Identifying and Manipulating the Aging of Social Networks M2 Report

Chenxi Li
Carnegie Mellon University
Electrical and Computer Engineering
chenxili@andrew.cmu.edu

Pengying Wang
Carnegie Mellon University
Electrical and Computer Engineering
pengyinw@andrew.cmu.edu

Abstract

Social media like Reddit plays a crucial role in citizens' participation in politics. Researching on potential ways of attack and counter-attack of the social networks has significance in understanding how "trolls"[3] try to guide the public opinions and what are the potential solutions to mitigate such influence. In this paper, we first collected four months of data from June to September from the Hong Kong subreddit to analyze the aging social network patterns corresponding to the beginning to the end of 2019 Hong Kong's protests. Then we simulate the Sybil attack[1] directly to the original network. Finally, we deploy the counter-attack method by applying the partition algorithm[6] to the attacked network. Based on our findings, the behaviors of trolls in social networks are further understood and the effect of the partition strategy to resist such influence is analyzed.

1. Introduction and Motivation

Recently, the growth of the social network has influenced citizens' life in various aspects. For instance, Hong Kong protest regarded as highly engaged politics activities shortly started in June, and the trend for the protest goes even hotter in mid-September. People use the social network to present politic opinions. As the popularity of online social networks has increased, the risk of attacking has also grown as well. Malicious activity issues in online social networks(such as Sybil attacks and malicious use on fake identities)[5] can severely affect the social activities when users participate in an online discussion."Troll"[3] is a virtual account presenting offensive and opposite opinions to rigging popularity to get a rise out of network users. It performs various types of illegal online actions and breaks trust in online association[5]. Reddit[4] is a social entertainment and academic forum powered by registered users' contents. Compared with Facebook and Twitter, it focuses more on the connection of posts and allows decentralized users to vote and discuss on topics or comments in subreddits. In this article, the main subject is identifying different life stages through building periodically social networks on subreddit(Hong Kong). We simulate Sybil attack to alter the social network scheme in subreddit of Hong Kong and propose a new method to defense or counter-attack the troll using partition strategies. By analyzing the network properties when altering the structure of the network, we try to figure out how the network scheme reacted to the Sybil attack and propose an effective method to defense the impact of the attack on the real online social network. The paper is organized as follows. In section 2, the study for previous work and background is given. In Section 3, we present a detailed description of approaches to achieving a Sybil attack and defense simulation. Section 4 proposes the setup of the experiment and results. Finally, we come up with a conclusion and discuss a short-term plan in the next milestone.

2. Previous Work

Sybil attack is first proposed by John R. Douceur, which is used to describe an attack that aims to gain as much influence as it can to control a considerable fraction in a peer-to-peer network [1]. "Trolling" is one such attack[5]. In social networks, a troll may create multiple fake accounts to specifically give a large number of supportive comments only on the posts or opinions published by this troll, which would finally end up as a huge community and continuously attract and influence other real identities. The Russian "troll farm" interference in the 2016 United States elections on Facebook is one of the typical cases of Sybil attack in the social network [2]. Lots of potential Sybil defense mechanisms specifically in the social network such as "SybilLimit"[7] and "Gatekeeper" [8] are done in recent years. The main idea is to detect the Sybil attack by community detection (not fast-mixing) and counter the attack by decentralizing the network through topological approaches to control the admission of the suspected nodes [7][8]. Inspired by the previous work, we first simulate the Sybil at-

tack to the original network and apply community detection to prove the success of attack deployment. Then the random partition algorithm is used to counter this attack.

3. Approach

There are three stages in presenting Sybil attack and defense. The first step is building the origin network without troll attack. The next step is simulating a Sybil troll to attack the origin network. In this step, an abnormal network schema is proposed. The final stage is utilizing partition algorithms to alter network pattern and trying to recover the network after effective defense.

3.1 Network Construction

To study the connections between different users' reactions with special topics, we build two kinds of weekly-changing networks: Karma network and User network. For the Karma network, the nodes are Karma points and the edges are directed from post to its comments or comment to its sub-comments. For the User network, the nodes are user id in Reddit. We connect the same users who comment on different topics or comments. Besides, we connect nodes who comments on the same parent nodes. Because it shows the trend that users would comment on more related topics. It can be presented in network properties. Each network represents new topics and user comments within seven days. Since Hong Kong protests become severer in July, we assume that the number of nodes and edges would on the increasing trend overall. The clustering coefficient would increase while the average path length would decrease after manipulation.

3.2 Attack simulation

To simulate the Sybil attack, we first create 100 fake accounts into the original user network. As these fake accounts are assumed to specifically give supportive comments on the troll's posts, these fake accounts should be fully connected in the original user network, which formalizes a community. In reality, as these fake accounts all are created by the troll under very low cost, these accounts usually have very low karma points like 1 point. Considering that some other honest users may also be attracted to comment on the troll's posts, we filtered the honest accounts from the karma network who only comment on the "1 karma point" posts out as they are more likely to comment on the troll's posts. We then rank the frequency of their comment times and assume that the user who commented equal or more than 10 times on the "1 karma point" posts would also comment on the troll's posts. These honest users would also be attracted to the troll's community. After deploying the attack, we perform community detection to count the number of users in the largest community to see if it is much larger than the number of users in the largest community in the original network. If so, the attack is successfully deployed.

3.3 Defense strategies

When we propose community detection and found an abnormal network scheme, for instance, the number of one community members is much larger than other communities. Although an attacker can create an arbitrary community[6], the identities are relatively isolated to other honest accounts. We use partition strategies as defense scheme. Based on connecting all users in the same layer, we shuffle each layer and make a partition to decrease the risk of high rank of communities. We set the cutoff as a threshold for each community. After implementing the partition algorithm, we try to recover the network pattern to origin status. We detect the community again and calculate the community with the largest number of members and compare it with original data properties. If the number of largest community members can be significantly decreased by defense manipulation, we could prove the counter-attack is successful.

4. Experimental Setup and Results

Based on the statistic data provided by Reddit, the number of the posts and comments per day increased sharply when the Hong Kong protests broke out. And the increasing trend still keeps going after September.

4.1 Karma Network

As Karma points can only be earned by actively posting a high-standard topic or commends, we have reasons to believe that the user with high Karma points would have a higher possibility to be commented and supported by other people. Here we set up 4 directed karma networks which represent 4 months of data from June to September to observe the network pattern over time. The node in this network is Karma points. If a user commented on a topic or a comment, then a directed edge between this user's comment and the target topic/comment would be added. Based on our assumption, people should tend to comment on people with higher Karma. And this assumption is proved to some extent in figure 2. As mentioned in the previous section, the number of posts and comments is increasing from June to September. The corresponding network property should be the continuous growth in the average path length in each network. Since the path length here represents how "deep" a topic is commented. So roughly speaking, the "hotter"

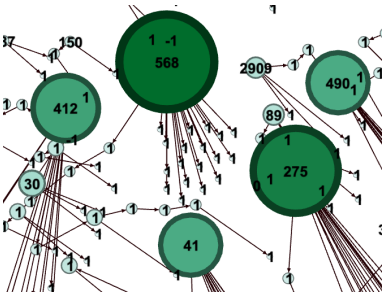


Figure 1: Portion of Karma Network

Table 1: Average Path Length for Four Month Karma Networks

	June	July	August	September
Avg. Path Length	2.112	2.246	2.519	2.501

a topic is, the "deeper" should it be commented. And this property is shown in table 1.

4.2 User Network

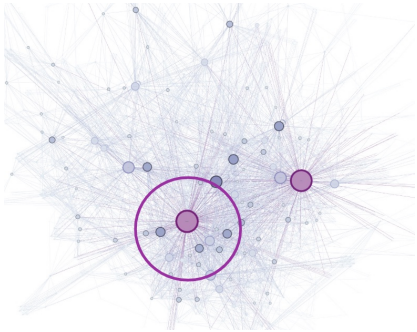


Figure 2: Sample User Network in One Week

We set up 16 independent undirected user networks that present 4 months from June 20, 2019, to Oct 7, 2019(regarded as protest summit) to observe the behaviors of the network through clustering coefficient and average path length change. As we assumed before, the clustering coefficient increased after the troll attack while the average path length decreased.

4.3 Sybil Attack Simulation

As mentioned in the approach section, we created 100 fake accounts to form the troll community. And honest accounts that only comment on the "1 karma point" posts/comments with the highest frequency (higher than 10 times) are filtered out as they are most likely to be attracted

Table 2: Network properties after manipulation

	origin	attack once	attack twice
Cluster coefficient	0.767	0.763	0.785
Avg. Path Length	3.308	3.385	3.495

by the troll posts. The filtered accounts are shown as follow:

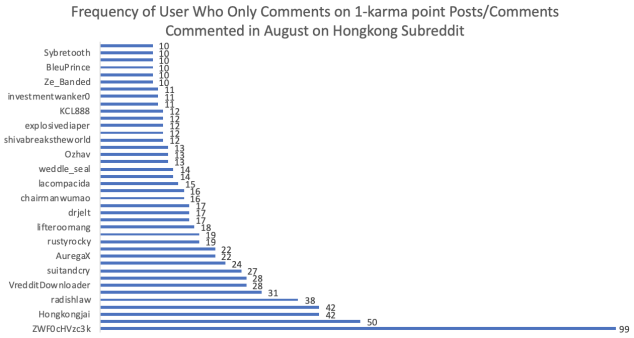


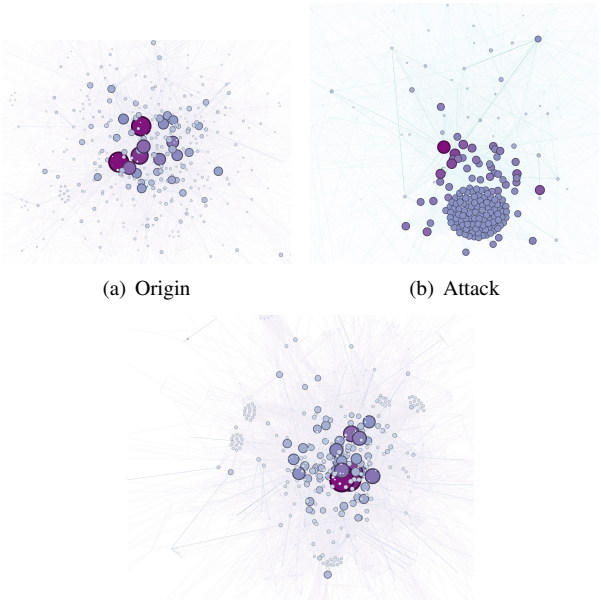
Figure 3: Honest Accounts that are Most Likely to be Attracted by Troll Community

We select the network constructed by August subreddit data to deploy the attack and counter attack. The original August user network is shown as follow:

After inserting the troll and fake accounts into the original network, the attacked user network is shown in figure 5. It is clear to see that a large community exists in the attacked user network. Comparing to the original user network, the number of users in the largest community changes from xxx to xxx, which means the troll forms a much larger community that has much greater influence in the whole network. A huge centralized community is the sign of the existence of the troll and Sybil attack, which means the attack is successfully deployed.

4.4 Partition Defense

Figure 4 shows the process of network pattern when the origin network receives attack and recover to the origin stage after manipulating partition defense strategies. We can conclude from the figure that there is a large fake community when the network suffers a Sybil attack. In Table 3, the origin network's largest community has 117 nodes while the attacked network with Sybil fake community has 722 nodes. The average length decreased as well and the clustering coefficient increased, which proves that the network has become denser after the attack. In the final stage, the size of the largest community has significantly decreased and close to the original network pattern. The average path length and cluster coefficient have returned to nearly the origin value.



(a) Origin (b) Attack

(c) Defense

Figure 4: Network graph in three stages

It has proved that the partition strategy can effectively defend against cyberattacks.

Table 3: Network properties change in three stages

	origin	attack	defense
largest community	117	722	156
average path length	3.331	3.095	3.279
cluster coefficient	0.729	0.776	0.755

5. Conclusion and Short-term Plans

Based on the karma and user network constructed by Reddit data, we could see that the network properties like average path length and cluster coefficient could reflect the trend and popularity of social issues. Simulating network troll to do Sybil attack to the original network shows the unique pattern of the attacked network like the existence of the huge community. The result of the deployment of the partition algorithm strategy to the attacked network proves such a topological method is effective in countering the Sybil attack.

In this stage, the attack and counter-attack only perform on the one-month network. In the next stage, we would like to introduce the time parameter to analyze the network topology change over time. Besides, we propose a weighted network in M2. We plan to

In the report, Chenxi contributed on Abstract, 2, 3.1, 3.2, 4.1, 4.3, 5 and Pengying contributed on 1, 3.1, 3.3, 4.2, 4.4, 5. For M3, Chenxi and Pengying will try to deploy attack and counter-attack over months on original monthly networks to get preliminary results before Nov. 7st and finish the analysis before Nov.15.

References

[1] J. R. Douceur et al. (2002) *The Sybil attack*, in Peer-to-Peer Systems, P. Druschel, Ed. Heidelberg, Germany: Springer, pp. 251–260.

[2] Doug Thompson *Everything you wanted to know about trolls but were afraid to ask* ShareAmerica. U.S. State Dept. Bureau of International Information Programs. 4 November 2015

[3] *Trolls and Their Impact on Social Media* https://unlcms.unl.edu/engineering/james-hanson/trolls-and-their-impact-social-media

[4] Giannis Haralabopoulos, Ioannis Anagnostopoulos (2015) *Lifespan and propagation of information in Online Social Networks: a Case Study* Journal of Network and Computer Applications,Volume 56, Pages 88-100

[5] M. Al-Qurishi, M. Al-Rakhami, A. Alamri, M. Alrubaian, S. M. M. Rahman and M. S. Hossain (2017) *Sybil Defense Techniques in Online Social Networks: A Survey* in IEEE Access, vol. 5, pp. 1200-1219.

[6] Viswanath, Bimal and Post, Ansley and Gummadi, Krishna P. and Mislove, Alan (2010) *An Analysis of Social Network-Based Sybil Defenses*, SIGCOMM Comput. Commun. vol. 40, pp. 363-374

[7] Haifeng Yu, Phillip B Gibbons, Michael Kaminsky, and Feng Xiao. (2008) *Sybillimit: A near-optimal social network defense against sybil attacks*, In IEEE Symposium on Security and Privacy (SP), pages 3–17,

[8] Nguyen Tran, Jinyang Li, Lakshminarayanan Subramanian, and Sherman SM Chow. (2011) *Optimal sybil-resilient node admission control*, In INFOCOM, pages 3218–3226. IEEE.