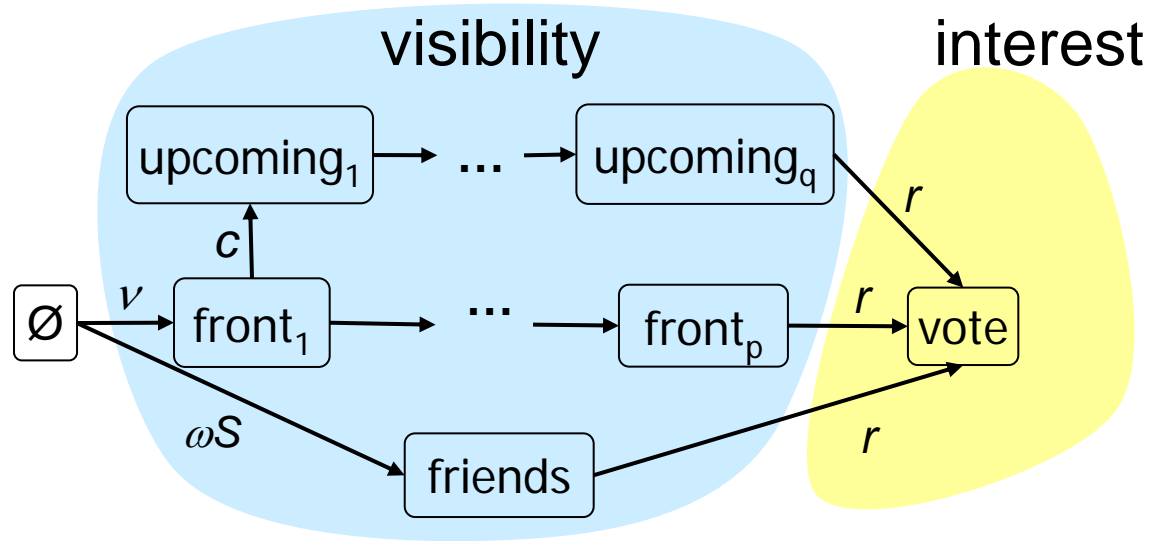
User submits a story;
her fans can see it via
friends interface

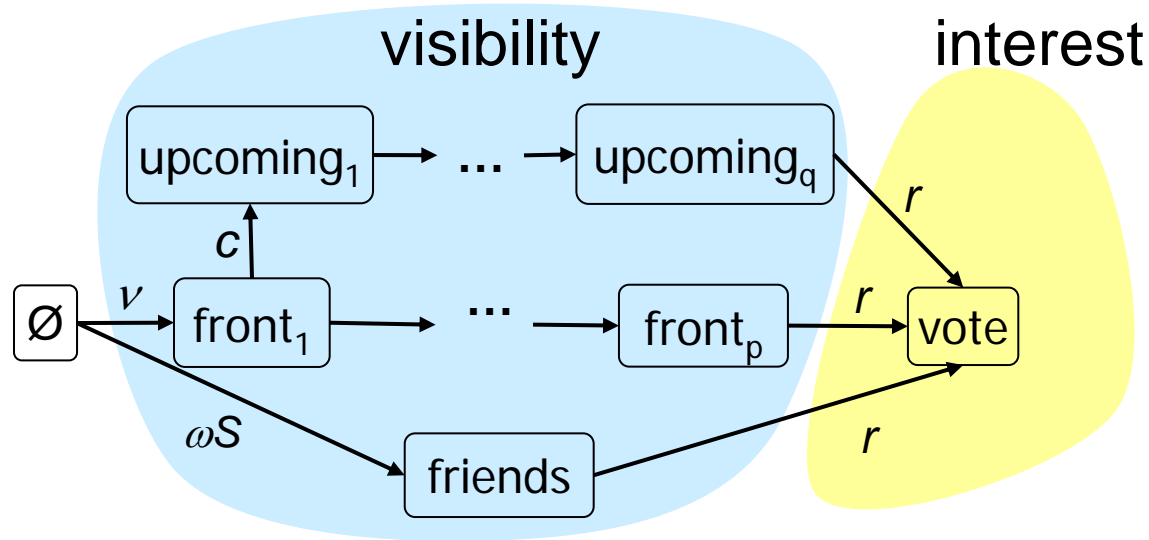
fan sees and votes on story;
her own fans can then see story
via friends interface



network of fans

User behavioral model





Story specific parameters	
r	'interestingness' – prob. story will receive a vote if seen
S	number of submitter's fans
General parameters	
v	rate users visit Digg
c	fraction of users viewing upcoming pages
ω	rate fans visit Digg

- How number of votes $N_{vote}(t)$ for a story changes

$$\frac{dN_{vote}(t)}{dt} = r \overbrace{(v_f(t) + v_u(t) + v_{friends}(t))}^{\text{visibility}}$$

- v_f - rate users find story on the front page queue
 - v_u - rate users find story on the upcoming stories queue
 - $v_{friends}$ - rate users find story through the friends interface
- r – fraction of users who see the story who choose to vote for it

- **Need model parameters for**
 - Story visibility
 - Story interestingness

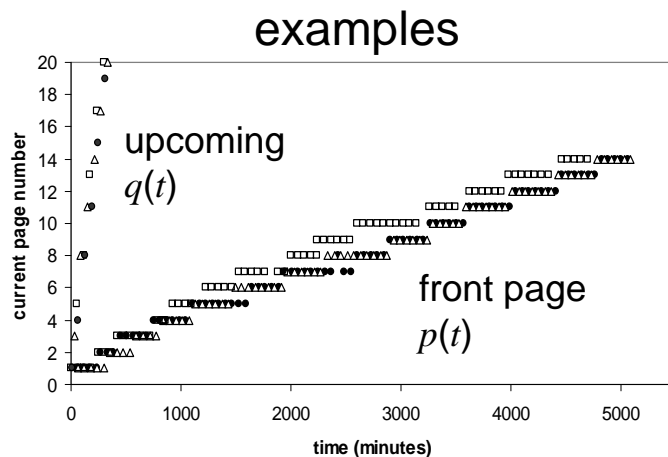
- **Estimate from behavior of sample of users**

- **Stories from front and upcoming pages**
 - number of votes vs. time since submission
 - for several days in May 2006
 - prior to availability of Digg API
- **Number of fans for active users**
- **2152 stories with at least 4 observations**
 - submitted by 1212 distinct users
- **510 of these stories promoted to front page**

- **User viewing behavior not available:**
 - which stories users look at
 - how they find stories
 - *front page, friends interface, ...*
- **Estimate indirectly from models & data**

- **Story location**
- **Navigating web sites**
- **Number of fans**

- For upcoming and front page lists:
 - location on page (1 to 15), which page (1st, 2nd, ...)
 - distance from top of list increases linearly with time



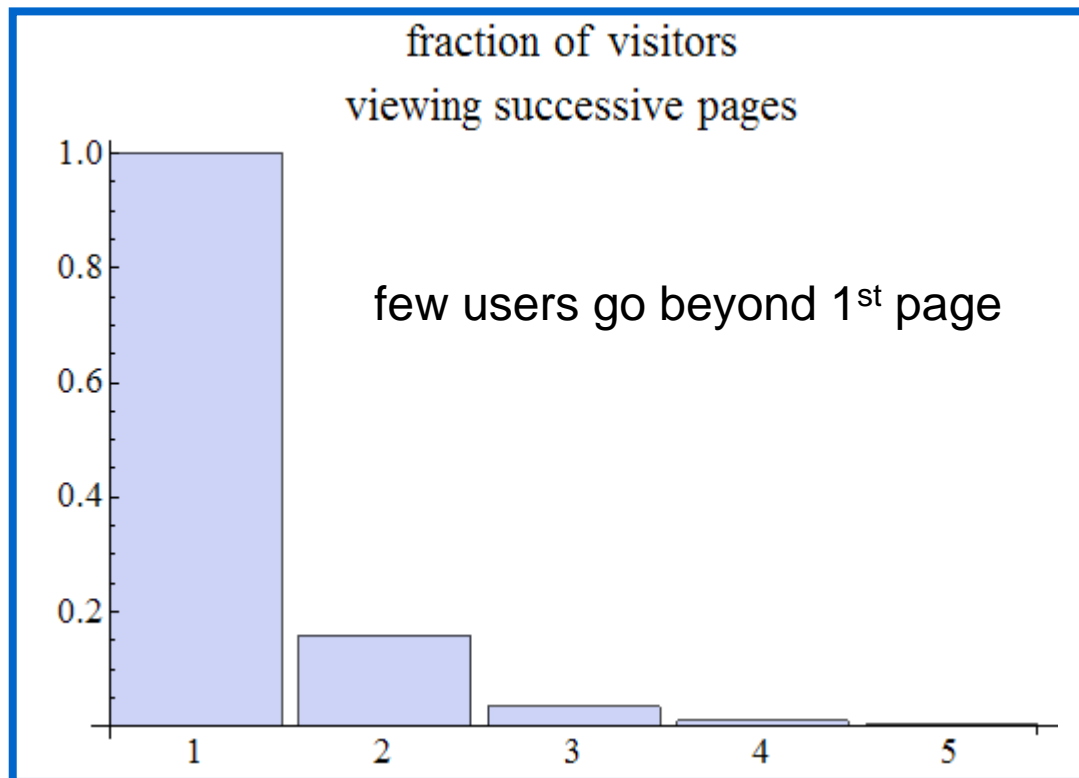
- Rate story position increases:
 - front page: ~ 0.2 pages/hr
 - upcoming: ~ 4 pages/hr
- 1/15th the rates new stories are
 - *promoted to front page* (~ 3 /hr)
 - *submitted as new stories* (~ 60 /hr)
 - since each page holds 15 stories
- Averages over hourly variation
 - [Szabo & Huberman 2008]

- **Digg promotion decision algorithm not public**
 - based on popularity expressed by user votes
- **Approximation from data:**
 - story promoted if
 - *at least 40 votes within 24 hours of submission*

- Story location
- Navigating web sites
- Number of fans

- **Empirical model of user following links on a Web site**

- “law of surfing” [Huberman et al. 1998]
- Inverse Gaussian distribution of #pages viewed before leaving web site



parameters
estimated from
Digg data & model

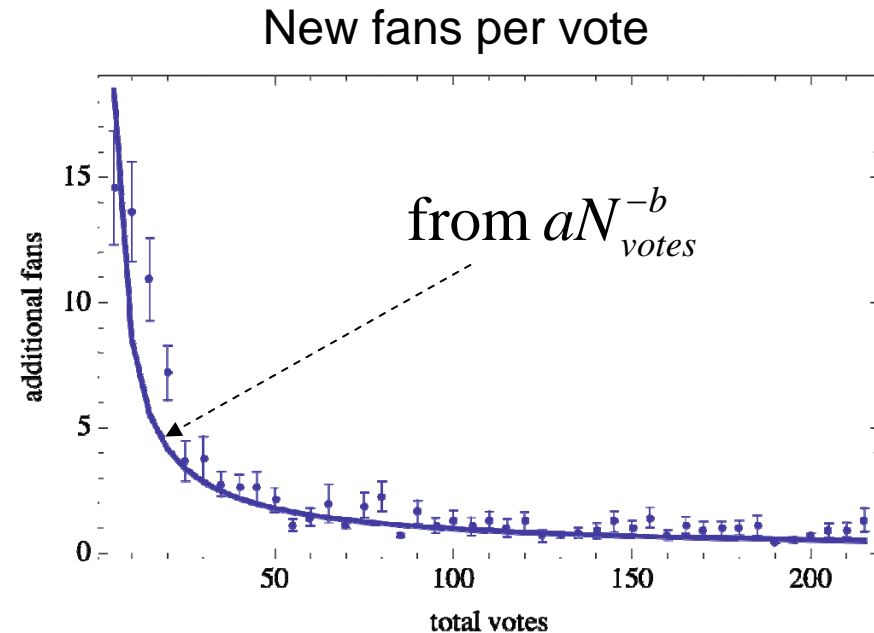
- Story location
- Navigating web sites
- **Number of fans**
 - At submission time, story visible to submitter's fans
 - Each voter enables their fans to see story

- **Model of number of fans not yet viewing story, $s(t)$**
 - based on number of votes on the story
 - story visible to submitter's fans at submission time: $s(0)$

$$\frac{ds}{dt} = -\omega s + aN_{\text{vote}}^{-b} \frac{dN_{\text{vote}}}{dt}$$

fans of prior voters visit Digg

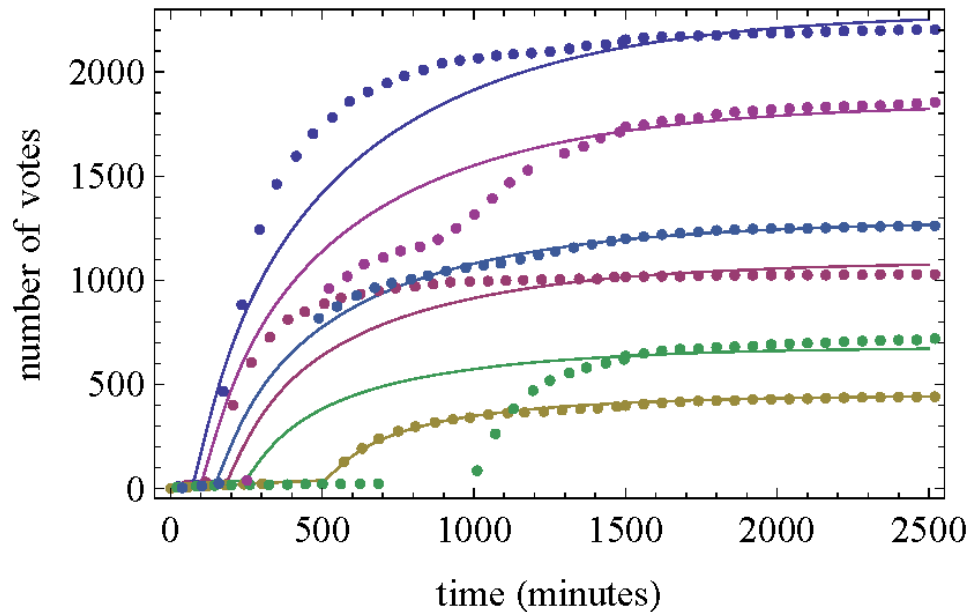
new fans from new votes



Parameterize how interesting the story is

- **Reasons users vote for story not available, e.g.,**
 - topic
 - novelty [Wu & Huberman 2007]
 - popularity (determining interest, not just visibility)
 - *e.g., "cool" fashion or gadgets*
 - ...
- **One approach: web-based experiments**
 - e.g., [Salganik et al. 2006]
- **Estimate from models & data**
 - from vote history after accounting for visibility

model vs. observations for 6 stories



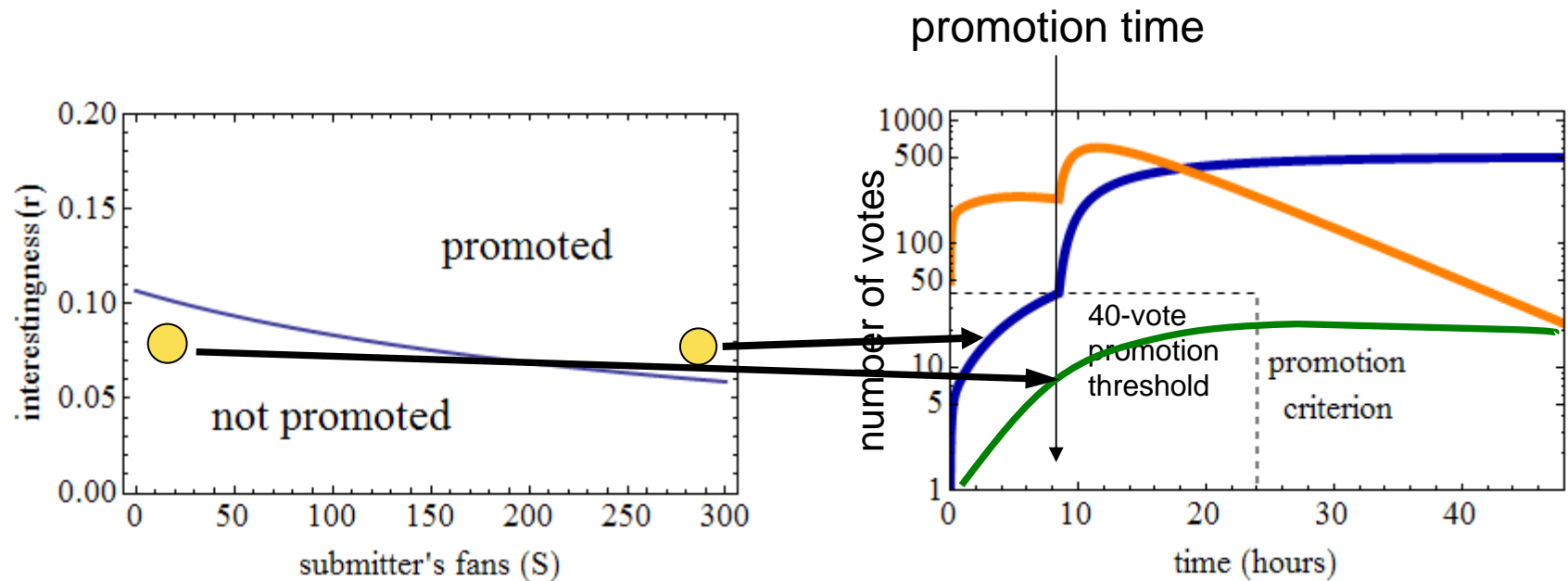
S	r	Final votes
5	0.51	2229
5	0.44	1921
40	0.32	1297
40	0.28	1039
160	0.19	740
100	0.13	458

model captures qualitative features

- slow growth initially
- influence of fans on promotion
- rapid growth after promotion (greater visibility)

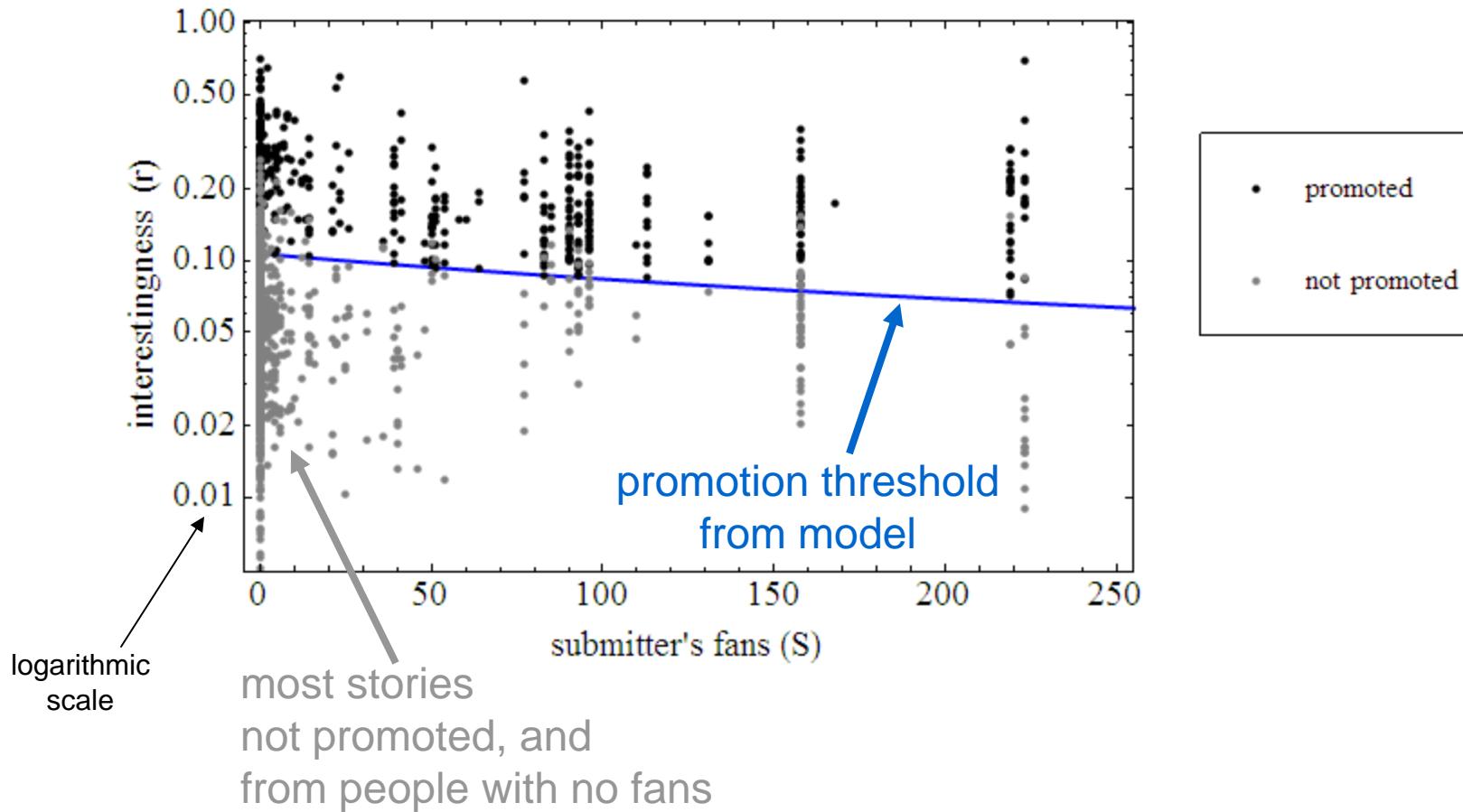
- Less interesting stories by highly connected users will be promoted

Values of S and r to get the story on front page

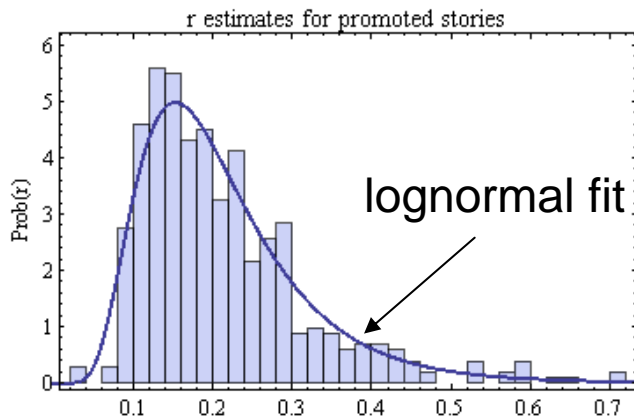


Predicting Promotion to Front Page

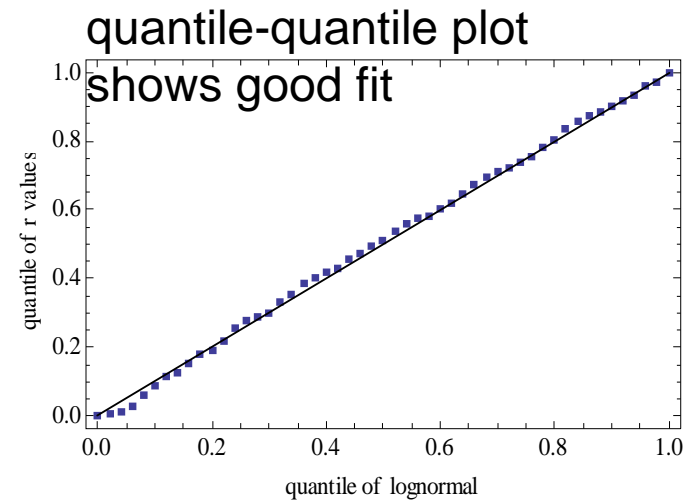
model prediction vs. data: 95% accurate



- **Long-tail distribution (lognormal)**
 - a few stories much more interesting than average
 - *after accounting for visibility via user interface part of model*



distribution of estimated interestingness values

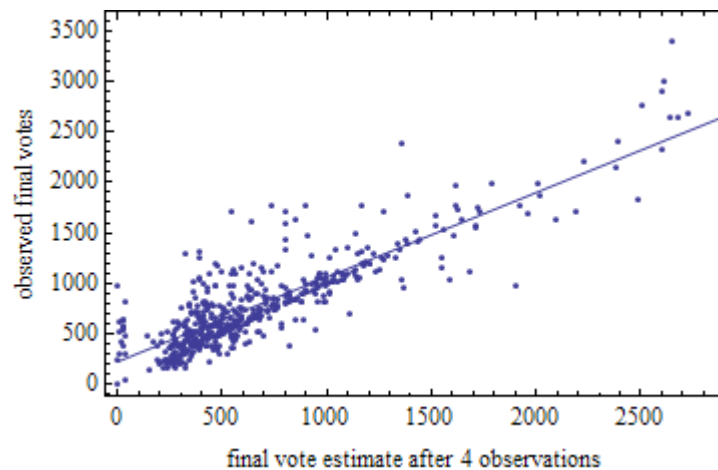


good fit with Kolmogorov-Smirnov test

- **Open question: why?**

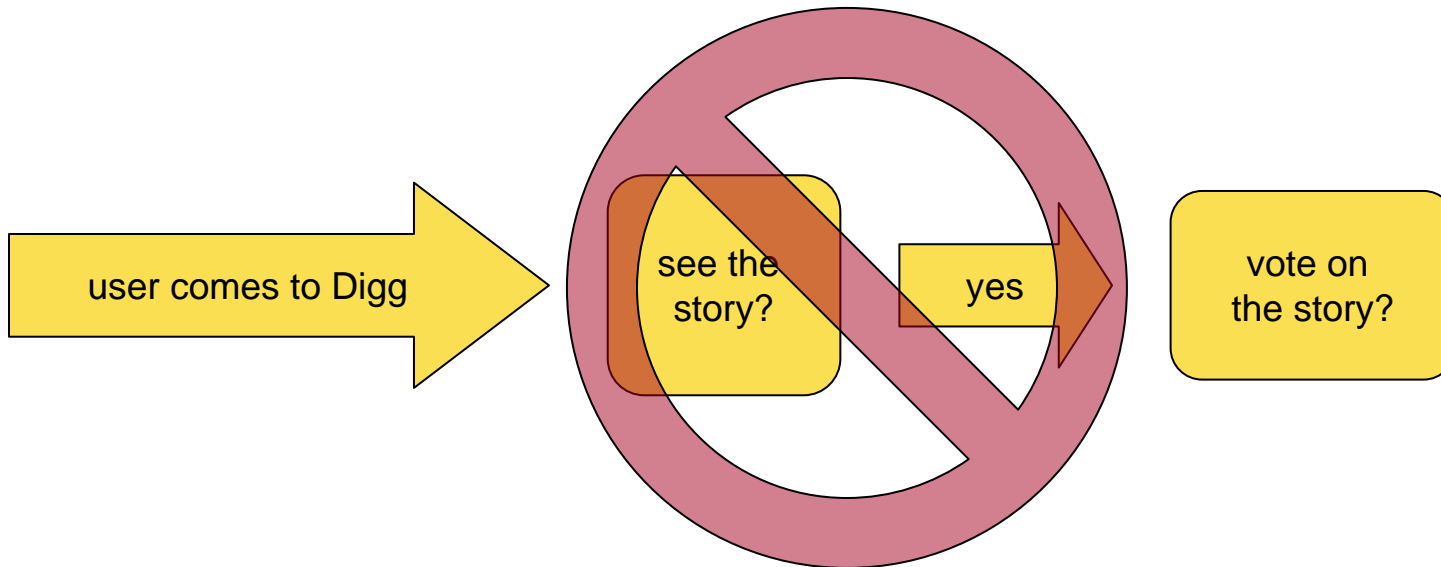
- **Behavior predictable from early reaction to story**
 - Initial votes (first hour) correlate with long-term popularity [*Szabo & Huberman 2008*]
 - also with YouTube, Slashdot
[Kaltenbrunner, Gomez & Lopez 2007; Crane & Sornette 2008]
- **Model-based prediction can beat early methods**
 - Estimate story interestingness r
 - *using initial votes, as few as 4 (few minutes)*
 - *network diffusion*
 - Use this value of r to predict long-term popularity

- Estimate story interestingness r using initial votes
 - Example: use first 4 observations
 - r estimates correlate 0.9 with those based on full history
 - prediction of final votes account for 75% of variance
 - rms prediction error: 244 votes



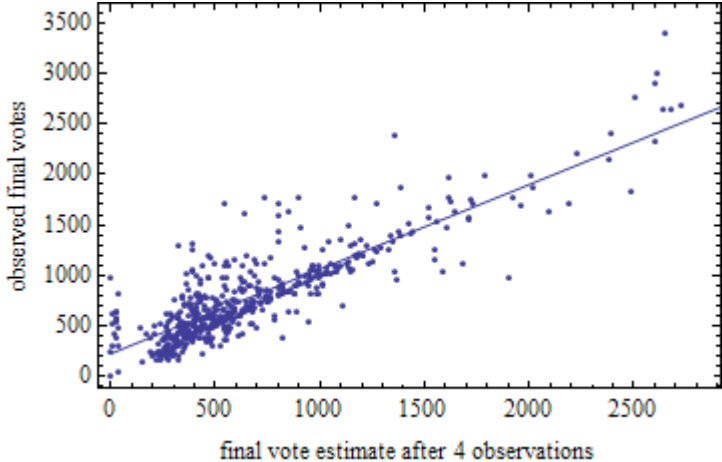
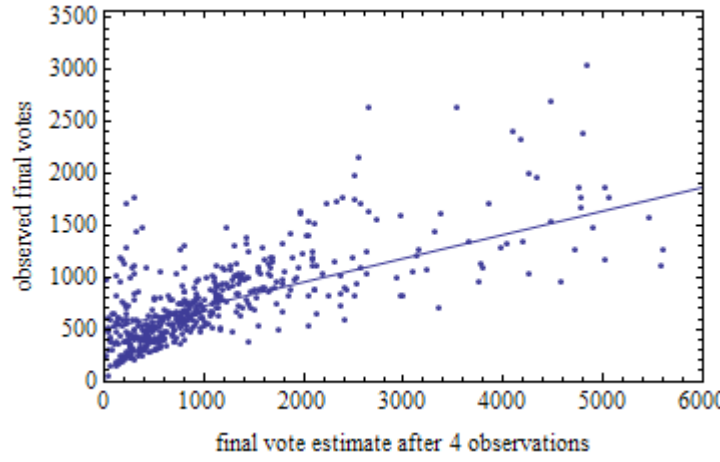
Model based on votes only?

- **Estimate based on initial votes only**
 - not including visibility model
 - *i.e., ignore effects of 'law of surfing' and social network*
cf [Szabo & Huberman 2008]



Model based on votes only?

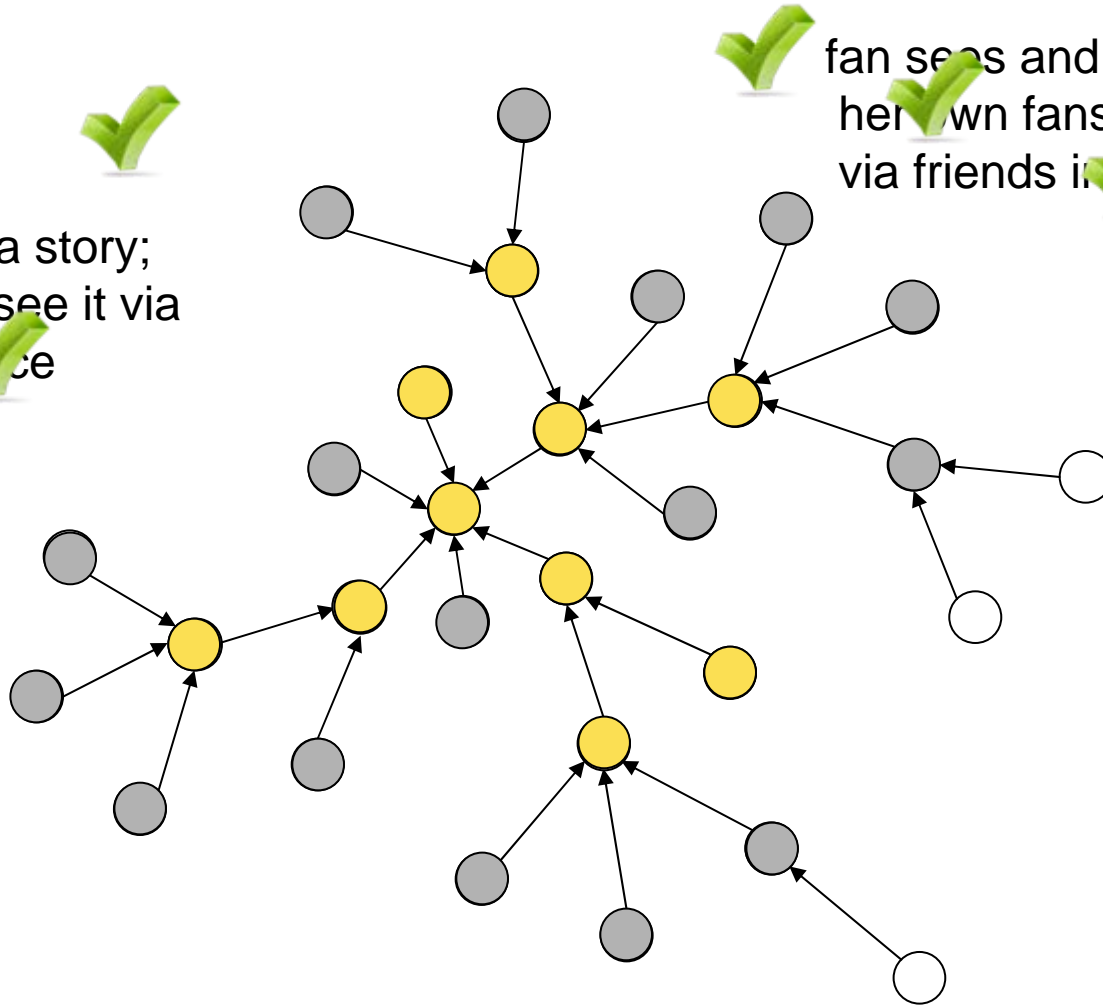


	Full model	Votes only
		
	<p>full model is better than not including visibility (differences significant, p-value $<10^{-4}$)</p>	
variance accounted for	75%	56%
rms prediction error	244	327

Visibility through Friends Interface

User submits a story;
her fans can see it via
friends interface

fan sees and votes on story;
her own fans can then see story
via friends interface

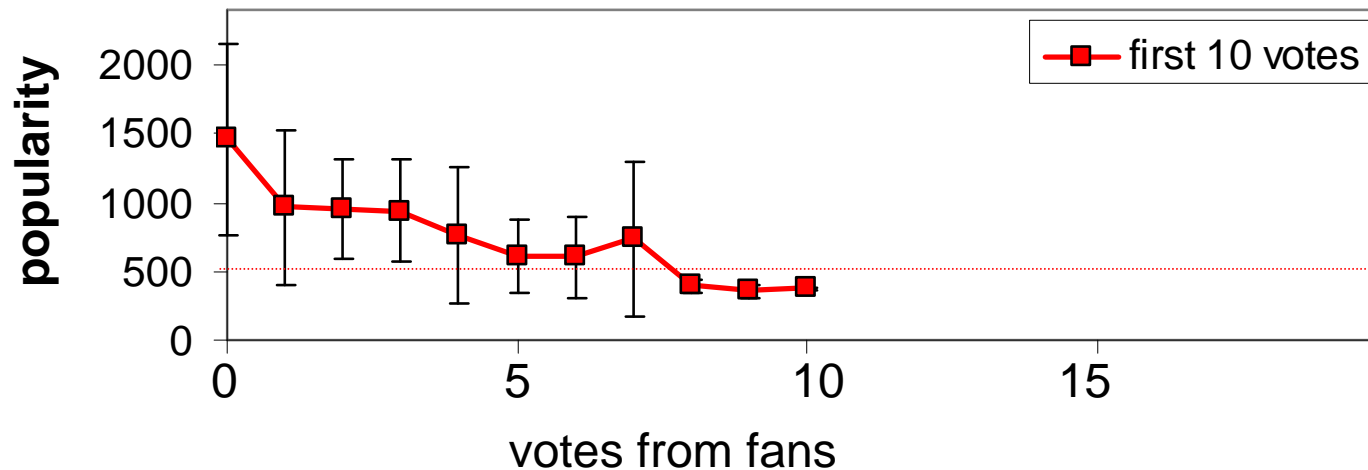


network of fans

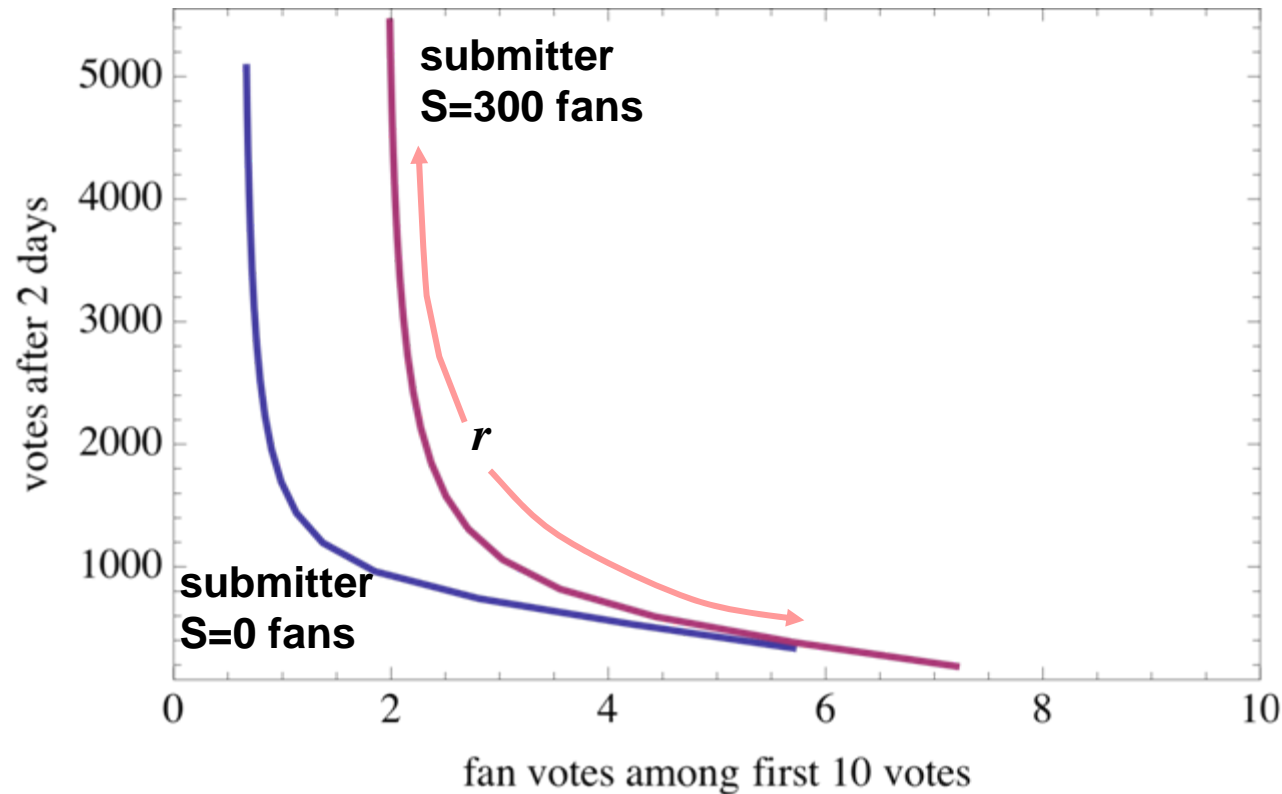


- **Stories that become popular initially receive fewer fan votes**
 - Fan votes = votes from fans of submitter and previous voters
 - Trend strong enough to predict whether a story will become popular (receive > 500 votes) [Lerman & Galstyan 2008]

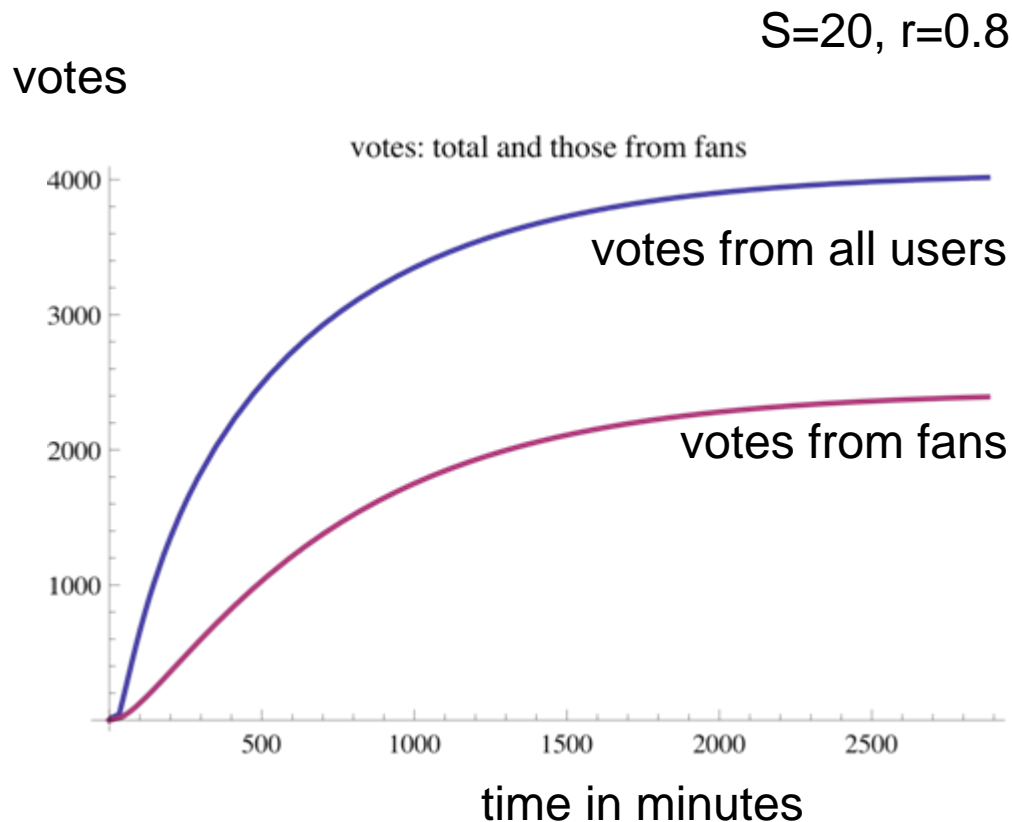
Final story popularity vs fan votes within the first 10 votes



- Model predicts qualitatively the same early reaction to the story from the network



- Estimate how interesting the story is, r , based on the network's early reaction to the story
- Use r value to predict dynamics of votes



- **Extensions to the model**
 - Different “interestingness” for fans and general users
 - Beyond simple Markov processes – incorporate memory of past votes
- **Model distribution of behavior, not just average**
 - Ubiquitous long-tail distributions
 - A few users contribute most content and activity
 - Growth of niche interests
 - *[Wilkinson 2008; Wu and Huberman 2007]*
- **Parameter estimation**
 - Model has some poorly determined parameters
 - *[Brown et al. 2003]*

- **Stochastic process approach**
 - connect user and system behaviors
- **Applicability:**
 - users have limited information and actions
 - limited use of personalized history
 - e.g., user communities on the web
 - *not face-to-face small group interactions*
- **Example: news aggregator Digg**
 - votes from visibility + interestingness
 - user model from info and actions provided by Digg UI
 - use the model to predict story popularity from the community's early reaction to it