

# Announcements

- Project 2nd milestones are graded.
  - Please let me know if your grade is different from your partner's so I can correct.
- No particular template about the project presentations
  - Expecting to see
    - Problem you are solving or analyzing
    - Dataset (if you did not use a readily available dataset, what did you do to make it work for your case)
    - Some of the analysis results
  - You can skip background/related works in the presentation. Include it only if it is germane to the discussion (e.g. you are modifying an existing algorithm or comparing against it)



# Wisdom In The Social Crowd: An Analysis Of Quora

**Gang Wang**, Konark Gill, Manish Mohanlal,  
Haitao Zheng and Ben Y. Zhao

University of California at Santa Barbara  
[gangw@cs.ucsb.edu](mailto:gangw@cs.ucsb.edu)

# Asking Questions on the Internet

- Systems to answer user questions on the Internet
  - Google - general information
  - Wikipedia - factual knowledge



Q: What is the population of Rio?



6.323 million (2010)  
Rio de Janeiro, Population

- But we often have questions that require...
  - Domain-specific knowledge
  - First-hand life experiences



Q: What is the most interesting souvenir you can buy in Rio?

# Online Q&A Services Today

- Question and Answer (Q&A) sites
  - Web services where people ask and answer questions
  - A *crowd-sourced* way to search information
- 

## Google Answers decides to close up shop

Google is closing its doors on Google Answers, but what's the real reason ...

## Microsoft Shutting Down QnA, Questions & Answers Site

## U.S. traffic to Yahoo Answers dropped 23%

- As the Q&A systems grow to massive scales...
  - More difficult for users to locate useful answers or interesting questions
  - Low-value questions (spam) overwhelm the system



# Quora - Social Q&A

- “Hottest” (most successful) today
  - First social network based Q&A
  - 350% traffic growth in 2012
  - Many answers are returned as top answers to Google queries
- Quora’s advantages
  - High-quality questions and answers
  - True domain experts participation politicians, actors, startup founders, etc.

The screenshot shows a mobile application interface for Quora. At the top, there's a red header bar with the text "Back", "Question", and "Follow". Below the header, the question title is "How is being a billionaire better than being a millionaire?". Underneath the title, a sub-question "What Does It Feel Like To Be X?" is visible. The main text of the question asks about the perks/benefits of being a billionaire compared to a millionaire. Below the text, there are buttons for "1+ Comments" and "Share", along with a gear icon for settings. Further down, it shows "14 Answers · 369 Followers · 46319 Views". A user profile for David S. Rose is displayed, including his title as an entrepreneur and angel investor, and his role as a mentor. It also shows 984 votes by various users like John Jeffrey Mardlin, Paul Lehman, and Kris Haamer. A large block of text from one answer discusses the changing nature of wealth and net worth.

AT&T 11:59 AM

Back Question Follow

Billionaires Sergey Brin Social Class

What Does It Feel Like To Be X?

**How is being a billionaire better than being a millionaire?**

How are the perks/benefits different? Are there distinct "classes" at the upper end of the wealth scale?

1+ Comments Share

14 Answers · 369 Followers · 46319 Views

David S. Rose, Entrepreneur, Angel Investor, Mentor,...  
984 votes by John Jeffrey Mardlin, Paul Lehman, Kris Haamer, (more)

Being a millionaire ain't what it used to be :-(. In thinking about net worth, it's helpful to consider everything using a common denominator such as your potential annual

How does Quora’s internal structures contribute to its success?

# A Measurement Study of Quora

- Limited understanding of Quora
  - Size of site (questions, users), growth rate
  - Mechanisms for content discovery, quality control
- Questions we asked in our study
  - How does Quora grow over time?
  - What's the impact of social graph on Q&A activities?
  - How does Quora direct users to the valuable content?  
Match experts w/ questions, and seekers w/ answers



# Outline

- Introduction
- Characterizing Quora
- Analyzing Graph Structures
- Implications

# A Typical Question Page

Quora  Topics

Movie Making: Acting | Breaking into Hollywood | Casting | Hollywood | Movies  
Television [Edit](#)

**What determines whether an actor is cast in a part in Hollywood?** [Edit](#)

I see talented actors in small parts & then someone will come out of nowhere & get cast in a starmaking part. Talent doesn't seem like a very important criteria or maybe it's because acting is an art & open to subjective judgment. [Edit](#)

1 Comment · Share (1) · Options

13 Answers

Rishabh Chahal [Add Bio](#) · [Make Anonymous](#)

Votes

Ashton Kutcher, I'm an actor and producer of film tv ... [Answer](#)

666 votes by Charlie Cheever, Jack Stahl, Joshua Forman, (r)

Talent, experience, relationships, directors, awards, and financial risk reduction.

Talent is subjective. (after all I've been accused by some of having very little) but you have to have some level of showmanship. That being said some of the best actors I know have a hard time finding work.

Experience is key. Back ground in theater, improv troops, formal training is very

Write · Home · Rishabh

## Related Questions

Related Questions

**What determines whether a celebrity is on the A-List, B-List, C-List or D-List?**

**Which modern actors would be best cast in a Hollywood remake of Ayn Rand's "The Fountainhead"?**

**Breaking into Hollywood: What is the best place to find actors willing to work for free in a short "how it works" web video?**

[More Related Questions](#)



### Share Question

Twitter Facebook LinkedIn

### Question Stats

Latest activity 27 Mar

This question has 1 monitor with 777351 topic followers.

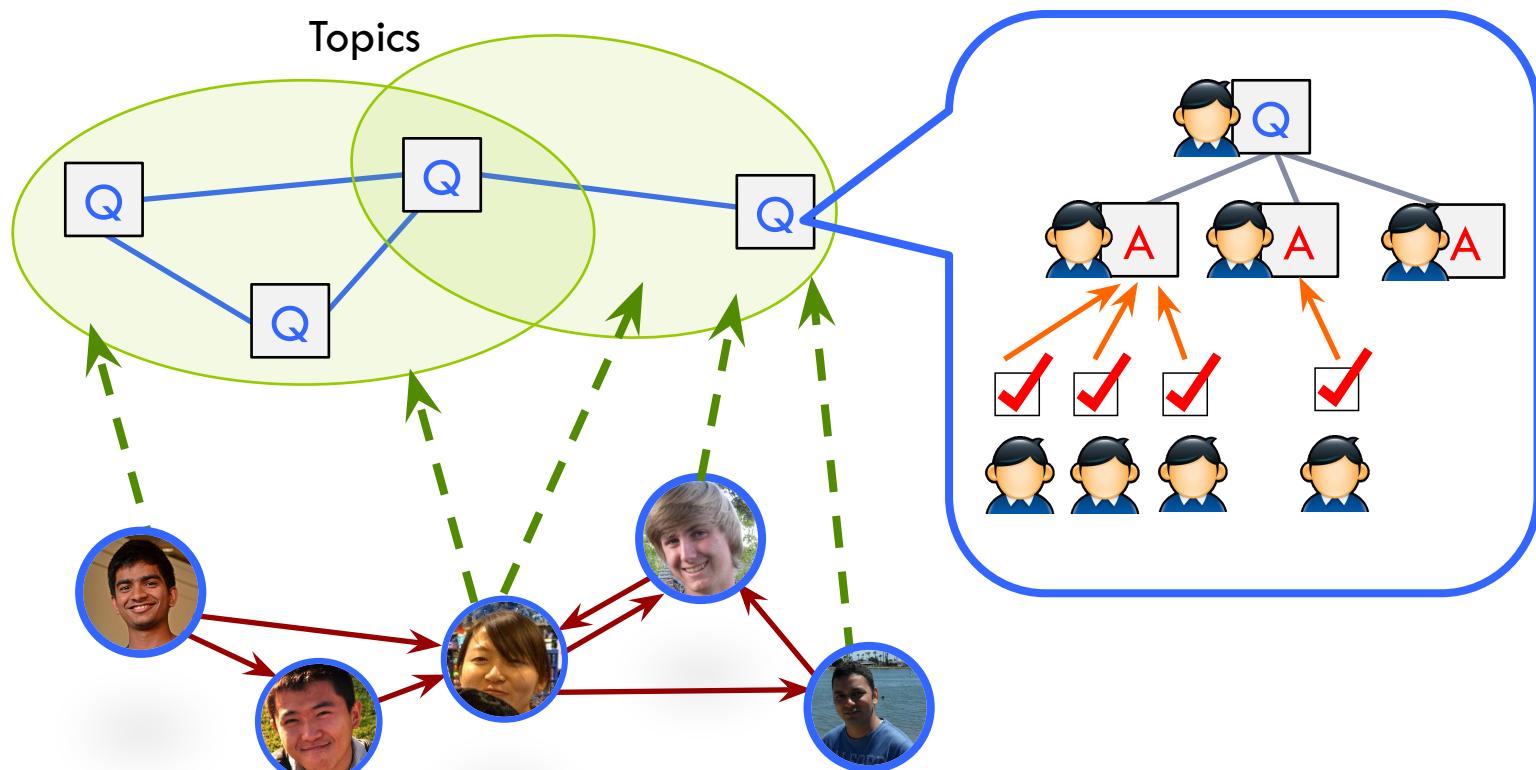
11384 people have viewed this question.

130 people are following this question.



# Graphs, Graphs, More Graphs

- *User-topic graph*: user following topics
- *Social graph*: user following other users
- *Related question graph*: connecting related questions



# Data Collection

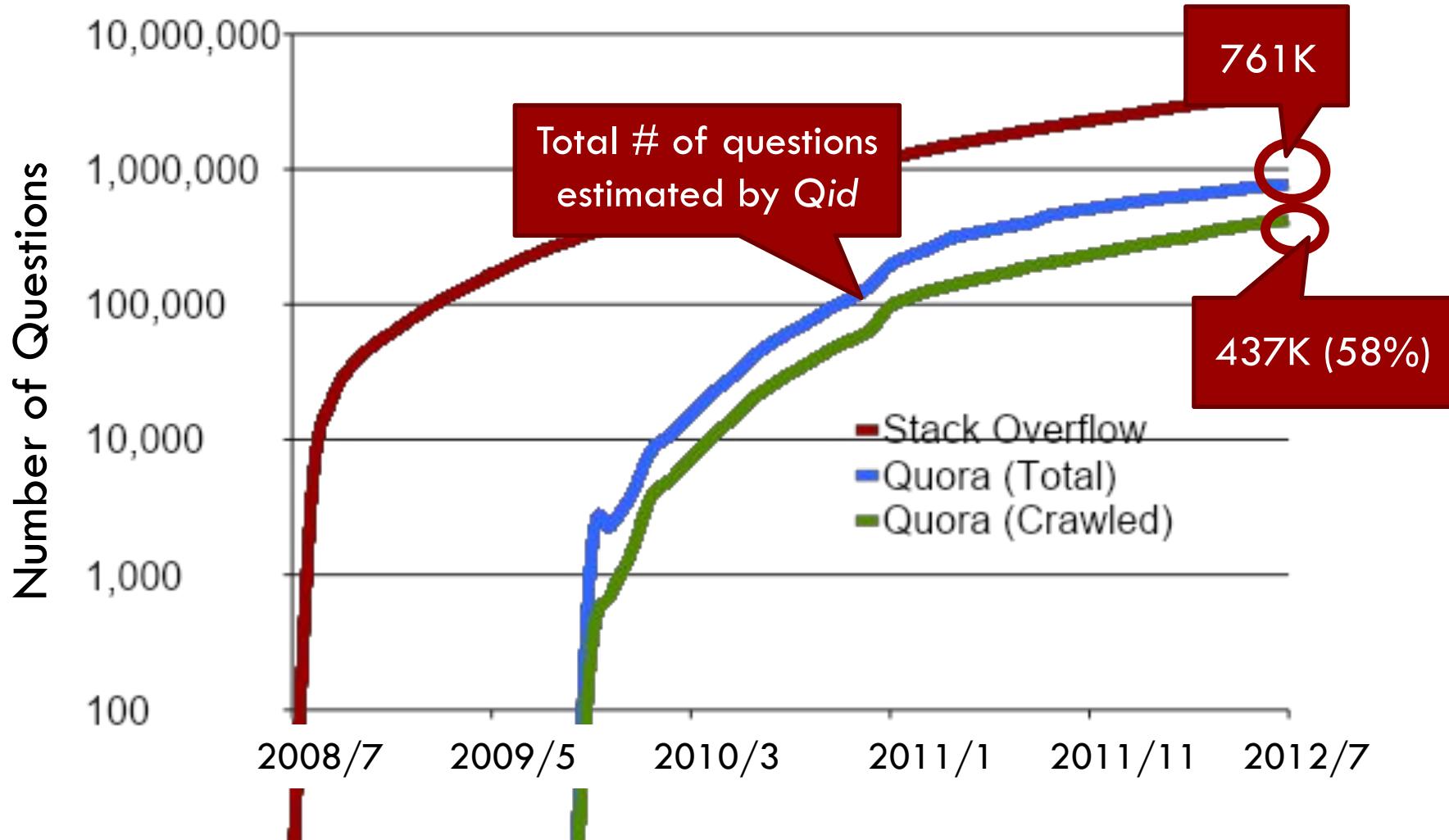
Website	Data Since	Total Questions	Total Topics	Total Users	Total Answers	Question Coverage
Quora	Oct. 2009	437K	56K	264K	979K	58%
StackOverflow	Jun. 2008	3.45M	22K	1.3M	6.86M	100%

- Crawling Quora
  - Snowball-crawled related question graph (August 2012)
  - Obtained the largest connected component
  - Slow speed, minor impact to the site
- Using the dataset of StackOverflow as a comparison



# Growth Over Time

Similar growth trend with StackOverflow



# Outline

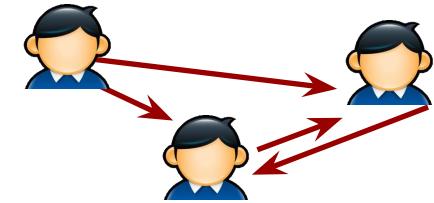
- Introduction
- Characterizing Quora
- Analyzing Graph Structures
  - Social Graph
  - Related Question Graph
- Implications



Details on User-Topic  
Graph in the paper!

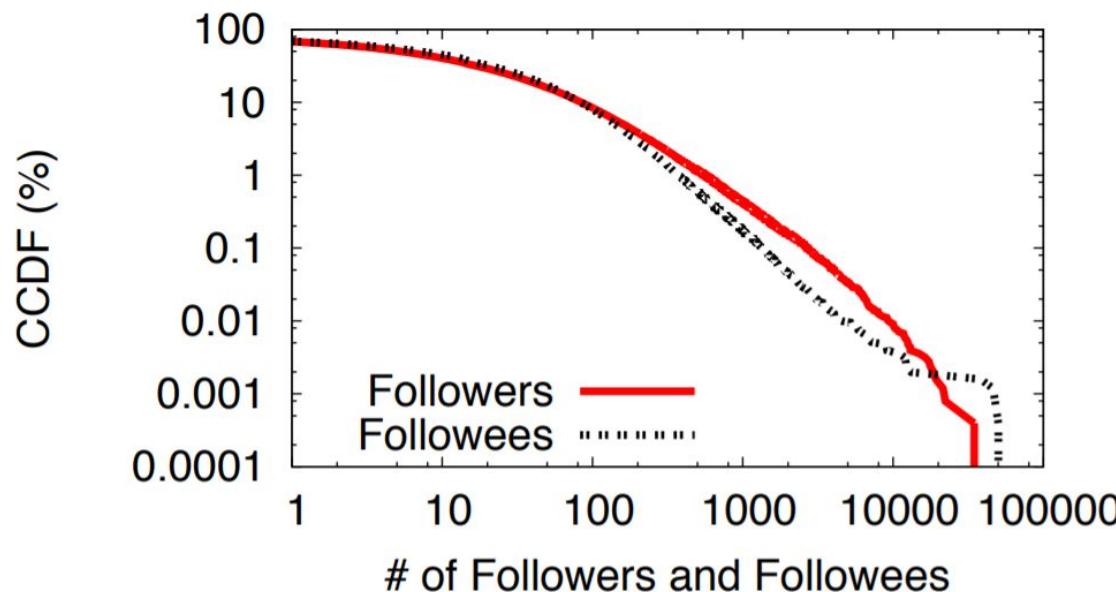
**How do social ties impact Q&A activities?**

# Social Graph Structure

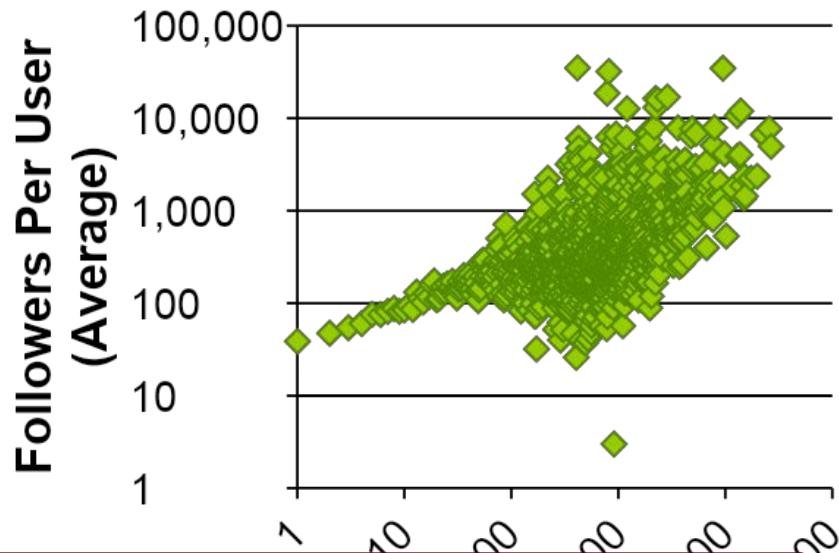
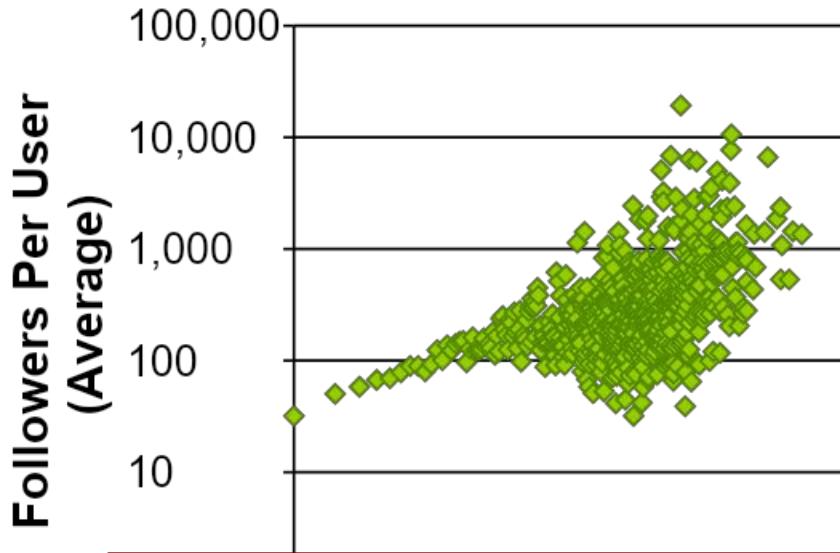


- Users can follow other users to build social connections
  - Asymmetric social graph
  - Users receive items in their newsfeed from people they follow

Social degree has power-law distribution



# Is the Social Graph Meaningful?

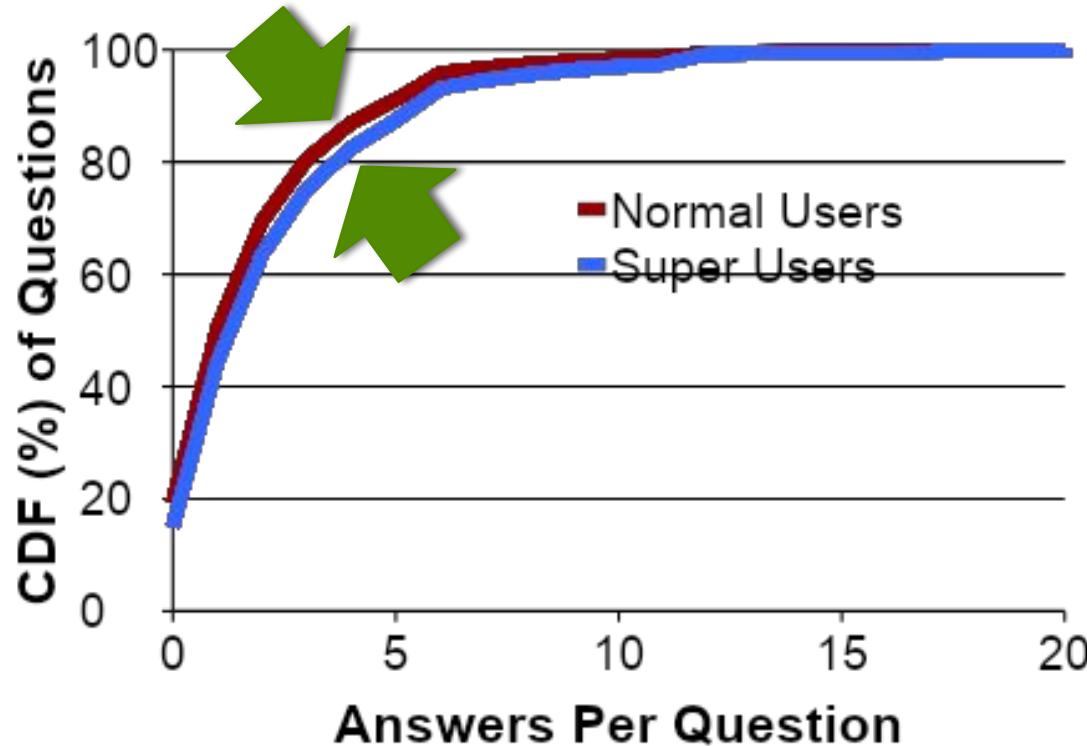


- More answers or high-quality answers == more followers
- Social structure could indicate content quality
- Correlation between user's # of followers and
  - # of total answers the user wrote
  - # of votes the user ever received

# Using Social Ties to Attract Answers

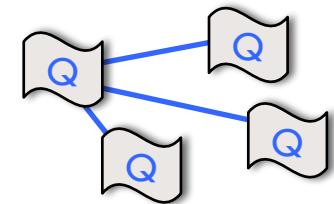
- Would social ties help to attract answers?
- Defining “super-users”
  - Top 5% users sorted by # of followers

Social ties have no effect on attracting answers

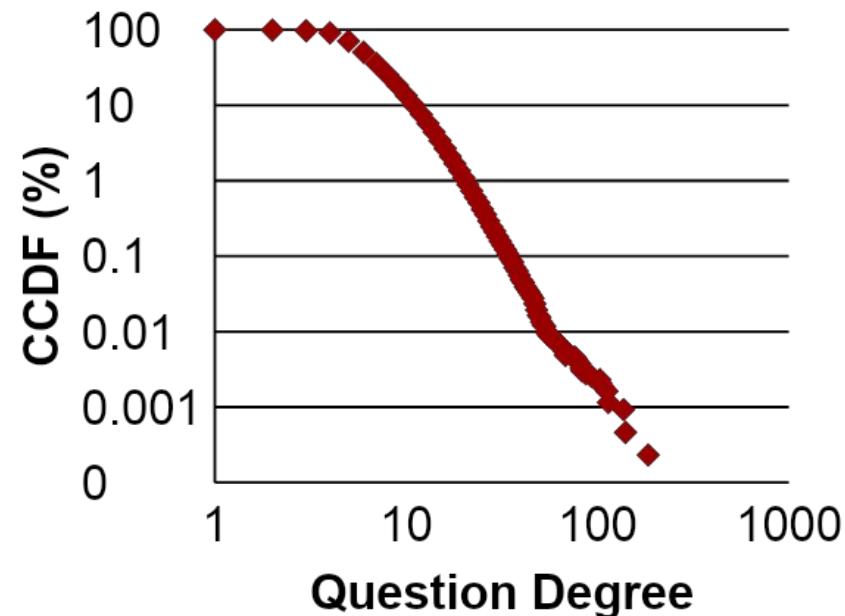


How does Quora direct users to  
“interesting” questions?

# Related Question Graph

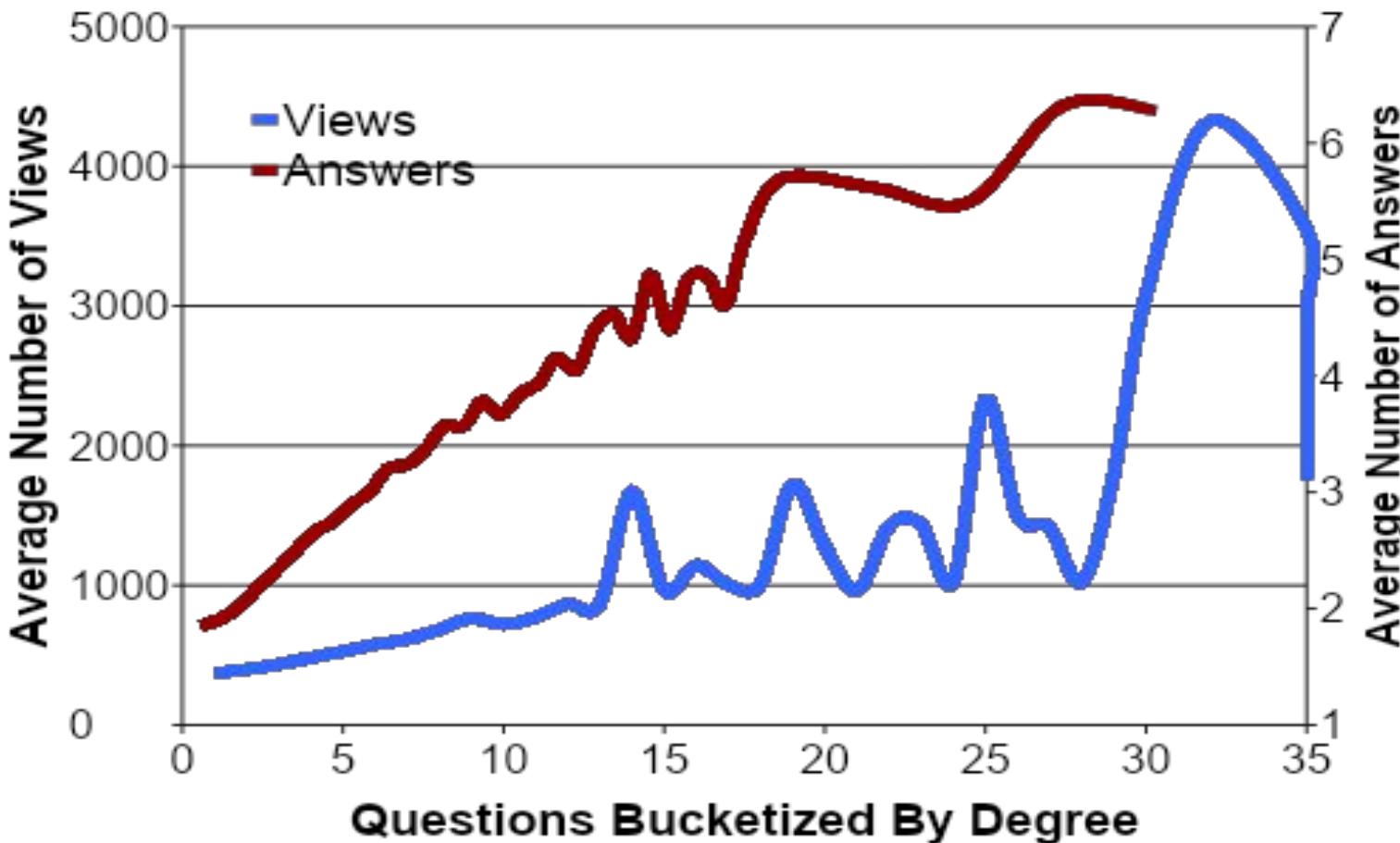


- Related question feature
  - Allows users to browse a series of related questions
- Related question graph
  - Questions as nodes, edges indicates “related” relationships
- Graph properties
  - Power-law structure
  - A small set of “core” questions inside each topic



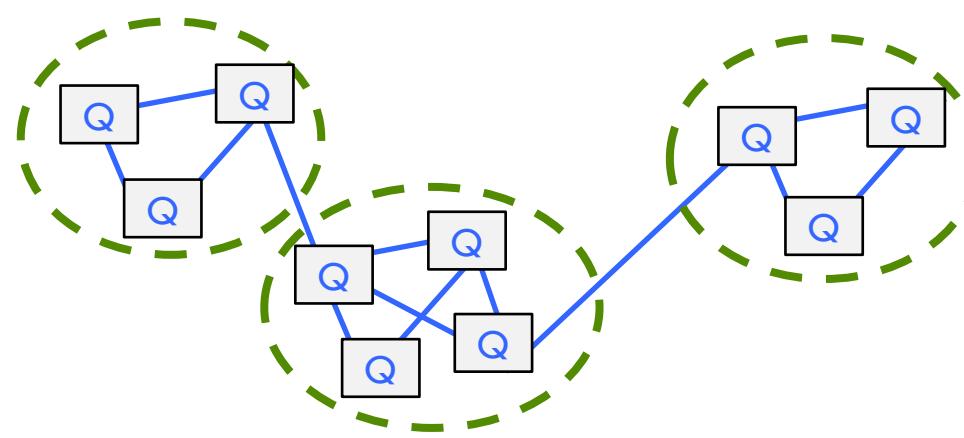
# Impact of Question Degree

- Strong correlation: question degree & user's attention on the question
- Question graph drives users to “core” questions



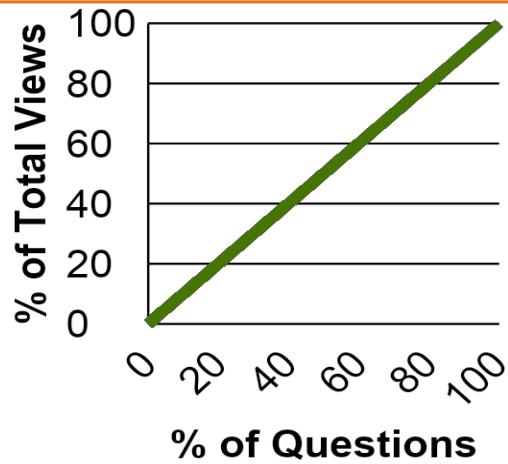
# User Attention on Similar Questions

- **Similar questions in Quora**
  - Questions around very close (same) subjects
  - Redundant questions asked by different users
- **Do users pay equal attention to similar questions?**
- Locating similar questions by partitioning question graph
  - METIS, produce clusters, each contains similar questions

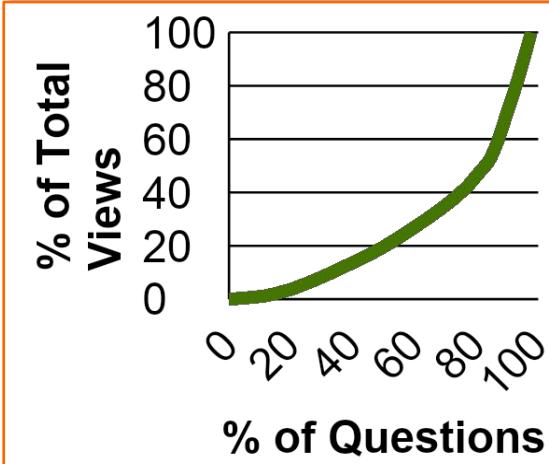


# Equal Attention on Similar Questions?

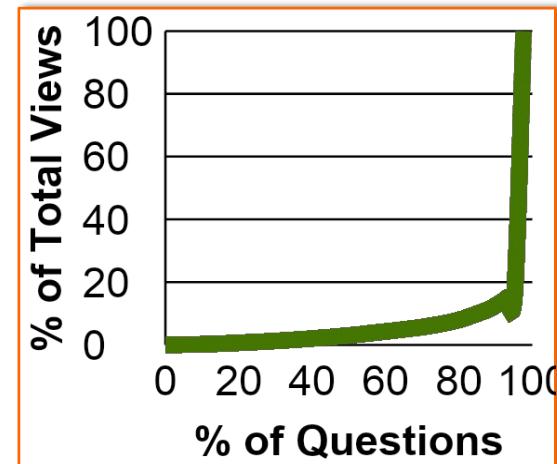
- Is user attention evenly distributed in each cluster?
  - Gini coefficient ( $G$ ): evaluate the uniformity of distribution
    - $G=0$ : perfect equality
    - $G \sim 1$ : extremely skewed distribution



$G=0$



$G=0.4$



$G=0.9$

- User attention is highly skewed in each cluster
- Excellent! users are not distracted by similar questions

# Implication and Conclusion

- Implication for crowdsourcing content sites
  - Q&A sites
    - Users attention is “skewed” to top questions
    - Avoid distraction, encourage contribution
  - Other sites such as *Yelp*, *TripAdvisor*
    - Drive enough reviews to key venues
    - Ensure reliable rating
- The **first** large-scale measurement study on Quora
- Graph structures contribute to effective content discovery
  - Social graph indicates content quality
  - Question graph focuses user attention

# Whispers in the Dark: Analysis of an Anonymous Social Network

Gang Wang, Bolun Wang, Tianyi Wang, Ana Nika,  
Haitao Zheng, Ben Y. Zhao

UC Santa Barbara  
[gangw@cs.ucsb.edu](mailto:gangw@cs.ucsb.edu)

IMC'14



# Concerns of Using Online Social Networks

- Users cannot **speak freely** in online social networks
  - User profile is linkable to real-world identity
  - Online actions can cause serious consequences



Justine Sacco And The Self-Inflicted Perils Of Twitter



Election 2010: Labour sacks candidate Stuart MacLennan in Twitter row



SOCIAL NETWORKING

## Georgia Bus Driver Fired for Facebook Post

It's not the first time someone's been canned for publicly embarrassing the boss.

By Anjani Trivedi | June 06, 2013 | 6 Comments

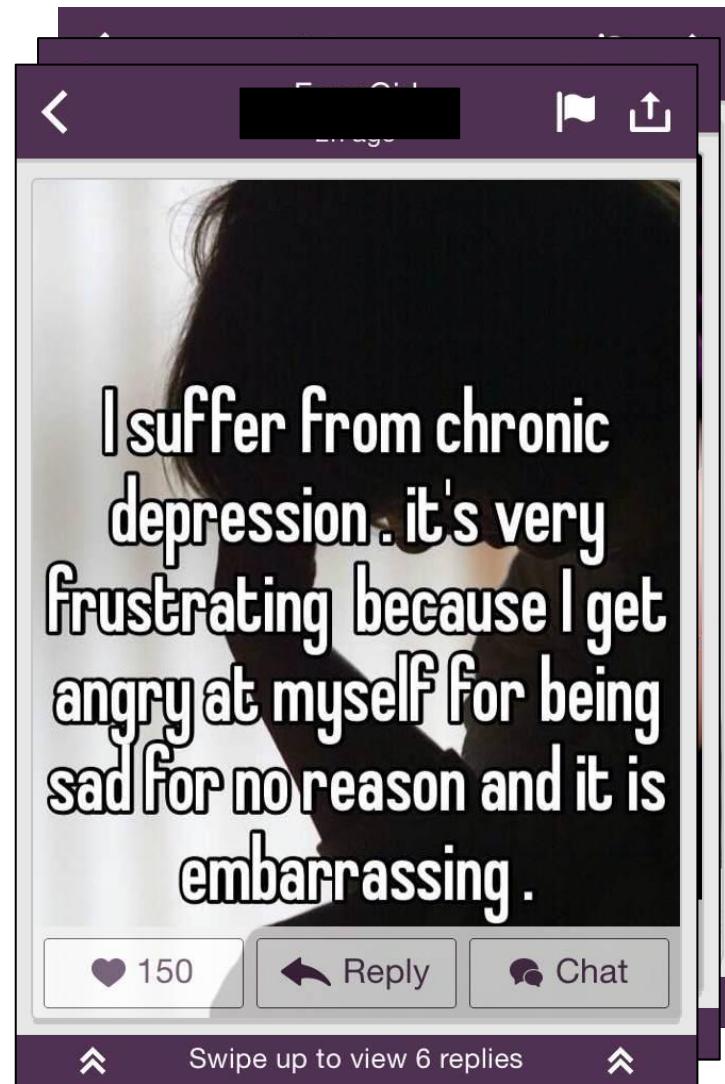
# Whisper Anonymous Social Network

- Whisper, an anonymous social app
  - Online profile **unlinkable** to real identity
  - Express freely without fear of retaliation or abuse
    - Share stories, seek advice, express complaints
    - Whistleblowers, teenagers avoiding bully
  - Interact with people anonymously
  - > 3 billion monthly page views, 2014
- Part of wave of new, anonymous social networks
  - SnapChat, Secret, Yik-yak, Wickr, Rooms (Facebook)



# w Key Features

- No personally identifiable information
  - No **real names**, only nicknames
  - No **user profiles** (phone#/email)
  - No explicit **social links**
  - Moderate content to make sure users don't reveal their identity
- Post whisper messages
  - Topics including relationships, family, work, religion, politics, sex, etc.
  - Secrets, confessions



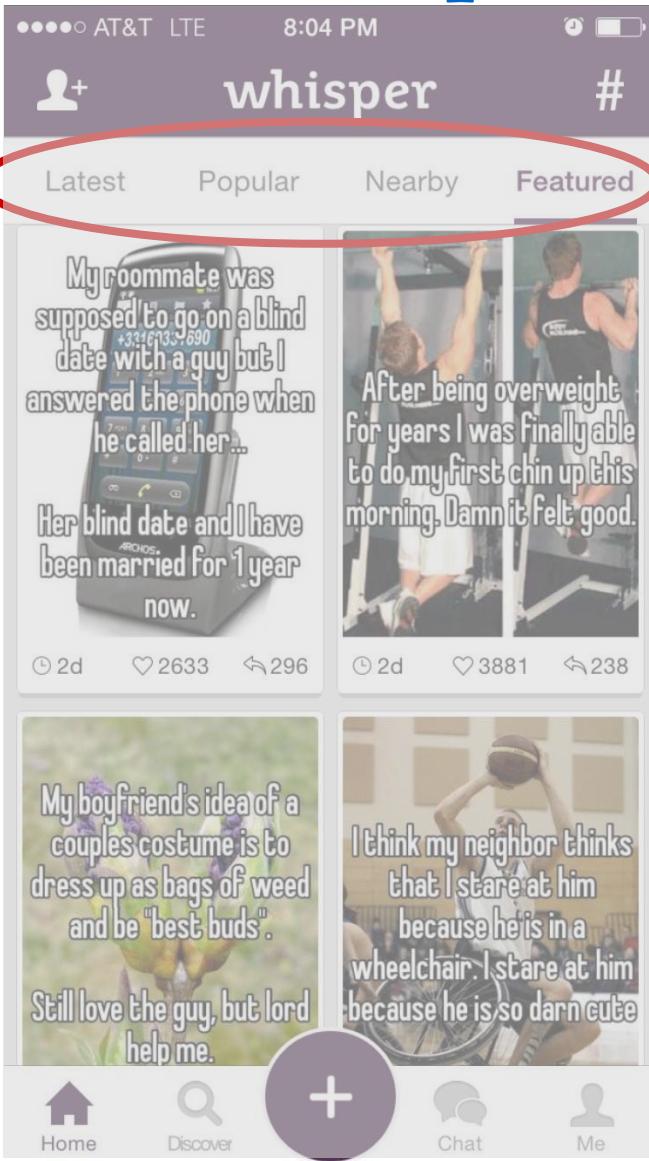
# Our Goals

- Understand how anonymity affects **user behavior** in anonymous social networks
  - How is Whisper's network structure different from existing networks like Facebook and Twitter?
  - How does anonymity impact the friendships between users and user engagement over time?
  - Implications on user anonymity and privacy

# Outline

- Motivation
- Dataset and Whisper Network
  - Data Collection
  - Basic Network Structure
- User Engagement and Stickiness
- Anonymity and Privacy in Whisper
- Conclusion

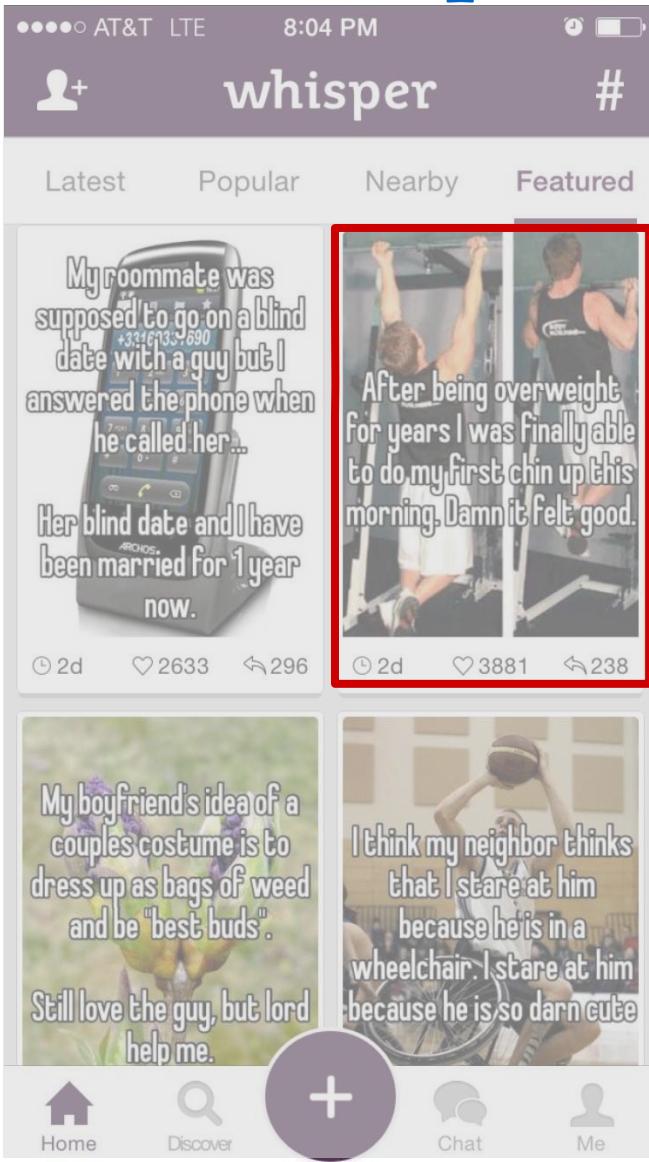
# Whisper Functions and Data



## Public whisper lists

- **Latest:** all recent whispers in the network
- **Nearby:** whispers in local area < 40 miles
- **Popular:** whispers received many replies
- **Featured:** editor-picked whispers

# Whisper Functions and Data



# Data Collection



- Crawled the “latest whisper” stream for 3 months\*
  - All public messages from February to May 2014
  - 9,343,590 original whispers, 15,268,964 whisper replies
  - 1,038,364 unique userIDs
- Interacted frequently with Whisper
  - In-person meetings to get data collection permission
  - Whisper removed GUID in June 2014

Global universal identifier (GUID) □  
Link the same user's data over time

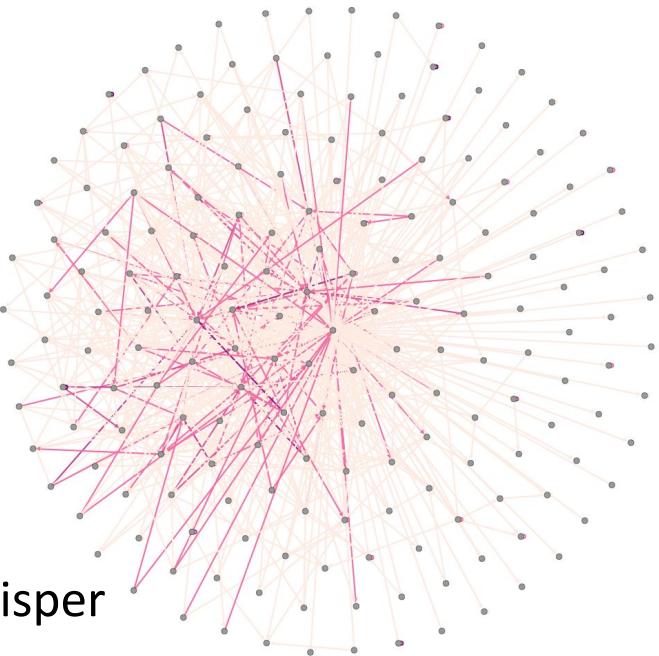
\*Data collection with Whisper's permission, IRB approved

# Basic Analysis: Interaction Graph

- How do users interact with each other with no explicit social links?
- **Interaction graph:** Whisper vs. Facebook and Twitter
  - Users are nodes, edges represent user interaction
  - 3-month time window for all three graphs

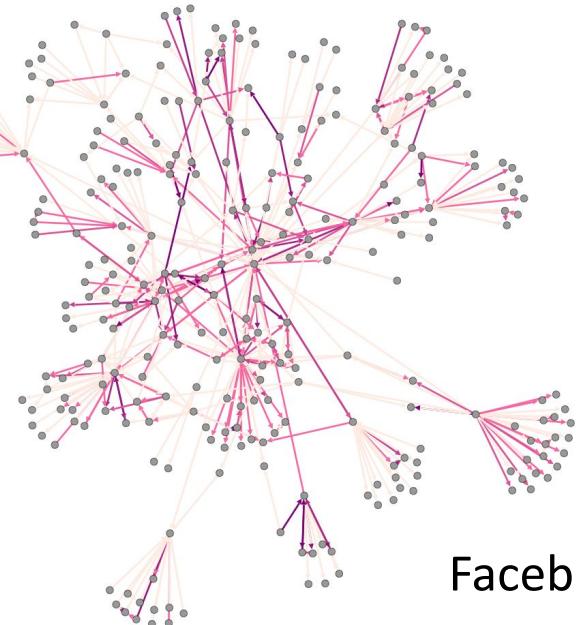
Graph	Interact. Event	Nodes	Edges	Avg. Degree	Clustering Coefficient	Avg. Path Length	Assort.
Whisper	Replies	690K	6531K	<b>9.47</b>	<b>0.033</b>	4.28	-0.011
Facebook	Wall Posts	707K	1260K	1.78	0.059	10.13	0.116
Twitter	Retweets	4,317K	16972K	3.93	0.048	5.52	-0.025

Whisper graph has *high dispersion*  
Interact with a wide range of strangers

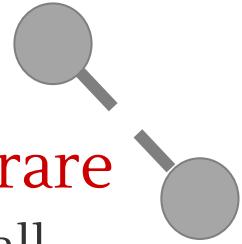


Existing social networks:  
Interact with a fixed set of friends

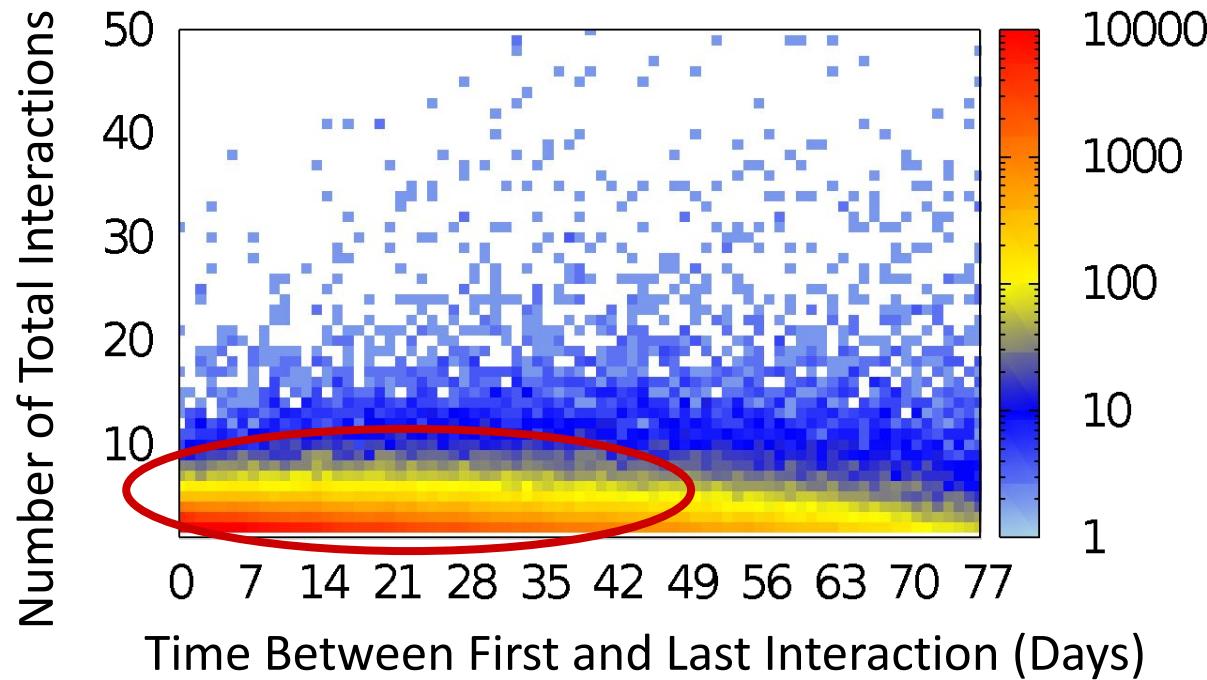
VS.



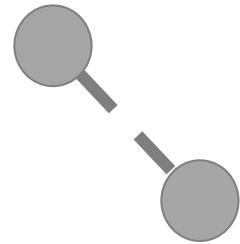
# Persistent Friendship



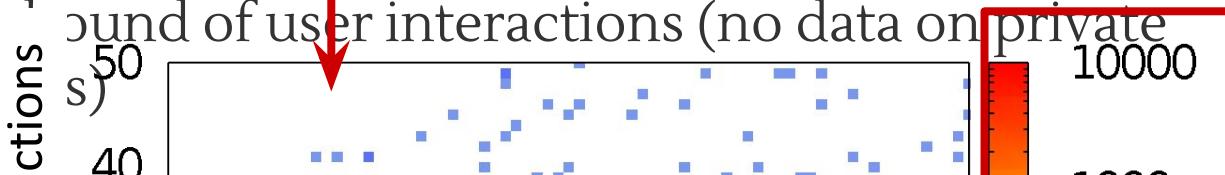
- Persistent user pairs (strong ties) are **extremely rare**
  - Only **7.7%** user pairs interacted multiple times (out of all edges)
  - Majorities are weak ties, talked once, never again
  - Lower bound of user interactions (no data on private messages)



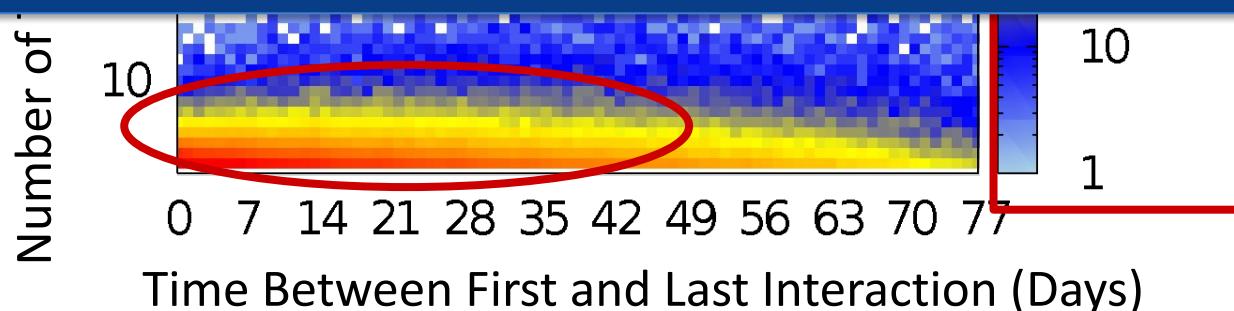
# Persistent Friendship



- Persistent user pairs (strong ties) are **extremely rare**
  - Only **7.7%** user pairs interacted multiple times (out of all edges)
  - Majorities are weak ties, talked once, never again
  - Low **count** of user interactions (no data on private mess...  
mess...  
s)  
s)



Majority of user-pairs have weak relationships:  
short-lived, with few interactions



# Do Communities Exist?

- Community detection on Whisper interaction graph
  - Modularity-based approaches: *Louvain* and *Wakita*
  - Resulting modularity: Louvain (0.492), Wakita (0.409)
- Modularity  $> 0.3 \square$  community structure
  - Facebook (0.63), Youtube (0.66), Orkut (0.67) [IMC'09]
  - Whisper has weak community structures

Even though users don't have persistent friends, they still form communities

# Why Do Users Form Communities?

- **Intuition:** users interact with nearby users (via nearby list)
- **Validation:** whether community membership correlates with geographic location
  - Example community of 28,342 users, its top 4 regions are
    - California (62%), Texas (1.5%), England (1.2%), Arizona (0.9%)
- Users within a community likely from the same region

# Why Do Users Form Communities?

- **Intuition:** users interact with nearby users (via nearby list)
- **Validation:** whether community membership correlates with geographic location
  - Example community of 28,342 users, its top 4 regions are
    - California (62%), Texas (1.5%), England (1.2%), Arizona (0.9%)
- Users within a community likely from the same region

Percentile of Communities	1 <sup>st</sup> Region	2 <sup>nd</sup> Region	3 <sup>rd</sup> Region	4 <sup>th</sup> Region
50-percentile	52%	3.9%	1.5%	1.4%
70-percentile	45%	1.4%	1.3%	1.3%
90-percentile	32%	0.9%	0.9%	0.8%

Users form communities based on geolocation

# Outline

- Motivation
- Dataset and Whisper Network
- User Engagement and Stickiness
  - User Engagement Over Time
  - Predicting Future Engagement
- Anonymity and Privacy in Whisper
- Conclusion

# From Network Ties to User Engagement

- **Background:** social ties impact network “stickiness”
  - Strong ties: close friends, weak ties: strangers
  - Strong ties help keep existing users from leaving
    - a more “sticky” network
- **Our question:** with a network of strangers, how well can Whisper maintain user engagement over time?
- Evaluate **per-user** engagement over time
  - How long do users stay active?
  - Do users turn dormant quickly?

# From Network Ties to User Engagement

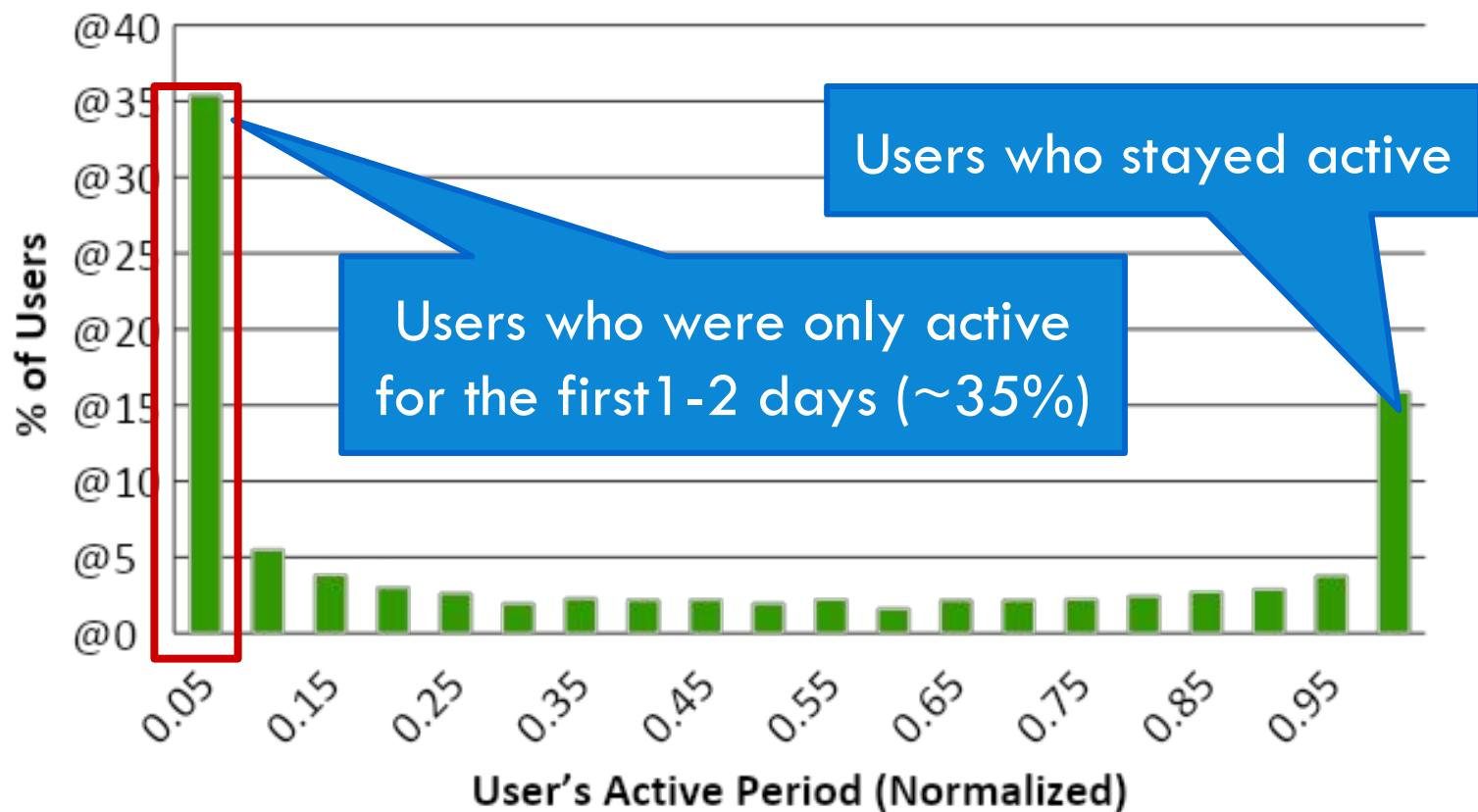
- **Background:** social ties impact network “stickiness”
  - Strong ties: close friends, weak ties: strangers
  - Strong ties help keep existing users **stranger**eaving



- Other examples of social ties in real life? Well, me?
- Every time you log in to a social media platform, you're renewing your connections with other users. This is what keeps you engaged and coming back for more.

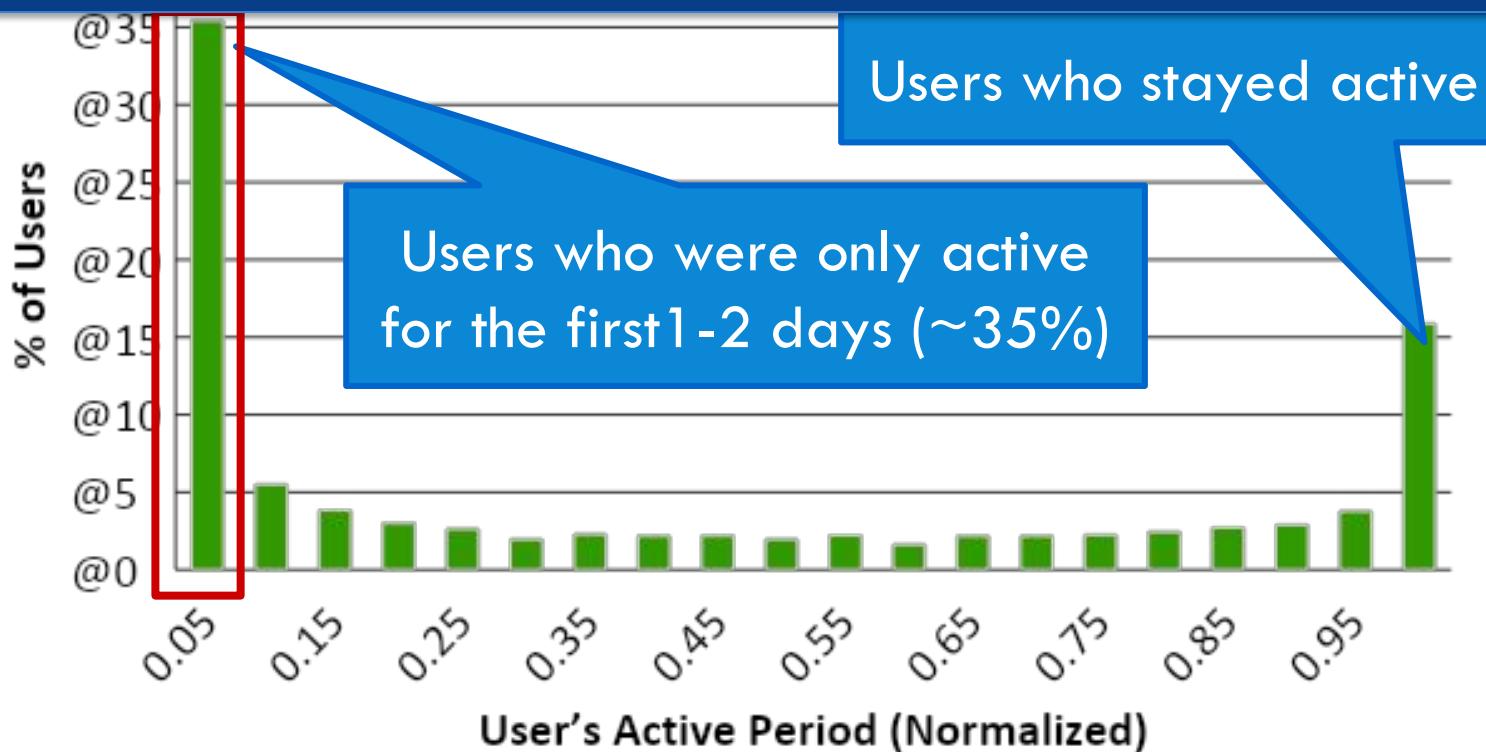
# How Long Do Users Stay Active?

- User's active period (normalized)
  - “Active” means users still generate new content
  - User's active period / our monitoring period of that user



# How Long Do Users Stay Active?

- User's active period (normalized)
  - “Active” means users still generate new content
- Significant portion of users quickly turn dormant
- Bimodal distribution □ predict users stay or not?

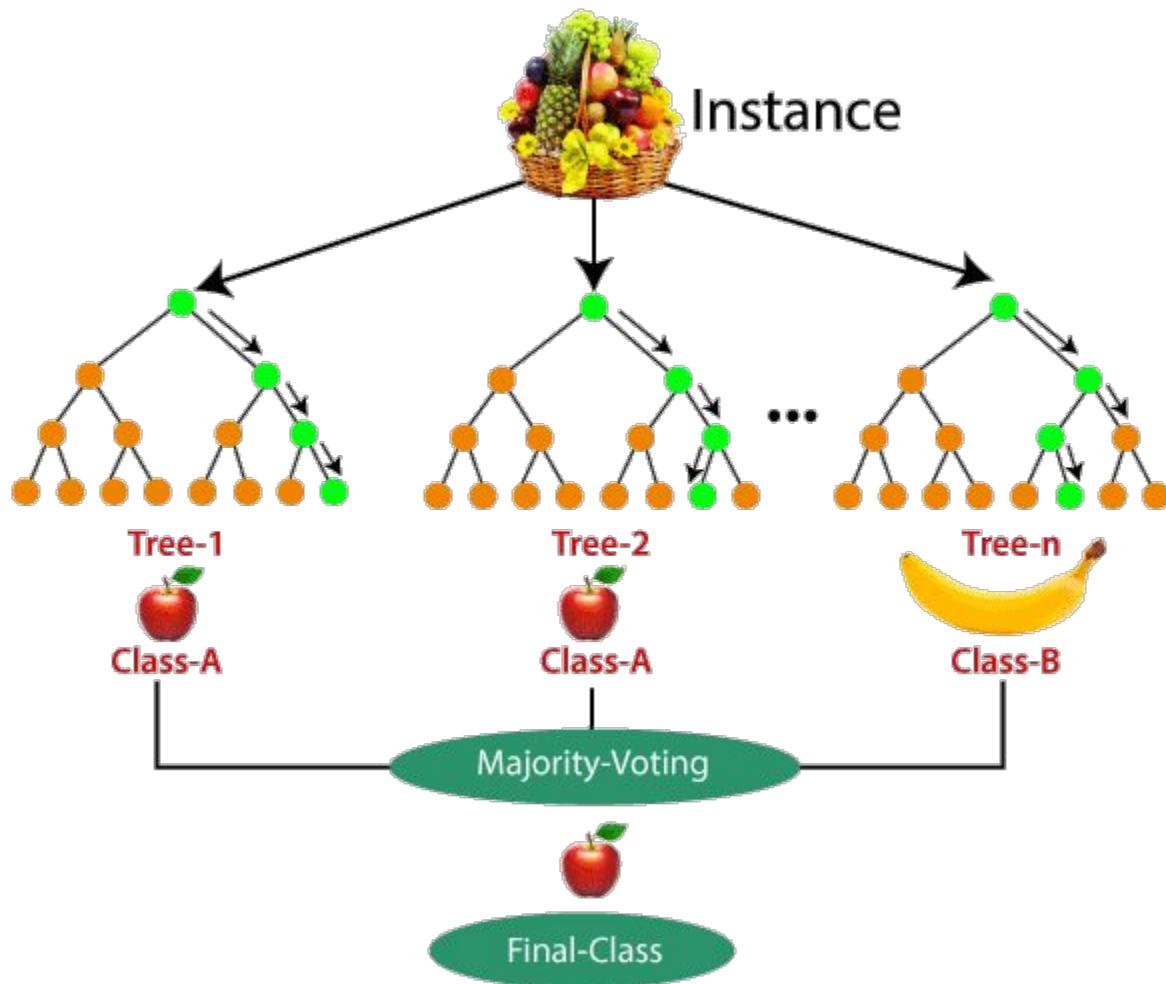


# Predicting User Engagement

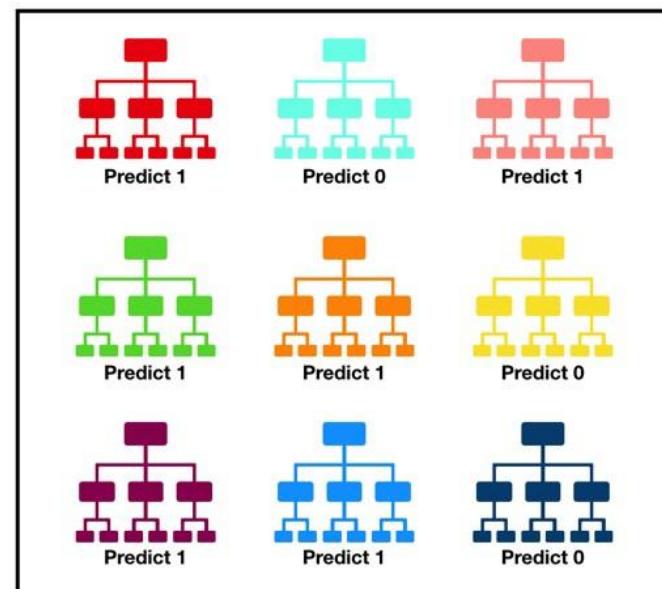
- Binary prediction, whether disengage quickly or not
  - Input: user's data during initial  $X$  days
  - Random Forest ML Classifier
- Features (20)
  - Content posting volume, frequency (7)
  - Social interactions (8)
  - Temporal features (2)
  - Activity trend (3)

A extensive list of features, can be further trimmed

# Quick Info on Random Forest



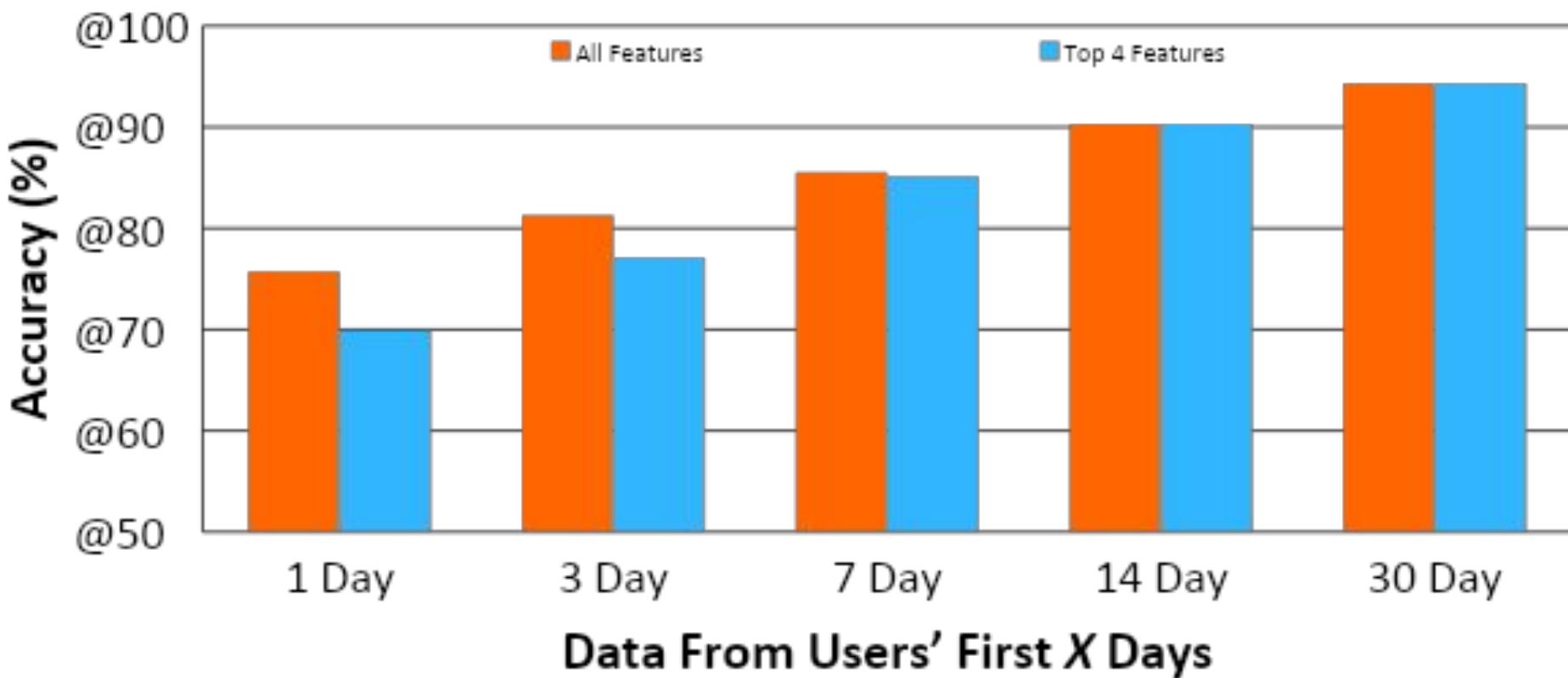
*A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.*



Tally: Six 1s and Three 0s  
Prediction: 1

# Prediction Result (Random Forest)

- Classify users using their first  $X$  days of data



# Prediction Result (Random Forest)

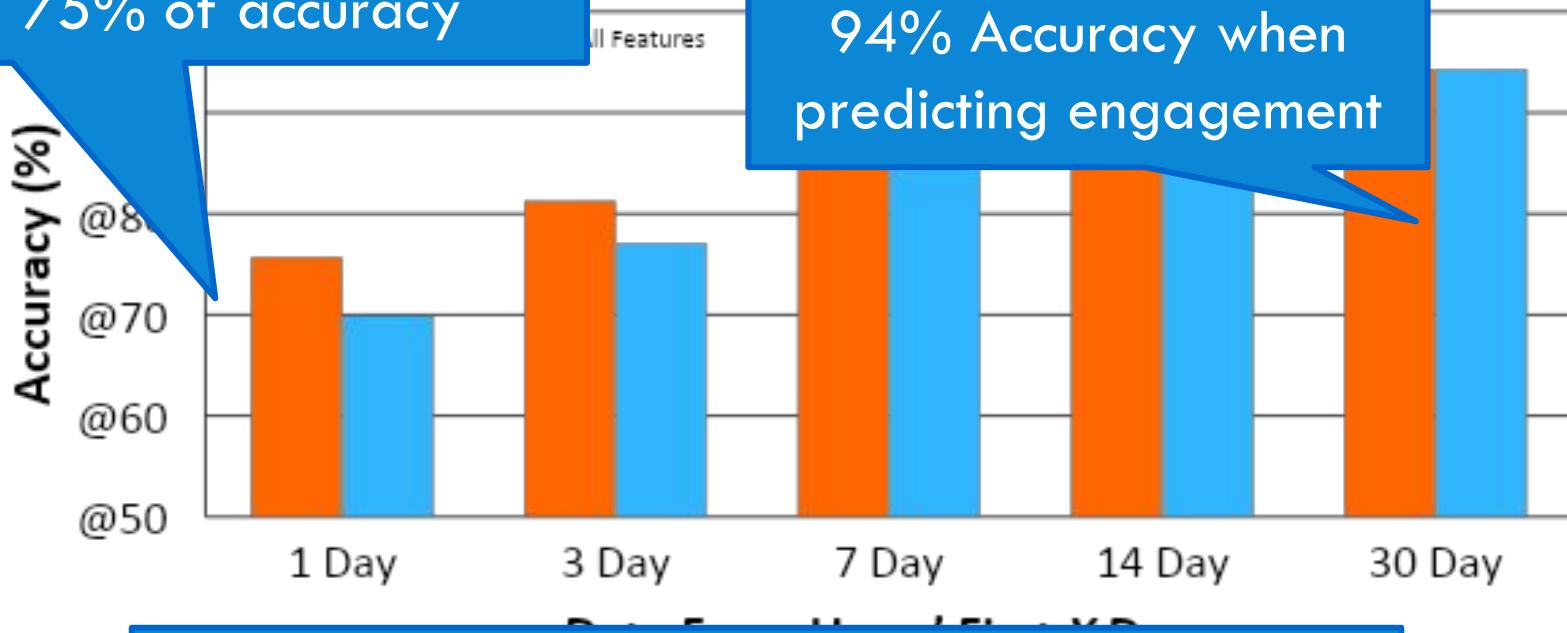
- Classify users using t

Top 4 Features produce accurate results

- # of days with > 1 whisper
- # of days with > 1 reply
- Is posting volume decreasing?
- # of total posts

1-day data already has  
75% of accuracy

94% Accuracy when  
predicting engagement



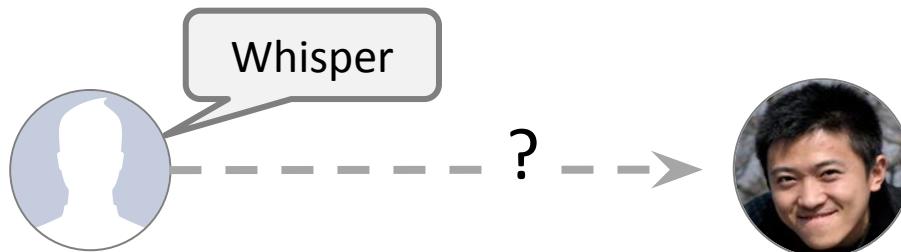
- Whisper can identify users likely to leave
- Increase user engagement using other tools

# Outline

- Motivation
- Dataset and Whisper Network
- User Engagement and Stickiness
- Anonymity and Privacy in Whisper
- Conclusion

# Privacy and Anonymity in Whisper

- Existing mechanisms to prevent PII leakage
  - No personal information is collected (no real name, phone# or email address)
  - Server only stores public whispers, private chats stay on the phone
  - Noise is added to user GPS **before** sending to Whisper's servers
- **Worst case:** attacker compromises servers and obtains data
  - Much more external data needed to de-anonymize users



# Location Tracking Attack

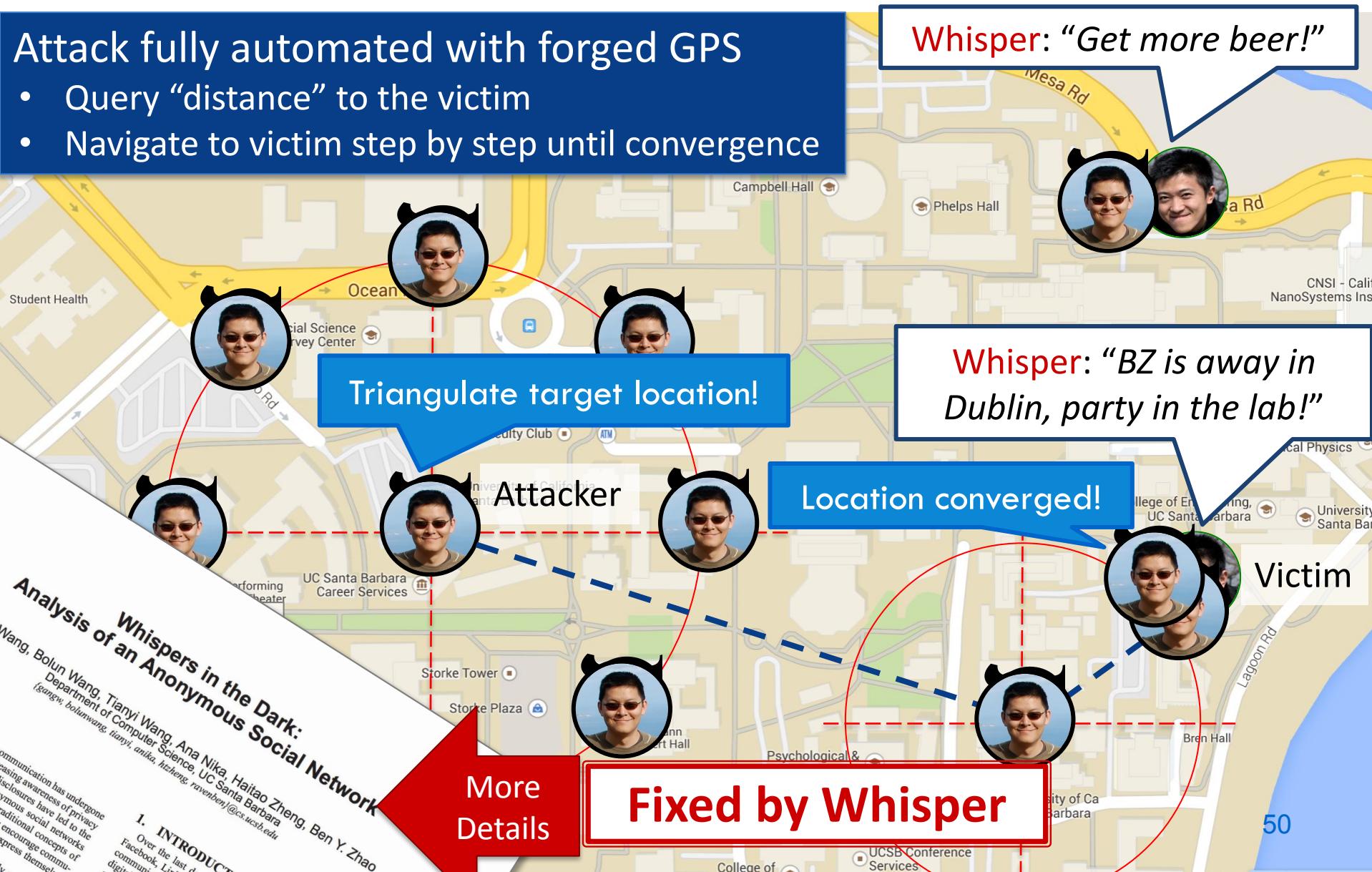
- Tracking whisper users' locations
  - Pinpoint current location: error < 0.2 miles
  - Allow attackers to follow (stalk) users
- How to attack
  - “Nearby list” shows whispers by **distance**
  - **Triangulate** user location using distance measure
  - **Reverse-engineer** Whisper’s noise function
- **Key problem:** lack of GPS authentication
  - Unlimited # of queries from any location (fake GPS input)
  - Use statistics to overcome noise



# An Example Attack

Attack fully automated with forged GPS

- Query “distance” to the victim
- Navigate to victim step by step until convergence



# Summary

- The first large-scale measurements on Whisper
- User interaction has high dispersion, difficult to build persistent friendship
- User engagement shows bimodal distribution, future engagement can be predicted by early-day data
- Anonymous apps can still leak personal information
  - Location: once shared with the app, has the risk of leaking
  - No reliable GPS authentication, attacker can query any locations



Thank You!  
Questions?

# References

- [COSN'13] GARCIA, D., MAVRODIEV, P., AND SCHWEITZER, F. Social resilience in online communities: The autopsy of friendster. In Proc. of COSN (2013).
- [IMC'09] KWAK, H., CHOI, Y., EOM, Y.-H., JEONG, H., AND MOON, S. Mining communities in networks: a solution for consistency and its evaluation. In Proc. of IMC (2009)