A Meme is not a Virus:
the Role of Cognitive Heuristics in Information Diffusion

USC Viterbi
School of Engineering

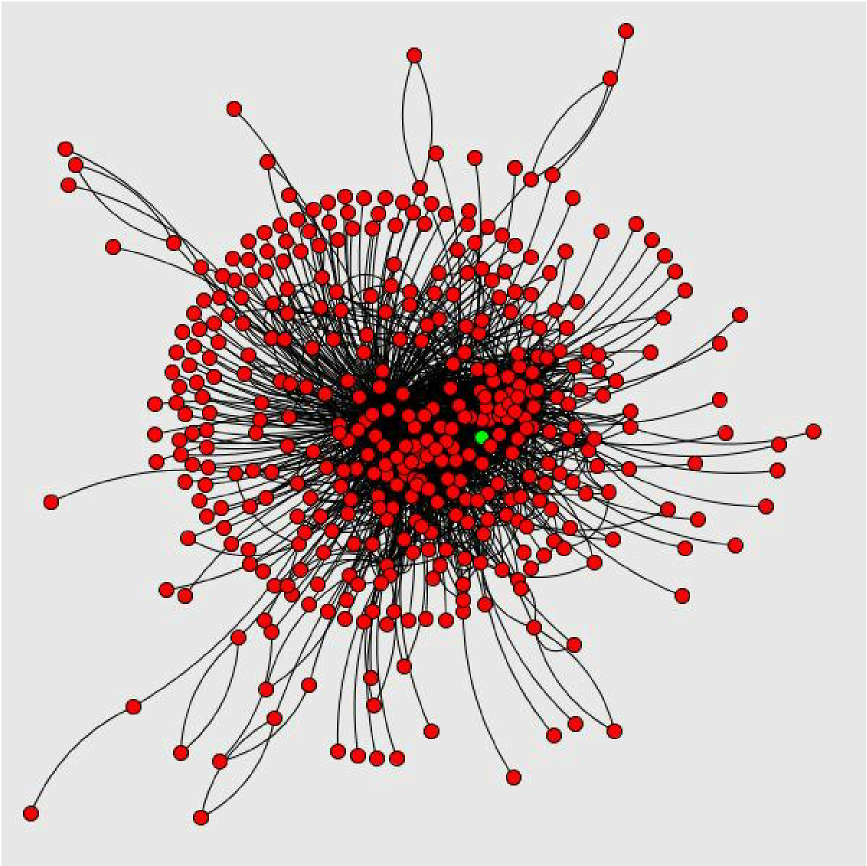
Kristina Lerman

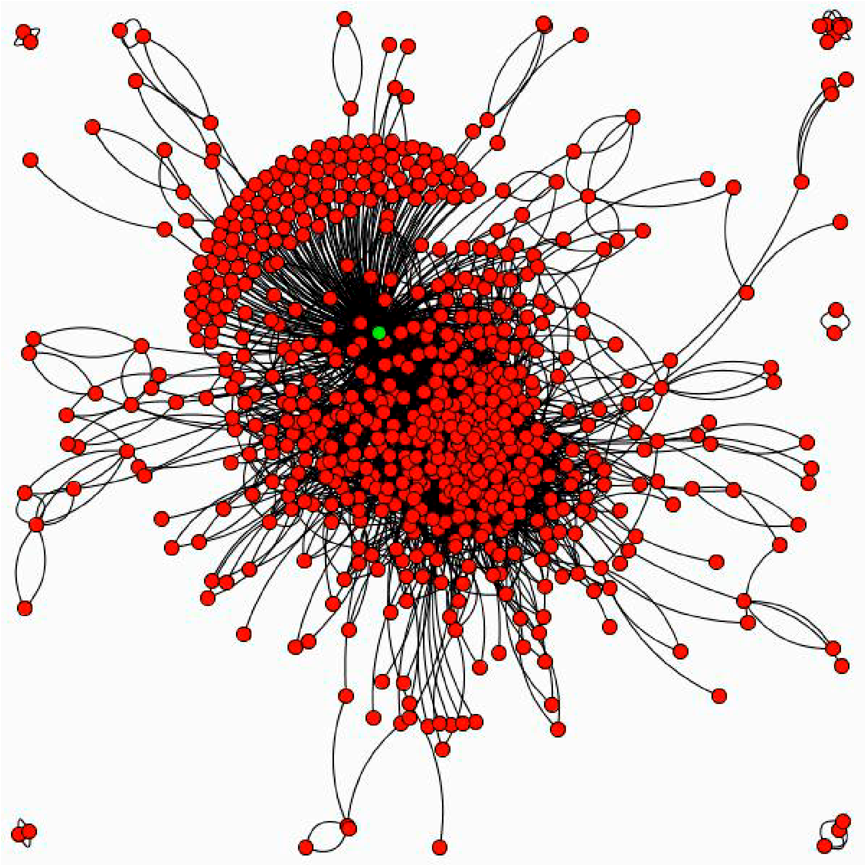
USC Information Sciences Institute

http://www.isi.edu/~lerman

ACM Hypertext Conference, Prague, Czech Republic, July 2017

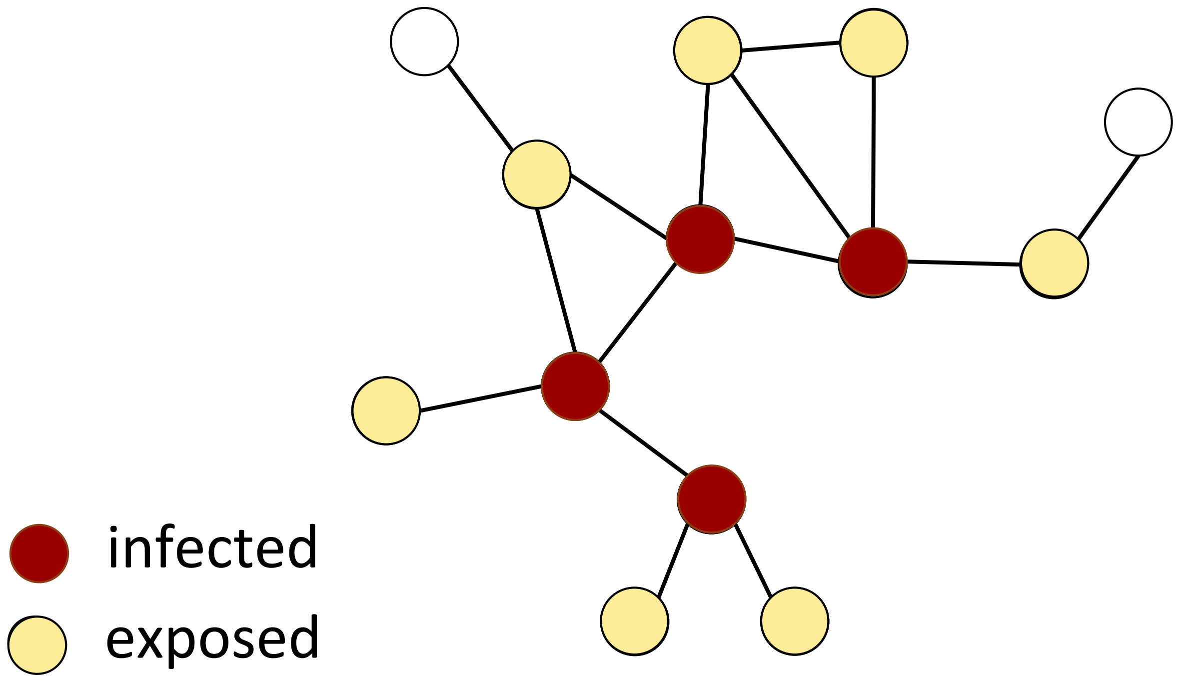
The spread of information in social networks





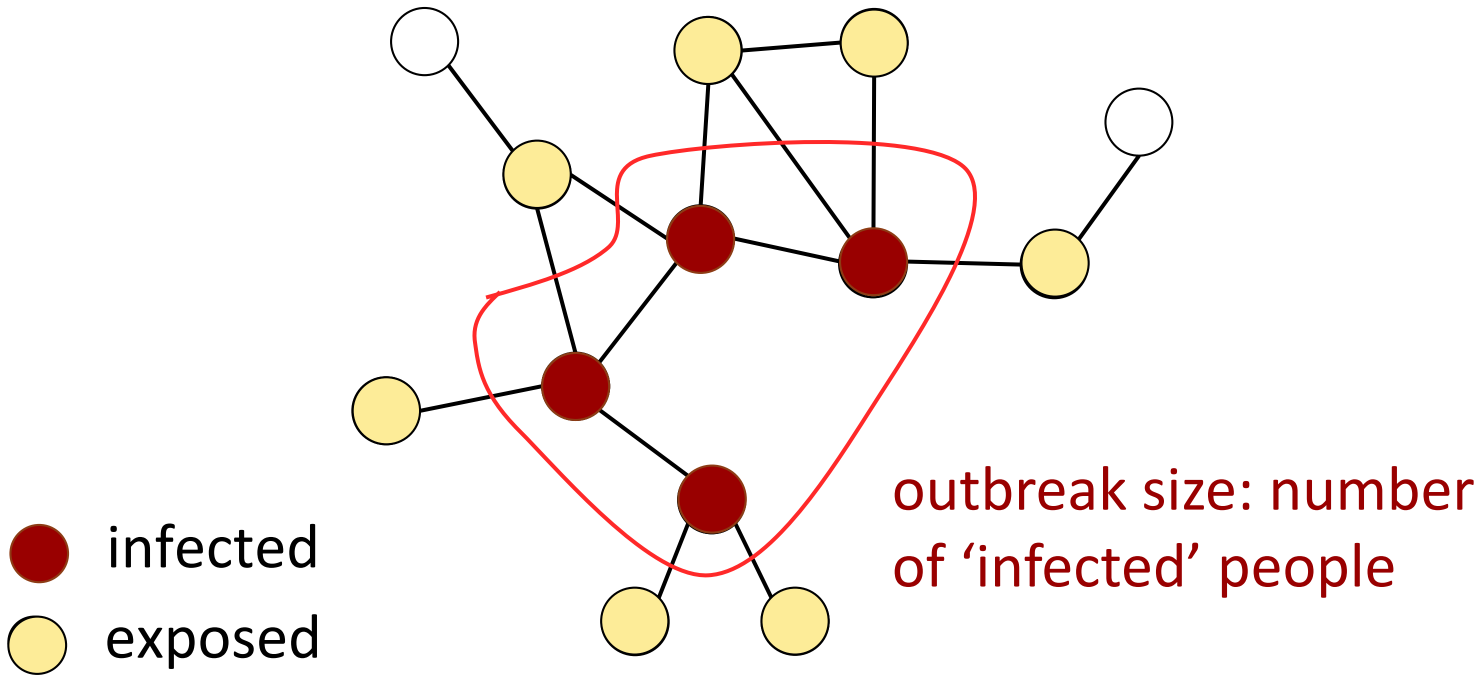
Information spread as social contagion

Standard model of contagion: "A meme behaves like a virus, with each exposure of a naïve individual by an informed friend potentially resulting in an 'infection' (meme transmission)" - M. Gladwell



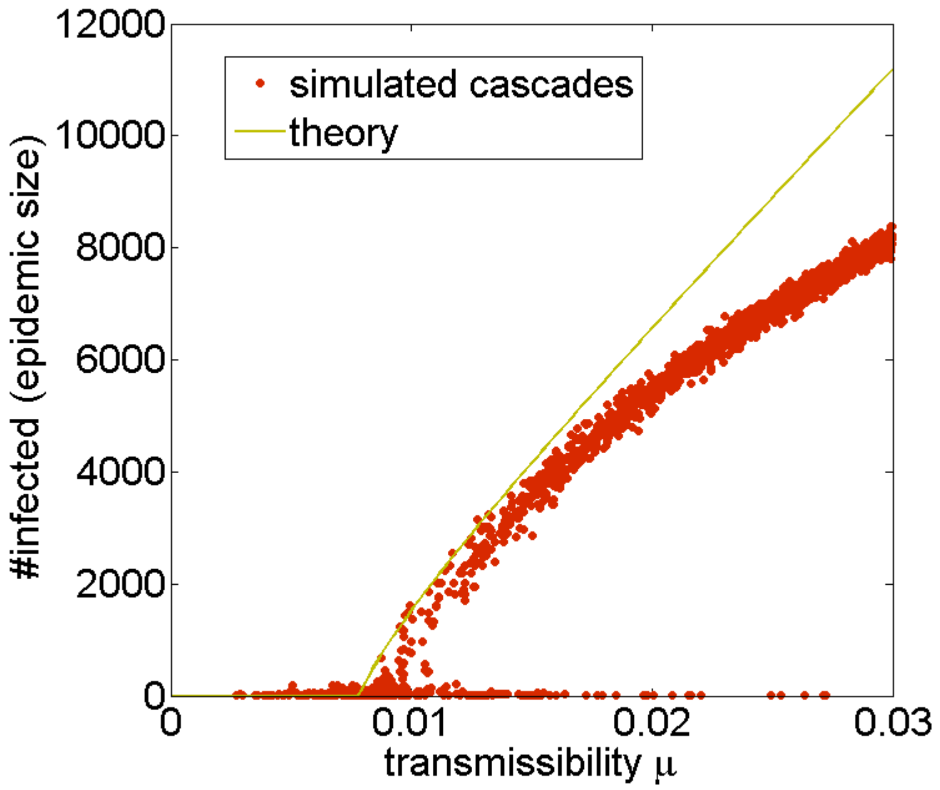
Information spread as social contagion

Standard model of contagion: "A meme behaves like a virus, with each exposure of a naïve individual by an informed friend potentially resulting in an 'infection' (meme transmission)" - M. Gladwell



How large are outbreaks?

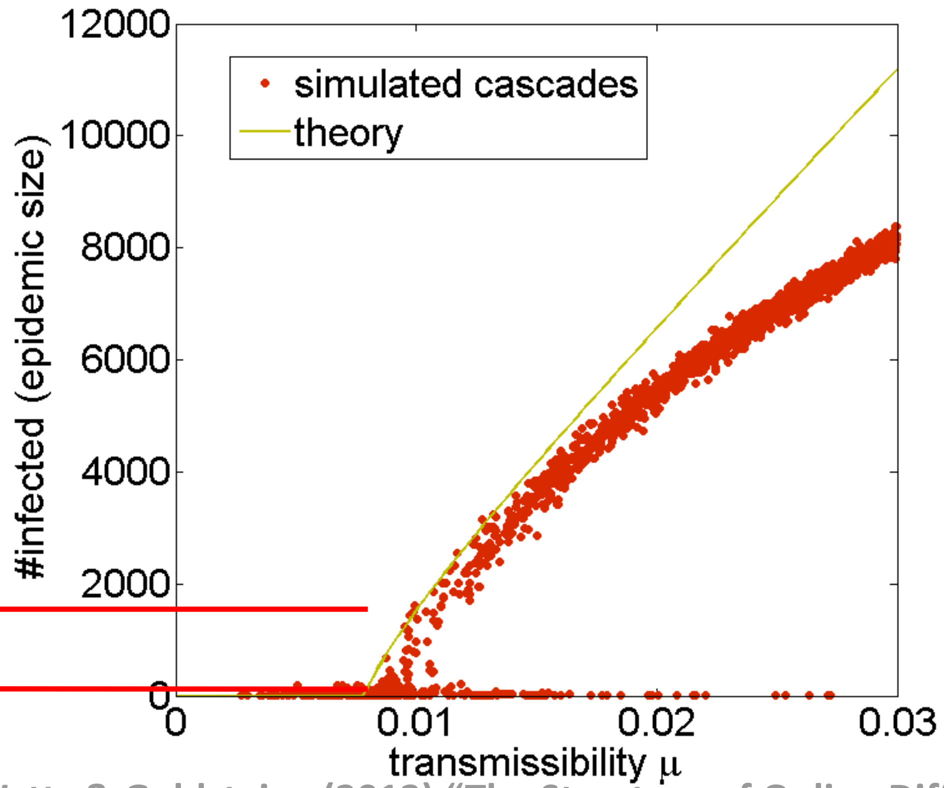
Standard model of contagion (independent cascade model) predicts large outbreaks above some value transmissibility



[Ver Steeg, Ghosh & Lerman (2011) “What stops social epidemics?” in *ICWSM*]

How large are outbreaks?

Standard model of contagion (independent cascade model) predicts large outbreaks above some value transmissibility

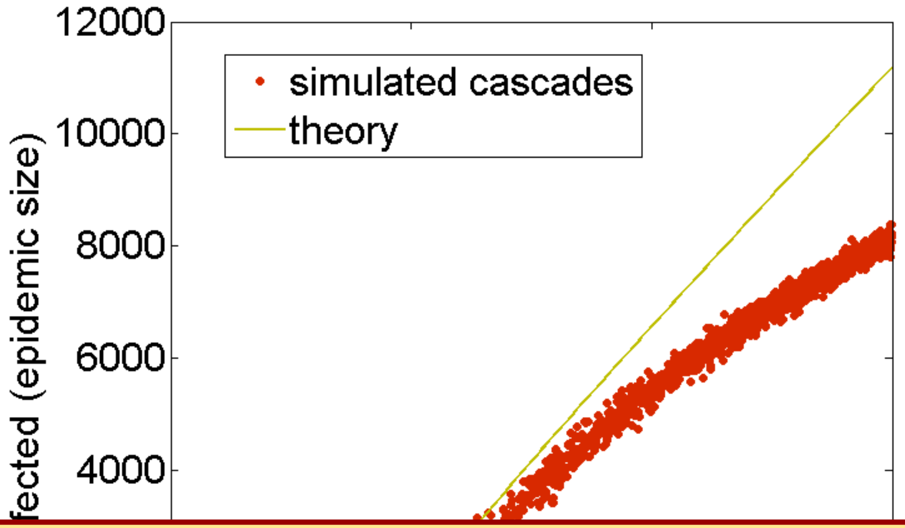


Most social media cascades fall in this range

[Goel, Watts & Goldsteing (2012) "The Structure of Online Diffusion Networks" in *EC.*]
[Ver Steeg, Ghosh & Lerman (2011) "What stops social epidemics?" in *ICWSM*]

How large are outbreaks?

Standard model of contagion (independent cascade model) predicts large outbreaks above some value transmissibility



Puzzle: There are few "viral" outbreaks in social media; even largest ones reach less than 5% of the network.

[Ver Steeg, Ghosh & Lerman (2011) "What stops social epidemics?" in *ICWSM*]

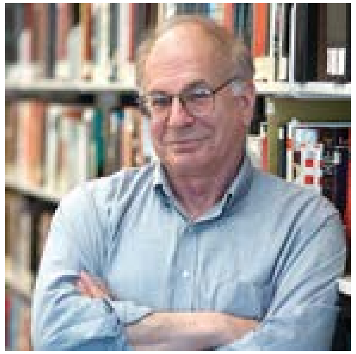
Roadmap

To understand information diffusion – and online behavior in general – we must account for cognitive factors

1. What are cognitive heuristics and biases?
2. How do we measure their impact on online behavior?
   - Empirical analysis of social media
   - Experimental study on MTurk
3. How do we model cognitive biases?
   - Accounting for cognitive heuristics simplifies models of information diffusion
4. Cognitive biases in applications

Bounded rationality (aka “thinking is hard”)

Herbert A. Simon



Daniel Kahneman

Amos Tversky

Bounded rationality

Constraints of available time, information, and cognitive capacity limit human ability to make rational decisions

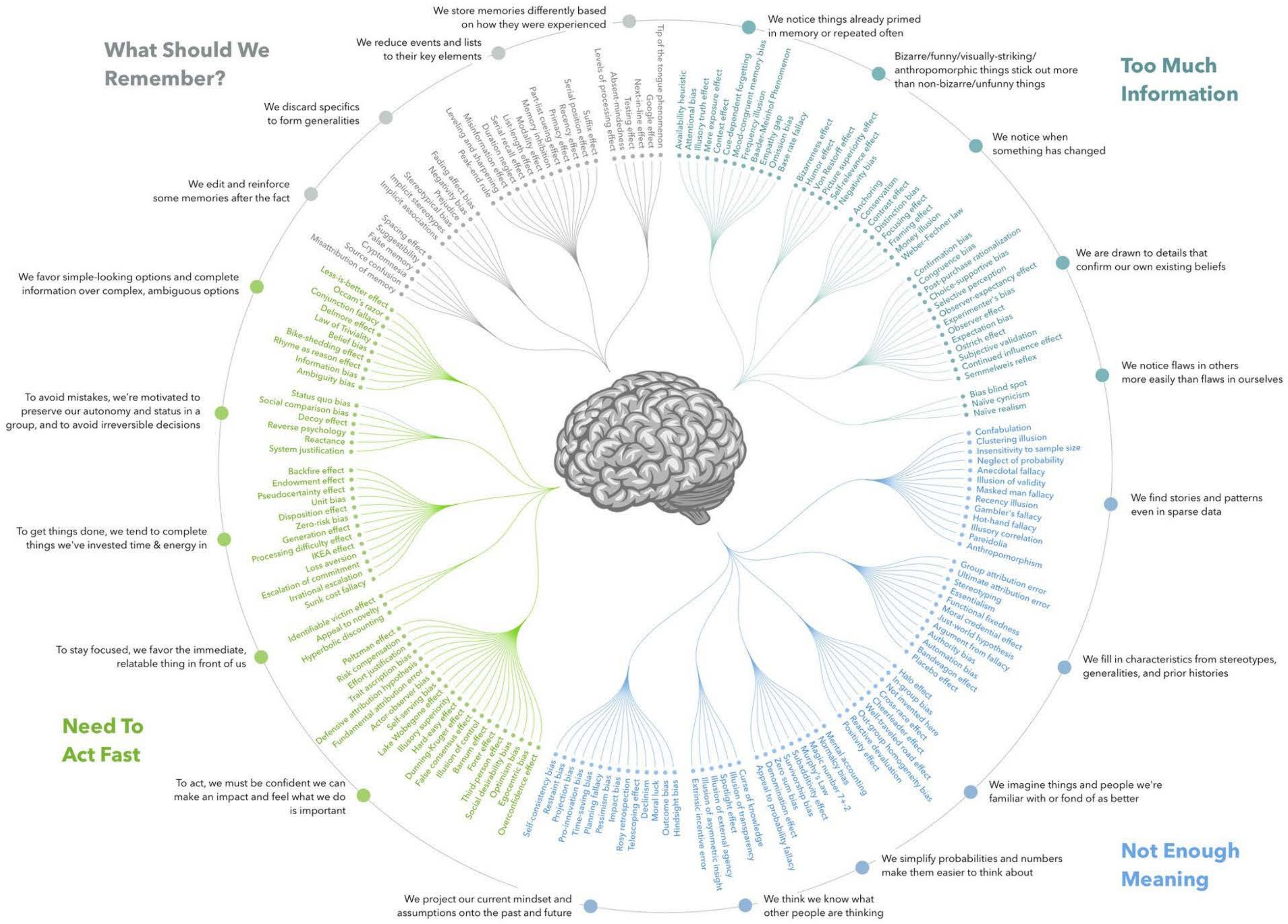
[Simon (1957). "A Behavioral Model of Rational Choice", in Mathematical Essays on Rational Human Behavior in a Social Setting. New York: Wiley]

Heuristics and biases

Mental shortcuts that help people make quick, but less accurate decisions, by focusing brain’s limited resources on the most salient information

[Tversy and Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science*
Kahneman (2011) *Thinking Fast and Slow*.]

COGNITIVE BIAS CODEX, 2016



ALGORITHMIC LAYOUT + DESIGN BY JM3 · JOHN MANOOGIAN III // CONCEPT + METICULOUS CATEGORIZATION BY BUSTER BENSON // DEEP RESEARCH BY WIKIPEDIANS FAR + WIDE

Types of cognitive biases we measured

Position bias: People pay more attention to items at the top of the screen or a list of items [Payne 1951]

[Buscher et al, CHI'09]

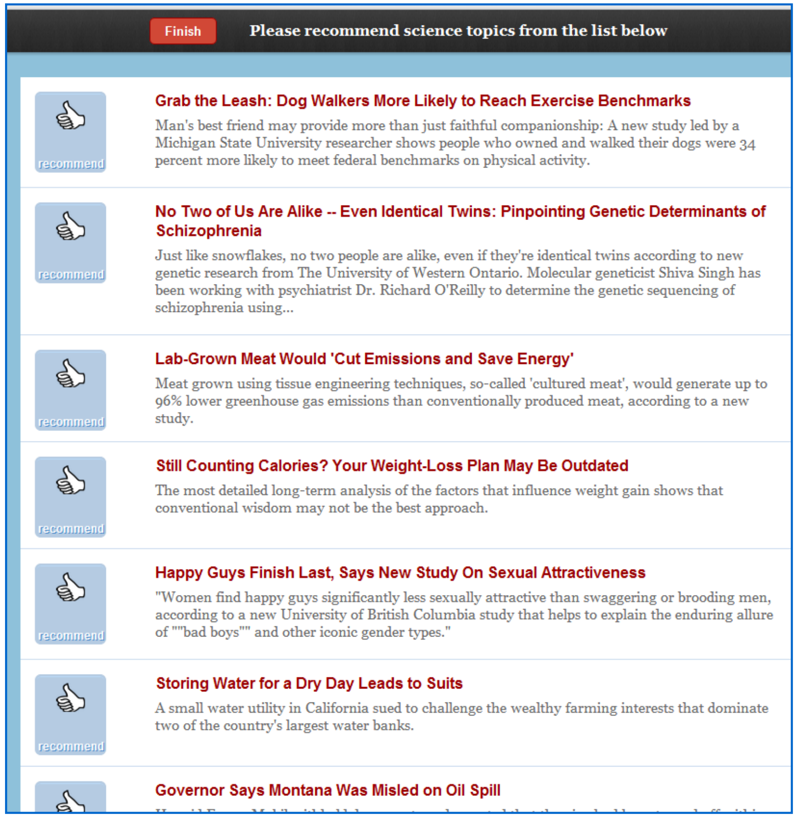
Social influence bias: People pay more attention to the popular choices

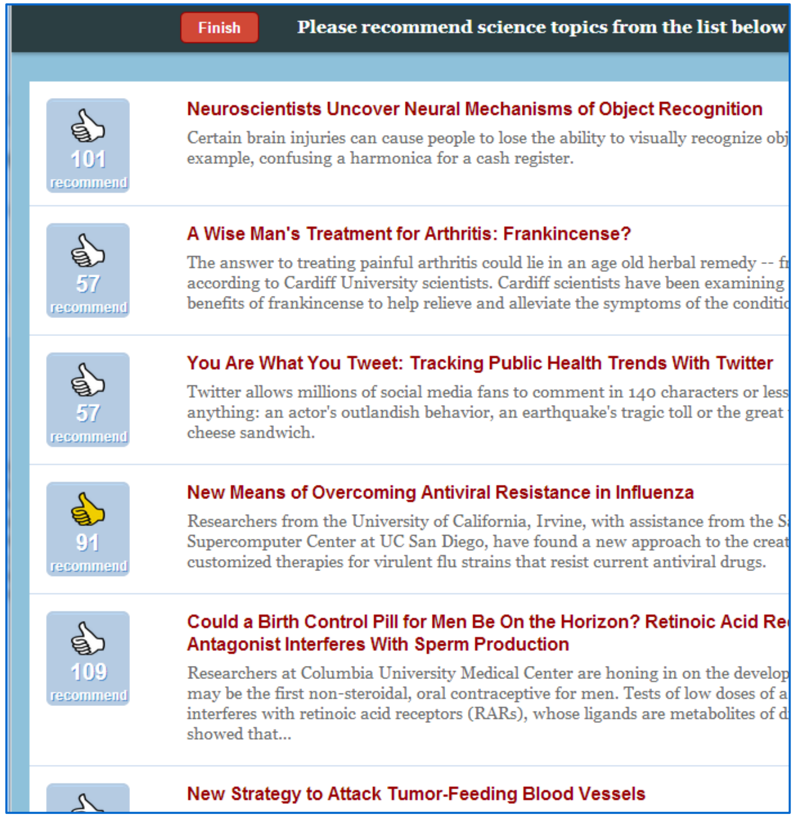
Other biases:

- Availability bias
- Primacy effect
- Confirmation bias

Measuring cognitive biases

- Controlled experiments on Amazon Mechanical Turk
- Asked people to recommend science stories they liked,
  - we varied the order stories were presented, and whether social signals were shown.

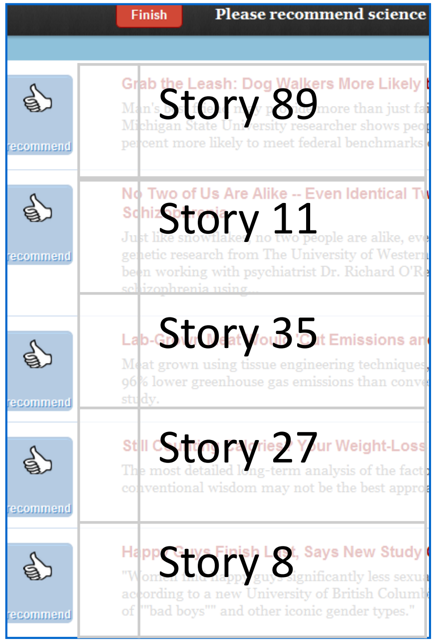




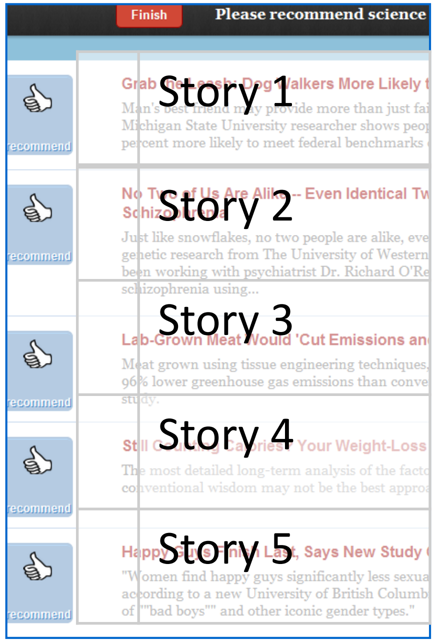
[Lerman & Hogg "Leveraging position bias to improve peer recommendation" in *PLoS One* (2014) arXiv:1202.3162]

Experimental design

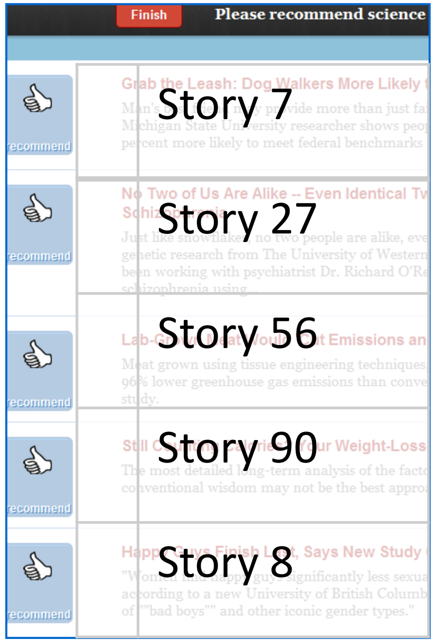
- Turkers asked to recommend stories from a list 100 science stories
- Vary ordering → measure outcomes (# recommendations)
- No direct social influence (users not shown # recommendations)
- Parallel worlds design, inspired by MusicLab experiment [Salganik et al., 2006]



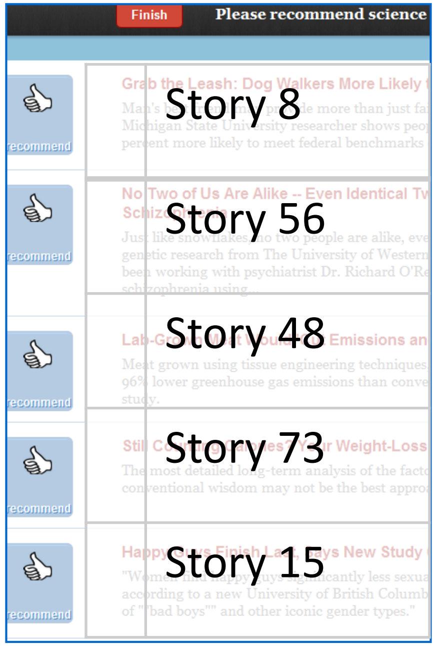
[random order]
control



[fixed order]



[by popularity]
 # recommendations

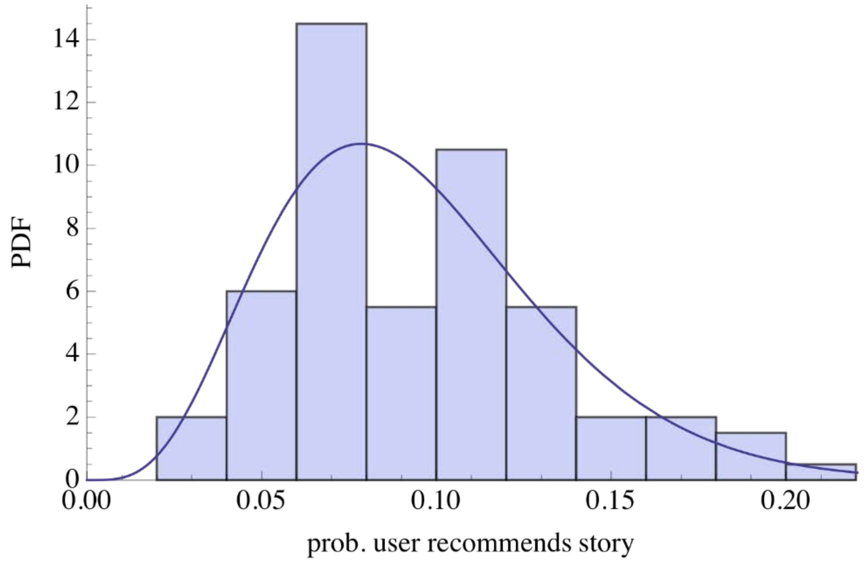


[by recency]
of recommen.

[Lerman & Hogg “Leveraging position bias to improve peer recommendation” in *PLoS One* (2014) arXiv:1202.3162]

"Quality"

Fraction of recommendations in the random ordering

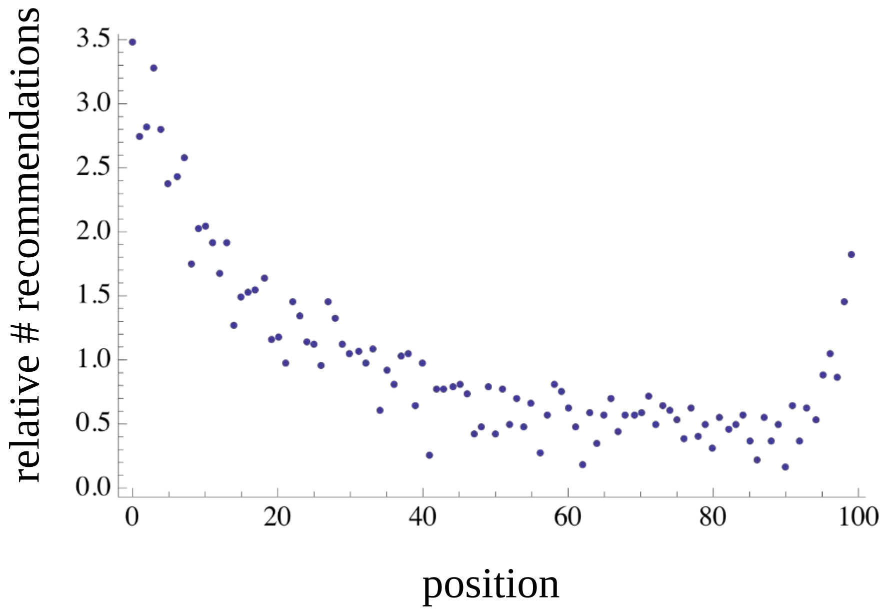


Large variation in how appealing stories are to users

[Lerman & Hogg (2014) "Leveraging position bias to improve peer recommendation" in *PLoSOne*]

Position bias

Accounting for quality, the number of recommendations a story receives simply due to its position gives position bias



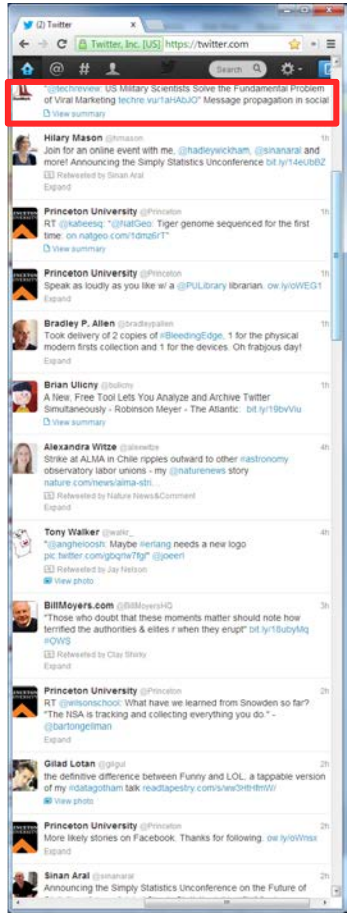
Items in top positions receive 4x as much attention as items in lower positions

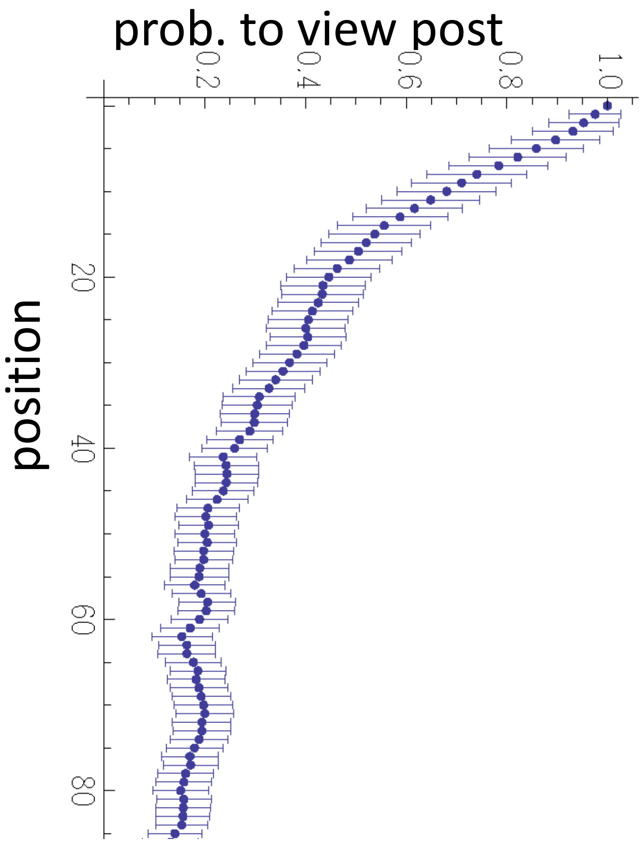
[Lerman & Hogg (2014) "Leveraging position bias to improve peer recommendation" in *PLoSOne*]

Position bias in social media

new post at the top of user's screen

post near the top is most likely to be seen

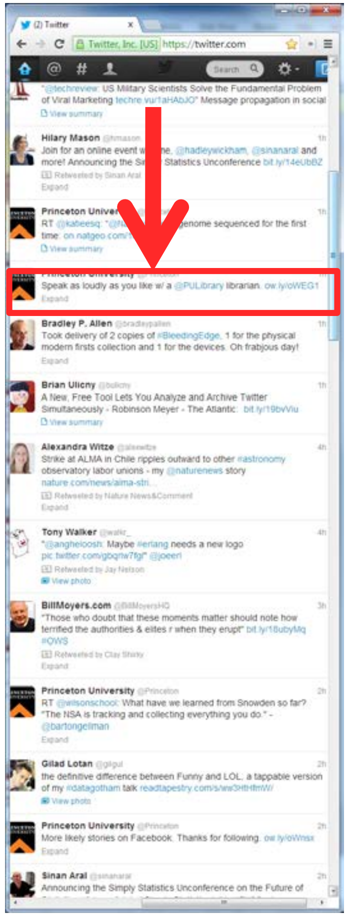


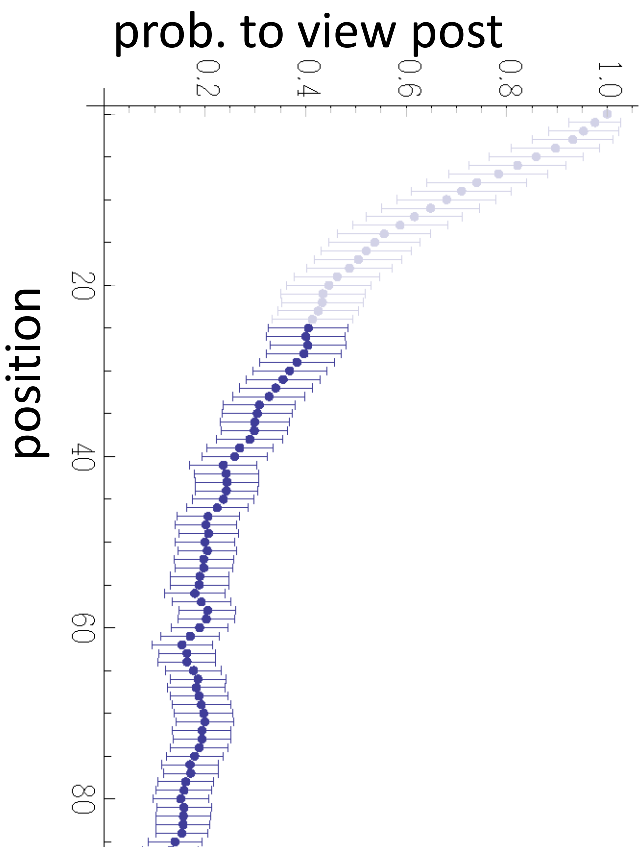


Position bias in social media

… later: newer posts from friends appear at the top

post is less likely to be seen



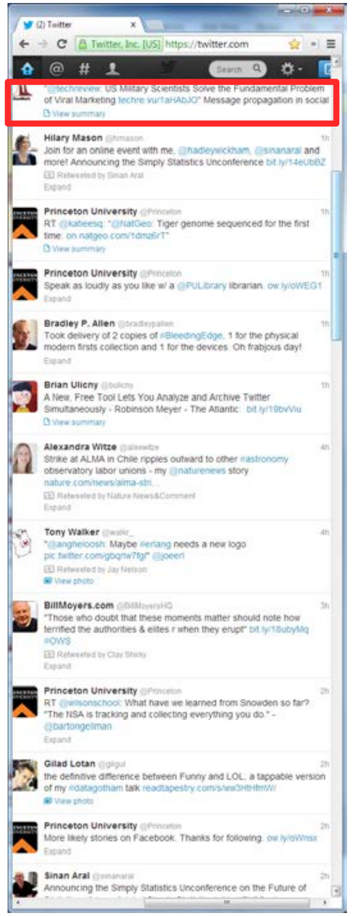


Users divide attention over all incoming posts

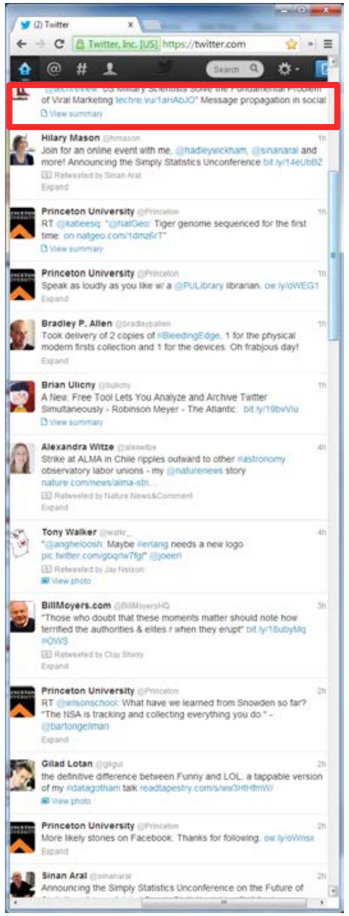
new post at top of user's screen

post near the top is most likely to be seen

few friends



many friends



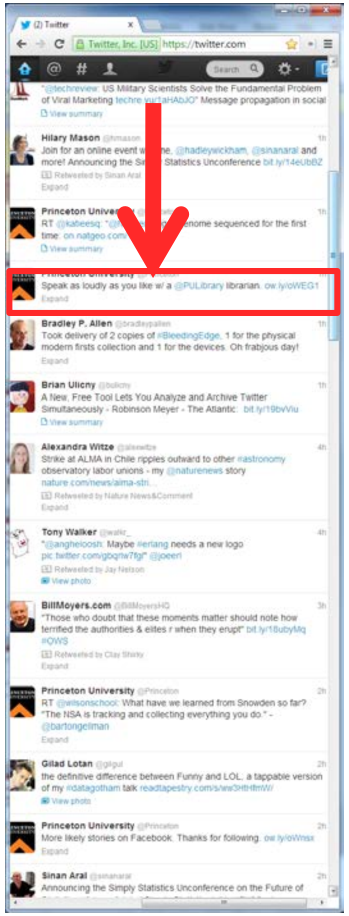
Users divide attention over all incoming posts

few friends

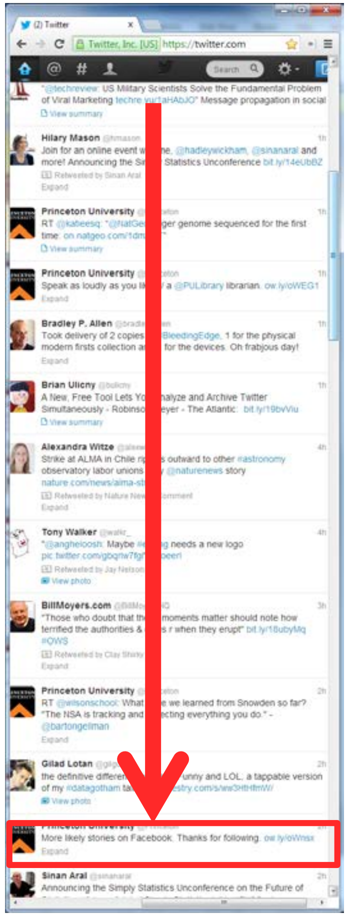
many friends

… later: newer posts from friends appear at the top

post is less likely to be seen

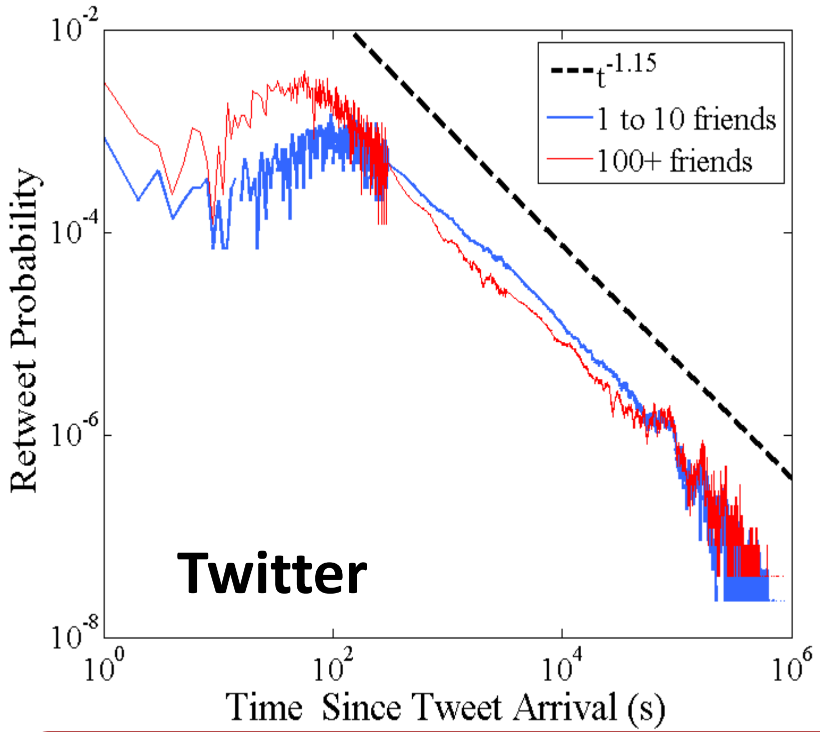


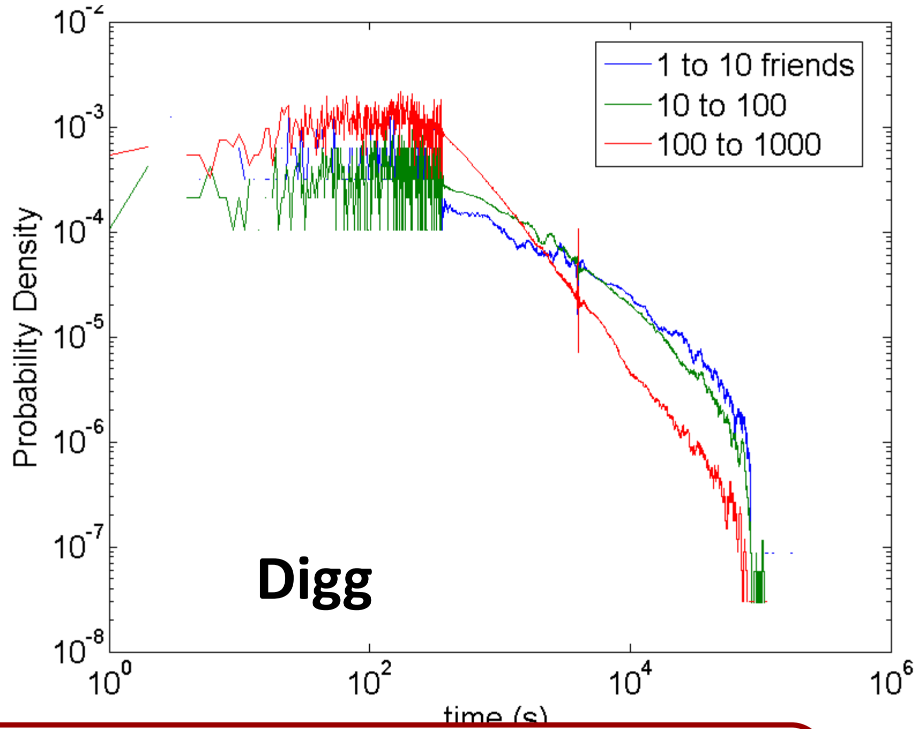
same age post is even less likely to be seen by a well-connected user



Position bias in social media: Empirical evidence

Retweet probability decreases with time since post's arrival



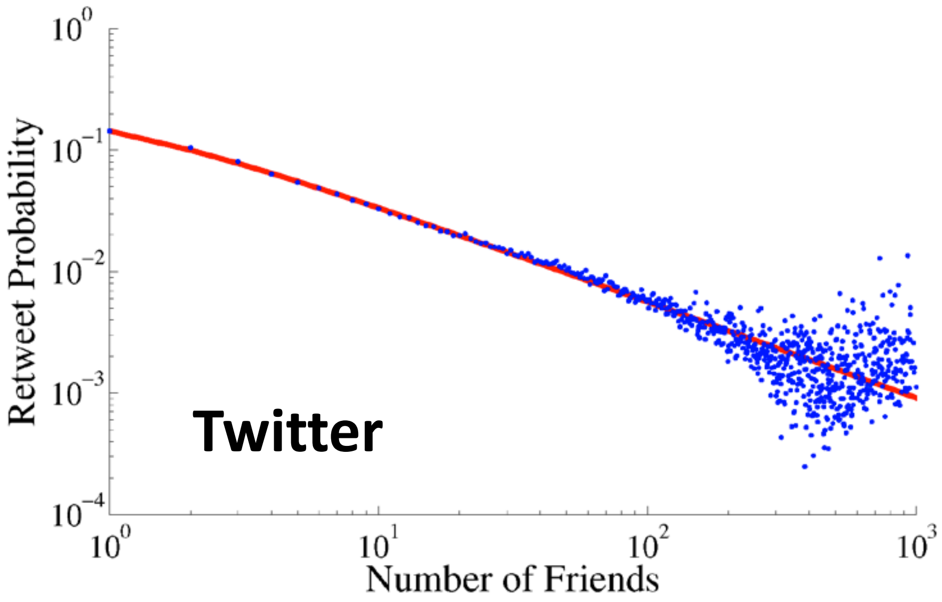


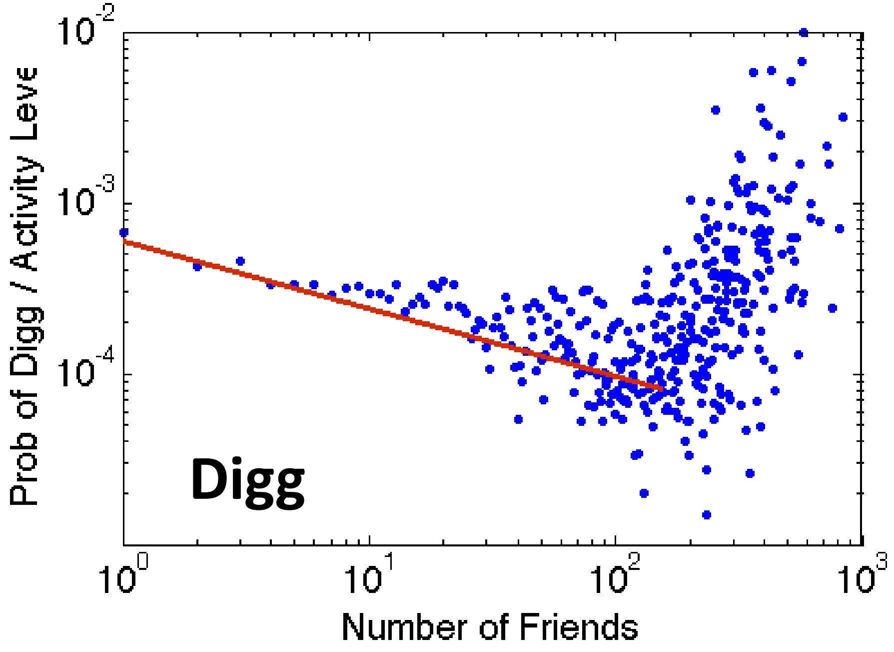
Observation: Well-connected hubs (i.e., those following many others) are less likely to retweet older posts.

[Hodas & Lerman "How Limited Visibility and Divided Attention Constrain Social Contagion" in *SocialCom-2012.* arXiv:1205.2736]

Users divide attention over all incoming posts

Retweet probability decreases with connectivity



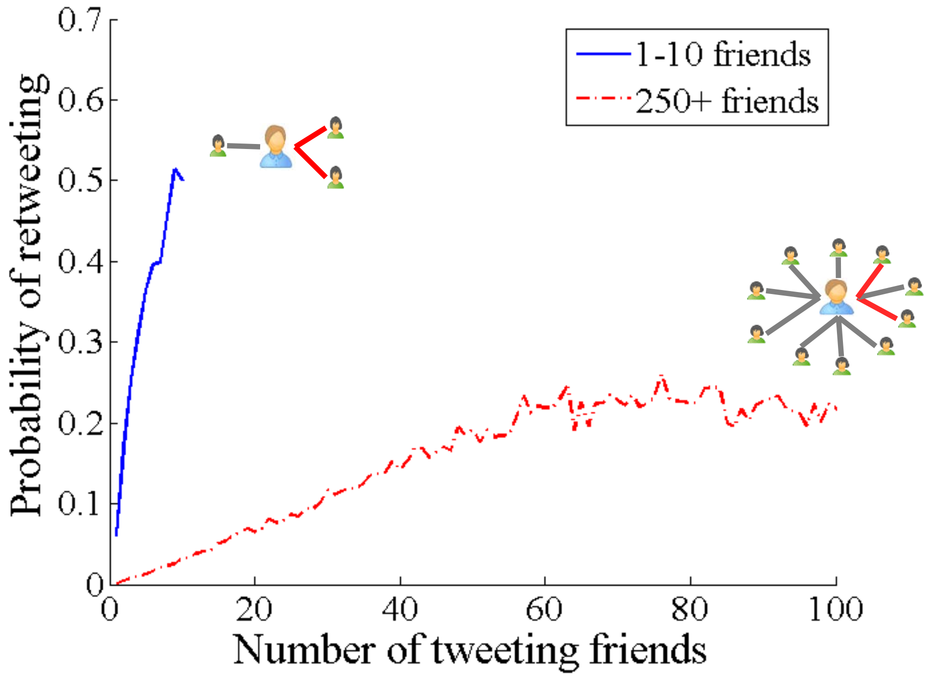


Observation: Well-connected people (i.e., those following many others) are less likely to retweet a post.

[Hodas & Lerman (2012) "How Limited Visibility and Divided Attention Constrain Social Contagion" in *SocialCom.* arXiv:1205.2736]

Exposure response

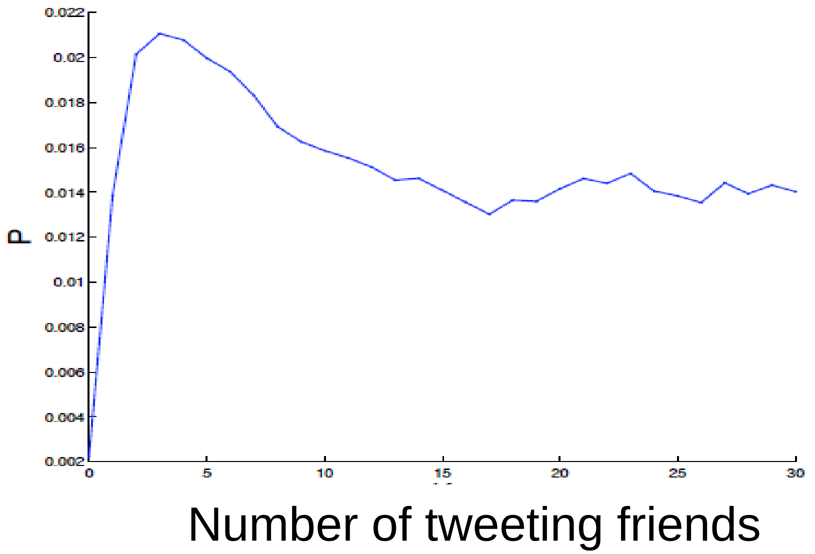
Highly connected people (i.e., hubs) are less susceptible to infection, due to their increased cognitive load



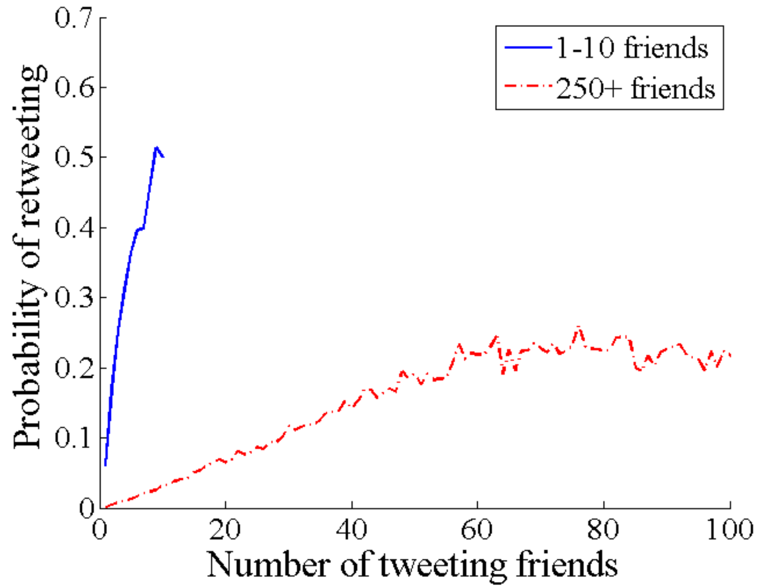
[Hodas & Lerman (2012) "How Limited Visibility and Divided Attention Constrain Social Contagion" in *SocialCom.* arXiv:1205.2736]

Complex vs simple contagion

Exposure response in social media: Additional exposures by friends appear to suppress response (probability to use a hashtag)[1]



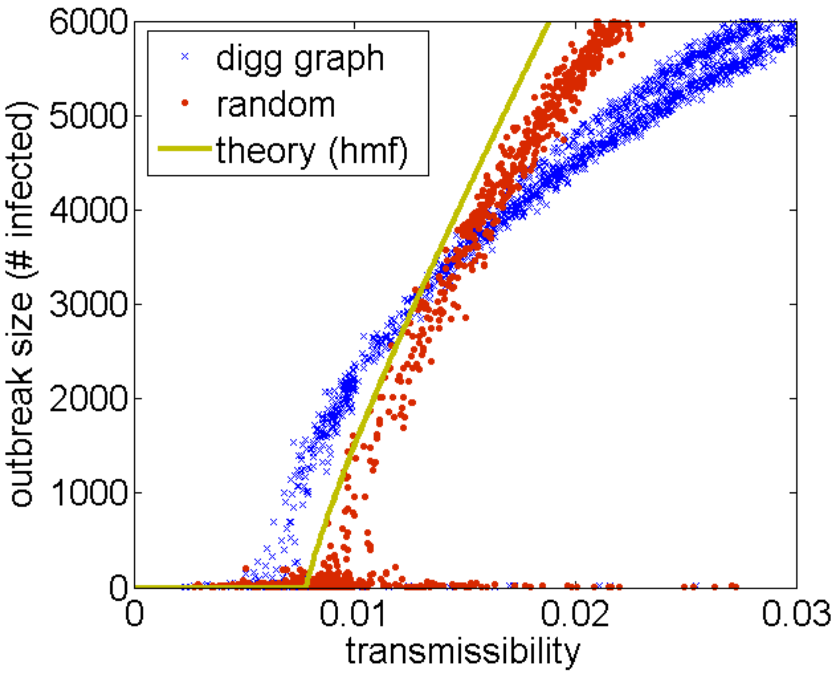
Exposure response in social media: When disaggregated by cognitive load, additional exposures amplify response (probability to retweet)



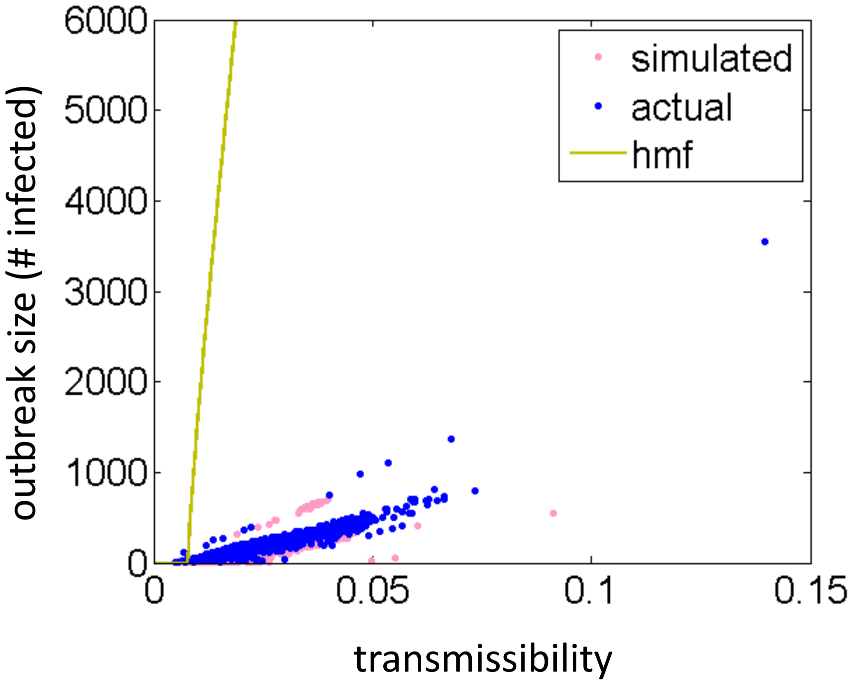
1. Romero, Meeder & Kleinberg (2011) "Differences in the Mechanics of Information Diffusion Across Topics" in *WWW*.
2. [Hodas & Lerman (2012) "How Limited Visibility and Divided Attention Constrain Social Contagion" in *SocialCom.*

Weak response of hubs suppresses outbreaks

Uniform susceptibility



Decreased susceptibility of hubs



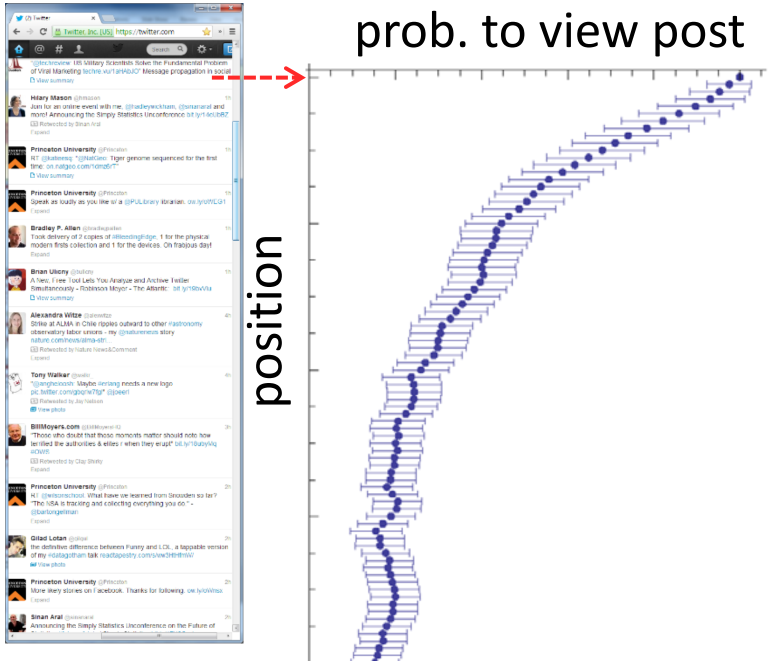
[Ver Steeg, Ghosh & Lerman (2011) "What stops social epidemics?" in *ICWSM*]

Modeling social contagion

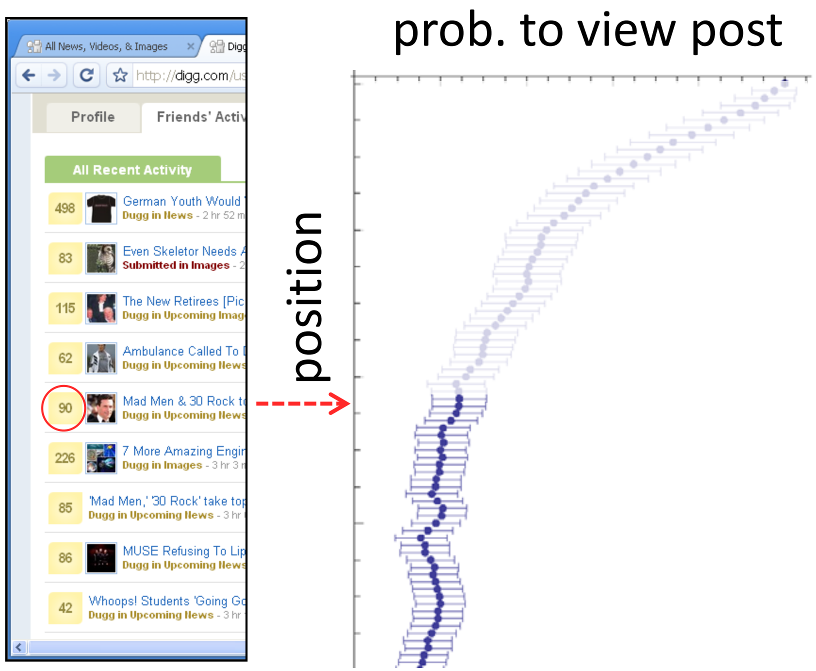
User must first see an item and find it interesting before he/she decides to retweet it

How do users respond to multiple exposures?

Twitter visibility: each retweet moves the post to top position in follower's stream



Digg visibility: a vote does not change position, but increments the social signal for followers



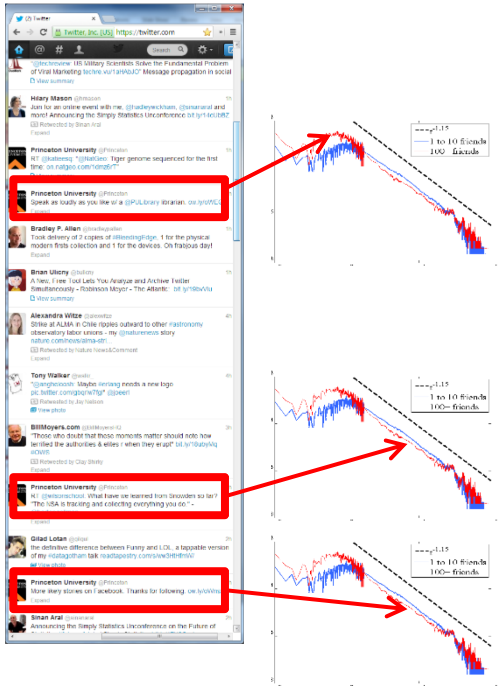
→ **web site's user interface affects salience of information, but social signals matter too**

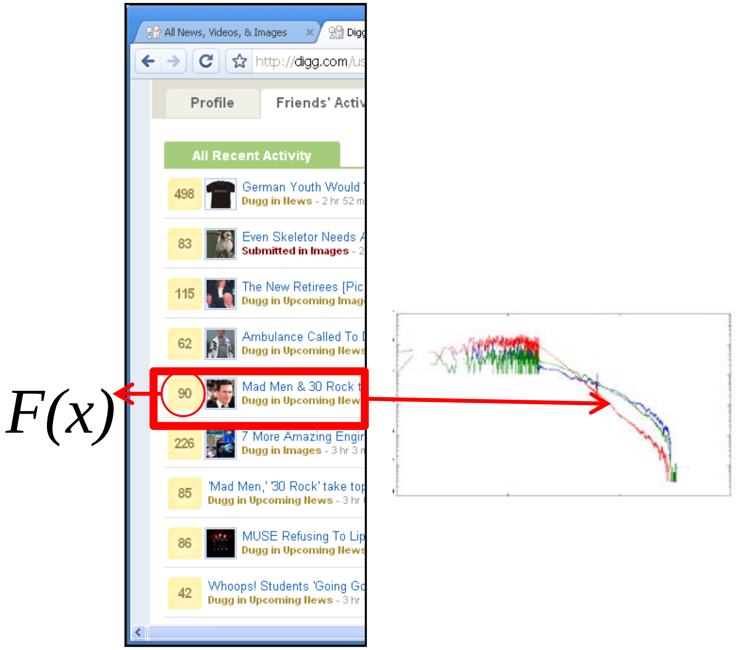
User response to multiple exposures

Probability that a user following n_f friends will retweet a post at time t after x exposures, depends on the **visibility** of exposures and **social influence** factor $F(x)$

$$P_{Twitter}(t; x, n_f) = F(x)\left(1 - \prod_{n=1}^{x} 1 - \mathcal{T}(t - t_n, n_f)\right)$$

$$P_{Digg}(t; x, n_f) = F'(x)\left(\mathcal{T}'(t - t_1, n_f)\right)$$

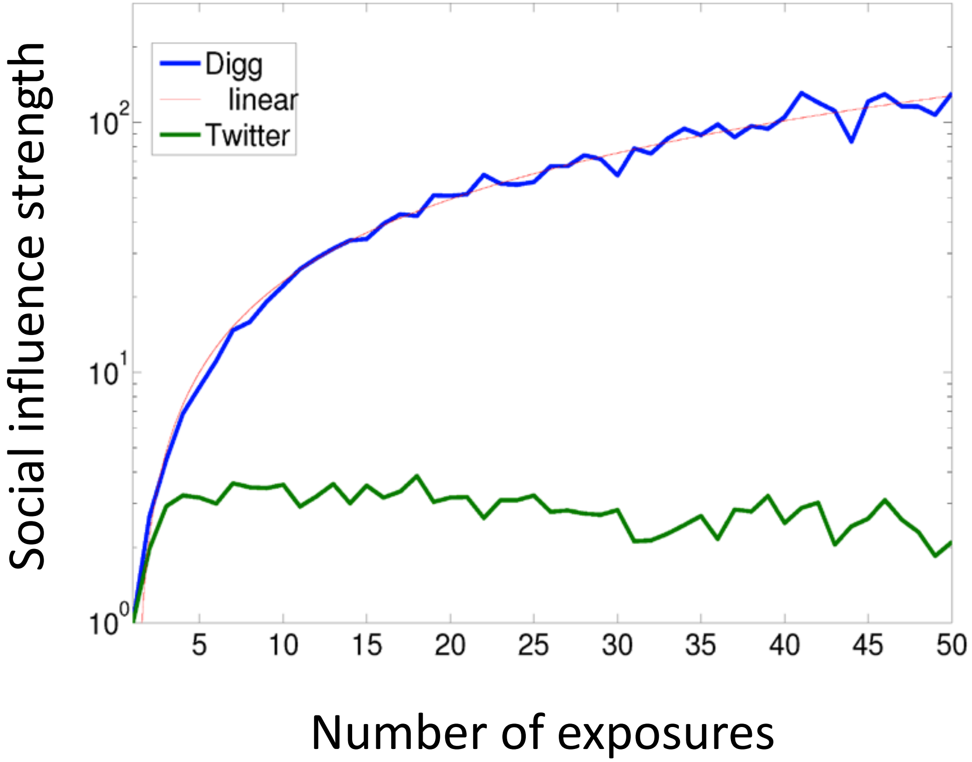




[Hodas & Lerman (2014) "The Simple Rules of Social Contagion" *Scientific Reports* 4]

Social influence amplifies response

Inferred social influence strength



Digg shows number of infected friends

Twitter does not, but users may remember earlier exposures

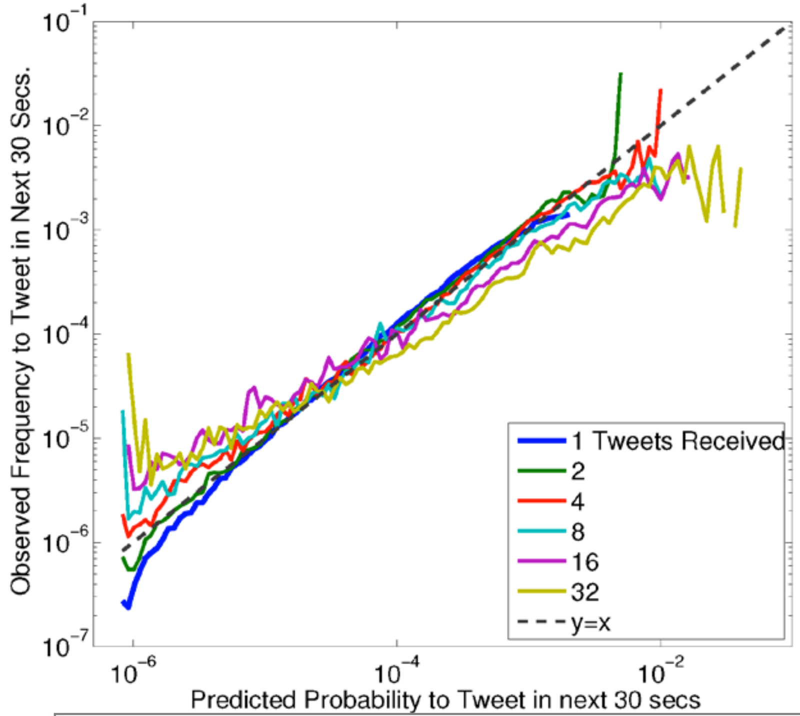
[Hodas & Lerman (2014) "The Simple Rules of Social Contagion" *Scientific Reports* 4]

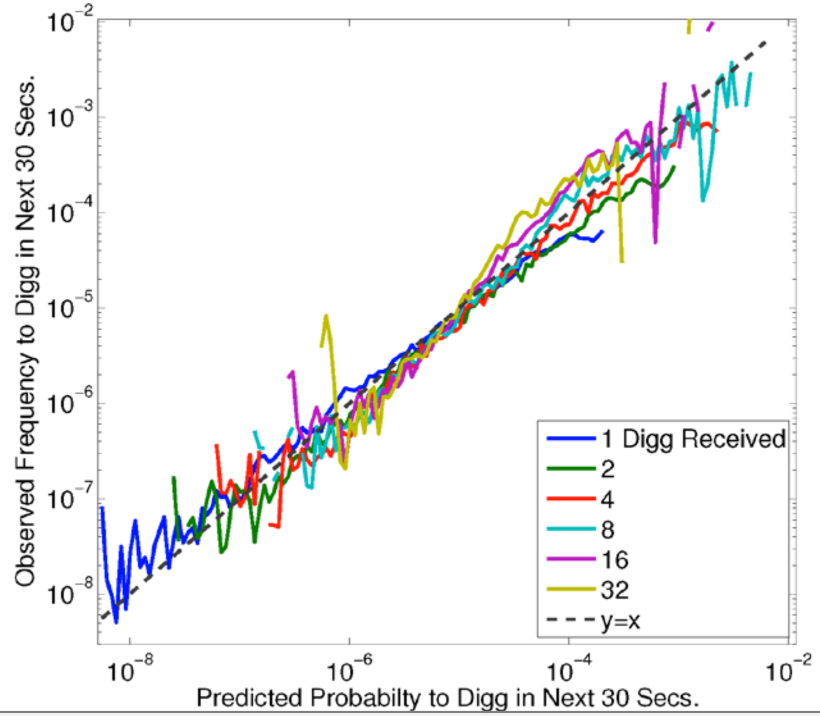
Predict user response to multiple exposures

Probability that a user following n_f friends will retweet a post at time t after x exposures, depends on the visibility of the exposures and social influence factor $F(x)$

$$P_{Twitter}(t;x,n_f) = F(x)\left(1 - \prod_{n=1}^{x} 1 - \mathcal{T}(t - t_n, n_f)\right)$$

$$P_{Digg}(t;x,n_f) = F'(x)\left(\mathcal{T}'(t - t_1, n_f)\right)$$





→ Model accurately predicts response regardless of exposures

Cognitive heuristics and navigation in networks

Navigation in social networks

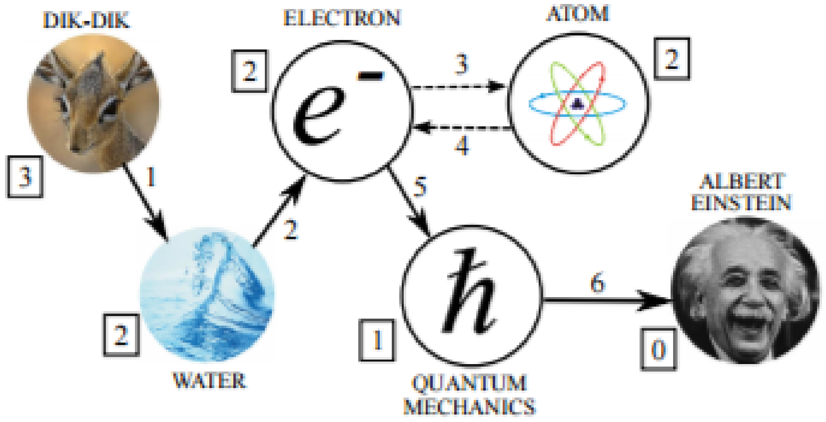
Stanley Milgram asked 160 random people in Kansas and Nebraska to deliver a letter to a stock broker in Boston. [Milgram, 1963]

“If you do not know the target, … mail this letter... to a personal acquaintance who is more likely than you to know the target.”

- Social networks are searchable!
  - Pairs of people are connected by short paths
  - People are remarkably good at finding short paths.

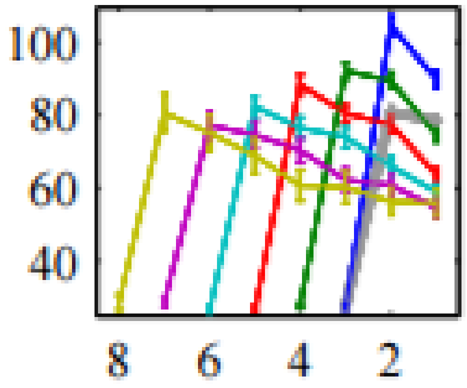
What makes online networks searchable?

- Wikispeedia game [West & Leskovec, 2012]
- On average, users reached a target in 3-4 hops



- Hubs are crucial, esp. initially
  - First hop gets user to a ‘hub’, i.e., a high-degree node, which is easily reachable from everywhere in a network

Average degree of a node reached in x hops



[West & Leskovec (2012) “Human Wayfinding in Information Networks”, in *WWW.*]

Navigation and page layout

- The layout of Wikipedia facilitates navigations
- Wikipedia page layout
  - Lead

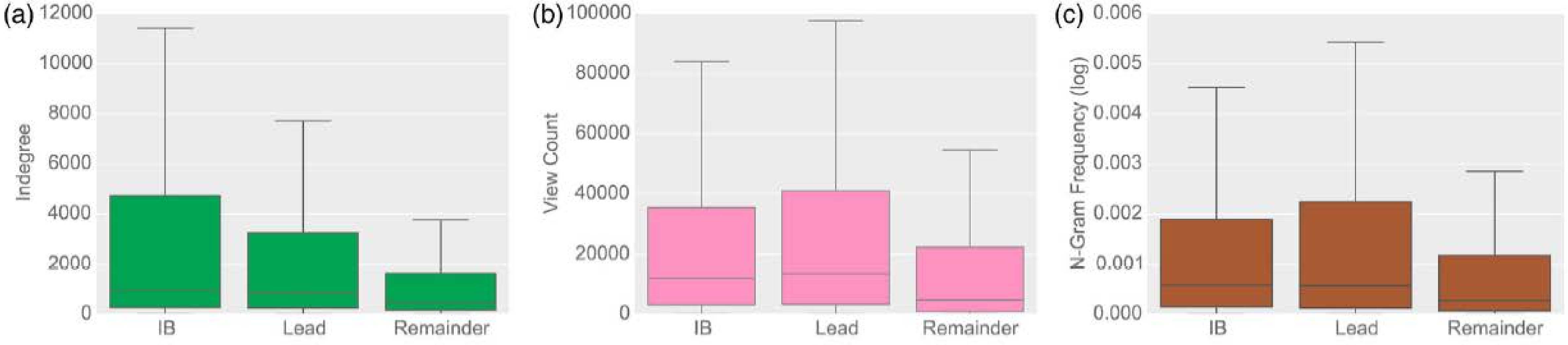
    First paragraph discusses general concepts
  - Infobox

    Section giving important statistics

[Lamprecht, Lerman, Helic & Strohmaier (2016) "How the structure of Wikipedia articles influences user navigation" in *New Review of Hypertext and Multimedia*]

Navigation and page layout

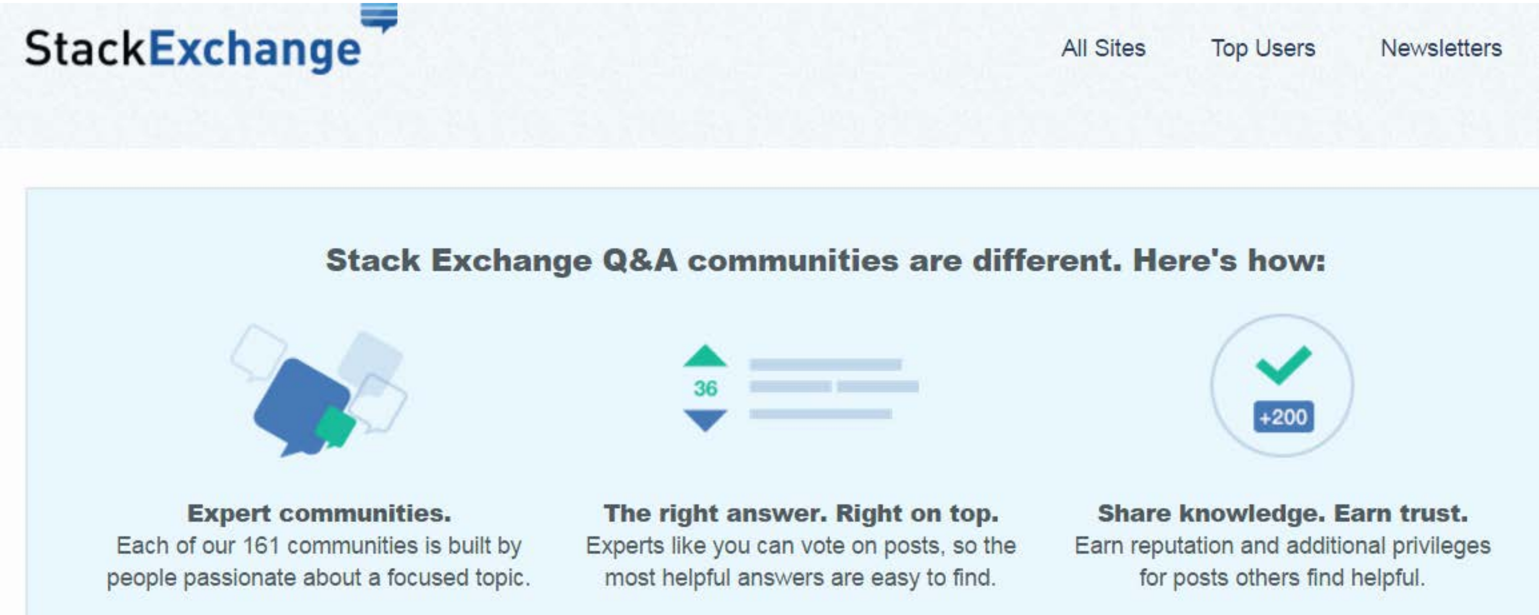
- People pay more attention to information in the lead and infobox sections (more views)
- Hyperlinks from these sections lead to hubs, i.e., pages
  - with higher degree (more links)
  - dealing with more general concepts (higher n-gram frequency)



Indegree (Wikipedia) View Count (Wikipedia) N-Grams (Wikipedia)

[Lamprecht, Lerman, Helic & Strohmaier (2016) "How the structure of Wikipedia articles influences user navigation" in *New Review of Hypertext and Multimedia*]

Cognitive heuristics and crowdsourcing

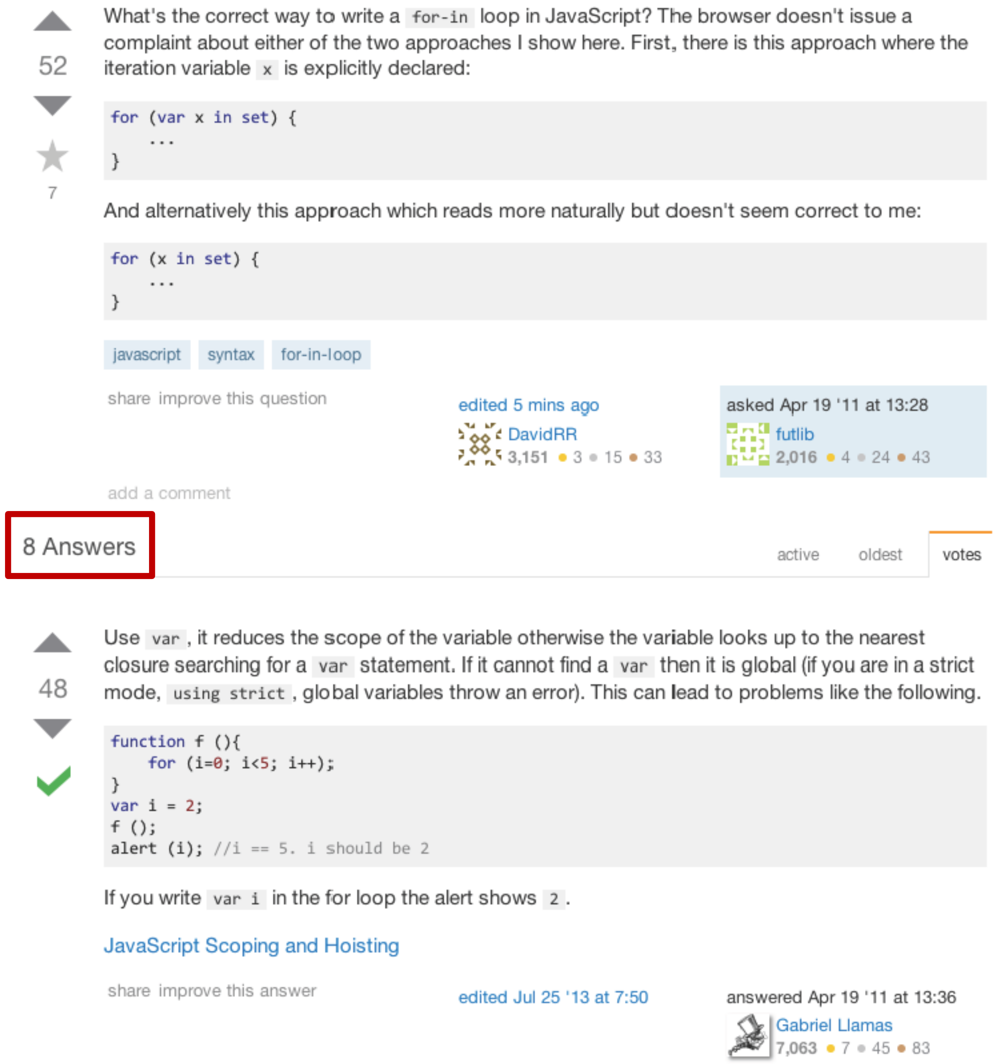


Anatomy of Stack Exchange

Question

Cognitive load
Number of answers to the question →

Answers



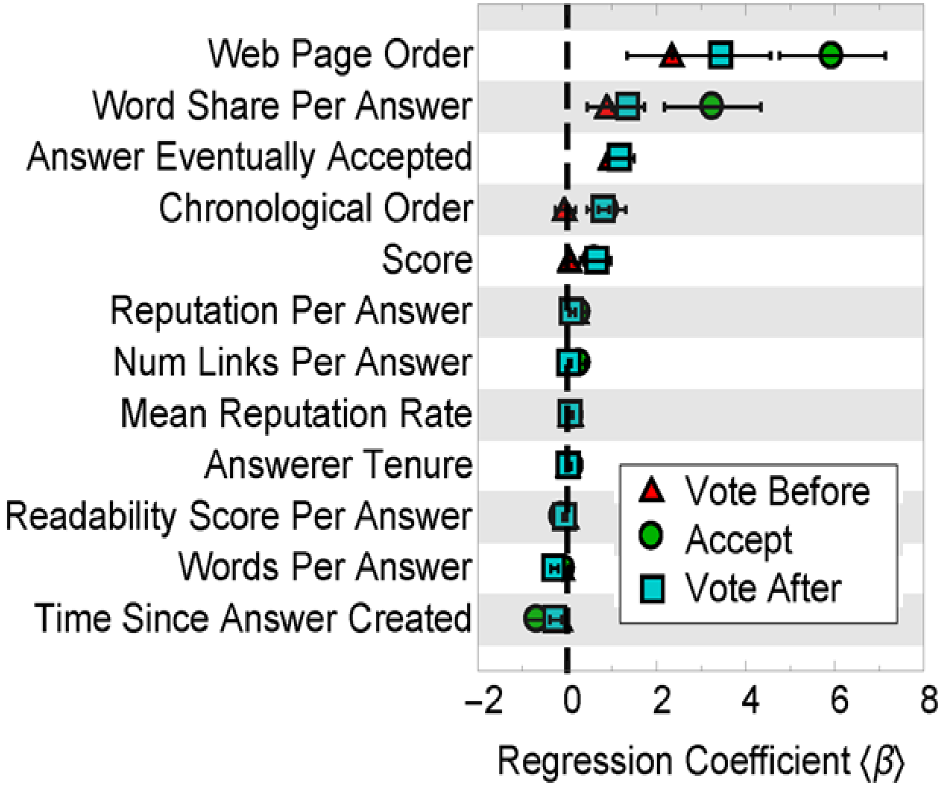
Answer features

- votes/score
- accepted?
- web page order
- chrono order

- num words
- word share
- hyperlinks
- readability
- age

- answerer reputation
- tenure

Regression coefficients highest for heuristics

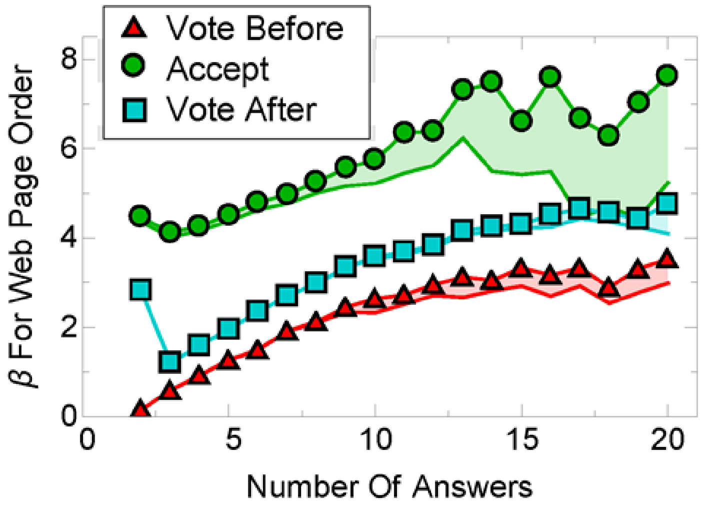


→ **Rather than evaluate all answers, people use simple heuristics to choose answers to vote for or accept.** Largest coefficients are:

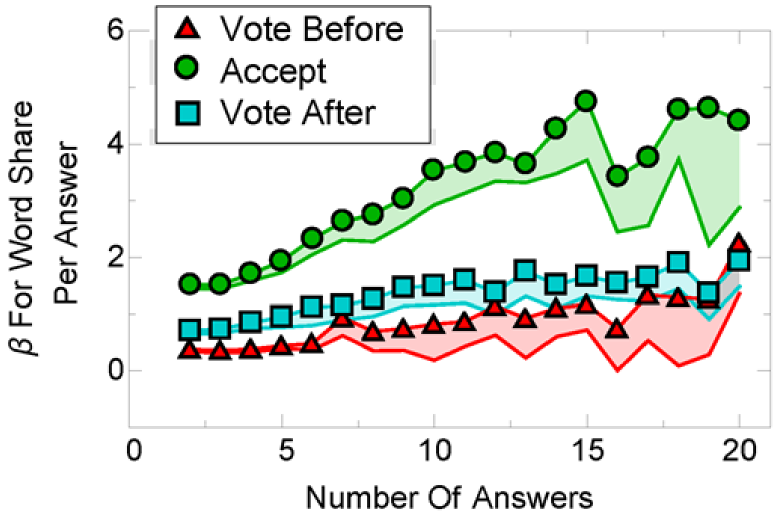
- Web page order → answer's rank (*cf* position bias)
- Word share → fraction of the screen it occupies (*cf* availability bias)
- Answer acceptance → social proof (*cf* social influence bias)

Cognitive load increases reliance on cognitive heuristics

Regression coefficient for web page order vs cognitive load*



Regression coefficient for for word share vs cognitive load*



* using number of answers available to a question as a proxy of cognitive load

[Burghardt, et al. (2017) The myopia of crowds: Cognitive load and collective evaluation of answers on Stack Exchange PloS one 12 (3), e0173610]

Summary

Availability of large-scale behavioral data has vastly expanded opportunities for discovery in the cognitive and behavioral sciences

- Evidence for bounded rationality in online behaviors
  - Rather than evaluate all available information and choices, people rely on simple cognitive heuristics
- Impact of cognitive heuristics on user choices and collective behavior
  - People rely on simple cognitive heuristics to make decisions, especially as their cognitive load increases
  - As a result, highly connected people suppress the spread of information online

Thanks to collaborators and sponsors

Tad Hogg

Greg Ver Steeg

Nathan Hodas

Rumi Ghosh

Farshad Kooti

Denis Helic

Markus Strohmaier

Daniel Lamprecht