Trust and Privacy Attacks in Online Social Networks

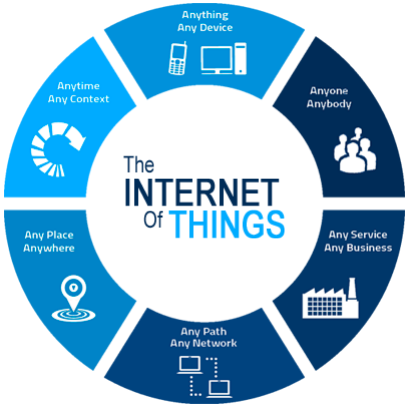
Yuhong Liu

Associate Professor, yhliu@scu.edu
Department of Computer Science and Engineering
Santa Clara University
2022/01/19

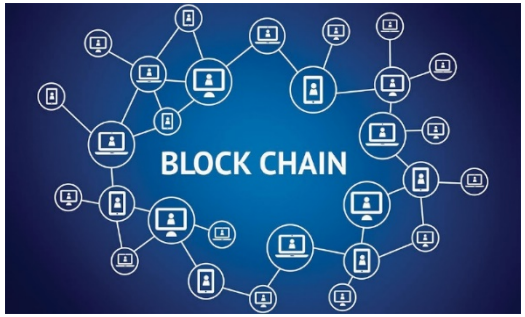
Santa Clara University

- Located at Santa Clara, CA
- A Jesuit private university
- The first high education operating institution in California (founded in 1851)

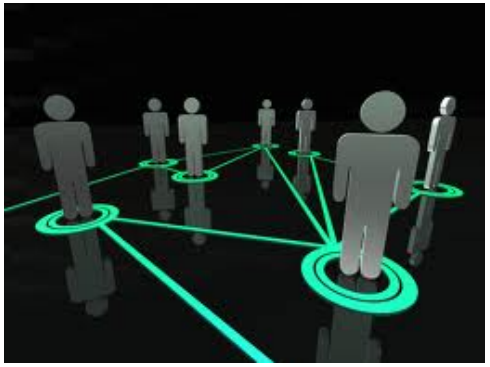
Research Interests – Trustworthy Computing



Energy efficient IoT security



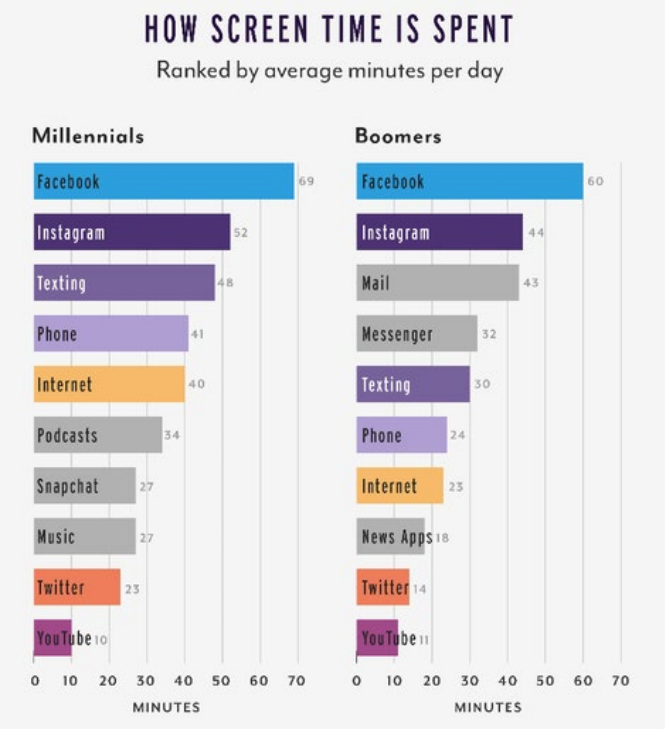
Secure and fair digital trading based on Blockchain



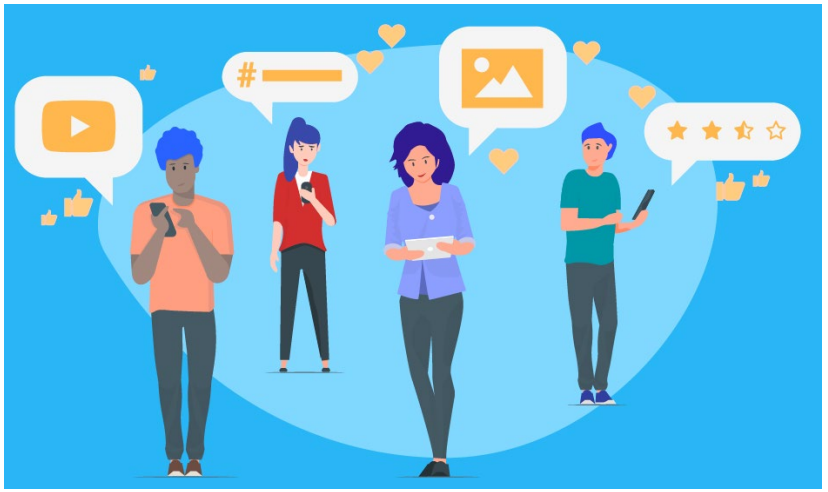
Social Network

- **Feedback based Online Reputation attack and defense**
- **Friendship Privacy**
- **Disinformation/misinformation propagation**

Popularity of Online Social Media



https://www.zdnet.com/article/americans-spend-far-more-time-on-their-smartphones-than-they-think/



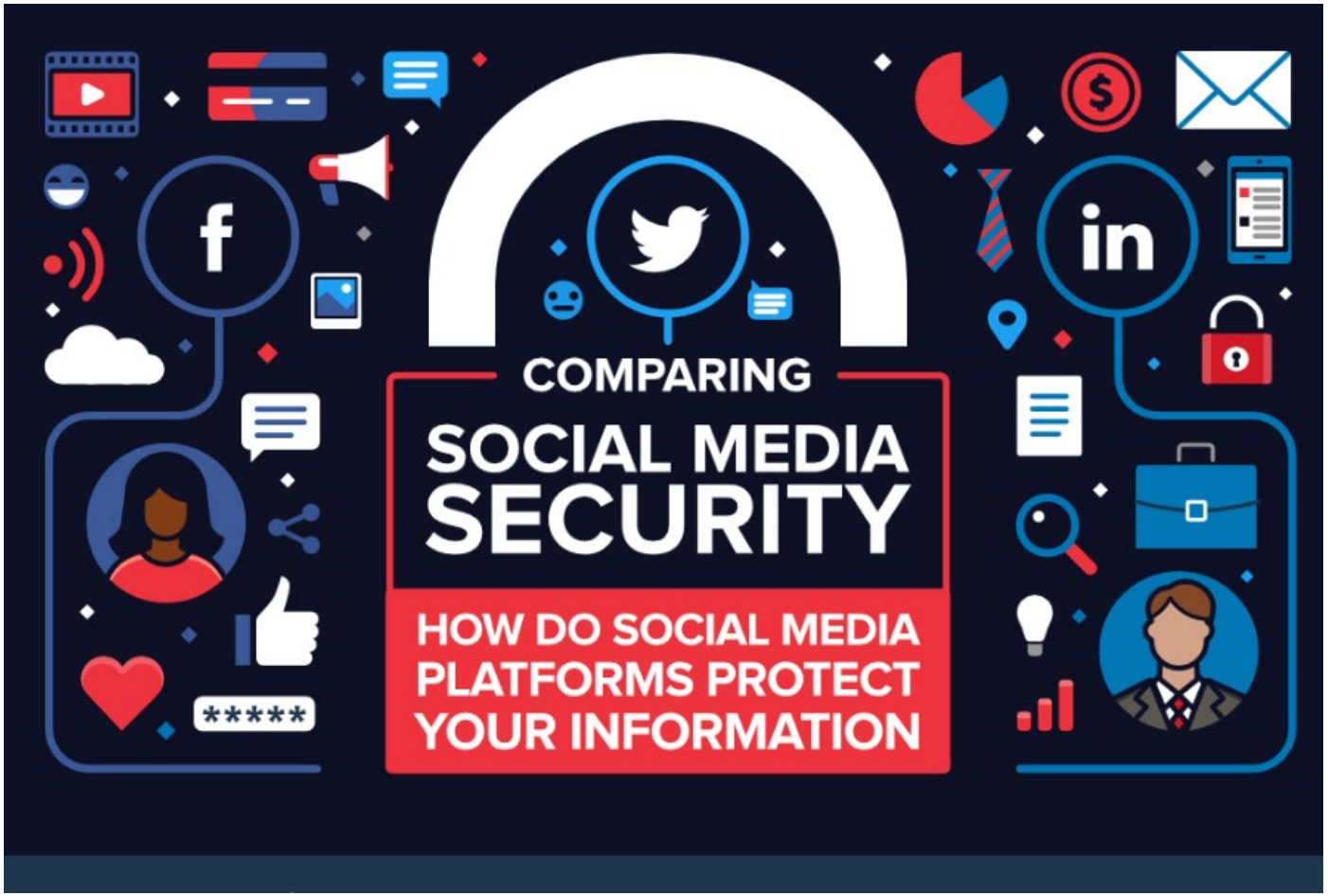
- YouTube users are uploading *500 hours of new videos every minute as of 2020, with more than 2 billion* logged-in monthly users
- Twitter: *187 million* monetizable daily active users, *59% users* regularly get news from Twitter
- Online Review systems: *89%* consumers worldwide read reviews before buying products.

Untrustworthy User Generated Content

Images from online sources

"When I realized that people believe what the Internet says more than reality, I discovered that I had the power to make people believe almost anything." - Andres Sepulveda, a political cyber hacker who digitally rigged elections across Latin America countries for eight years by spreading false information on online social media.

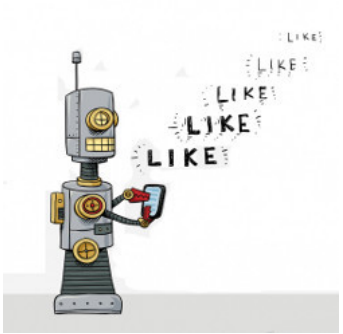
Arising User Privacy Concerns



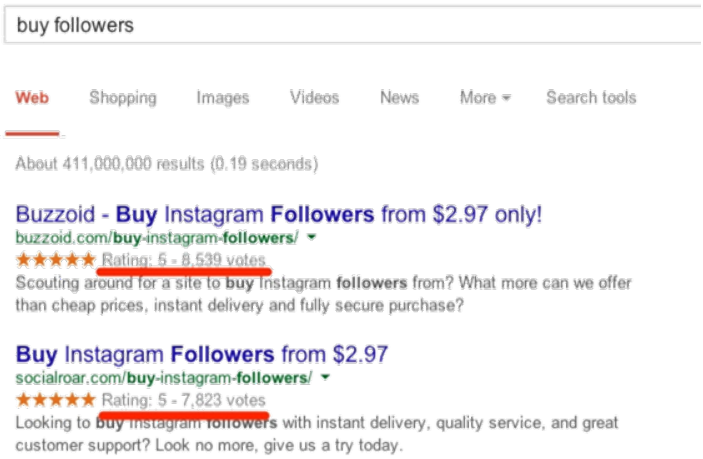
Images from online sources

Various Ways to Attack

Social bots



https://www.medpagetoday.com/infectiousdisease/covid19/91296



Promoting?

Example Projects

- Attack in Online Reputation Systems
  - *Efficiently Promoting Product Online Outcome: An Iterative Rating Attack Utilizing Product and Market Property*
- Friendship Privacy in Online Social Network
  - *Retrieving Hidden Friends: A Collusion Privacy Attack Against Online Friend Search Engine*
- Diversity of Social Groups V.S. Misinformation Propagation
  - *Correlating Diversity and Resistance to Misinformation in Social Media Groups*

Efficiently Promoting Product Online Outcome: An Iterative Rating Attack Utilizing Product and Market Property

Yuhong Liu, *Wenqi Zhou and Hong Chen, "Efficiently Promoting Product Online Outcome: An Iterative Rating Attack Utilizing Product and Market Property", IEEE Transactions on Information Forensics and Security, Vol. 12, No. 6, pp. 1444-1457, 2017.*

Online Reputation Systems

Significance:

- 79% of shoppers say they trust online reviews as much as personal recommendations
- Product pages with customer reviews bring 3.5 times more conversions than those without

Source: https://www.oberlo.com/blog/online-review-statistics

Challenging Issue

Online ratings can be easily manipulated!

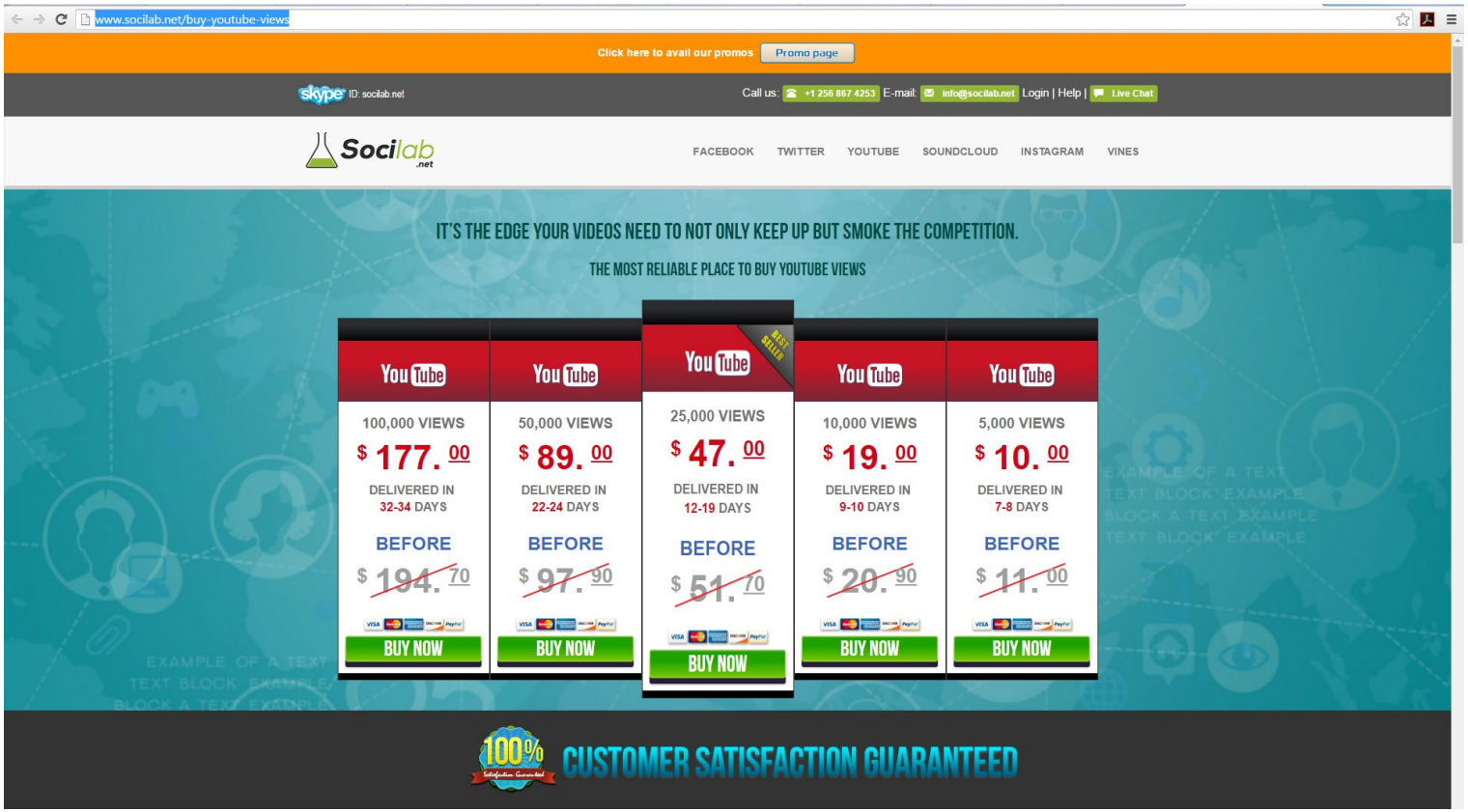
Right after Released

One Month Later

Images come from internet

Challenging Issue

There are companies providing fake ratings for very cheap price !



http://www.socilab.net/buy-youtube-views

Images come from internet

Existing Studies

- Diverse attacks:
  - Self-boosting
  - Alternative behavior
  - Bad mouthing
  - ….

Research Gap

- There are many rating attack studies mainly focusing on
  - How to boost their own products' rating as much as possible without being detected.
- Some aspects are underestimated:
  - Assume high rating values/volumes => increase product sales
  - Treat all products equally
  - Considering only external boost; not the market's "internal feedback"

Can we propose an advanced attack by re-examining these aspects?

Step 1: What factors influence product sales and how?

Influential factor for Online User Choices

Influential Factors	Rating Value	Rating Volume	Network Effect	Herding Effect
Impact	Non-linear	Larger volume => more downloa d/sales	Greater user base -> more download/s ales	Online consumers follows others' purchase
Additional Feature (Product Ranking**)**	The impact of each factor differs over products with different rankings			

The Proposed Quantile Regression Model

$$
\begin{aligned}
log(d_t^i)(\alpha) =& \beta_0(\alpha) + \beta_1(\alpha) * Rev\_u_{t-1}^i + \beta_2(\alpha) * \bar{r}_{t-1}^i \\
&+ \beta_3(\alpha) * \bar{r}_{t-1}^i * \bar{r}_{t-1}^i + \beta_4(\alpha) * R_{t-1}^i \\
&+ \beta_5(\alpha) * log(\widetilde{v}_{t-1}^i) + \beta_6(\alpha) * log(\widetilde{d}_{t-1}^i) \\
&+ \beta_x(\alpha) Controls_{x,i,t} + \xi_{i,t}(\alpha)
\end{aligned}
$$

Variable	Description
d_t^i	number of Weekly downloads of software i at week t
$\widetilde{d}_t^i$	total number of downloads of software i by week t
$\bar{r}_t^i$	average user rating of software i by week t
$\widetilde{v}_t^i$	total number of user rating of software i by week t
R_t^i	the rank of software i at week t measured by weekly sales/downloads
$Rev\_u_t^i$	a binary variable to measure if software i is reviewed by users at week t

Data Set

- In particular, we collect weekly data of software downloads and online user ratings from CNETD over 26 weeks in four categories.
  - Anti-virus
  - Download Managers
  - File Sharing
  - Web Browser

Impact of Rating Volume

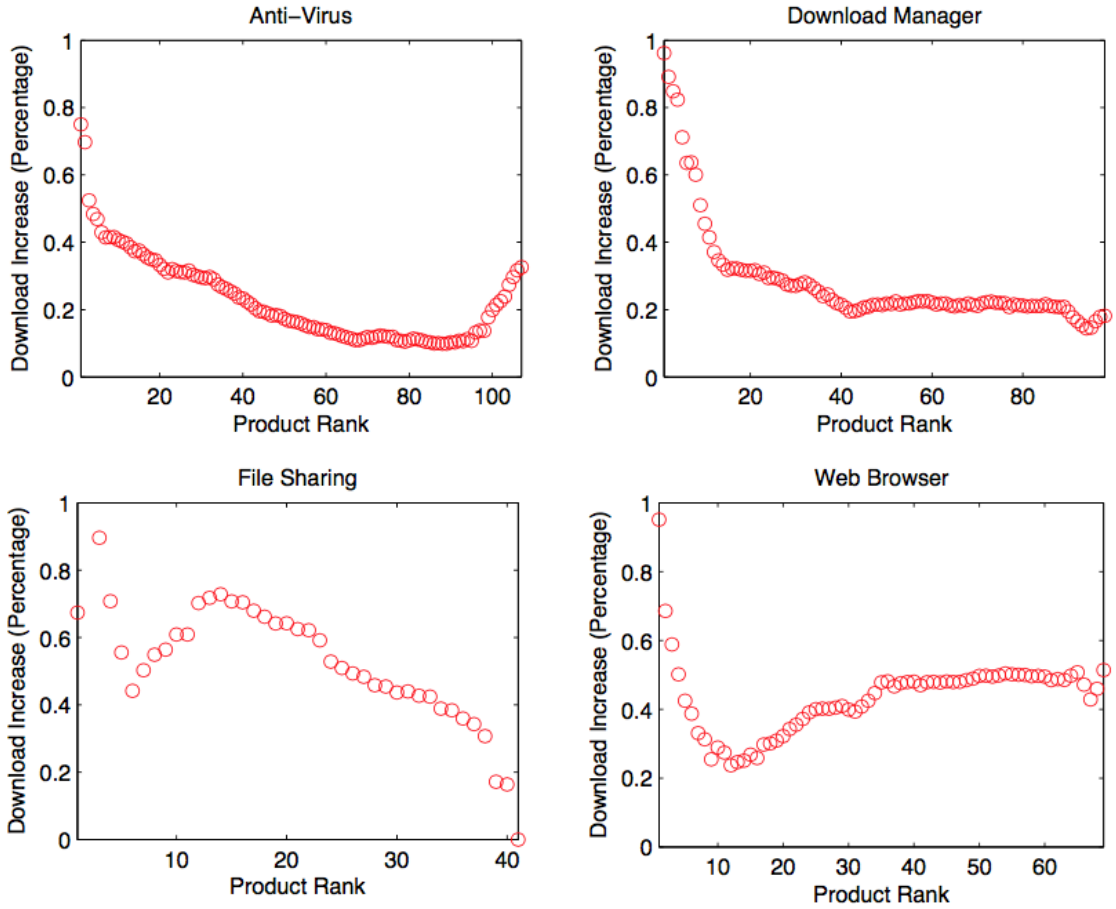


Fig. 2: Impact of Rating Volume Change

Observations:

1. the increase of rating volumes will always lead to positive impact on improving future downloads.

1. the downloads of top ranked products are influenced most by rating volume changes. the impact of rating volume change dramatically drops for lower ranked products.

Impact of Rating Value

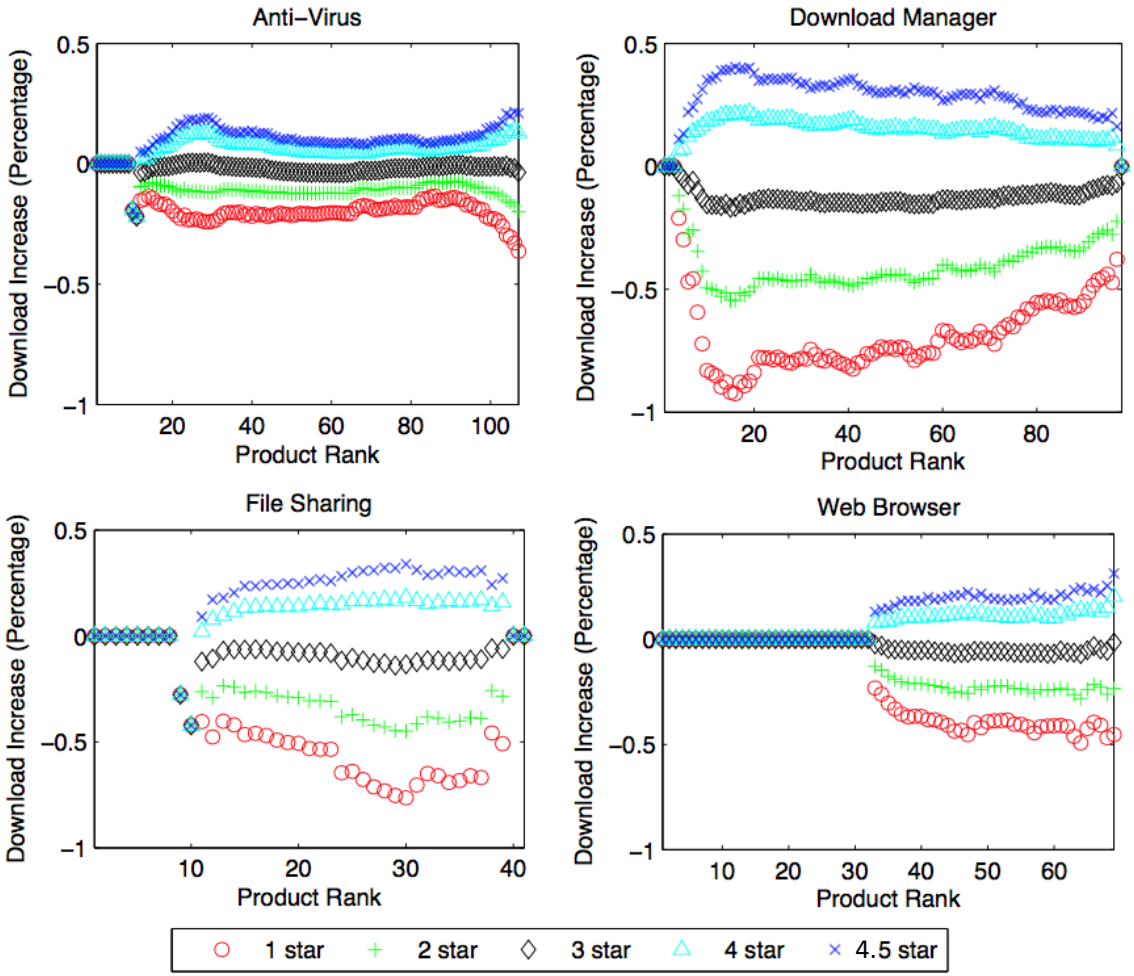


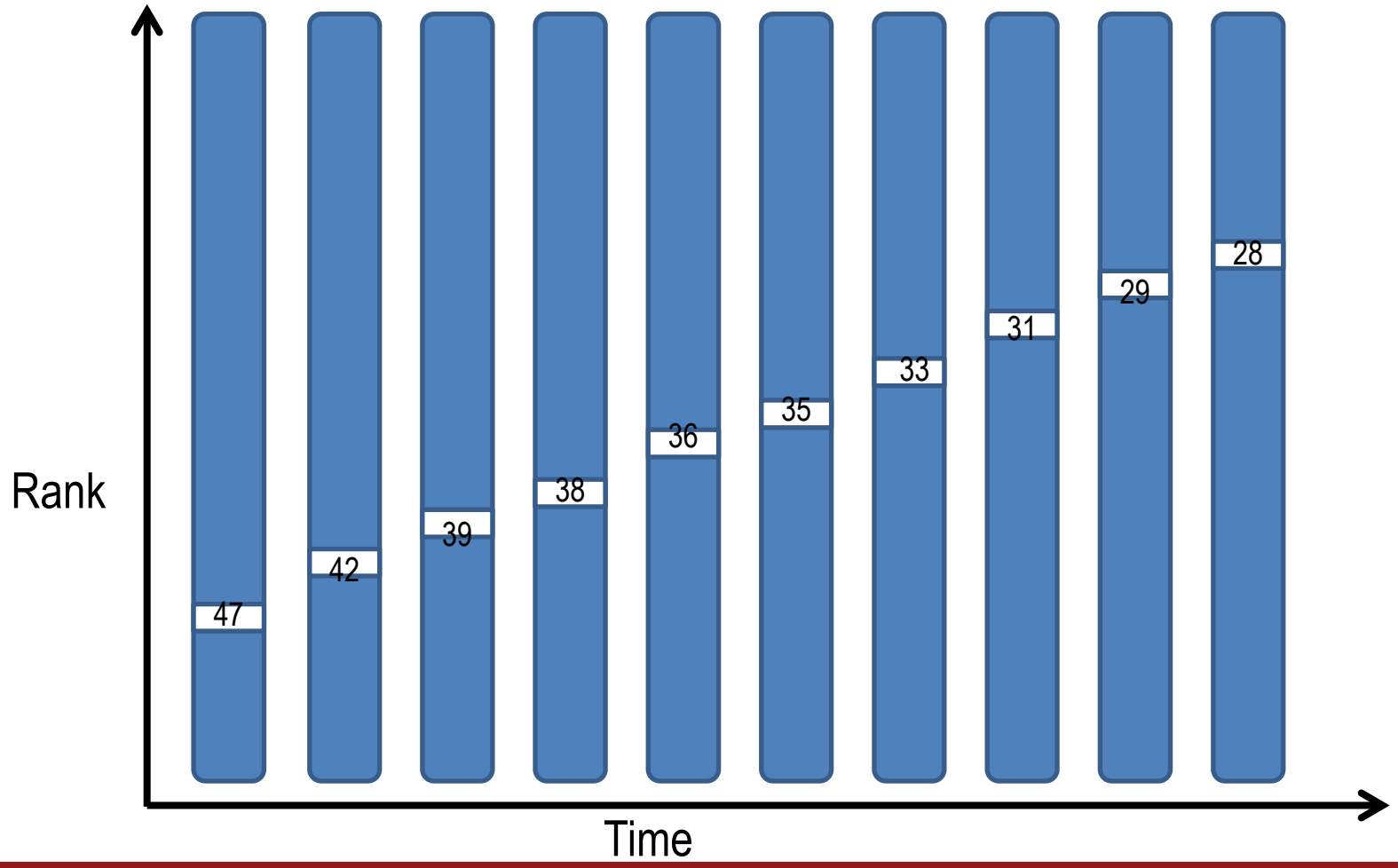
Fig. 1: Impact of Rating Value Change

Observations:

- top ranked products will not be influenced by the rating value changes.
- Overly inflating the rating value of a product with low original ratings may even hurt its next week's downloads.

An Example of "Internal Self-exciting"

The software in the File Sharing market named "BadBlue Personal Edition 2.7" was originally ranked as 47.



Observations

Product sales increase can be caused by

- External manipulations - rating value and volume change
- Internal consecutive self-exciting power

Step 2: Proposed Attack

Attack Design

In particular, we design the attack strategy by considering two aspects:

(1) how to determine the unfair rating values and volume

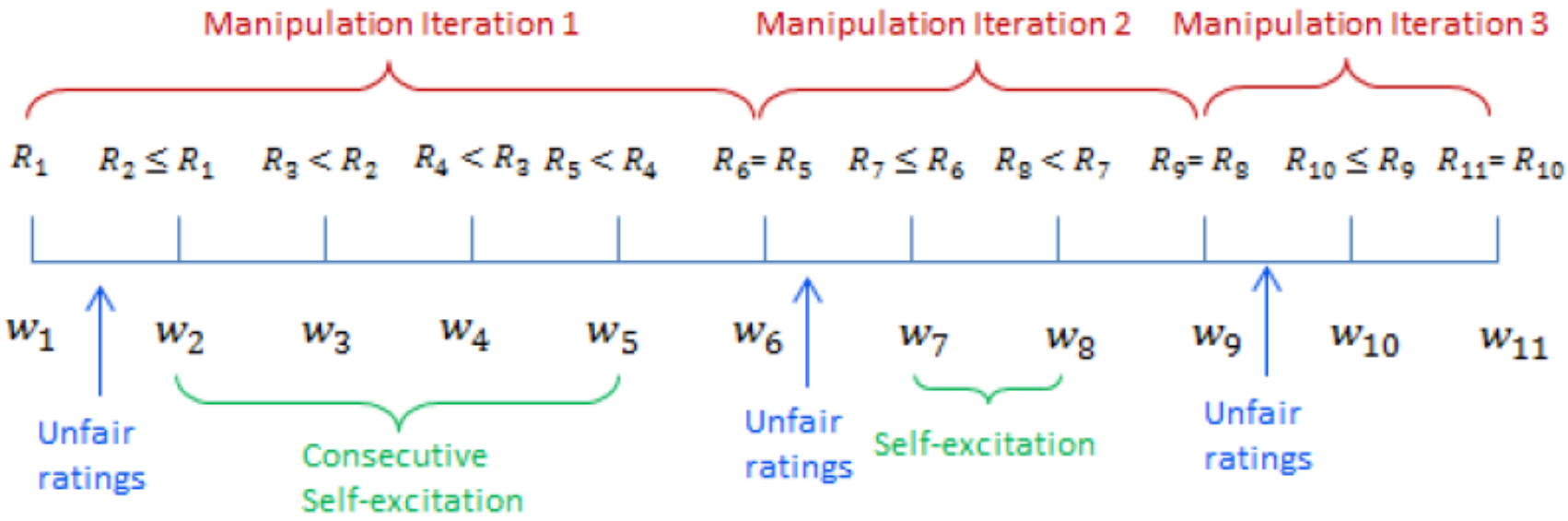
(2) how to take advantage of the market's internal self-excitation power

Proposed Attack – Rating Value & Volume

Hence, we propose a rating value $r_c(\alpha)$, calculated as $r_c(\alpha) = -\frac{\beta_2(\alpha)}{2\beta_3(\alpha)}$, as the **critical rating value**, which changes across different product quantiles.

	Feasible Attacks	Optimal Attack
Type I ($\bar{r} > r_c(\alpha)$)	$\bar{r}_{mal} \in (\bar{r}_o - \lambda,\ 5],\ (\lambda > 0)$ $v_{mal} > 0$	$\bar{r}_{mal} = 5$ $v_{mal} = N$
Type II ($\bar{r} = r_c(\alpha)$)	$\bar{r}_{mal} =$ any value $v_{mal} > 0$	$v_{mal} = N$
Type III ($\bar{r} < r_c(\alpha)$)	$\bar{r}_{mal} \in [1,\ \bar{r}_o - \lambda),\ (\lambda < 0)$ $v_{mal} > 0$	S1: $\bar{r}_{mal} = \bar{r}_o$ $v_{mal} = N$
		S2: $\bar{r}_{mal} = 1$ $v_{mal} = N$

Proposed Attack – Consecutive Market Self-excitation



Comparison to All-together Strategy

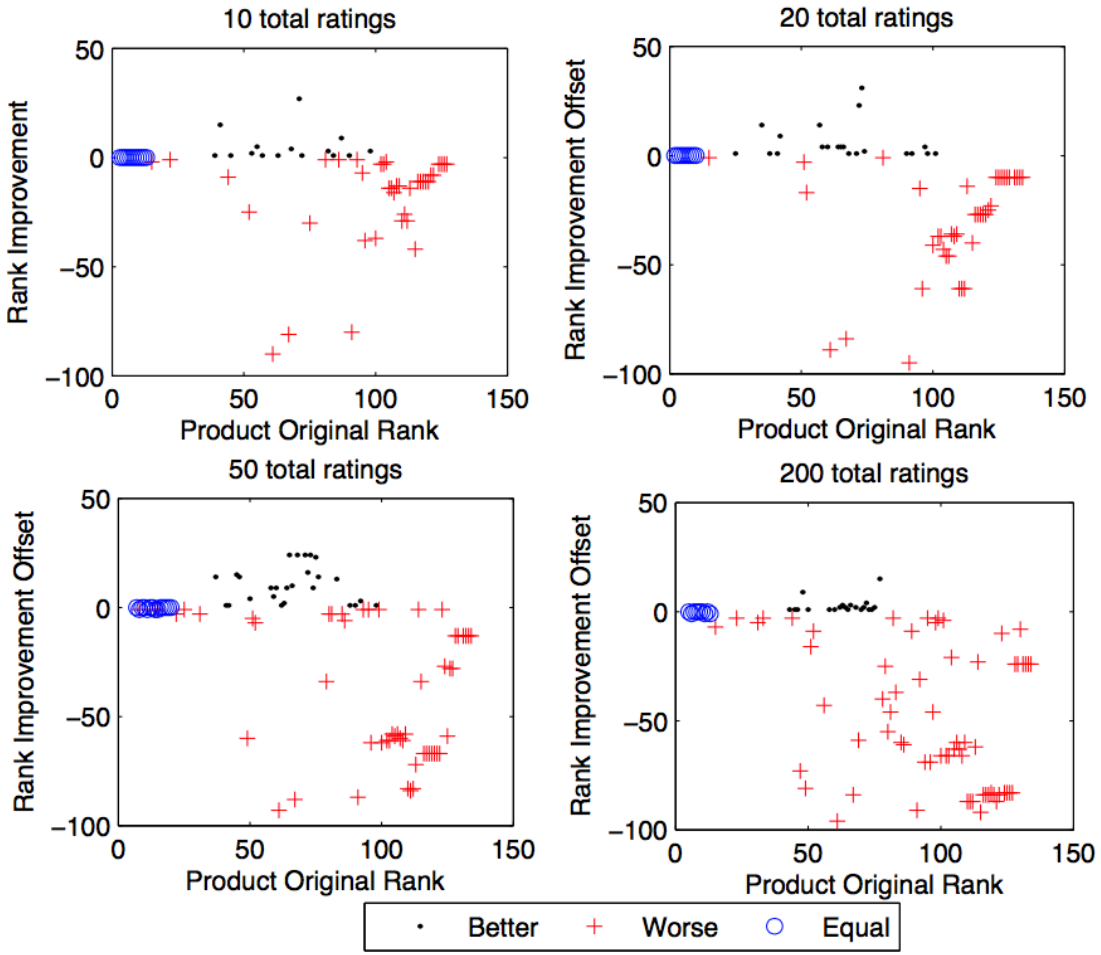


Fig. 11: Comparison between S_{iter} and S_{all}

Key Takeaway

- Customized attacks are more effective
- Market's internal self-excitation power can be utilized

Retrieving Hidden Friends: A Collusion Privacy Attack Against Online Friend Search Engine

Yuhong Liu, *Na Li, "Retrieving Hidden Friends: A Collusion Privacy Attack Against Online Friend Search Engine", IEEE Transactions on Information Forensics and Security, 14, no. 4 (2018): 833-847.*

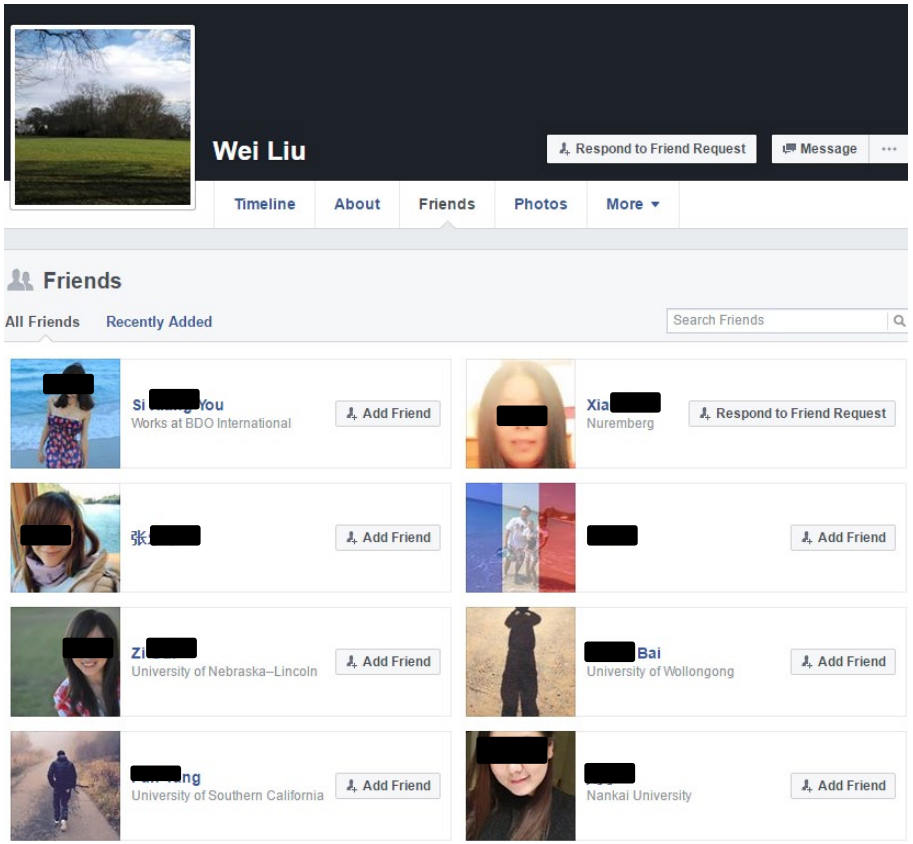
Facebook Privacy Scandal

The Facebook data privacy scandal 2017 centers around the collection of personally identifiable information of "*up to 87 million people*" by the political consulting and strategic communication firm Cambridge Analytica. That company—and others—were able to gain access to personal data of Facebook users.

Friend Search Engine

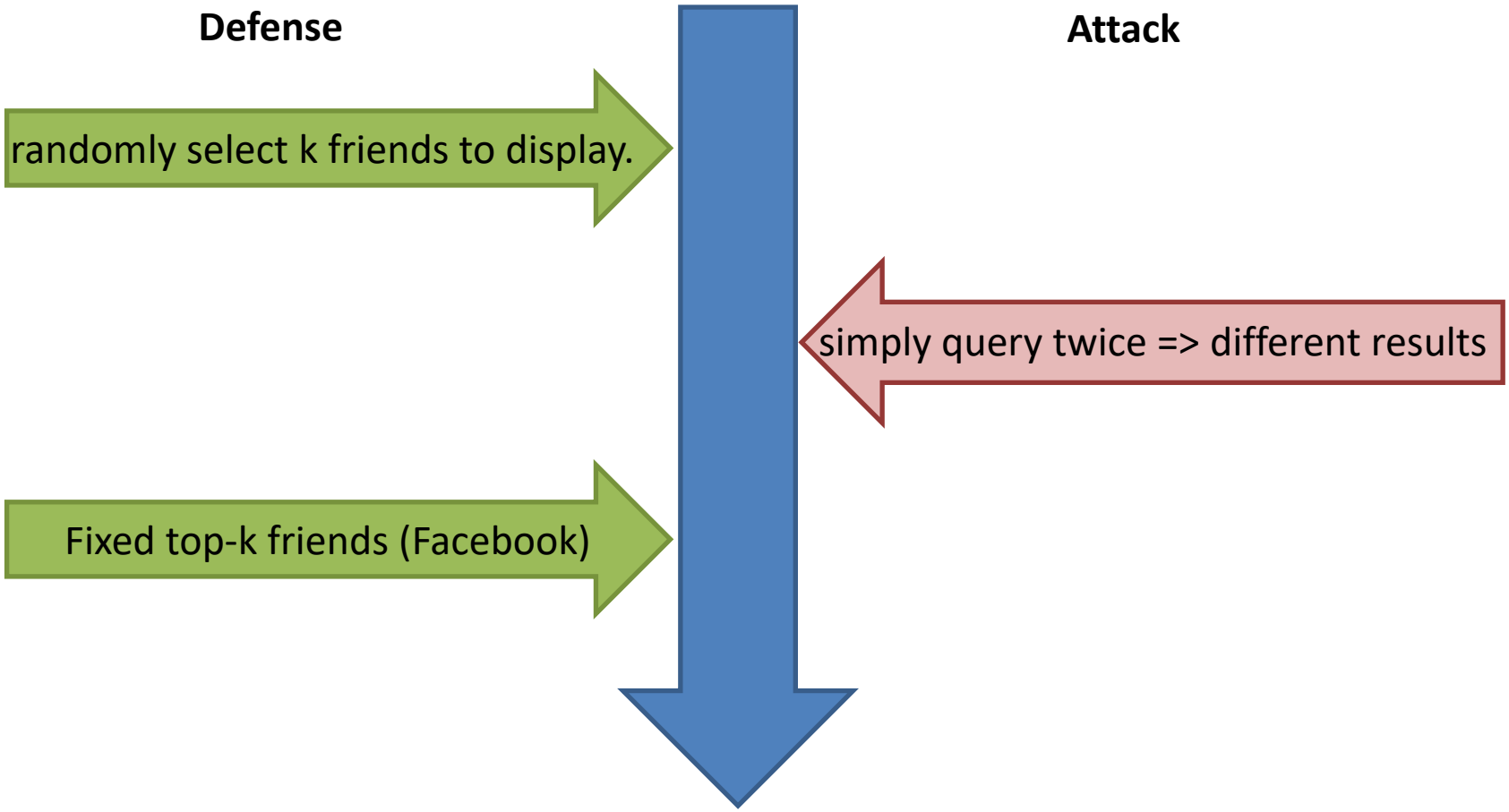
Various online social network applications are developed for people to interact with family, friends and colleagues.

Friend Search Engine



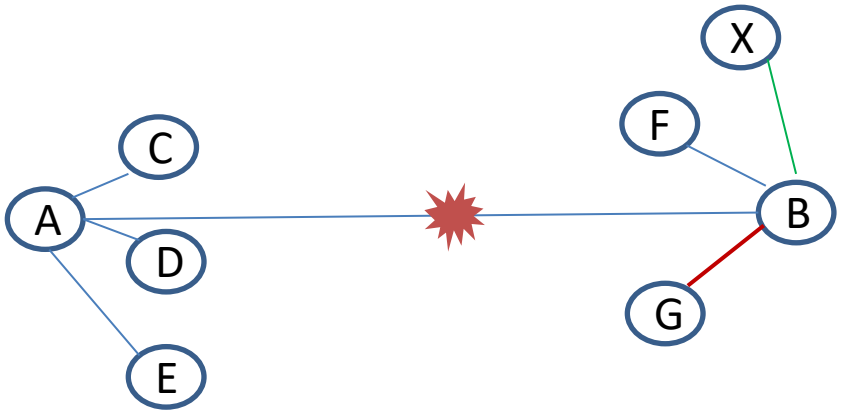
Control # of Friends To Release

- How about users do not want to share all of their friends?



Which Friend to Release?

Mutual Effect:



Defense: keep track of the released friends' privacy

It works effectively against any *individual* malicious attacker.

This Work: Collusion Attacks

Attack Goal: multiple malicious requestors coordinately query the system => mislead the system to leak additional friends.

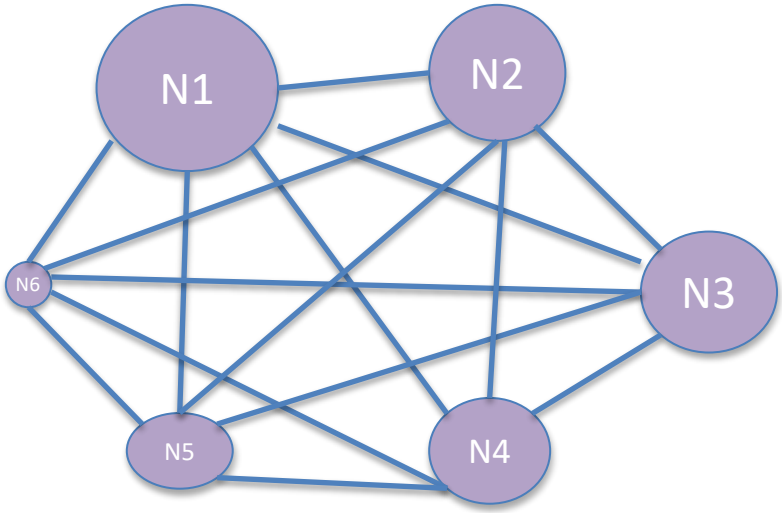
Step 1: A Toy Example – Attacks against a social clique

Attacks in a Social Clique (Dense Network)

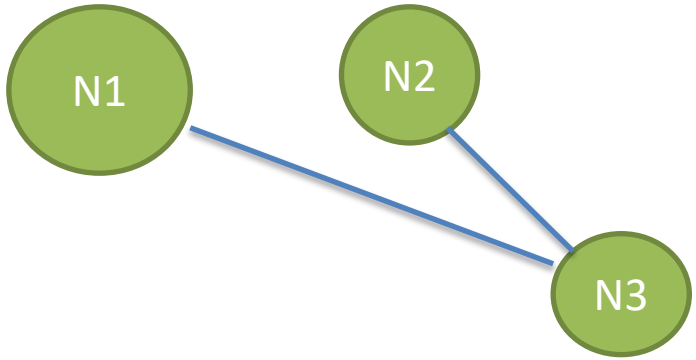
$$I_{N_1} > I_{N_2} > I_{N_3} > I_{N_4} > I_{N_5} > I_{N_6}$$

Assume:

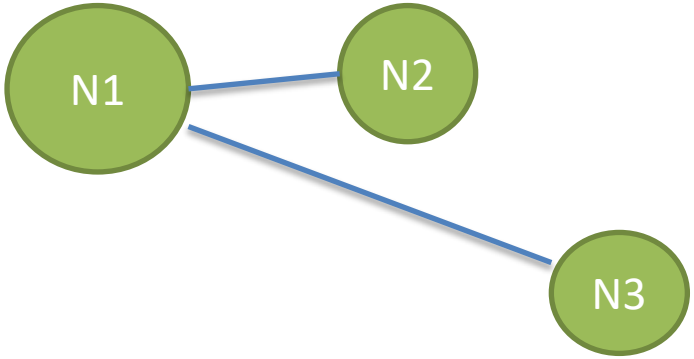
- N3 is victim node
- K = 2



Malicious requestor MR1



Malicious requestor MR2



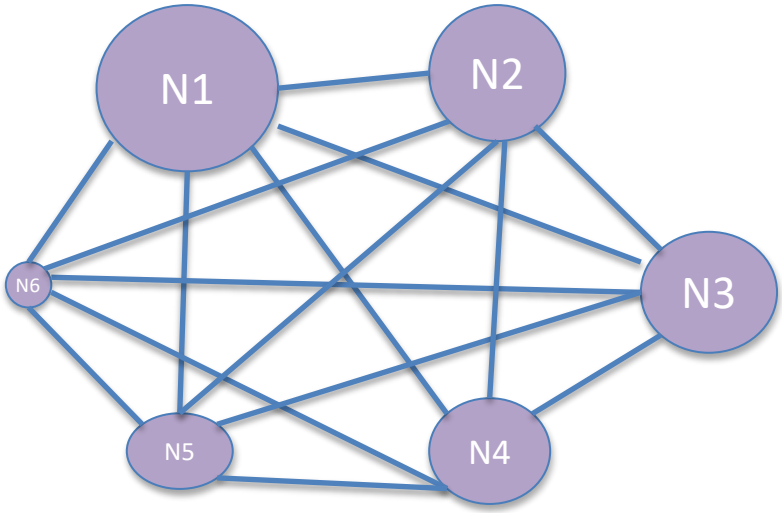
Violation fails !

Attacks in a Social Clique

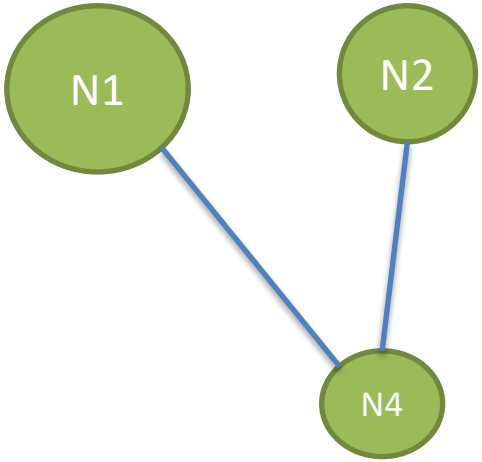
$$I_{N_1} > I_{N_2} > I_{N_3} > I_{N_4} > I_{N_5} > I_{N_6}$$

Assume:

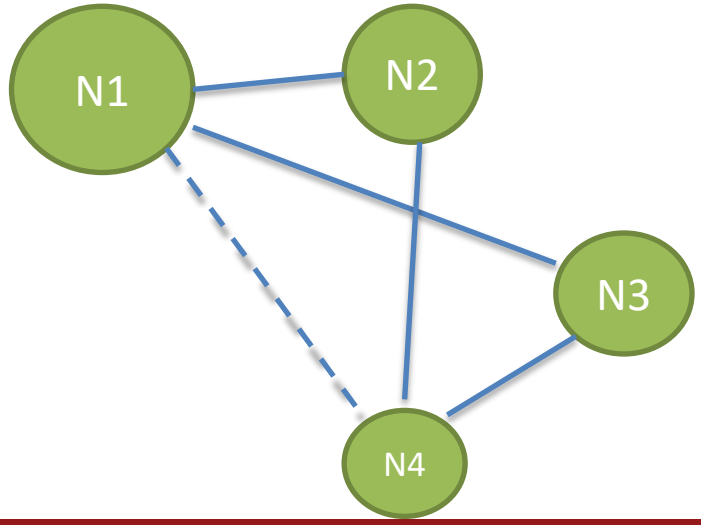
- N4 is victim node
- K = 2



Malicious requestor MR1



Malicious requestor MR2



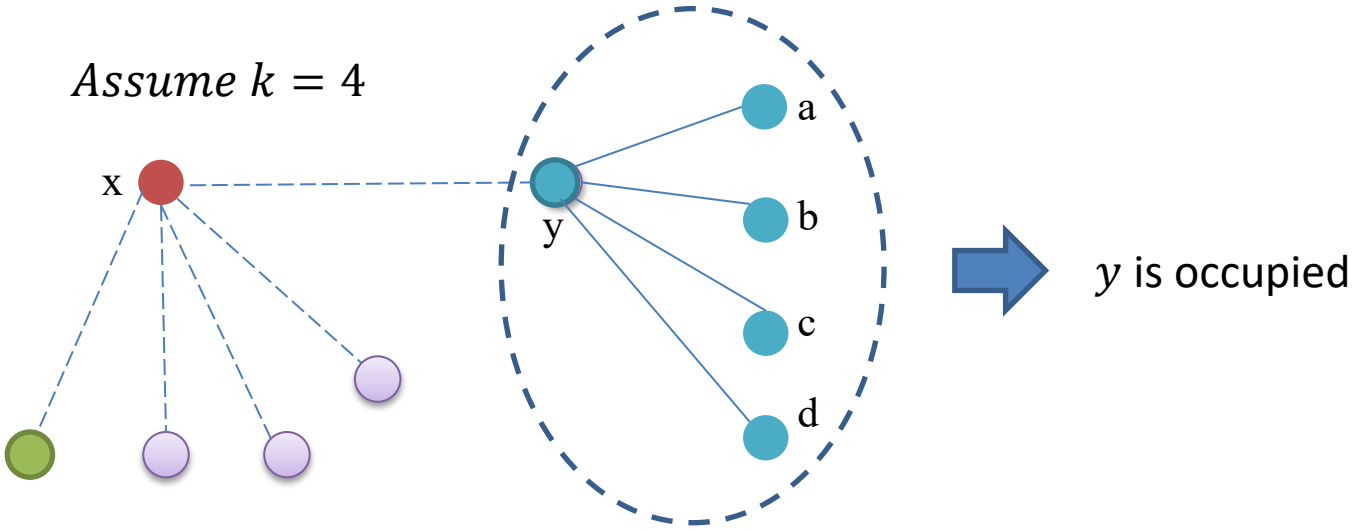
Violation succeeds!

Observations

In a social clique,

- Top k+1 nodes: cannot be directly violated
- Other nodes: can be violated by occupying at least one of its friends.

Inspiration – violating privacy through occupation



Node Types

Definition IV.2. Popular Node: A node that is on the top k influential friend list of **ALL** its top k influential friends.

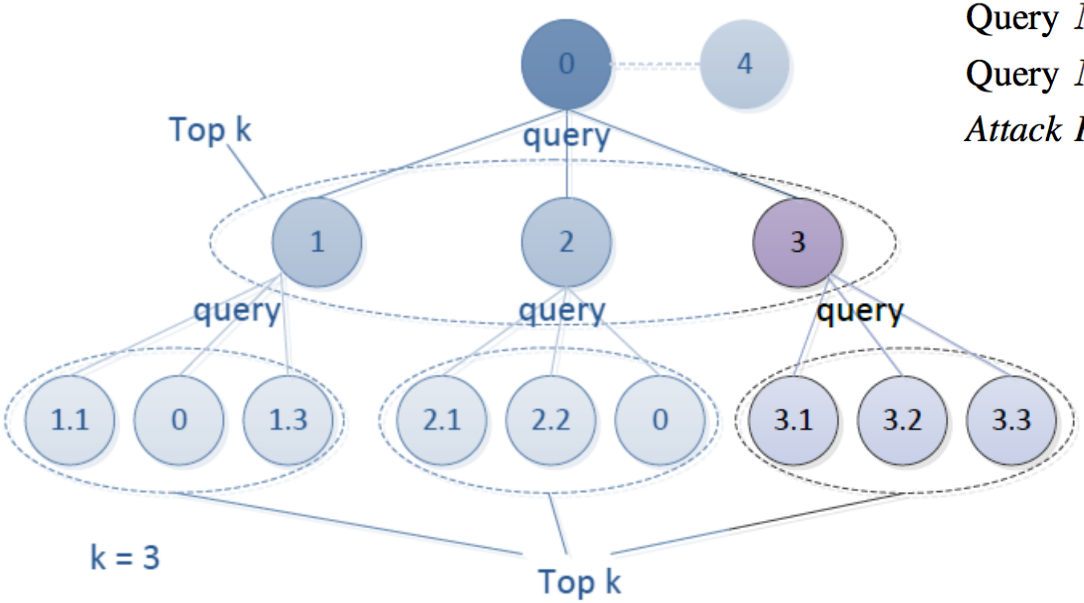
Definition IV.3. Non-Popular Node: A node that is **NOT** in the top k influential friend list for **AT LEAST ONE** of its top k friends.

Non-Popular Node: can always be violated through occupation.

Popular Node: may be violated by recursively occupying friends' friends.

Step 2: Attacks in General Networks

Attack against a Non-popular node in a General Network



$\mathbf{MR_1}$:

Query $N_0 ->$ retrieve $E_{(N_0,N_1)}, E_{(N_0,N_2)}, E_{(N_0,N_3)}$

$\mathbf{MR_2}$:

Query $N_1 ->$ retrieve $E_{(N_1,N_{1.1})}, E_{(N_1,N_0)}, E_{(N_1,N_{1.2})}$

$\mathbf{MR_3}$:

Query $N_2 ->$ retrieve $E_{(N_2,N_{2.1})}, E_{(N_2,N_{2.2})}, E_{(N_2,N_0)}$

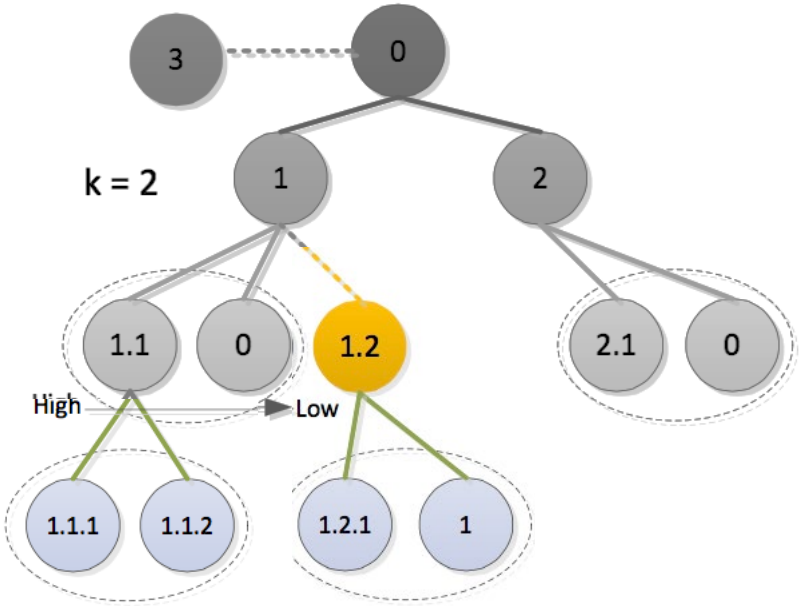
$\mathbf{MR_4}$:

Query $N_3 ->$ retrieve $E_{(N_3,N_{3.1})}, E_{(N_3,N_{3.2})}, E_{(N_3,N_{3.3})}$

Query $N_0 ->$ retrieve $E_{(N_0,N_1)}, E_{(N_0,N_2)}, E_{(N_0,N_4)}$

Attack Result: Succeed

Attacks against a popular node in a General Network



$\mathbf{MR_1}$:

Query $N_0 ->$ retrieve $E_{(N_0,N_1)}, E_{(N_0,N_2)}$

$\mathbf{MR_2}$:

Query $N_1 ->$ retrieve $E_{(N_1,N_{1.1})}, E_{(N_1,N_0)}$

$\mathbf{MR_3}$:

Query $N_2 ->$ retrieve $E_{(N_2,N_{2.1})}, E_{(N_2,N_0)}$

$\mathbf{MR_4}$:

Query $N_{1.1} ->$ retrieve $E_{(N_{1.1},N_{1.1.1})}, E_{(N_{1.1},N_{1.1.2})}$

Query $N_1 ->$ retrieve $E_{(N_1,N_{1.2})}, E_{(N_1,N_0)}$

$\mathbf{MR_5}$:

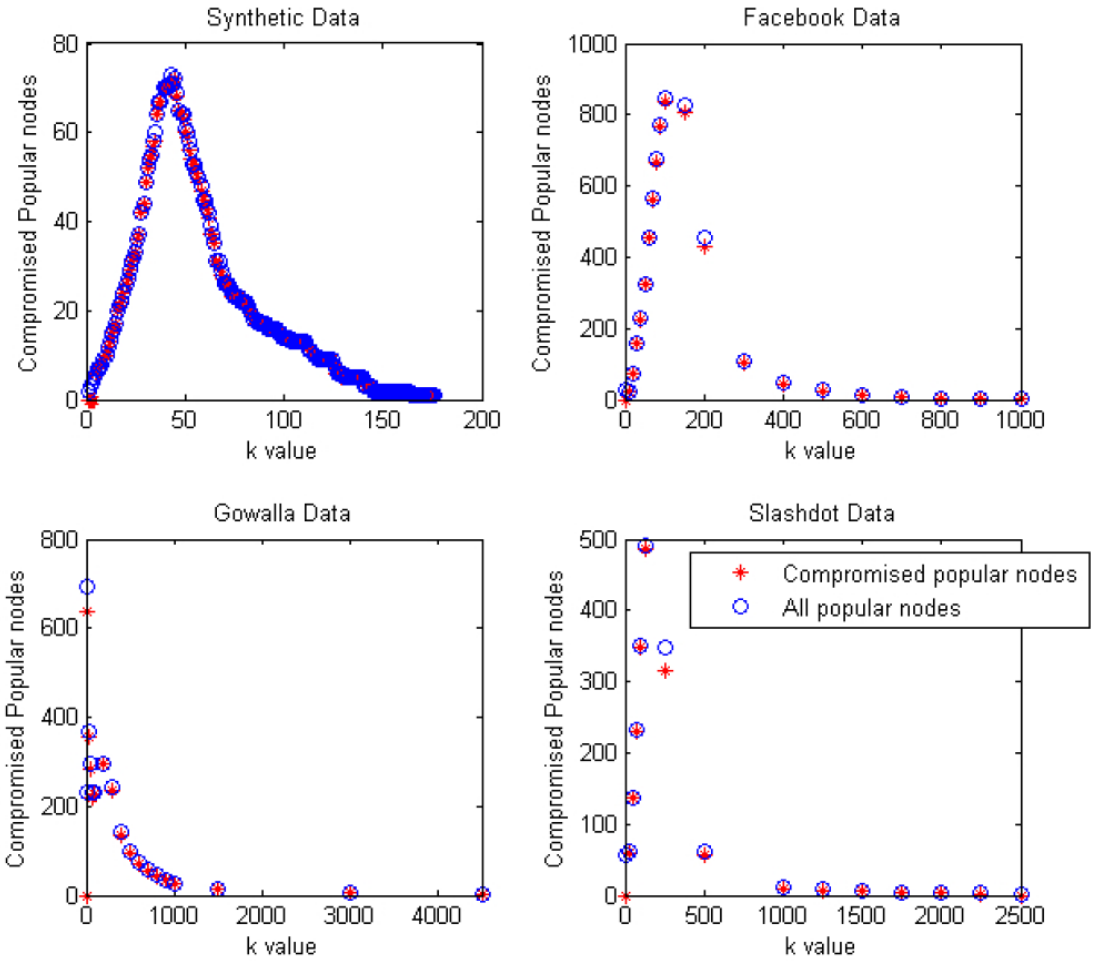
Query $N_{1.2} ->$ retrieve $E_{(N_{1.2},N_{1.2.1})}, E_{(N_{1.2},N_1)}$

Query $N_1 ->$ retrieve $E_{(N_1,N_{1.2})}, E_{(N_1,N_{1.1})}$

Query $N_0 ->$ retrieve $E_{(N_0,N_2)}, E_{(N_0,N_3)}$

Attack Result: Succeed

Attack Effectiveness



Summary

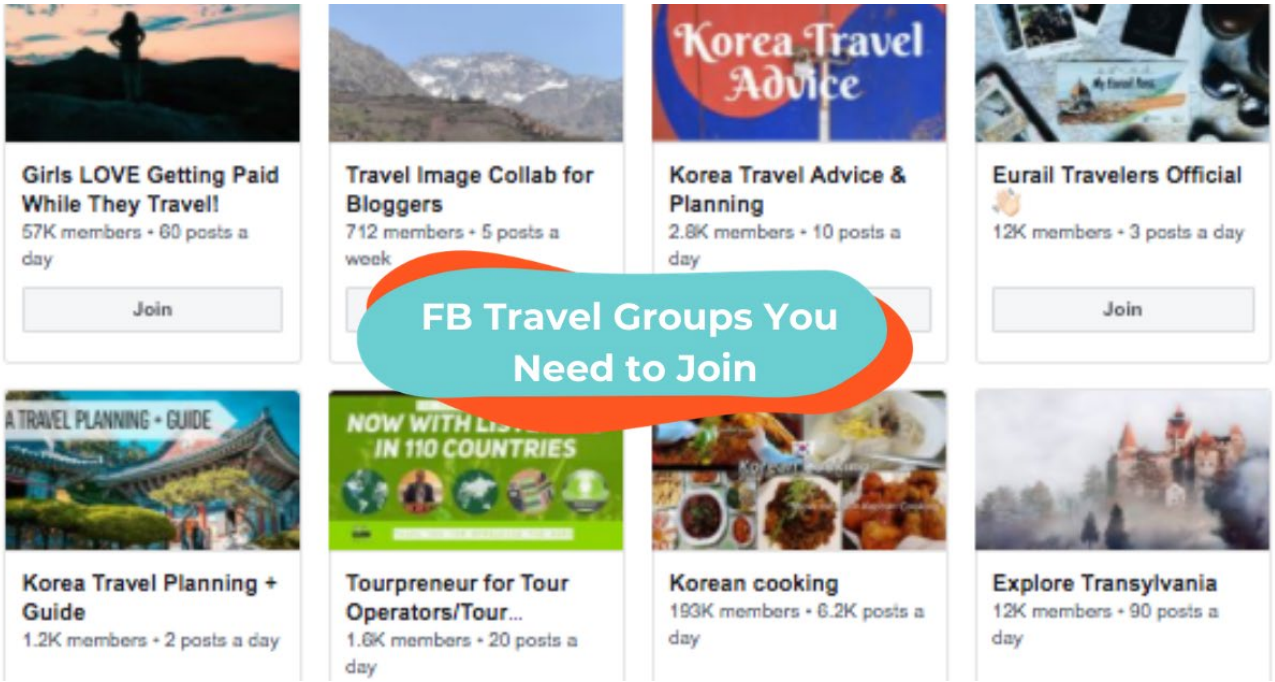
Privacy protection is challenging.

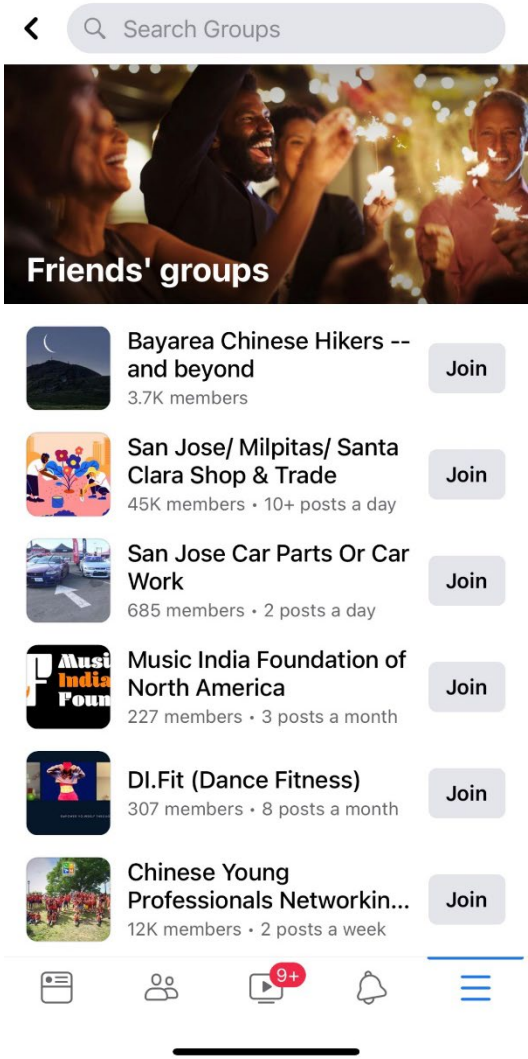
Your privacy may be in the hands of your friends

Correlating Diversity and Resistance to Misinformation in Social Media Groups

I Chang, Orion Sun, Jasper Ahn, Yuhong Liu, "Correlating Diversity and Resistance to Misinformation in Social Media Groups", IEEE Global Humanitarian Technology Conference (GHTC), 2021

Social Network Groups





Hypothesis

Greater diversity in social media groups correlates with *greater resistance* to misinformation

Diversity: the measure of *how much variety* is present in the characteristics of a group. (not include race/age, but in terms of vocabulary range, content engagement, interactions with other social groups)

Misinformation resistance: how likely a social media group will *internalize and interact* with sources of misinformation once being exposed

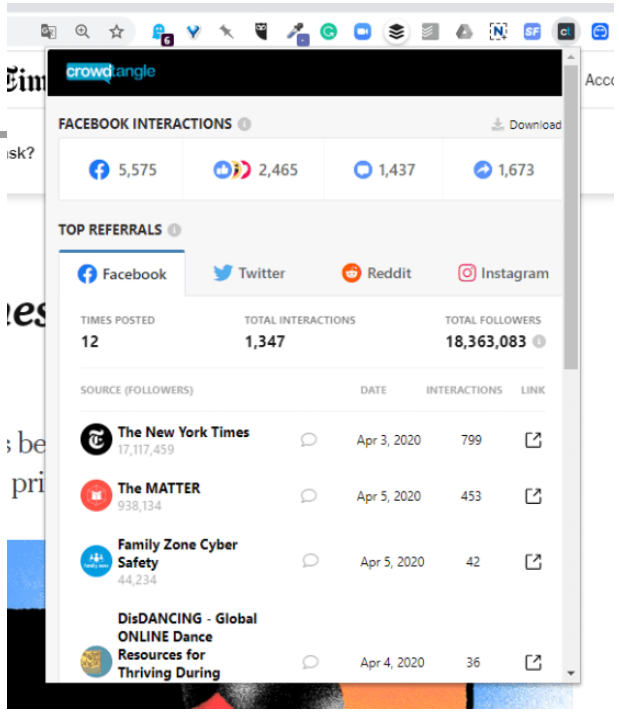
Ground Truth
Media Bias / Fact Check

Factual Reporting
Very High
High
Mostly Factual
Mixed
LOW
Very Low

3000 pairs of domain names and factual accuracy scores

Data Collection

FaceBook CrowdTangle



source	score
1	Low
2	High
…	…
3000	Very low

Fact Check

source	Post	Groups
1	Top 1	G1, G2…
…	…	…
1	Top 1000	G2, G9 …
2	Top 1	G5, G10…
…	…	…
2	Top 1000	G1, G20 …
…		…

FaceBook

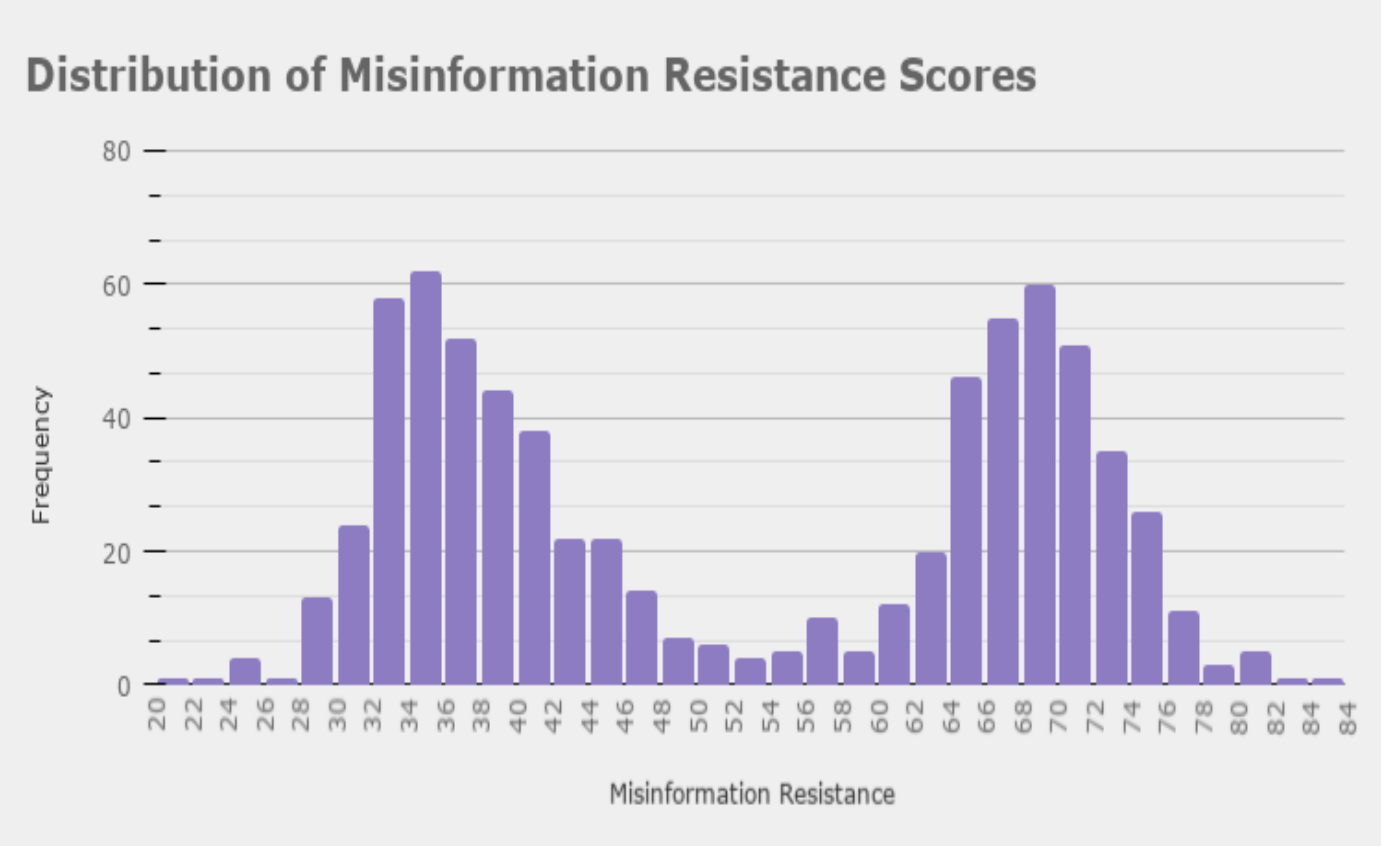
Groups >=50 sources

719 Facebook Groups

Each Group → Top 100 posts → Diversity Analysis

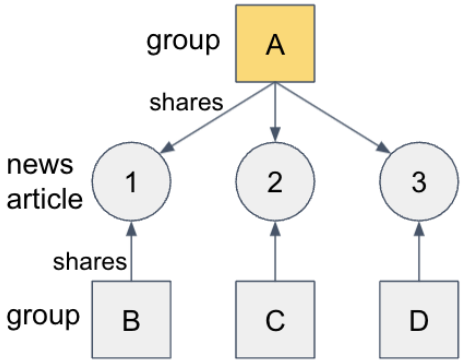
Group Resistance

- Average factual accuracy score of associated sources

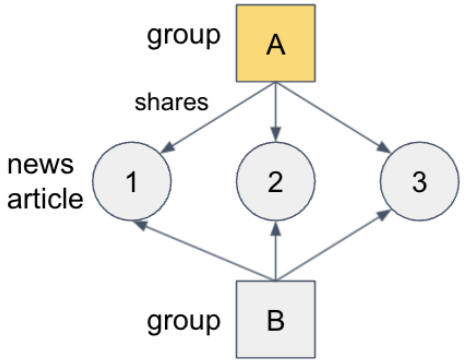


Proposed Entropy based Diversity Metrics

- Word entropy
  - Base form of words
- Post type entropy
  - Link posts, image posts, video posts, etc
- Average/Total reaction entropy
  - Like, Love, Haha, Angry, etc.
- Top level domain/domain entropy
  - .com, .org, .int, .edu, etc
  - cnbc.com
- Mutual Network Analysis (MNA) Entropy score



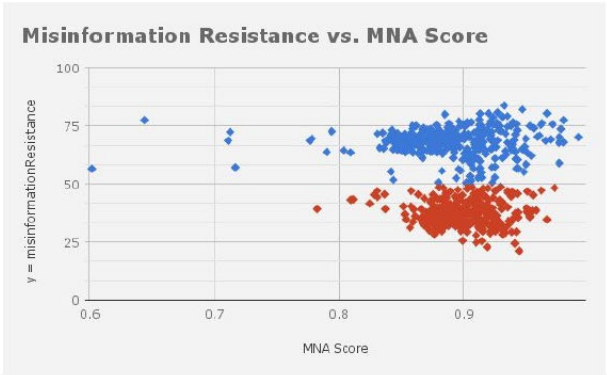
High MNA Score

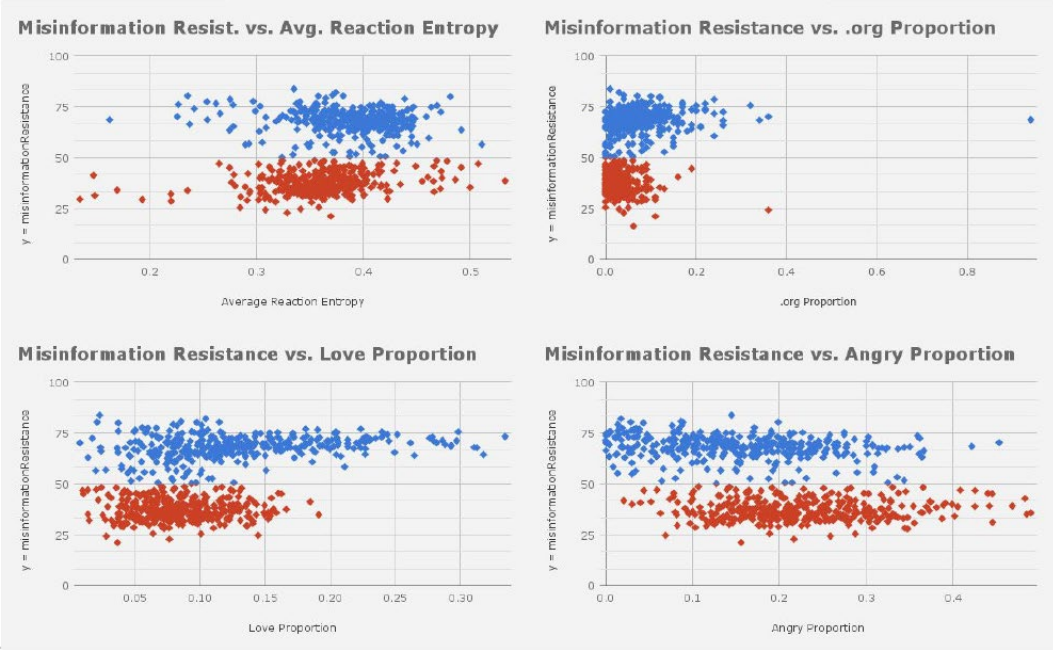


Low MNA Score

Key Metrics Correlated with Misinformation Vulnerability

- Diversity Metrics
  - MNA Score
  - Average Reaction Entropy
- General Metrics
  - Top Level Domains
    - .org
  - Reactions
    - Love and Angry





Key Takeaways

- The diversity of social groups is related to their resistance to misinformation.

Conclusion

- Online social networks are facing various security challenges
- The fast evolving arm race between attacks and defenses requires continuous future researches.

Thank You !

Questions ?

Yuhong Liu
Associate Professor
Computer Science and Engineering Department,
Santa Clara University
500 El Camino Real Santa Clara, CA, 95053

Email: yhliu@scu.edu
Tel: 408-551-3513

Students/Researchers/Alumni at Our Lab