

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

**Kristina Lerman**

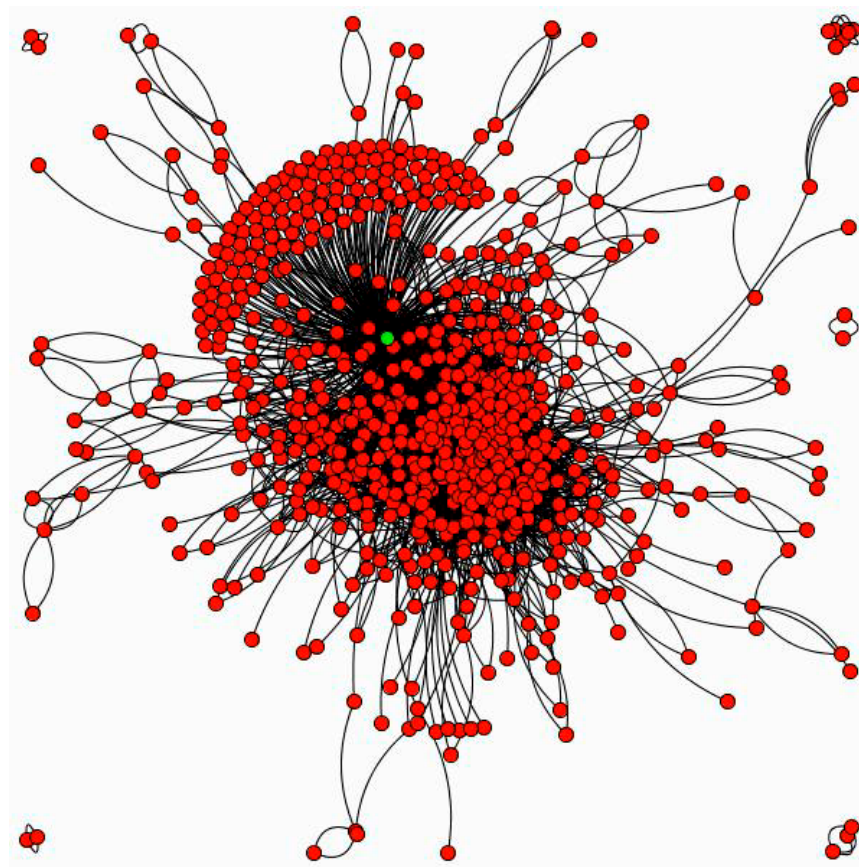
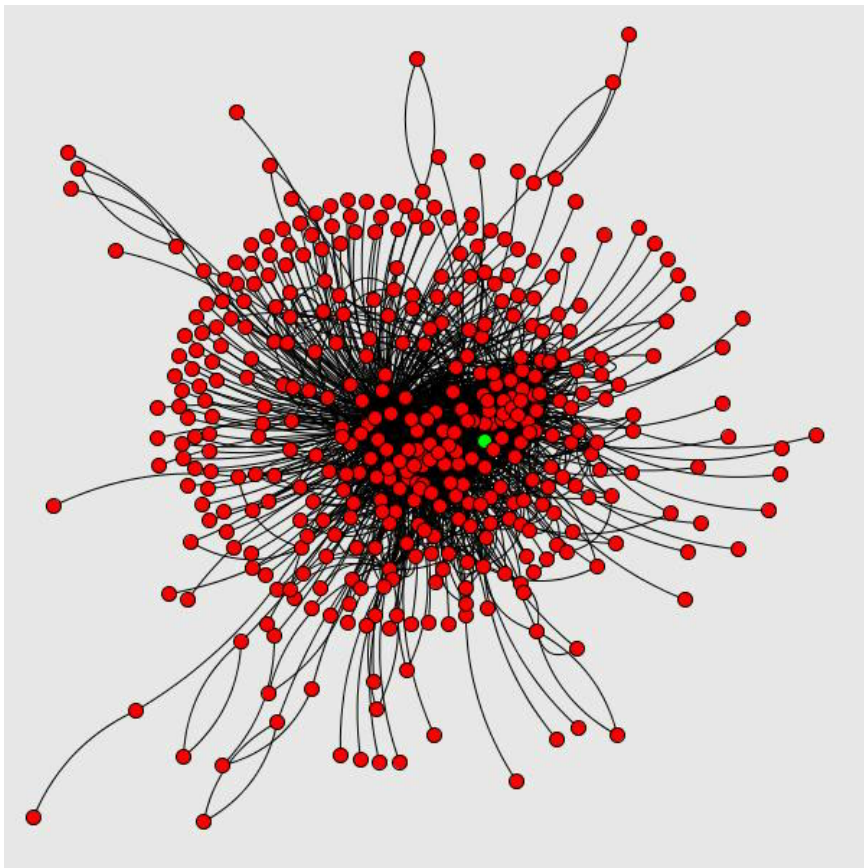
**USC Information Sciences Institute**

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

ACM Hypertext Conference, Prague, Czech Republic, July 2017

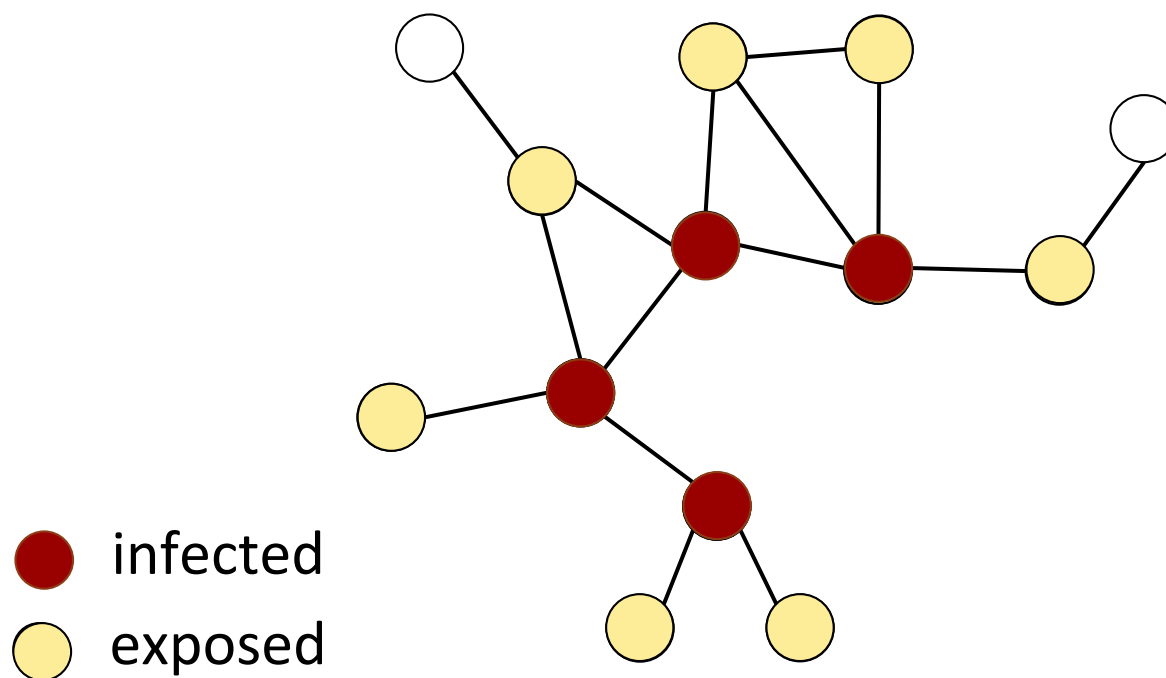


# The spread of information in social networks



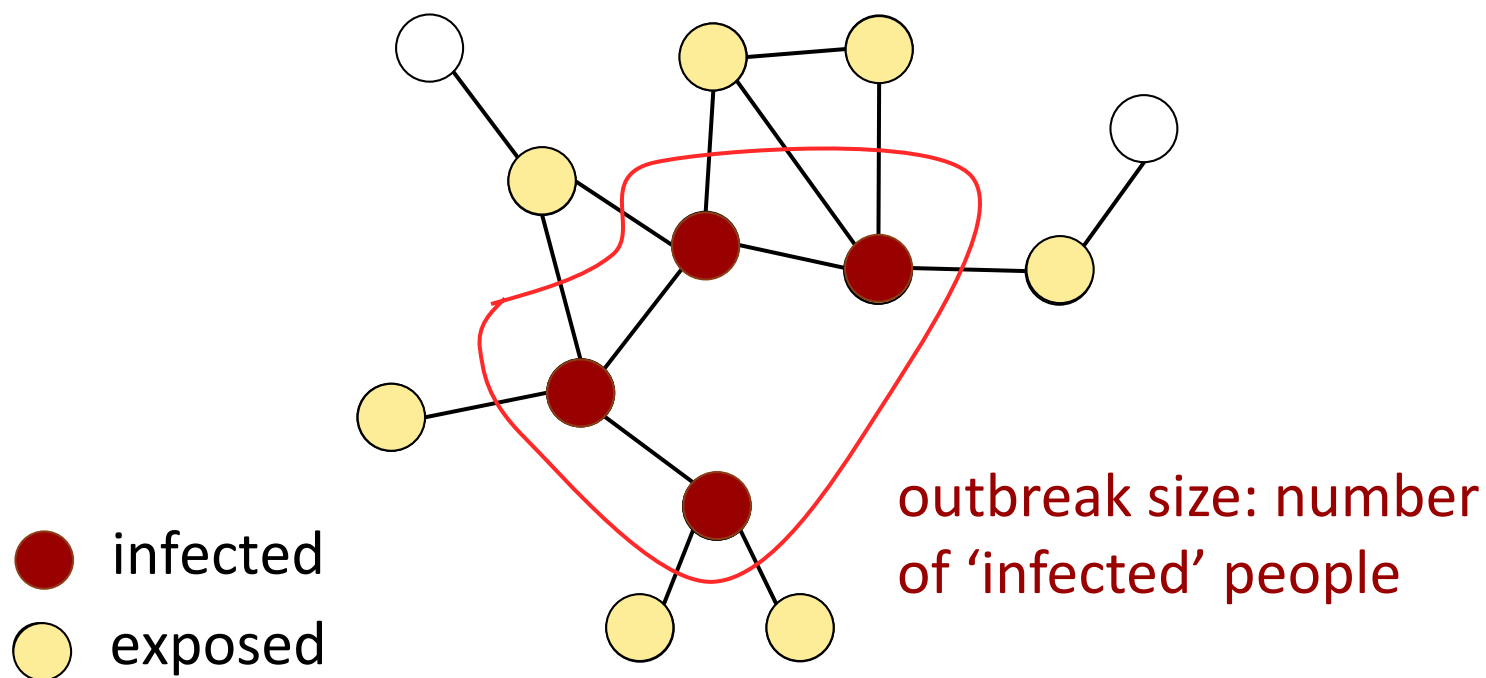
# Information spread as social contagion

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



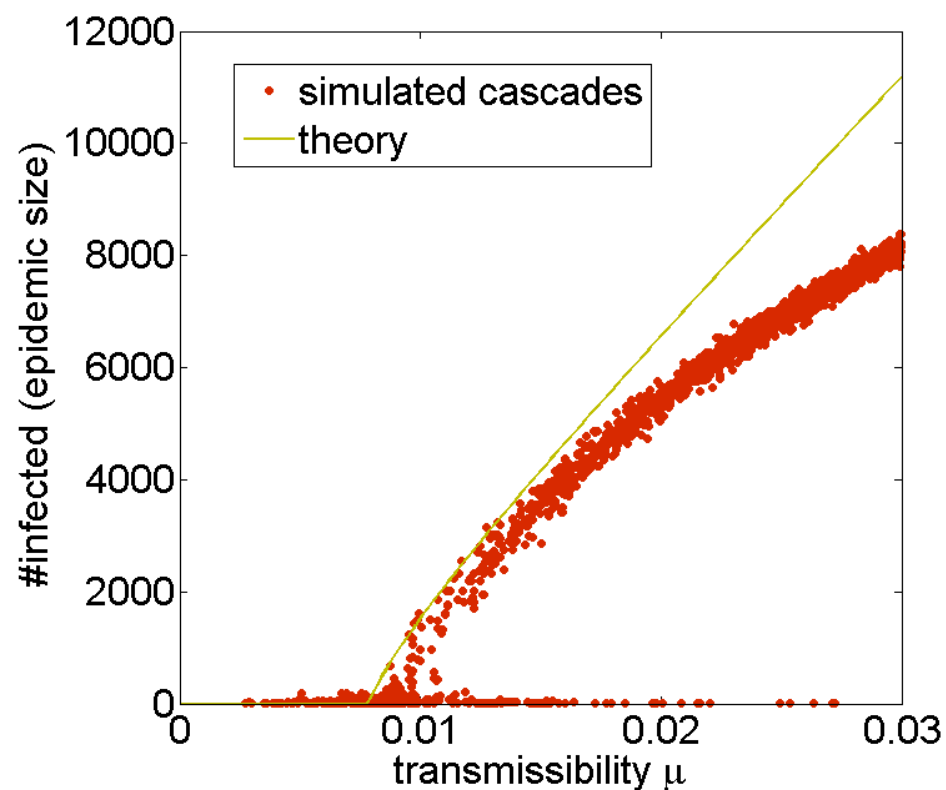
# Information spread as social contagion

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



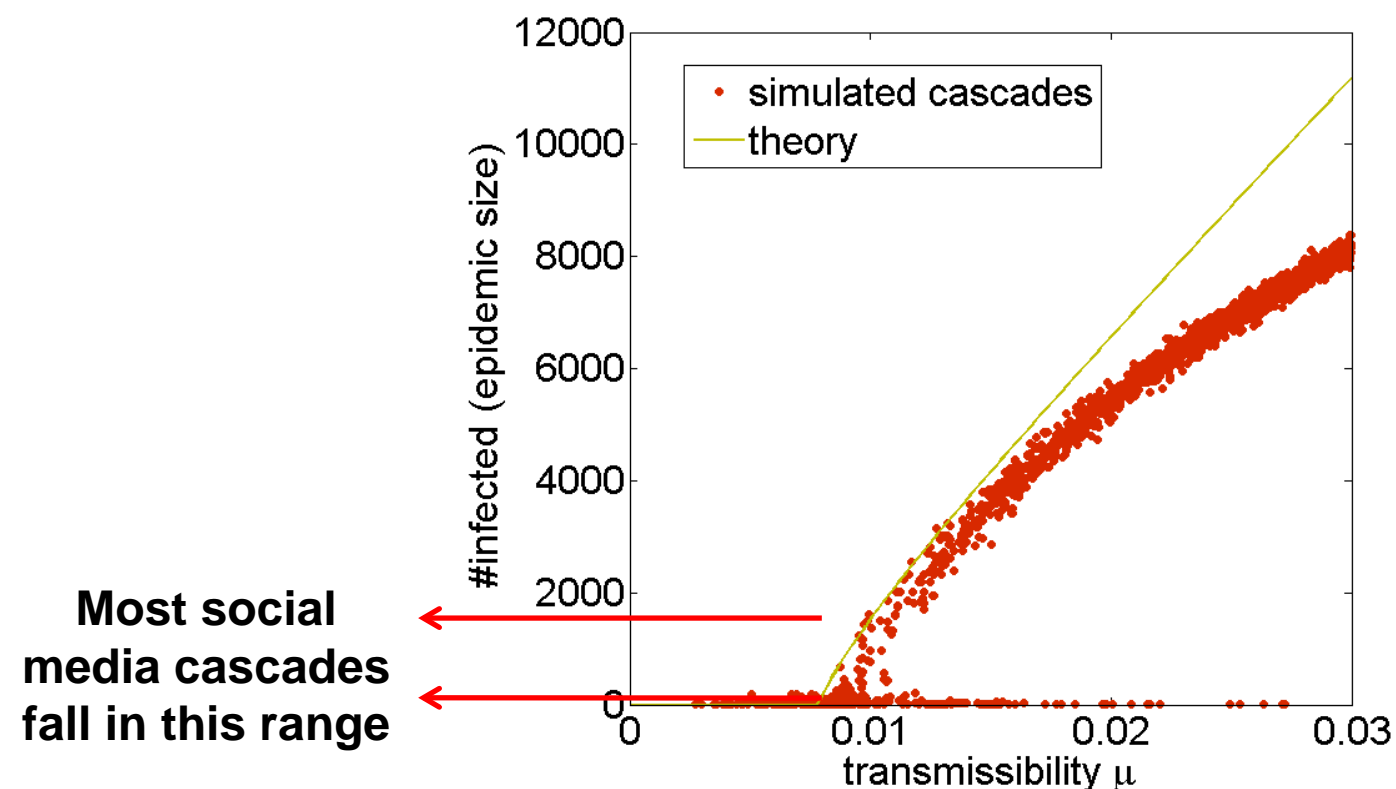
# How large are outbreaks?

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



# How large are outbreaks?

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

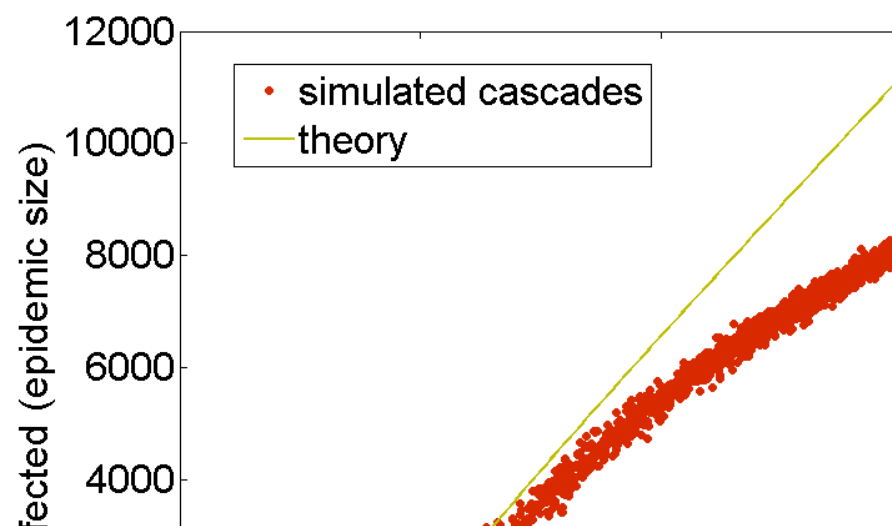


[Goel, Watts & Goldsteing (2012) "The Structure of Online Diffusion Networks" in *EC*.]

[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.**

# 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”)

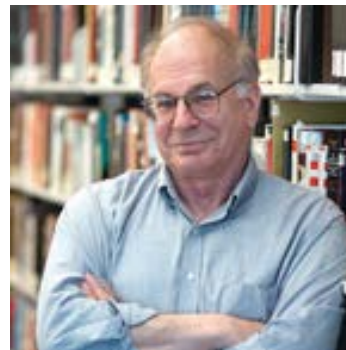


Herbert A. Simon

## 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]



Daniel Kahneman



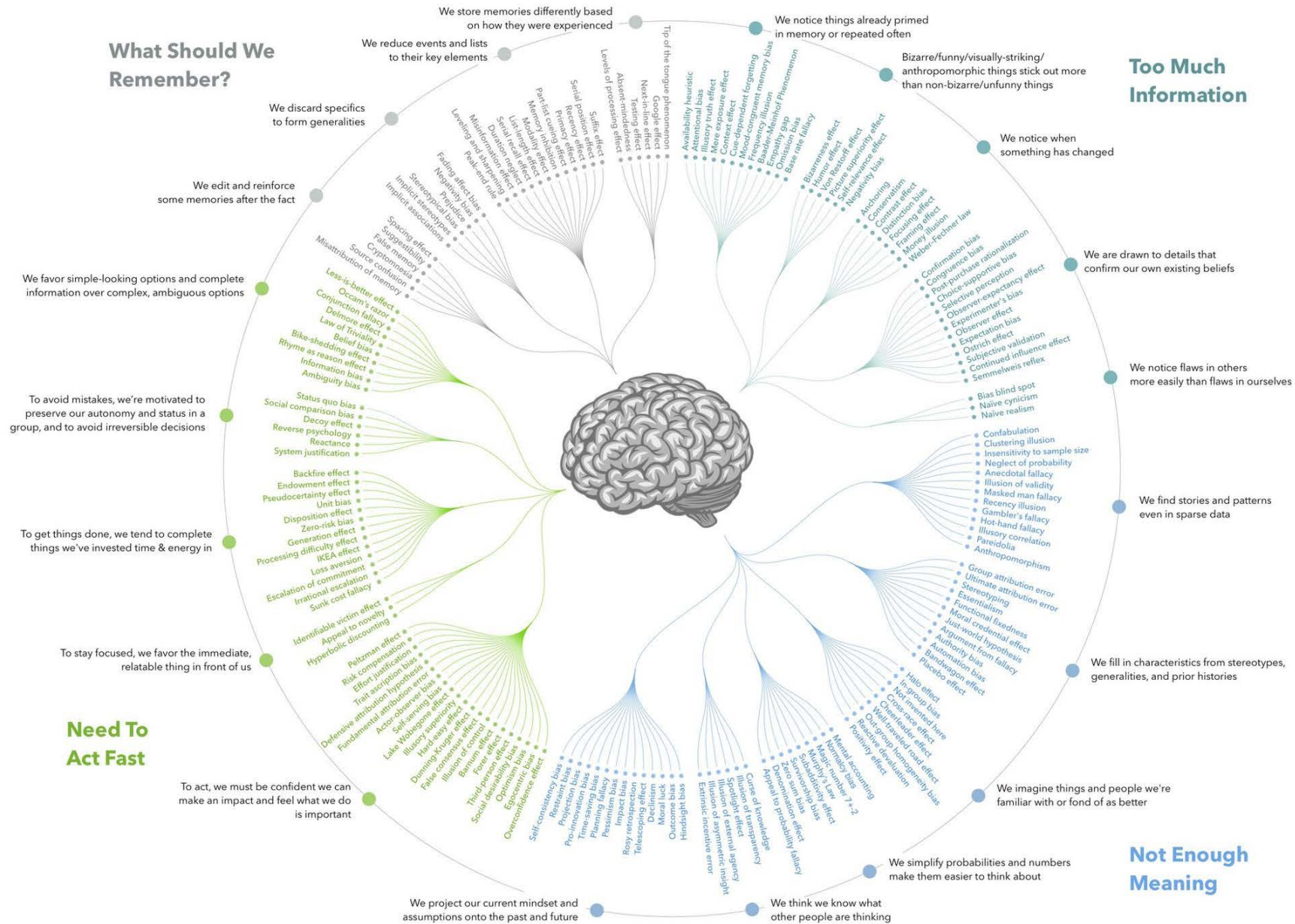
Amos Tversky

## 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

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

## COGNITIVE BIAS CODEX, 2016



# 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



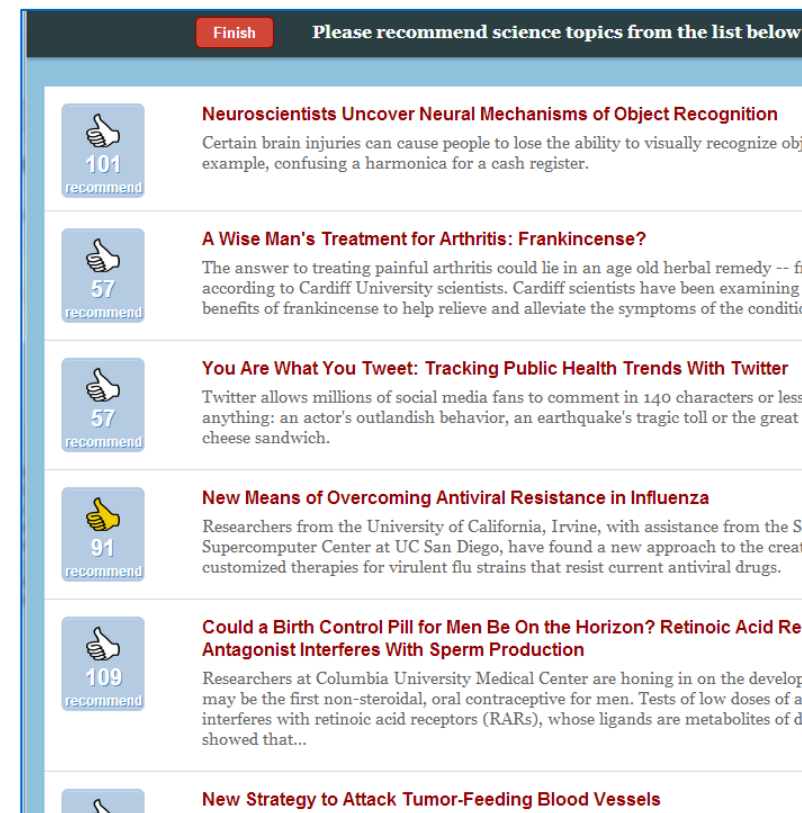
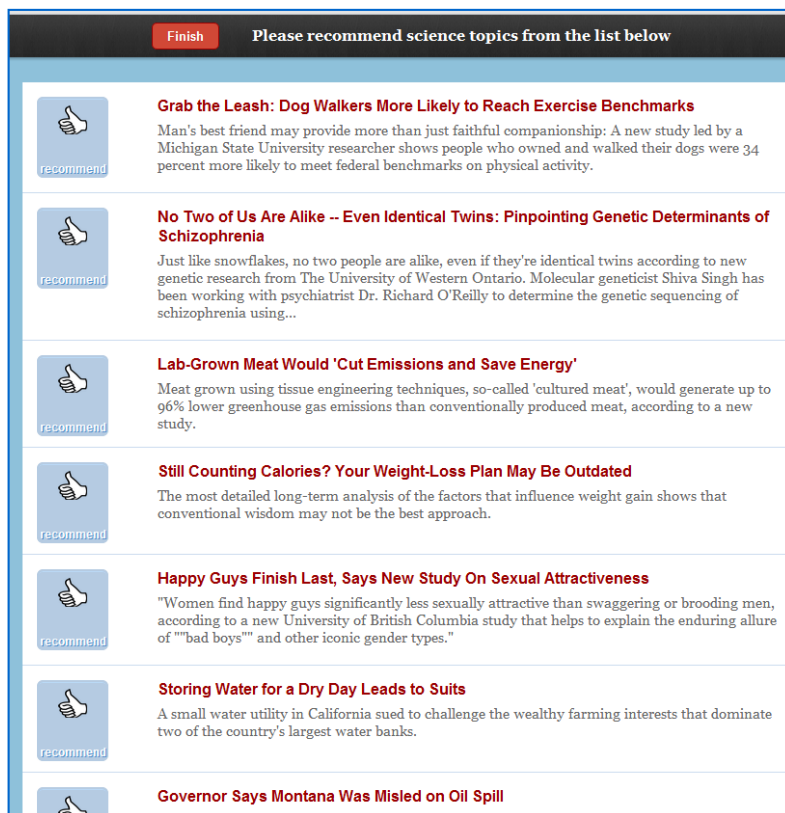
**Other biases:**

- Availability bias
- Primacy effect
- Confirmation bias



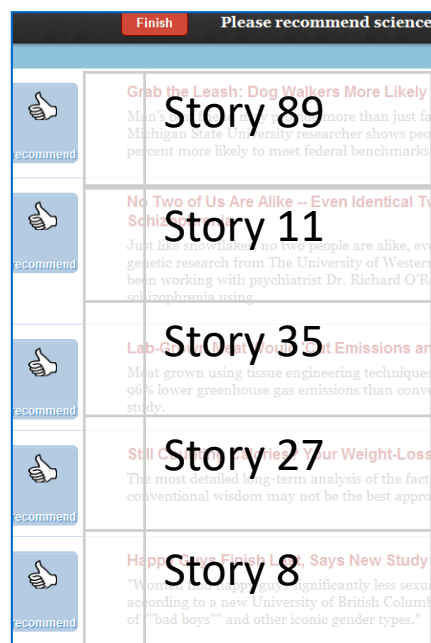
# Measuring cognitive biases

- 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.



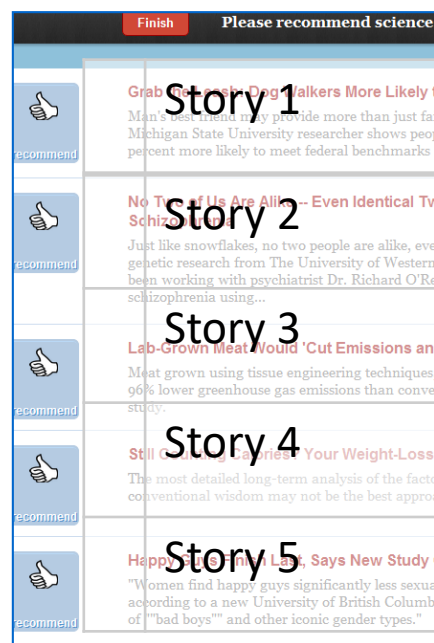
# Experimental design

- 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]

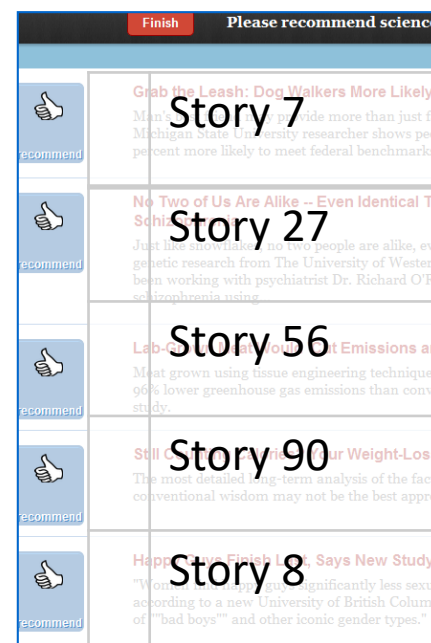


[random order]

**control**

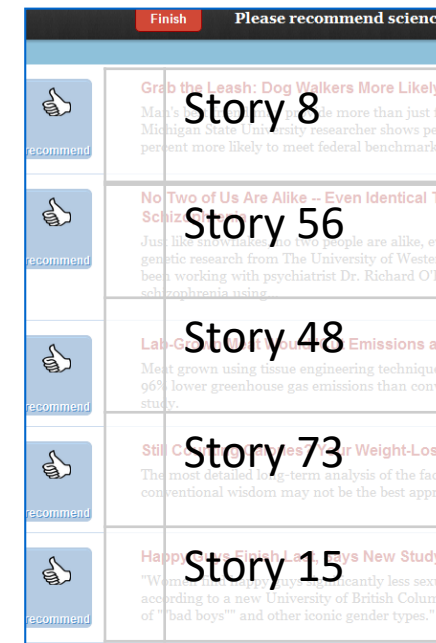


[fixed order]



[by popularity]

# recommendations

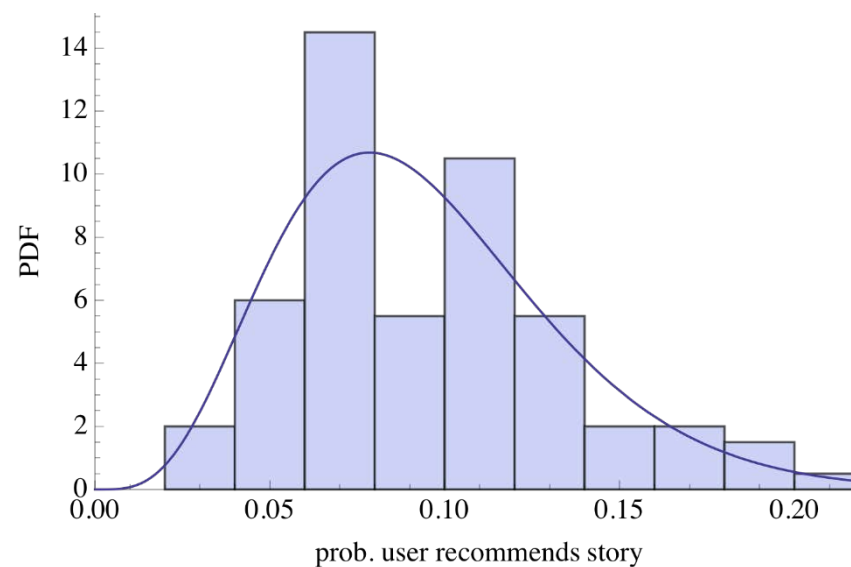


[by recency]

of recommen.

# “Quality”

**Fraction of recommendations  
in the random ordering**

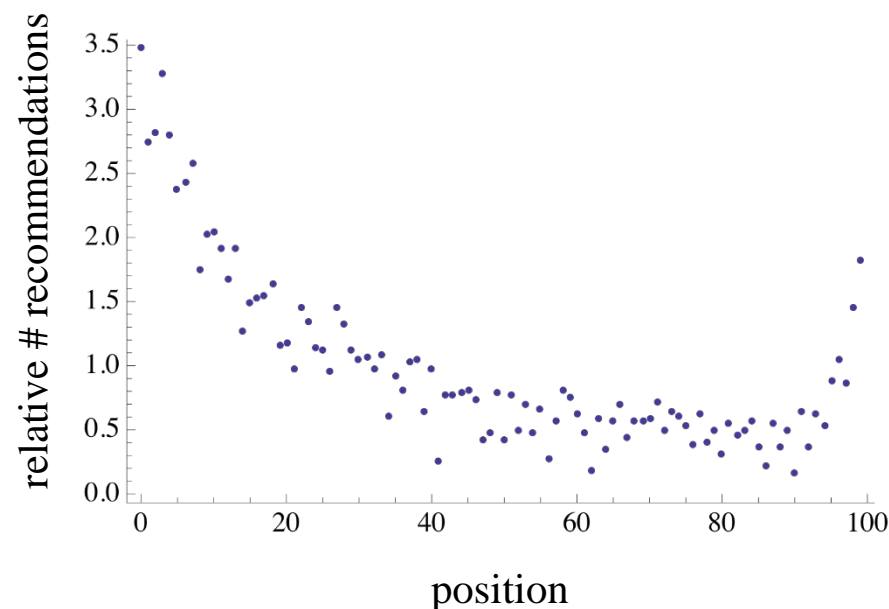


Large variation in how appealing stories are to users

[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**



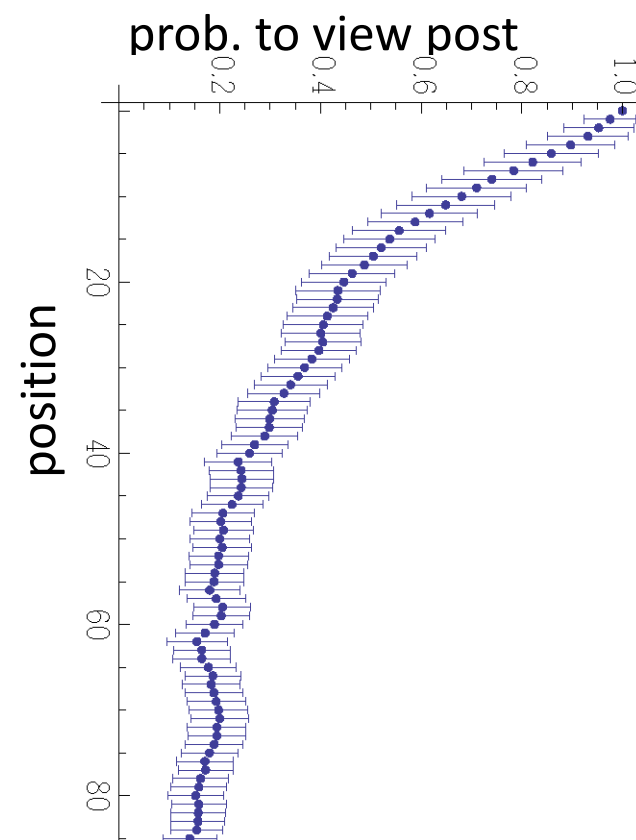
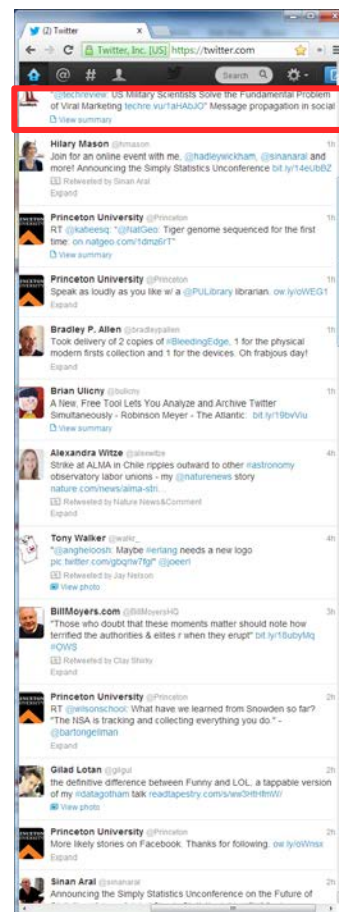
Items in top positions receive 4x as much attention as items in lower positions

[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

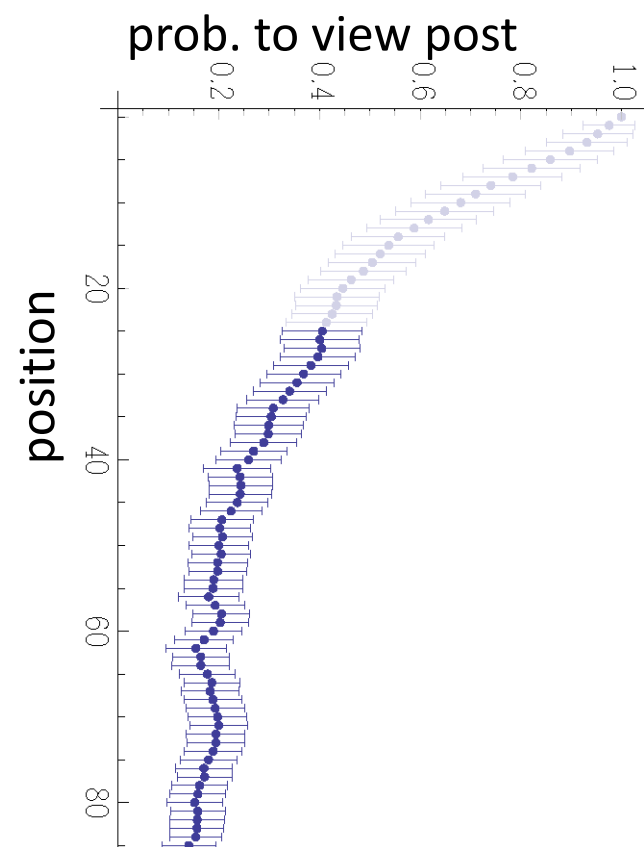
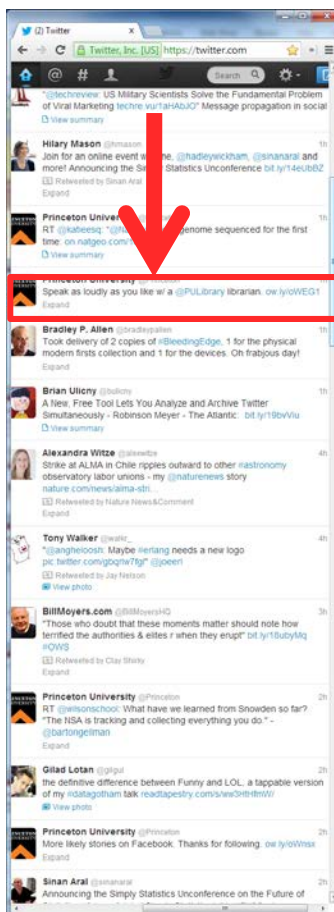




# Position bias in social media

... later: newer posts  
from friends appear  
at the top

post is less likely to  
be seen



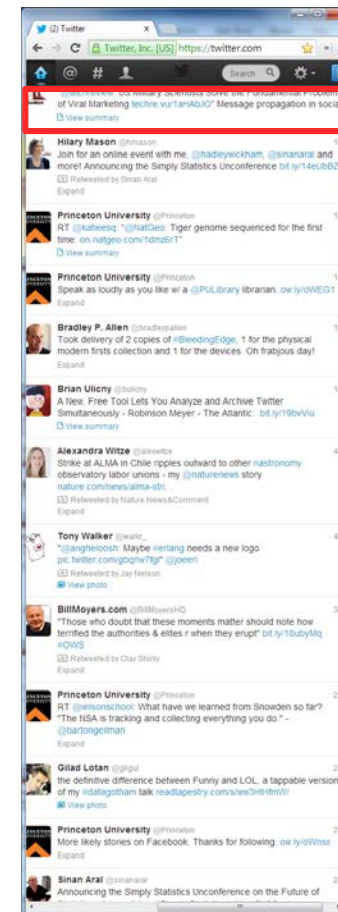
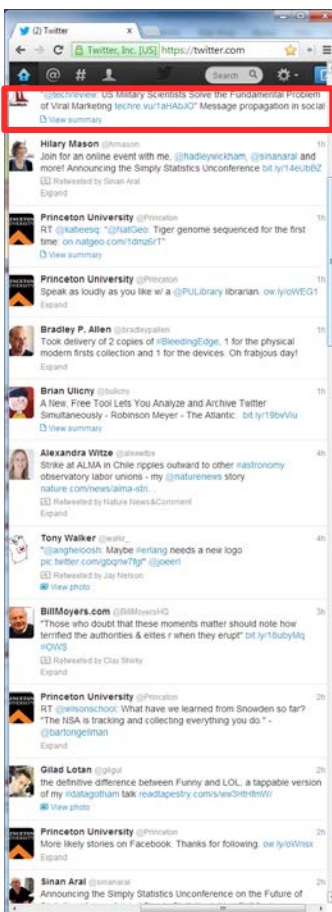
# Users divide attention over all incoming posts

few friends

many friends

new post at top of user's screen

post near the top is most likely to be seen



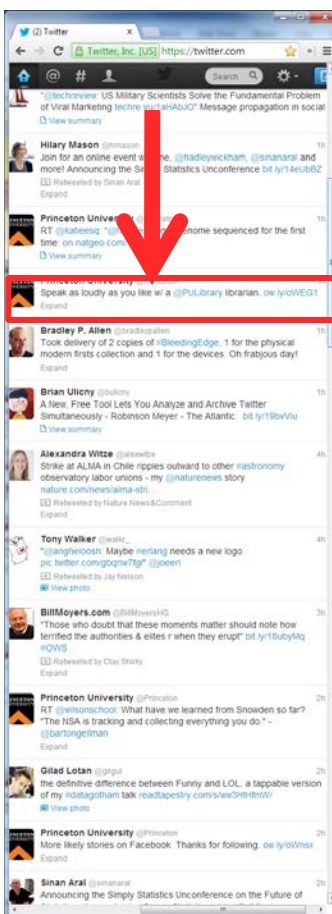
# Users divide attention over all incoming posts

few friends

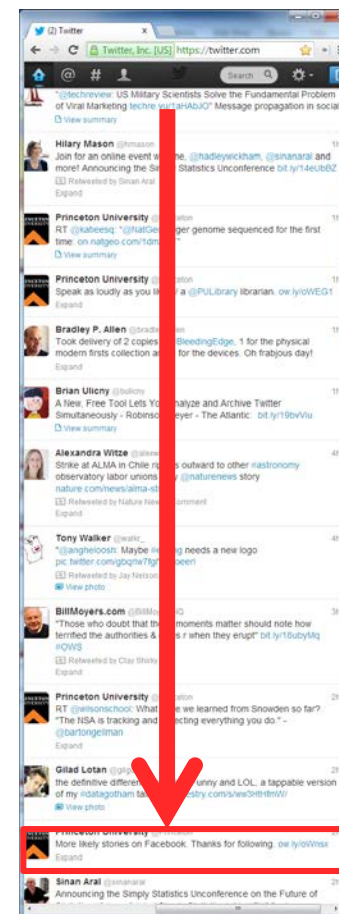
many friends

... later: newer posts  
from friends appear  
at the top

post is less likely to  
be seen

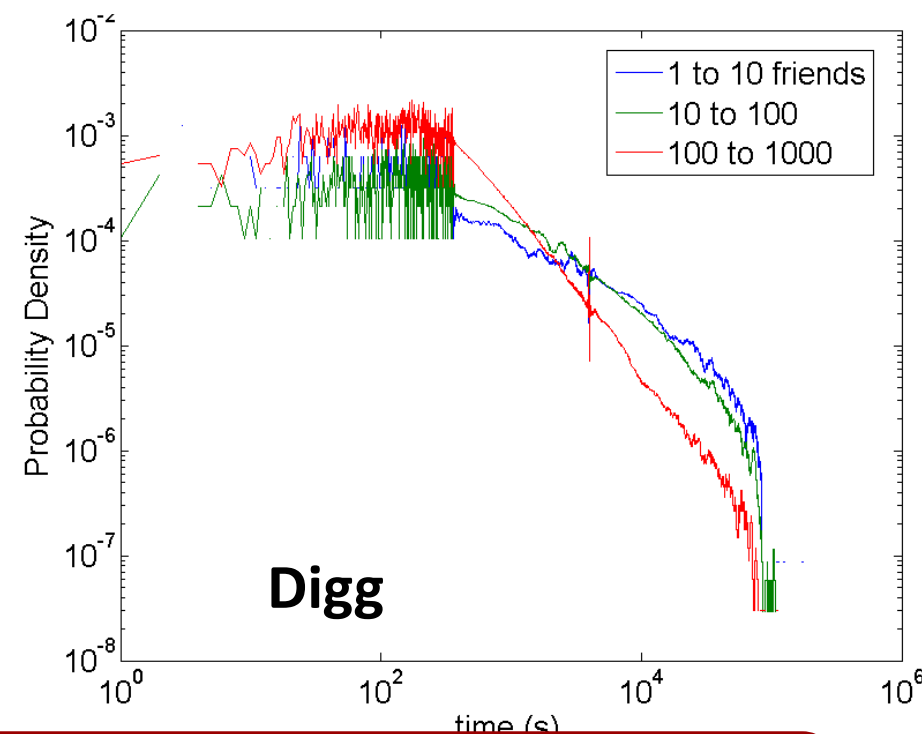
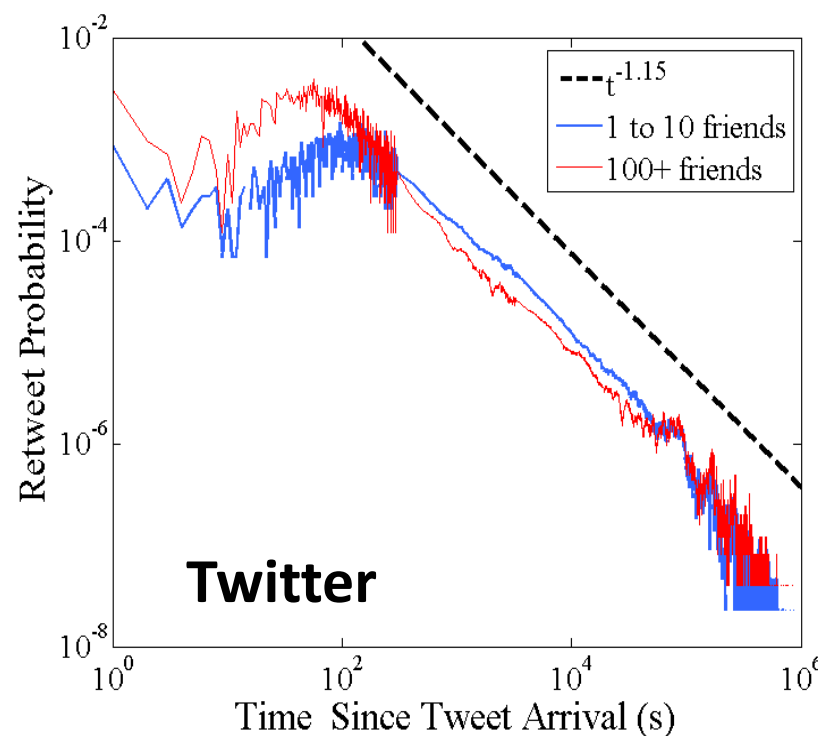


same age post is  
even less likely  
to be seen by a  
well-connected  
user



# Position bias in social media: Empirical evidence

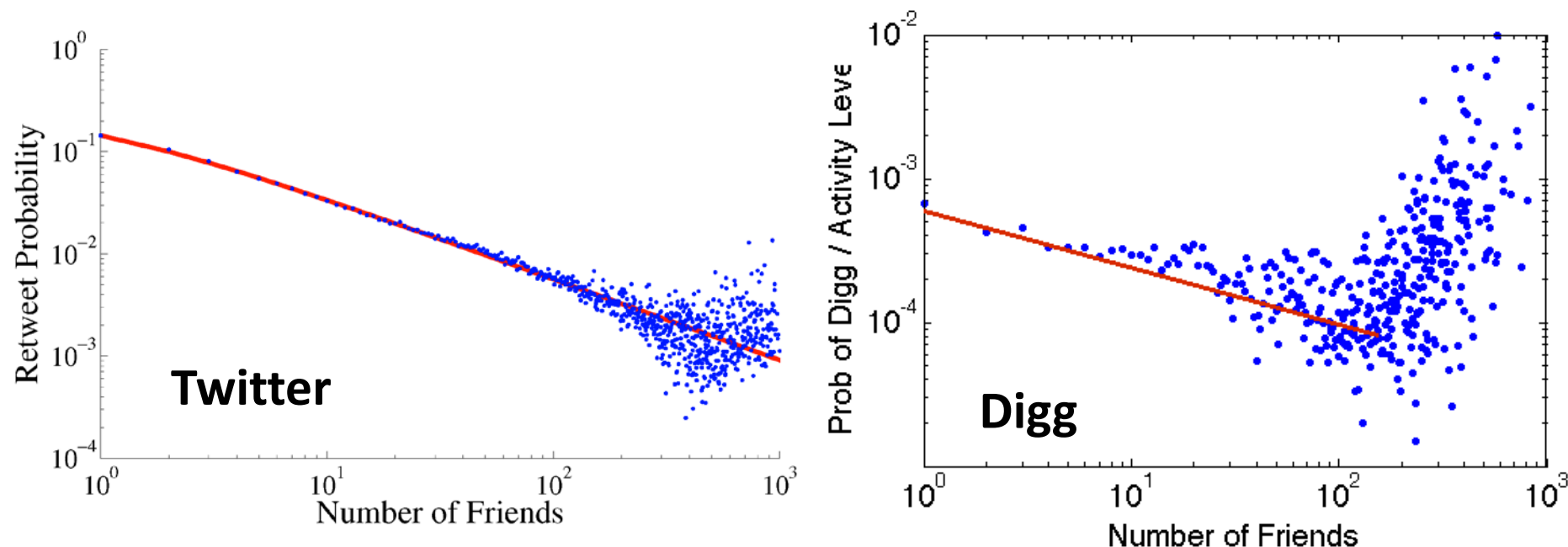
Retweet probability decreases with time since post's arrival



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

# Users divide attention over all incoming posts

Retweet probability decreases with connectivity

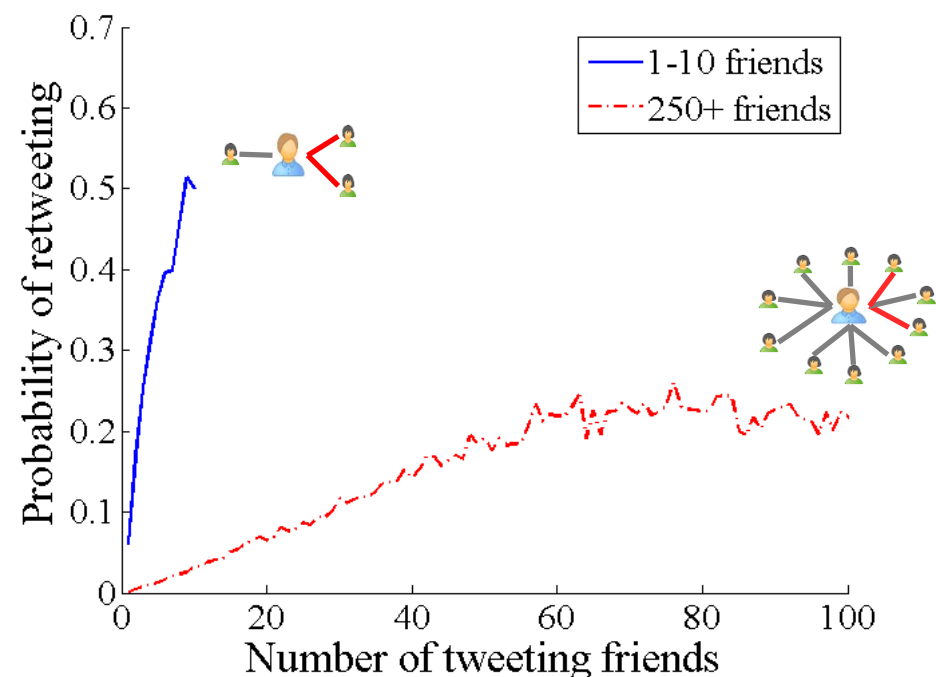


**Observation: Well-connected people (i.e., those following many others) are less likely to retweet a post.**

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

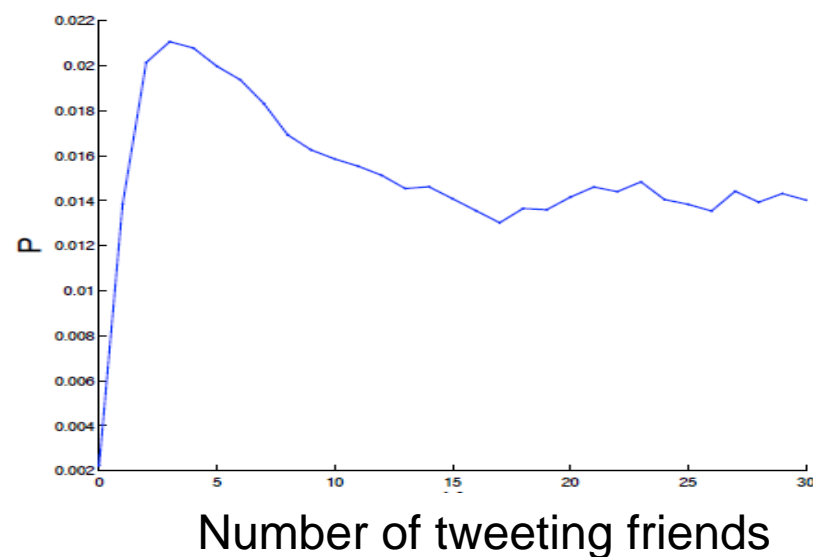
# Exposure response

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

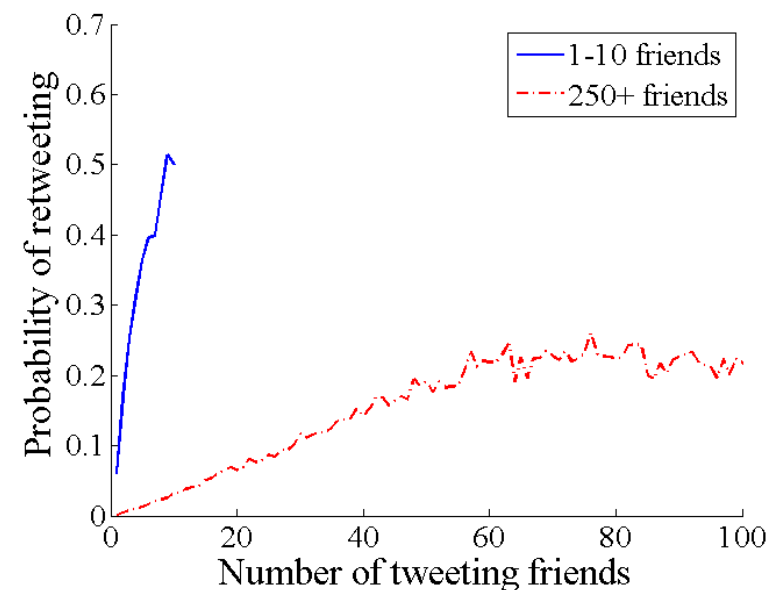


# Complex vs simple contagion

**Exposure response in social media:  
Additional exposures by friends  
appear to suppress response  
(probability to use a hashtag)<sup>1</sup>**



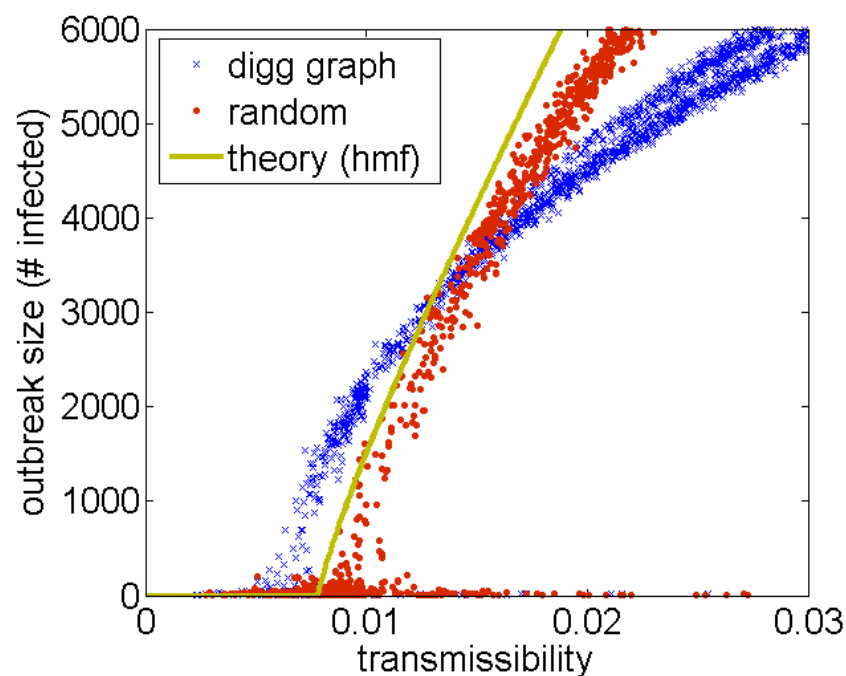
**Exposure response in social media:  
When disaggregated by cognitive  
load, additional exposures amplify  
response (probability to retweet)**



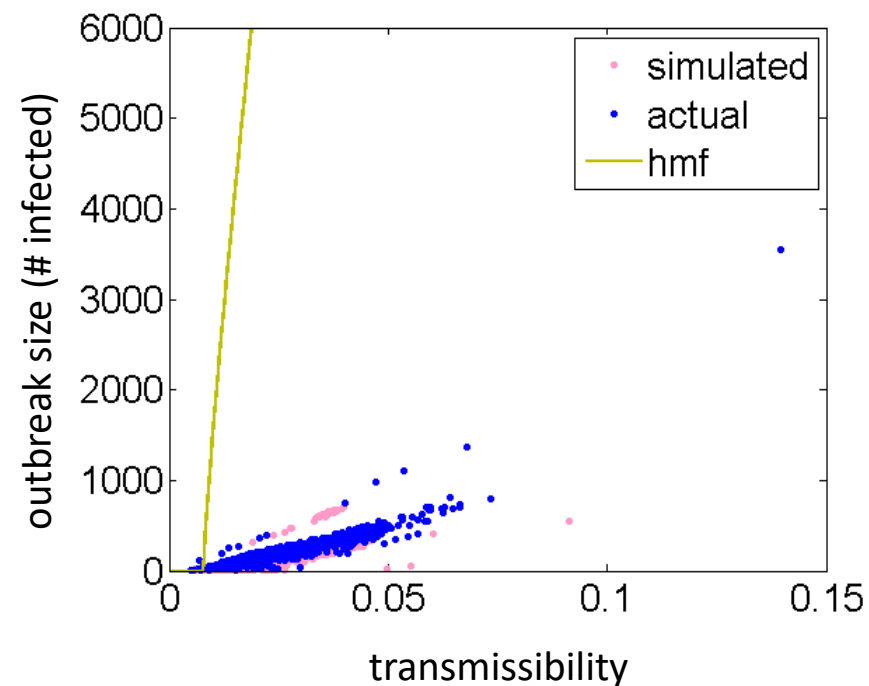
1. Romero, Meeder & Kleinberg (2011) “Differences in the Mechanics of Information Diffusion Across Topics” in *WWW*.
2. [Hodas & Lerman (2012) “How Limited Visibility and Divided Attention Constrain Social Contagion” in *SocialCom*.

# Weak response of hubs suppresses outbreaks

## Uniform susceptibility



## Decreased susceptibility of hubs

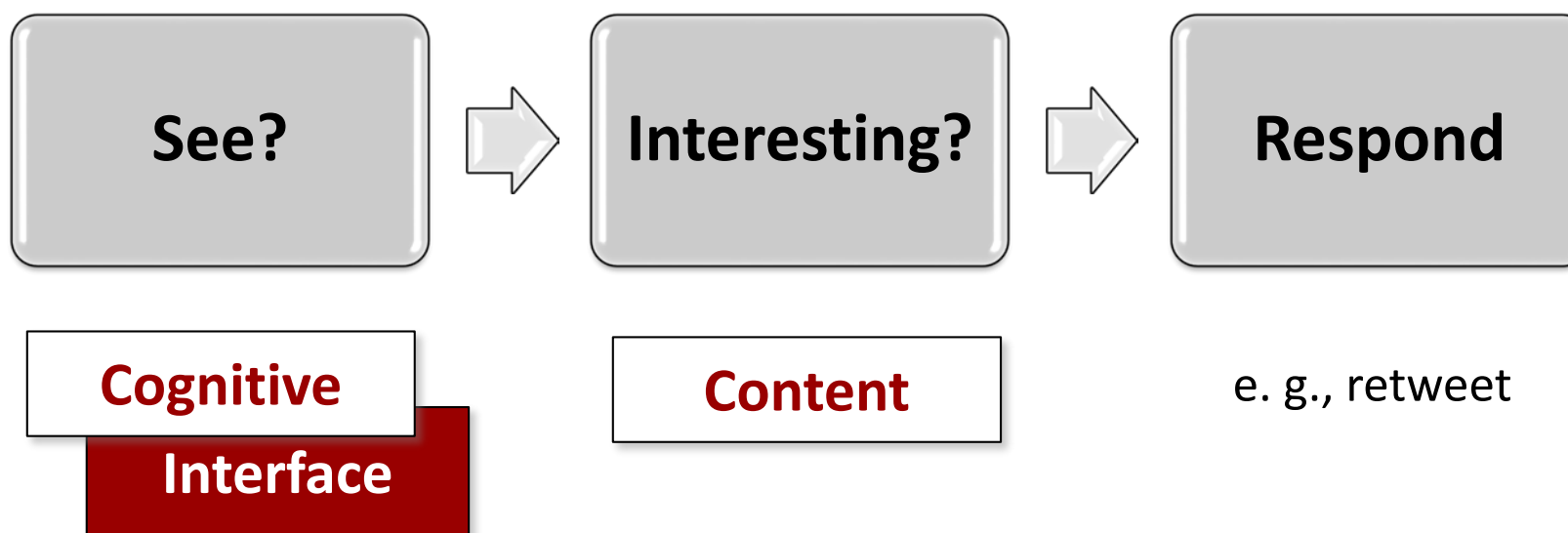


[Ver Steeg, Ghosh & Lerman (2011) "What stops social epidemics?" in *ICWSM*]



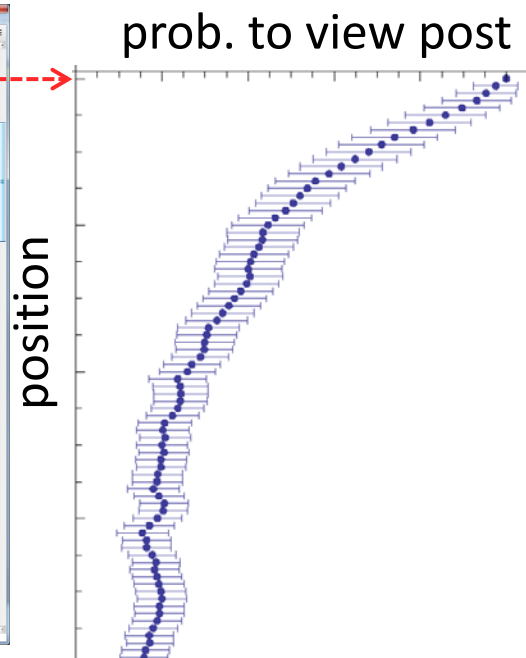
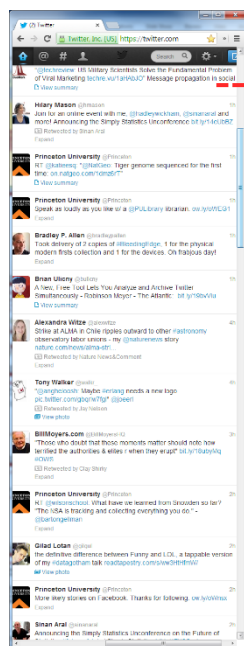
# Modeling social contagion

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

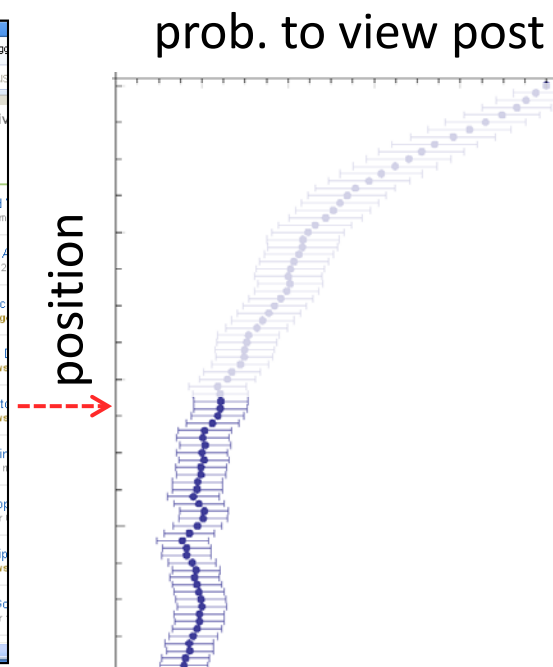


# How do users respond to multiple exposures?

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



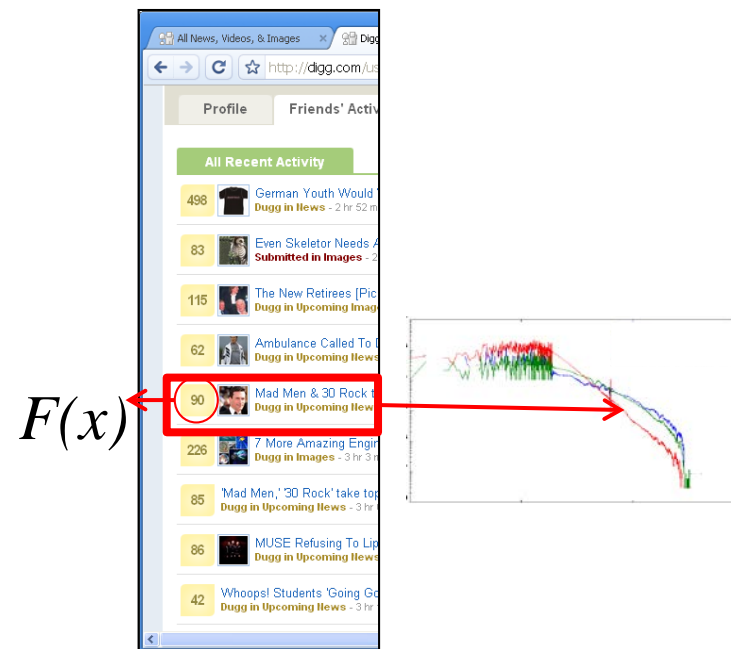
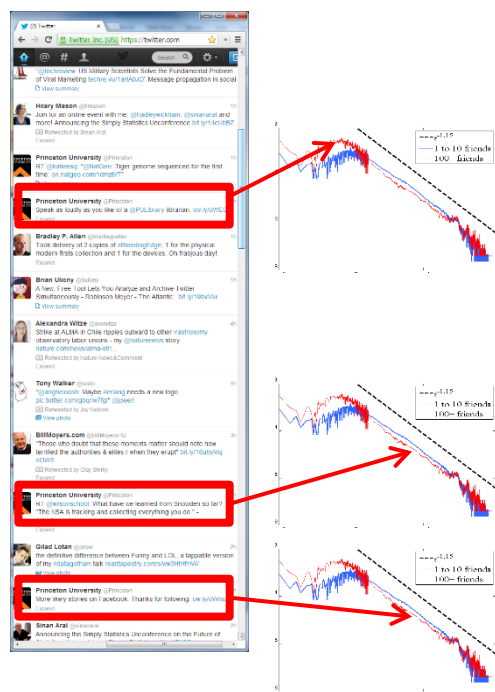
→ web site's user interface affects salience of information, but social signals matter too

# 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  $F(x)$

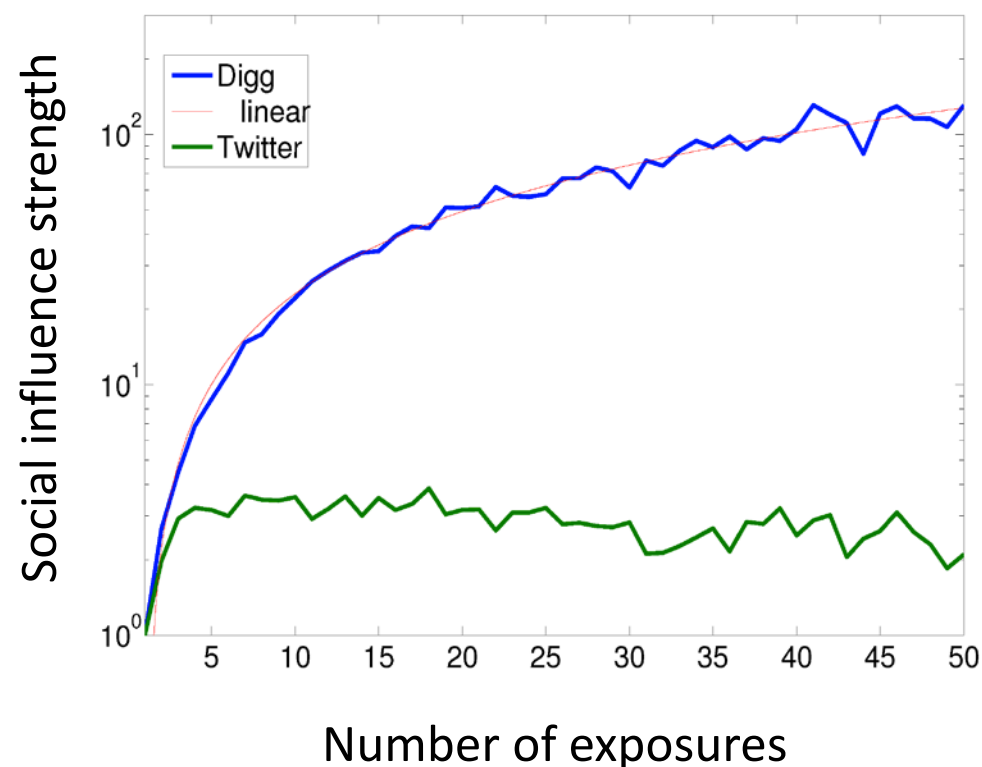
$$P_{Twitter}(t; x, n_f) = F(x) \left( 1 - \prod_{n=1}^x 1 - \mathcal{T}(t - t_n, n_f) \right)$$

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



# Social influence amplifies response

## Inferred social influence strength



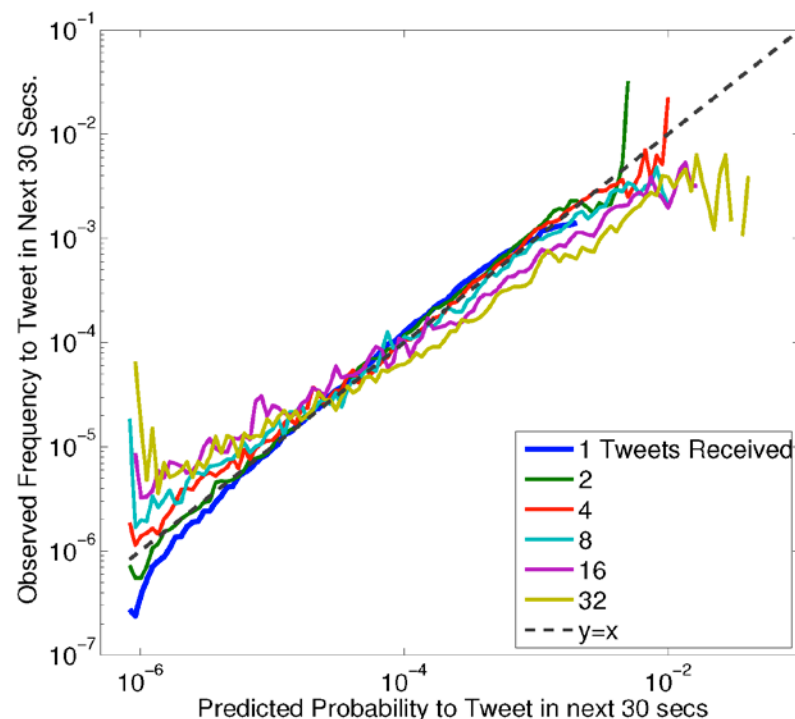
**Digg shows number of infected friends**

**Twitter does not, but users may remember earlier exposures**

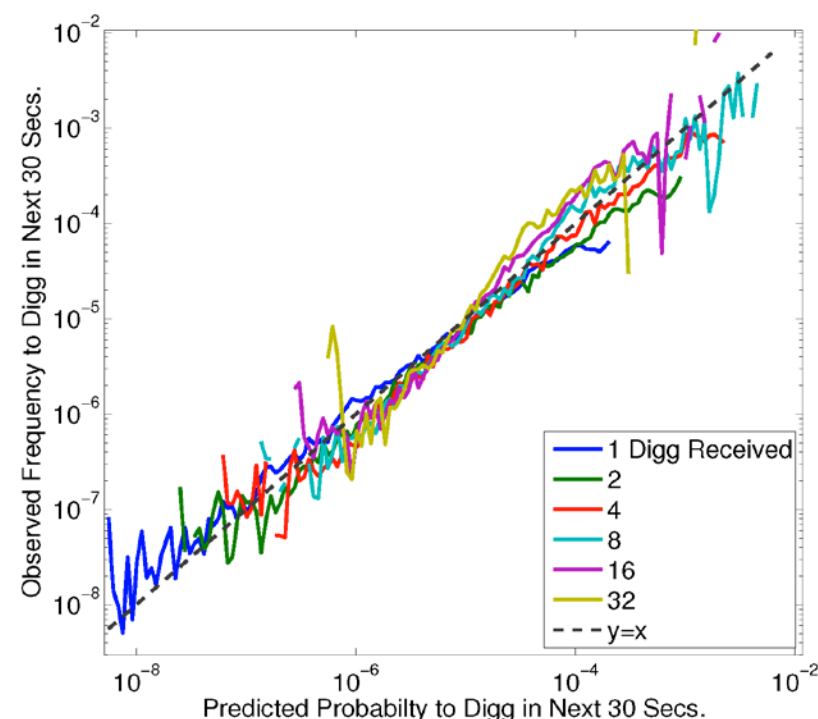
# Predict user response to multiple exposures

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  $F(x)$

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



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



**→ 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]

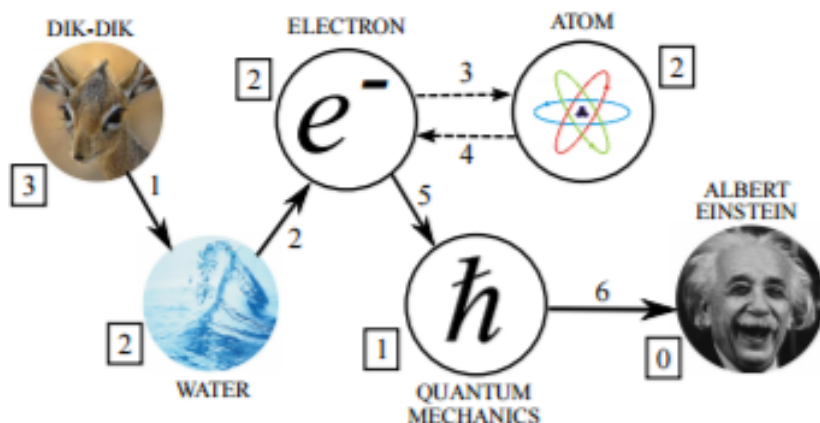
*"If you do not know the target, ... mail this letter... to a personal acquaintance who is more likely than you to know the target."*



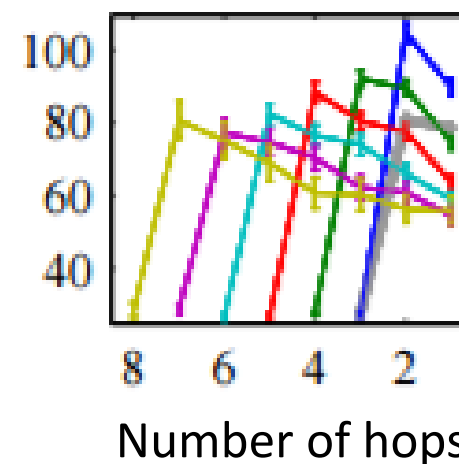
- **Social networks are searchable!**
  - Pairs of people are connected by short paths
  - People are remarkably good at finding short paths.

# 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





# Navigation and page layout

- The layout of Wikipedia facilitates navigations
- Wikipedia page layout
  - Lead
    - First paragraph discusses general concepts
  - Infobox
    - Section giving important statistics



The screenshot shows the Wikipedia article for Eyjafjallajökull. The page layout includes a sidebar on the left with navigation links, a main content area, and a search bar at the top right. The article title is "Eyjafjallajökull". Below the title, there is a "Lead" section (highlighted in blue) and an "Infobox" section (highlighted in red). The "Lead" section contains the first paragraph of the article, which discusses the glacier and its location. The "Infobox" section contains a table with key information about the glacier, including its location, elevation, and pronunciation. The "Contents" table of contents is also visible, listing sections such as Geography, Etymology, and Geology.

**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools  
What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Page information  
Wikidata item  
Cite this page

Print/export  
Create a book  
Download as PDF  
Printable version

Languages  
Afrikaans  
العربية  
Aragonés  
Asturianu  
Azərbaycanca  
Бân-lâm-gú  
Башҡортса  
Беларуская  
Беларуская

Article Talk Read Edit View history Search

## Eyjafjallajökull

From Wikipedia, the free encyclopedia

Coordinates: 63°37′12″N 19°36′48″W﻿ / ﻿

**Eyjafjallajökull** (pronounced [ˈejjaˌfjaˌtlaˌjœ̝kʏtʃ] ( listen); Icelandic for "glacier of Eyjafjöll") is one of the smaller ice caps of Iceland, situated to the north of Skógar and to the west of Mýrdalsjökull. The ice cap covers the caldera of a volcano with a summit elevation of 1,651 metres (5,417 ft). The volcano has erupted relatively frequently since the last glacial period, most recently in 2010.<sup>[2][3]</sup>

**Contents**

- Geography
- Etymology
- Geology
  - 1821 to 1823 eruptions
  - 2010 eruptions
  - Relationship to Katla
- See also
- References
- External links

**Lead**

**Infobox**

**Eyjafjallajökull**  
Guðnasteinn  
Hámundur

Gígjökull, Eyjafjallajökull's largest outlet glacier covered in volcanic ash

**Elevation** 1,651 m (5,417 ft)  
**Pronunciation** Icelandic pronunciation: [ˈejjaˌfjaˌtlaˌjœ̝kʏtʃ]

**Location**

**Location** Iceland  
**Coordinates** 63°37′12″N 19°36′48″W﻿ / ﻿

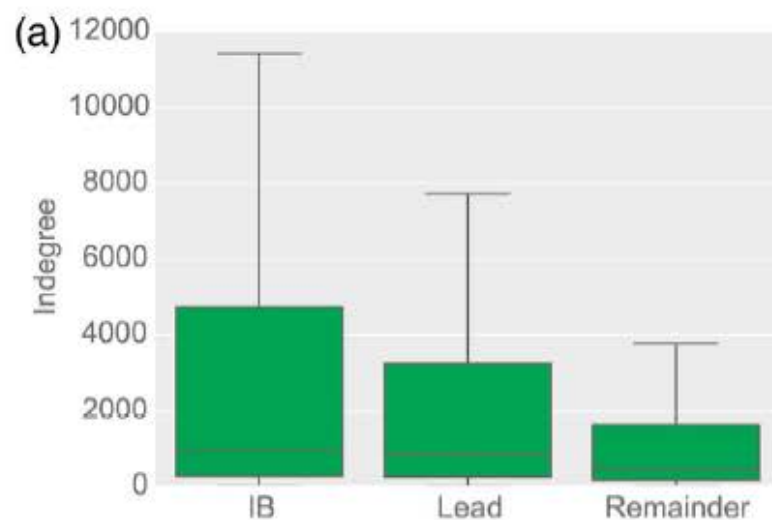
**Geology**  
**Type** Stratovolcano  
**Volcanic arc/belt** East Volcanic Zone  
**Last eruption** March to June 2010

**Geography**

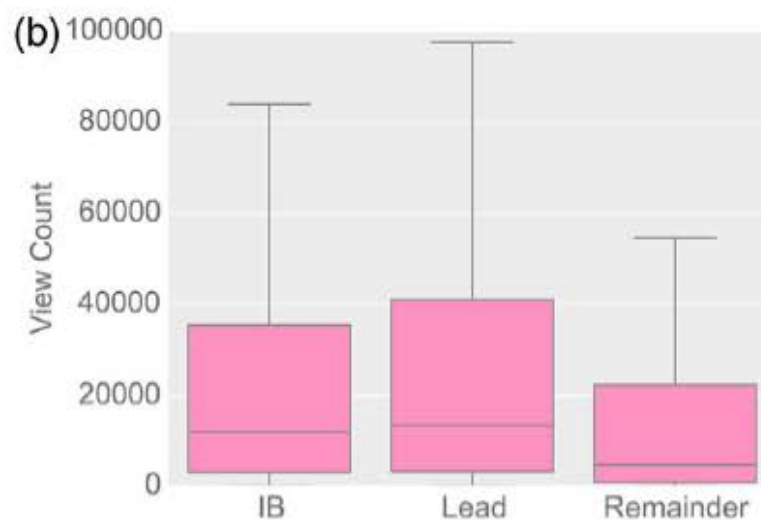
Eyjafjallajökull consists of a volcano completely covered by an ice cap. The ice cap covers an area of about 100 square kilometres (39 sq mi), feeding many outlet glaciers. The main outlet glaciers are to the north; Gígjökull, flowing into Lónið, and Steinholtjökull, flowing into Steinholtslón.<sup>[4]</sup> In 1967 there was a massive landslide on the Steinholtjökull glacial tongue. On 16 January, 1967 at 13:47:55 (or 1:47:55 PM) there was an explosion on the glacier. It can be timed because the seismometers in Gígjökull monitored the movement. When

# Navigation and page layout

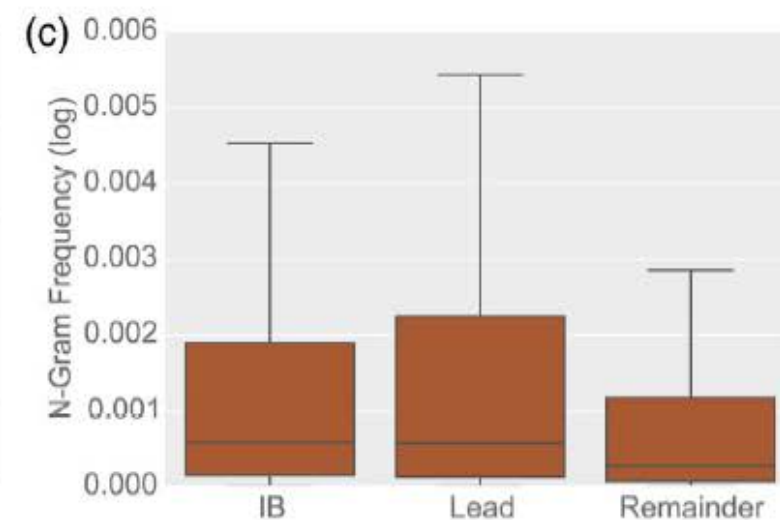
- People pay more attention to information in the lead and infobox sections (more views)
- Hyperlinks from these sections lead to hubs, i.e., pages
  - with higher degree (more links)
  - dealing with more general concepts (higher n-gram frequency)



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




[All Sites](#)
[Top Users](#)
[Newsletters](#)


**Stack Exchange Q&A communities are different. Here's how:**



**Expert communities.**  
Each of our 161 communities is built by people passionate about a focused topic.



**The right answer. Right on top.**  
Experts like you can vote on posts, so the most helpful answers are easy to find.



**Share knowledge. Earn trust.**  
Earn reputation and additional privileges for posts others find helpful.

# Anatomy of Stack Exchange

## Question

Cognitive load  
Number of  
answers to the  
question →

## Answers

▲ 52 ▼  
★ 7

What's the correct way to write a `for-in` loop in JavaScript? The browser doesn't issue a complaint about either of the two approaches I show here. First, there is this approach where the iteration variable `x` is explicitly declared:

```
for (var x in set) {
  ...
}
```

And alternatively this approach which reads more naturally but doesn't seem correct to me:

```
for (x in set) {
  ...
}
```

javascript syntax for-in-loop

share improve this question

edited 5 mins ago  
DavidRR  
3,151 ● 3 ● 15 ● 33

asked Apr 19 '11 at 13:28  
futlib  
2,016 ● 4 ● 24 ● 43

add a comment

**8 Answers** active oldest votes

▲ 48 ▼  
✓

Use `var`, it reduces the scope of the variable otherwise the variable looks up to the nearest closure searching for a `var` statement. If it cannot find a `var` then it is global (if you are in a strict mode, using `strict`, global variables throw an error). This can lead to problems like the following.

```
function f (){
  for (i=0; i<5; i++);
}
var i = 2;
f ();
alert (i); //i == 5. i should be 2
```

If you write `var i` in the for loop the alert shows `2`.

[JavaScript Scoping and Hoisting](#)

share improve this answer

edited Jul 25 '13 at 7:50

answered Apr 19 '11 at 13:36  
Gabriel Uamas  
7,063 ● 7 ● 45 ● 83

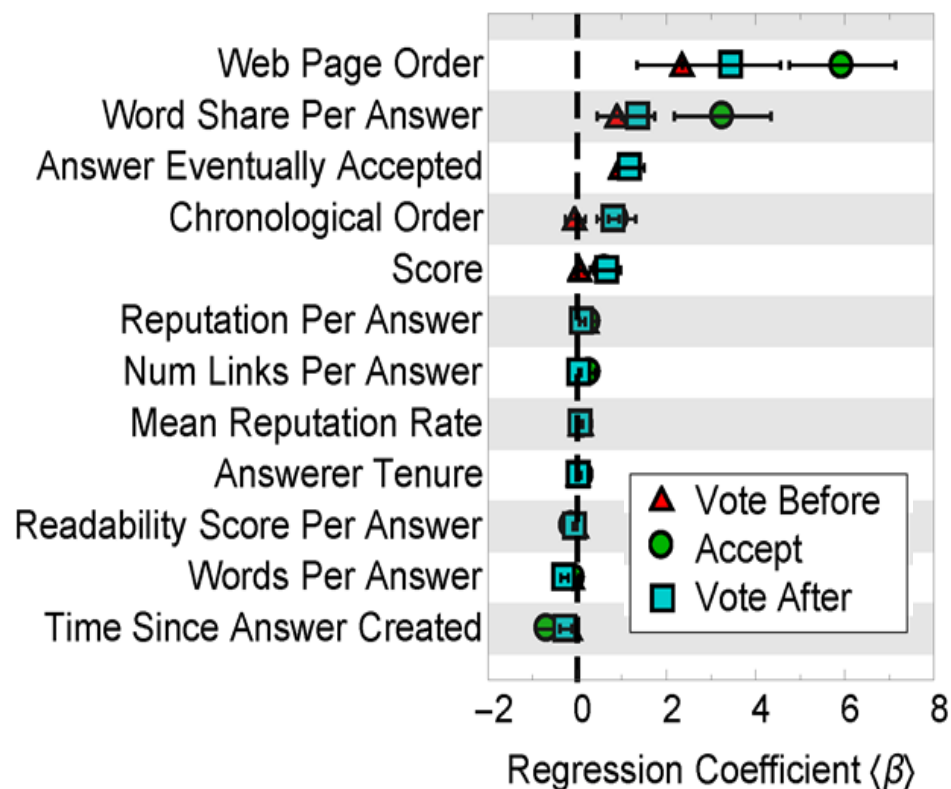
## Answer features

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

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

- answerer reputation
- tenure

# Regression coefficients highest for heuristics

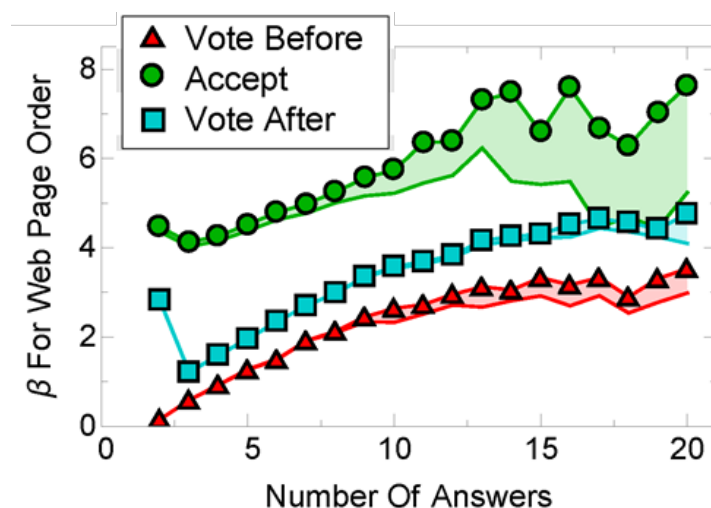


→ Rather than evaluate all answers, people use simple heuristics to choose answers to vote for or accept. Largest coefficients are:

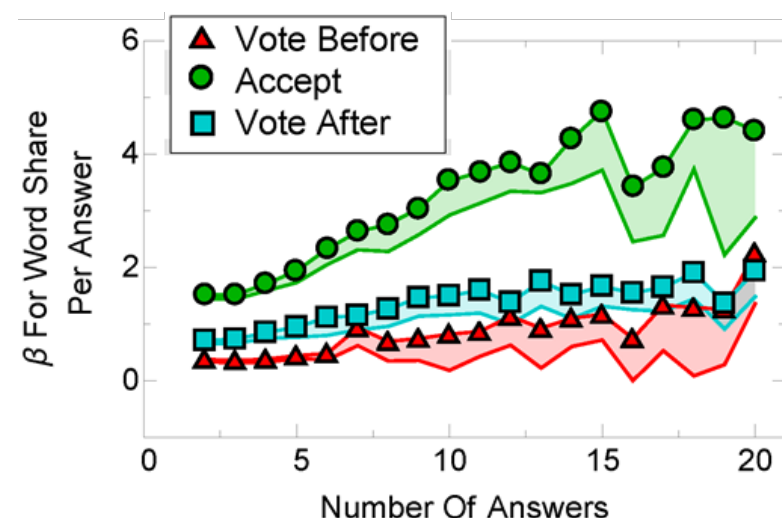
- Web page order → answer's rank (*cf* position bias)
- Word share → fraction of the screen it occupies (*cf* availability bias)
- Answer acceptance → social proof (*cf* social influence bias)

# Cognitive load increases reliance on cognitive heuristics

Regression coefficient for web page order vs cognitive load\*



Regression coefficient for word share vs cognitive load\*



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

# 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

