Network Destroy-Repair Game - Milestone 1

Runqi (Tony) Huang
Carnegie Mellon University
AndrewID: runqih
runqih@andrew.cmu.edu

Vikas Kashyap
Carnegie Mellon University
AndrewID: vhuvinah
vhuvinah@andrew.cmu.edu

***Abstract*—Robustness is an important functionality of real-world networks, as it provides the resilience to guard against external attacks. By designing different attack and defense algorithms, this project will analyze four real-world networks where resilience to external threats is essential to the network's users and operations. This report will present the motivation, overall approach and preliminary results for this project.**

I. Introduction and motivation

Real-world interactions and relations can often be formulated in terms of networks and graphs, and these networks can significantly influence people's daily lives. In this immensely complicated world, perfection is idealized but never achieved, and seldom is a network designed without flaw. Critical failures with severe consequences, sometimes with human lives at stake, can occur without a proper understanding of the underlying mechanism of the networks [1]. One important feature of network centers around the network's robustness, which measures the network's resilience and ability to withstand external attacks.

Given the important role of network's robustness, this project will investigate network's behavior under different circumstances, where safety is integral to the performance, operation and user experience of the networks. We will leverage real-world data sets that are highly relatable and design a variety of different attack and defense algorithms to examine the robustness of these networks. Through a careful selection of data sets and experiments, we study different network attack/defense scenarios and contribute to a better understanding of network robustness.

II. Previous work

Recent years have witnessed the rise of deep learning and deep learning on graph structures has also shown promising results. Deep graph networks are similarly susceptible to adversarial attack, and one method adopts reinforcement learning that learns to attack the network based on long-term rewards [4] [5]. Due to a lack of reward mechanism, this project will adopt an alternative heuristic-based approach. However, it can seen that by introducing the proper reward mechanisms, networks, including ones we will study, may potentially strengthen themselves through a reinforcement learning approach.

There are various metrics to measure network robustness [6] [7]. One of the metrics is the Molloy-Reed criterion [1] [3], which connects network integrity with its degree k and average degree $\langle k \rangle$. Specifically, the criterion states that the network has a giant component if $\langle k^2 \rangle / \langle k \rangle > 2$.

III. Approach and steps

We have divided the approach of this project into the following steps. First, we will select four networks from a reliable data source. Second, we will employ the concepts from lectures to develop a basic understanding of the networks. Third, we will choose a standard evaluation metric and then implement attack/defense algorithms on these networks. Lastly, based on our experimental result, we will hypothesize on the underlying reasons that cause the observed resilience/fragility.

A. Selecting networks and data sets

We will use the Stanford SNAP data set [2] for this project, because SNAP contains a variety of interesting network scenarios with a large number of nodes and edges, which can serve as a good representation of their real-world counterparts. We have selected four networks for the following reasons:

- Social network: A criminal can disguise themselves and connect with people on social network such as Twitter. These unfriendly connections can be fraudulent and lead to financial loss.
- P2P network: Peer-to-peer network is vulnerable to attacks. Malware can be transmitted to users and useless data (poison) can also be injected into the system.
- Autonomous system: Security poses a fundamental challenge to the internet. In some cases, national security can be at stake without a strong network integrity.
- Email network: Spams can become an annoying headache and detrimental to user experience. Phishing emails also poses security risk and exposure of personal information.

B. Understanding basic properties of the selected networks

In this step, we will analyze a set of attributes of the network, including the number of nodes, number of edges, average degree, diameter, and average and global clustering coefficient. These concepts will provide the basic structural information on the network.

C. Setting evaluation metrics and implementing algorithms

In this step, we will survey the appropriate metrics to measure the robustness of networks. The first metric that we will adopt is the Molloy-Reed criterion [1] [3], as discussed in the previous section.

We will then design an attack algorithm that removes edges in the network and the corresponding defense algorithm. Initially, we will design an algorithm that randomly selects the edges for removal and repair. Subsequently, we will improve our algorithms by incorporating heuristics on the structural information of the network so that the algorithm can more intelligently select edges that are more vulnerable.

In implementing the algorithm, we are assuming that the attacker and defender have complete access and knowledge to the networks. In other words, the actions of the attacker and defender can be directed to all the edges in the network.

D. Formulating hypothesis on network robustness

In the final step, we will formulate hypothesis on the observed experimental results. We will attempt to answers questions on why certain networks are more resilient to attack than others are and what are the countermeasures that can potentially strengthen the network. This step will mostly consist of qualitative analysis.

IV. Preliminary Results

For this milestone, we have completed the first and second steps outlined above. The basic properties of the four networks are summarized in Table I, and we have also visualized social network in Figure 1 (due to the limited space, we are only showing the visualization for the social network).

TABLE I
FUNDAMENTAL PROPERTIES OF THE FOUR NETWORKS

	Social	P2P	AS	Email
Is is Directed?	Yes	Yes	No	No
Number of nodes	81,306	62,586	10,900	36,692
Number of edges	1,768,149	147,892	31,180	183,831
Avg degree	43	4.73	5.72	10.02
Diameter	7	11	9	11
Avg clustering coef	0.40	0.0027	0.501	0.716
Global clustering coef	0.11	0.0039	0.0386	0.0213

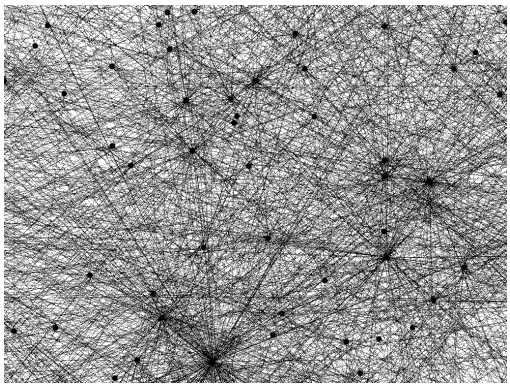

Fig. 1. Visualization of the social network data set.

As we can see from the table and figures above, the selected networks have a high number of nodes and edges, with one network having 1.77 million edges. As a result, we believe that these networks have sufficiently large size that is representative of real-world networks. In addition, as we can see from the degree distribution plot in Figure 2 (due to the limited space, we are only showing the degree distribution for the social network), the network has few individuals with higher degree, and common sense suggests that these may belong to public figure account. Therefore, we conclude that the preliminary results make sense in terms of their accuracy in representing a real-world scenario.

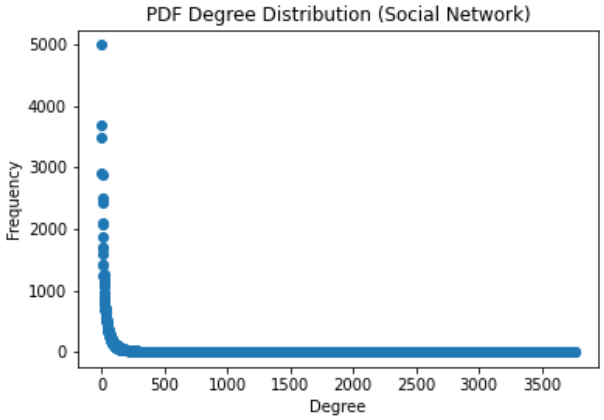

Fig. 2. Degree distribution of the social network data set.

V. Conclusion and Short-Term Plans

This project is motivated to investigate real-world networks and their robustness to withstand attacks. For the current milestone, we have selected four networks and analyzed their fundamental attributes. In milestone 2, we will follow our road-map and initiate the implementation of attack/defense algorithm, starting with a random algorithm. We will divide the work equally, where Runqi Huang will design the attack algorithm, while Vikas Kashyap the defense algorithm. We will then merge our algorithms to compete against each other, and monitor the behavior of the networks. Please stay tuned.

References

[1] A.-L. Barabasi. Network Science. Cambridge University Press. 2016. http://barabasi.com/f/619.pdf
[2] J. Leskovec and A. Krevl. SNAP Datasets: Stanford Large Network Dataset Collection. 2014. https://snap.stanford.edu/data/
[3] M. Molloy and B. Reed. A criticial point for random graphs with a given degree sequence. Random Structures and Algorithms, 6: 161, 1995.
[4] H. Dai, H. Li, T. Tian, X. Huang, L. Wang, J. Zhu, and L. Song. Adversarial attack on graph structured data. arXiv preprint arXiv:1806.02371, 2018.
[5] L. Sun and Y. Dou and C. Yang and K. Zhang and J. Wang and P. S. Yu and L. He and B. Li. Adversarial Attack and Defense on Graph Data: A Survey. IEEE. 2022.
[6] W. Ellens and R. E. Kooij. Graph measures and network robustness. arXiv. 2013.
[7] J. Wu, Y. Tan, H Deng, Y. Li, B. Liu, X. Lv. Spectral Measure of Robustness in Complex Networks. arXiv. 2008.