Adversarial perturbations to manipulate the perception of power and influence in networks

Mihai Valentin Avram*, Shubhanshu Mishra*, Nikolaus Nova Parulian *, Jana Diesner*
* *School of Information Sciences*
University of Illinois at Urbana-Champaign
Champaign, USA 61820
Email: {mihaia2,smishra8,nnp2,jdiesner}@illinois.edu

***Abstract*—Observed social networks are often considered as proxies for underlying social networks. The analysis of observed networks oftentimes involves the identification of influential nodes via various centrality metrics. Our work is motivated by recent research on the investigation and design of adversarial attacks on machine learning systems. We apply the concept of adversarial attacks to social networks by studying strategies by which an adversary can minimally perturb the observed network structure to achieve their target function of modifying the ranking of nodes according to centrality measures. This can represent the attempts of an adversary to boost or demote the degree to which others perceive them as influential or powerful. It also allows us to study the impact of adversarial attacks on targets and victims, and to design metrics and security measures that help to identify and mitigate adversarial network attacks. We conduct a series of experiments on synthetic network data to identify attacks that allow the adversarial node to achieve their objective with a single move. We test this approach on different common network topologies and for common centrality metrics. We find that there is a small set of moves that result in the adversary achieving their objective, and this set is smaller for decreasing centrality metrics than for increasing them. These results can help with assessing the robustness of centrality measures. The notion of changing social network data to yield adversarial outcomes has practical implications, e.g., for information diffusion on social media, influence and power dynamics in social systems, and improving network security.**

***Index Terms*—Social Network Analysis, Adversarial Attacks, Network Robustness, Centrality Measures**

I. Introduction

Social network analysis (SNA) is a common approach for studying complex systems constituting interaction between social agents. A typical starting step for a SNA is the construction of a social network based on observed data, with the assumption that this network reflects or closely approximates the true underlying network. This step is often followed by the identification of influential nodes in the network via centrality measures. The topmost influential nodes can be further analyzed and their network position interpreted based on the given research question and content domain.

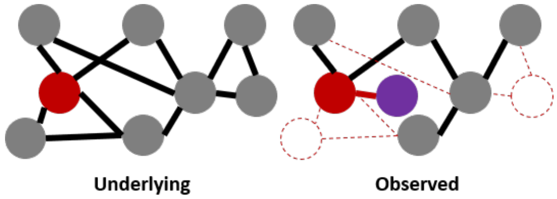

Fig. 1: Example: network analysis of an observed network. The adversarial node (red) adds a new node not present in the underlying network (e.g., creating a fake account), hides its ties to other nodes (e.g., unfriending accounts), and prevents two nodes from being observed (e.g., getting other accounts deactivated or deleted). Refer to figure 2 for legend.

In this paper, we study the scenario where an adversarial node in the underlying network is used to or aims to manipulate the observed network (see Figure 1) in order to change (increase or decrease) its observed ranking based on common centrality measures. Our problem formulation is motivated by two aspects. First, we aim to advance our understanding of the susceptibility of centrality measures to noisy measurements of a network [1]. Second, we leverage recent progress with studying adversarial attacks on automated systems, in particular prior work on fooling machine learning systems built for classifying images [2], [3], audio data [4], text data [5], and more recently network data [6], to examine the impact of adversarial attacks on the perception of the power and influence of individual nodes in networks. Our work focuses on studying an adversary's ability to manipulate its ranking with respect to other nodes in the network as opposed to its classification label. Likewise, our framework also allows to study the susceptibility of targets to these attacks. Manipulating the perception of a node's influence has real-world applications, e.g., boosting the ranking of a website. Similarly, there are use cases for appearing less influential, e.g., authorities may want to understand adversarial strategies for concealing influence in networks, vulnerable agents or organizations may wish to diminish their measurable relevance, and individuals may seek increased privacy protection through

https://doi.org/10.1145/3341161.3345026

obfuscating their power rankings [7].

We conduct a systematic set of controlled experiments, where we test the impact of parameter settings for network topology, adversarial moves, and rank change direction (increase or decrease) on centrality metrics. These experiments focus on finding single adversarial moves that result in moving a node from the bottom x (=10) percentile to appear in the top y (=10) percentile of centrality values and vice versa. Our work is most closely related to that of Waniek and colleagues [7], who focused on adversarial attacks that make nodes and communities appear less influential (called node and community hiding) through local network changes.

Our findings suggest that a small set of local and global moves can enable an adversary to change its rank drastically across various network types and centrality measures. We implemented our approach in a tool [8][1] that allows researchers to simulate adversarial attacks on their network data.

We believe our work can help in furthering research on the robustness of centrality measures against adversarial attacks. Our work also contributes to transferring concepts and advances from the area of adversarial attacks on machine learning systems to the SNA domain.

II. Related Work

A. Adversarial Attacks

Adversarial attacks have been popularized in the field of machine learning owing to the theory of Generative Adversarial Networks (GANs) [2]. GANs utilize a generator and a discriminator, each aimed at increasing the loss of the other by changing its output. GANs are capable of generating sample data that closely resemble the distribution of the input data. Work in this area has led to research on generating adversarial examples [9] by making manipulations to the input data that are imperceptible to humans (i.e., the change in features is small as measured by some distance metric). These adversarial examples can cause a machine learning classifier to predict a label of their choice. For example, [3] have shown that minimal modifications of a stop road sign, e.g., with stickers or graffiti, can trick a classifier to interpret the image as a speed limit sign. While this approach has been popular is the domain of computer vision, this general notion has also been applied to other fields, such as text classification [5], speech recognition [4], and node classification in networks [10]. For text classification models, adversarial attacks have been used to change text elements, such as (characters in) words, to fool a machine learning algorithm into flipping the sentiment label for a piece of text data, e.g., from positive to negative [5]. In speech to text recognition, adversarial attacks have been used to perturb an audio waveform to change the output of the model to any desired text [4]. These approaches are often called *targeted attack*. A comparison of various approaches is presented in Table I. In this paper, we build upon this prior research by searching for strategies for generating minimal modifications of network data that result in changing the assessment of the power and influence of individual nodes in networks.

B. Adversarial Attacks on Social Networks

The notion of changing social network data to yield various adversarial outcomes has practical implications, e.g., for information diffusion on social media and in offline networks, influence and power dynamics in social systems, recommender systems, link prediction, and network security.

Waniek and colleagues [7] used a method called ROAM (Remove One, Add Many) to hide nodes from detection based on various ranking measures in social networks. Our work is closely related to theirs in that both papers model the notion of node hiding via changes to the local neighborhood of a node. However, our approach is broader as we also consider changes to the whole network, as well as aim to make the adversarial node more prominent in the network. Our framework can also be extended to study strategies for making network metrics more robust to adversarial pertubations. A brief comparison of Waniek's and our approach is provided in Table II. Yu et al. [10] model a social network as a Stackelberg game between a defender and an attacker to represent targeted information propagation in a network. Zhang et al. [12] modify the PageRank algorithm to render it insensitive to collusion attempts. Wang et al. [11] propose GraphGAN, which uses a GAN to boost the performance of various graph analysis tasks, such as recommendation systems, link prediction, and node classification. Moreover, Tang et al. [13] used Topical Affinity Propagation to model topical social influence in large networks. Based on that, they created an analysis framework inspired by the design of viral marketing strategies to identify a set of individuals who can be targeted for spreading content to maximize influence in a social network.

C. Network Robustness

A social networks is represented as a graph (sets of vertices and edges) $G = (V, E)$. An adversary's goal can be assumed to be measured by some target function (e.g. $\Delta_{centrality} \geq \delta$), owing to some constraints (e.g. $\Delta_G \leq \epsilon$). The purpose of an adversary may be to change the perception of the power and influence of some (groups of) nodes and edges. Prior work has looked at the robustness of graph analytic metrics to such changes. Borgatti et al. [1] utilize random perturbations to the nodes and edges of a network to study how node centrality rankings change and which centrality measures are robust to this noise. We borrow ideas from this approach, but focus on creating targeted attacks to change a specific (adversarial) node's ranking. Valente et al. [14] found that centrality measures are strongly correlated on average but provide distinct information in symmetric networks. Karrer et al. [15] used an information-theoretic distance method called *variation of information* to test the robustness of network community structure to network perturbations. Author name disambiguation playes a key role in constructing networks from observed data. Kim et al. [16], [17] evaluated the impact

[1] https://github.com/uiuc-ischool-scanr/social-network-adversarial-perturbations

Domain	Case	Possible Attack	Goal
Image	Image generation	Generative Adversarial Networks [2]	Improve the image generation
Text	Sentence labeling	Change text elements (characters or words) [5]	Change target label
Audio	Speech recognition	Add noise to audio signal [4]	Change target words or phrases
Networks	Recommender system; Information diffusion; Social influence; Network security	GraphGAN [11]; Add/remove nodes and edges [7]	Improve link prediction; Change relevance of nodes and edges

TABLE I: Comparison of Adversarial attacks in different domains

	Waniek et al., 2018 [7]	Ours
Goal	Node hiding	Node hiding or node prominence
Allowed changes	Local edge changes	Local + Global edge changes, addition of nodes (e.g., introducing fake identities), removal of nodes (e.g., deleting accounts)
Attack success criteria	Decrease in ranking	Moving from top x (=10) percentile in centrality metric ranking to bottom y (=10) percentile (and vice-versa)
Ranking criteria	Centrality measures and models of influence	
Experimental costs of adversarial move	3	1 (stricter)
Experiments	Apply ROAM in multiple rounds	Exhaustive search over possible moves in synthetic small-scale networks

TABLE II: Comparison of our approach to Waniek et al. [7]

of insufficient or incorrect author name disambiguation on scholarly network metrics, the detection of key players and network topologies, and assumptions about underlying social processes to applicable theories of link formation in co-author networks. The idea was also extended to email networks in Diesner et al. [18]. Mishra et al. [19] have shown how flawed author name disambiguation can lead to wrong conclusions about gender bias in science. Our work is related to this area of research, which examines how flaws in network data construction and pre-processing can incorrectly inflate or discredit the influence of nodes.

III. Methods

A. Experimentation Design

To understand the effects of an adversarial attack on a network and finding the perturbations that manipulate the network according to given target criteria, we designed an experimental framework. On a high level, our experiments can be described as a) an adversary making changes to the network via a set of moves, b) evaluating changes in the adversary's centrality, and c) selecting the optimal (based on criteria described below) move set that can achieve the adversary's goal of either increasing or decreasing its centrality sufficiently. A more detailed description is provided next:

1) Adversarial Node Sampling: For a given network, identify an adversarial node by randomly selecting a node from a pre-specified percentile for a centrality metric of choice. For instance, we can select a node from among the top x (=10) or bottom y (=10) percentile based on betweenness centrality.

2) Adversarial Moves: In general, a network can be changed by either adding or removing nodes or edges. However, in practice, each of these changes may have associated costs that vary depending on the location of the node in the network, network structure, node evaluation metrics, and social context. In our experiments, we classify the set of allowed moves as **local** moves that can be performed within an ego network, or **global** moves that are performed anywhere in the network. Examples of possible local and global moves are shown in Figure 2. In practice, local moves may be less costly compared to global moves. For instance, removing an edge to an immediate friend may be cheaper than removing the edge between a friend and their other friend. An example of local move is deleting a friend on Facebook, which requires one click and does not need the permission of the other party or multiple users.

3) Evaluate Perturbed Network: Using the set of adversarial moves as a reference, the framework tries each move, and computes the evaluation metrics for the adversarial node. Each move results in the reduction in budget for future moves based on its move cost. Successive moves are applied to the updated network until the adversary's goal is achieved or the move budget is exhausted. The move set which optimized the evaluation metric the most is recorded. If multiple moves result in the same change in evaluation metrics, we keep all the moves. An important subtlety here is that at each move in the move set, the framework takes a greedy approach in carrying out that move. For example, on a move *self_remove_edge_friend* with the evaluation target being a decrease in closeness centrality, the framework will try all possible changes for the sampled nodes, and perform the move that is equivalent to removing an edge to the immediate friend of the target node. The move that satisfies the evaluation criterion (decreasing closeness centrality) was selected as the **optimal move**, and the adversarial perturbation was considered a success. This methodology extends to all of the moves and sets of moves that result in a perturbation, and to all perturbations found by our framework.

B. Experimental Setup

We aim to identify if cheap (cost=1) adversarial perturbations (across moves with uniform cost) can be identified

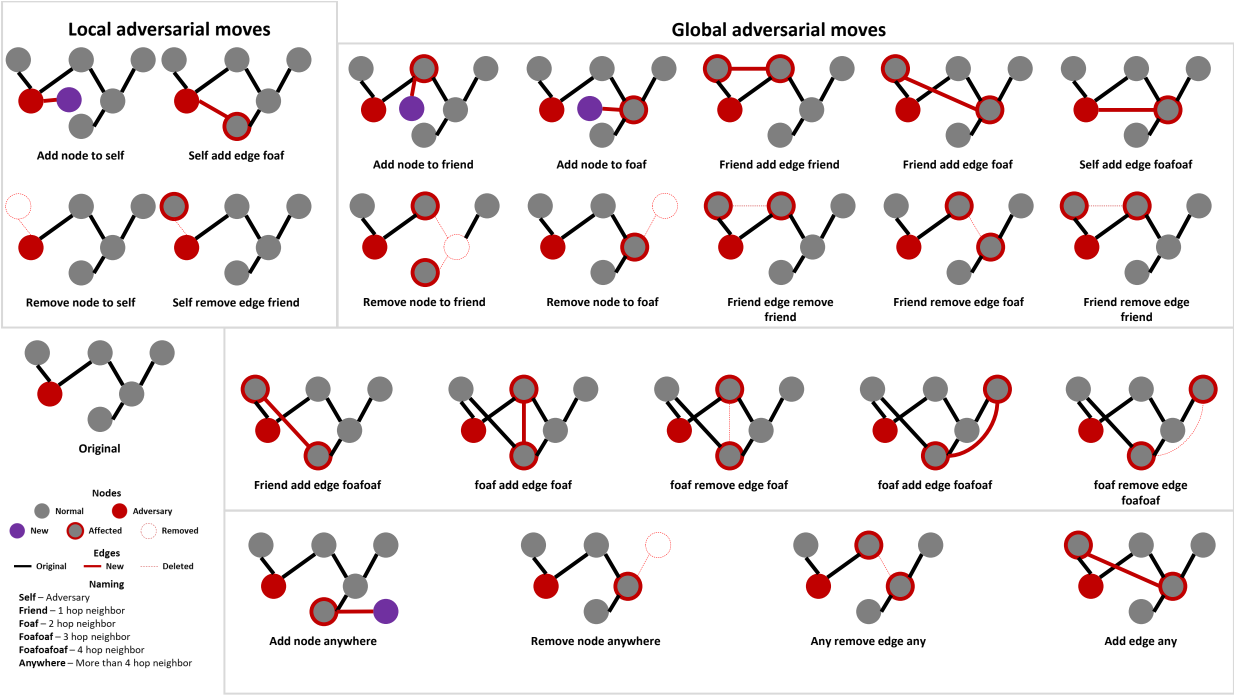

Fig. 2: Possible move set for an adversarial node (red) in a network

across a variety of network topologies. Our goal is to identify if we can successfully increase or decrease four commonly used centrality measures (degree, closeness, betweenness, and eigenvector) via cheap moves from the local and global move sets for a randomly selected adversarial node. Our generated experimental data is available from [20][2].

1) Network Generation: We assess our approach across 100 random networks based on common network topologies. Each network consists of 20 nodes (except for cellular network). We considered the following topologies:

- **Small-World Network**: A small-world network is parameterized via k and p. The parameter k defines the number of neighbors that a new node could be further connected to in the graph generation. The parameter p defines the probability of rewiring an edge. We used $p = 0.2$ and $k = 4$ for small-world network generation [21], [22].
- **Scale-Free Network**: Scale-free networks are parameterized via m. Here, m represents the number of edges which a newly introduced node will create. We used $m \in [3, 5]$ for generating scale-free networks [23].
- **Random Network**: The random network parametrized via p. Here, p is the probability of creating a new edge when generating the network. We used $p = 0.2$ for generating random networks [24].
- **Cellular Network**: We generate cellular networks by combining five random networks (with a $p = 0.2$ as discussed previously) with 20 nodes for each network (total 100 nodes). We combined the five random networks by allowing them to be connected via one edge. This means that one of the five clusters would have four randomly created edges that connected that cluster to the other four clusters.

2) Centrality Measurement: An adversary's power and influence in the network is measured using the following standard centrality measures:

- **Degree centrality**: number of neighbors per node;
- **Closeness centrality**: measures how close a node is to other nodes in the network. The more central a node, the closer that node to all other nodes;
- **Betweenness centrality**: measures the number of times a node is located on the shortest path between other nodes. This measure indicates which nodes can act as bridges between nodes;
- **Eigenvector centrality**: is recursively defined as being connected to other influential nodes with respect to node degree.

3) Sampling Tier: We used a random sampling strategy for each experiment to pick an adversarial node. We rank-ordered the nodes in a network based on their percentile ranks for a given centrality measure. The percentiles were divided into 10 *tiers*, and nodes were placed into related *tiers*. Nodes were randomly sampled from the specified tier. In this study, we experimented with performing adversarial perturbations for the lowest percentile 0-10%, and highest percentile 90-100% of nodes. The goal for the adversarial node was to either increase (from the bottom 10% to the top 10%) or decrease (from the top 10% to the bottom 10%) its rank.

4) Reach Type: We evaluated the effect of the adversarial attack for two different types of moves (Reach Type):

[2] Simulated data: https://doi.org/10.13012/B2IDB-2134305_V1

- **Local moves**: perform the adversarial attack on the adversary's ego network. This represents a low-cost move.
- **Global level moves**: perform the adversarial attack anywhere in the network, including moves to the adversary's ego-network. Modifications to nodes and ties outside of the ego network can represent high-cost moves.

IV. Results

Before delving into the analysis, it is important to note that the centrality measures we consider for our analysis are highly correlated with each other as shown in [14]. Hence, it is natural that if an adversarial move results in the change of a centrality measure in a specific direction, the change might be similar for other centrality measures.

A. Optimal moves across configurations

Our first analysis deals with identifying the frequency of various optimal moves across different combinations of network types, centrality measures, change direction, and reach type.

First, we consider random graphs (see Figure 3). We find that the *remove_node_to_self* and *self_remove_edge_friend* are the most frequently selected adversarial moves across all network types for decreasing influence. We also observe that this trend is same across local and global reach. The pattern is more diverse for increasing influence, where the most prominent move sets differ for local and global reach type. In-fact, for increasing influence, adding an edge outside of the ego network is most effective.

Second, we turn to scale-free networks (see Figure 4). Here, the *self_remove_edge_friend* is the most frequently selected adversarial move for decreasing most centrality measures except for degree. Furthermore, the possible adversarial moves that result in increased centrality scores are more consistent and limited to local moves compared to the random graphs case.

Third, for small world networks, the patterns for centrality decrease are similar to those for random graphs, except for eigenvector centrality, where the most prominent move is *remove_node_to_self.* The diversity in moves for increasing centrality is similar to those for random graphs.

Finally, for cellular networks, the patterns are similar to those for small world networks, with slightly more prominence of *add_node_to_self* for decreasing eigenvector centrality.

Overall, we find that the dominant moves for decreasing a node's centrality are removing a node connected to the adversary (ego), and removing an edge between ego and an immediate neighbor. This pattern is consistent across various repetitions on combinations of network types, centrality measures, change direction, and reach type. For an increase in centrality measures, we observe slight diversity in the best adversarial moves across our experimental settings. It is important to note that an adversarial move may not only affects the centrality of an ego and its neighbors, but also the scores of other nodes in the network. For example, removing the edge between two node not only affects their degree centrality, but may also affects closeness, betweeness, and eigenvector centrality of other nodes, thereby changing the overall ranking. We plan to investigate this kind of change to other nodes in future work.

B. Change in centrality scores after adversarial moves

For our second analysis, we look at the absolute change in the centrality score of the adversarial node after the adversarial move has been made. This helps to understand the impact of an adversarial strategy, and the sensitivity of graph metrics to these strategies. In Figure 7, we show the distribution of the change in the adversary's centrality scores after the adversarial move. The figure does not contain degree centrality changes as these change are either ± 1 for all cases. Similarly to the previous results, we observe similar patterns resulting from adversaries aiming at decreasing centrality scores. This outcome is also due to the fact that global moves are a superset of local moves, resulting in a local move being selected in both cases if this move is indeed optimal. However, for increasing centrality scores, significantly larger changes were observed due to global moves compared to local moves.

C. Adversarial Network Experimentation Tool

In order to allow other researchers to apply our adversarial attack framework to their networks, we have developed a tool [8] that can help with adversarial attack simulations on networks. Our tool can be utilized for running exhaustive search for adversarial moves with varying move costs on small to medium sized networks. The tool is implemented in Python, supports parallel processing, and leverages the NetworkX [25] library for loading graph data. The tool allows users to do the following: generate synthetic networks, import synthetic/real-world networks, configure graph change or perturbation criteria, sample nodes and experiments, find all possible graph changes based on the configuration criteria, save simulation steps and results to files, and visualize and plot results. This contribution enables the reproducibility of our results and can facilitate further research in this area.

V. Conclusions

In this work, we present a framework that allows us to experimentally identify patterns and insights related to manipulating network data such that node rankings are altered. This can represent the attempts of an adversary to boost or demote the degree to which others perceive them as powerful and influential in a network. It also allows us to study the impact of adversarial attacks on targets and victims, and to design metrics and security measures that help to identify and mitigate adversarial network perturbations. With this framework, an experiment can be executed to study how an adversarial node in a network can change its centrality ranking by perturbing the network via local or global moves. Finally, simulations are run to find the moves or patterns that optimize the evaluation criteria.

Empirical findings reveal that an adversary can manipulate a network using a limited set of moves across common

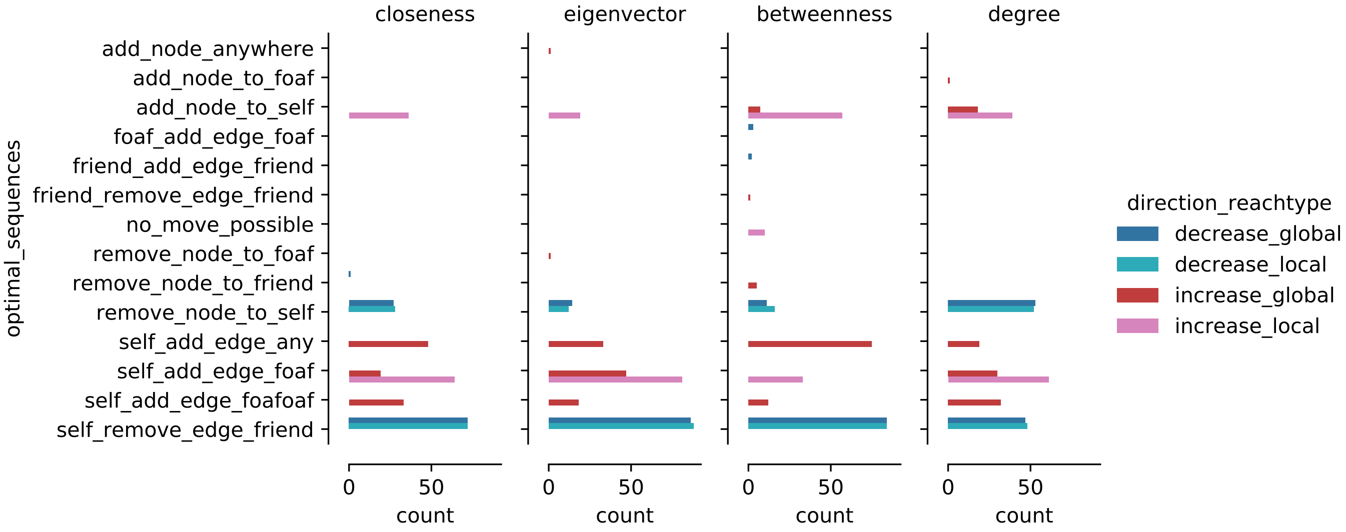

Fig. 3: Optimal move counts across 100 Random Graph networks for each experiment (color and columns)

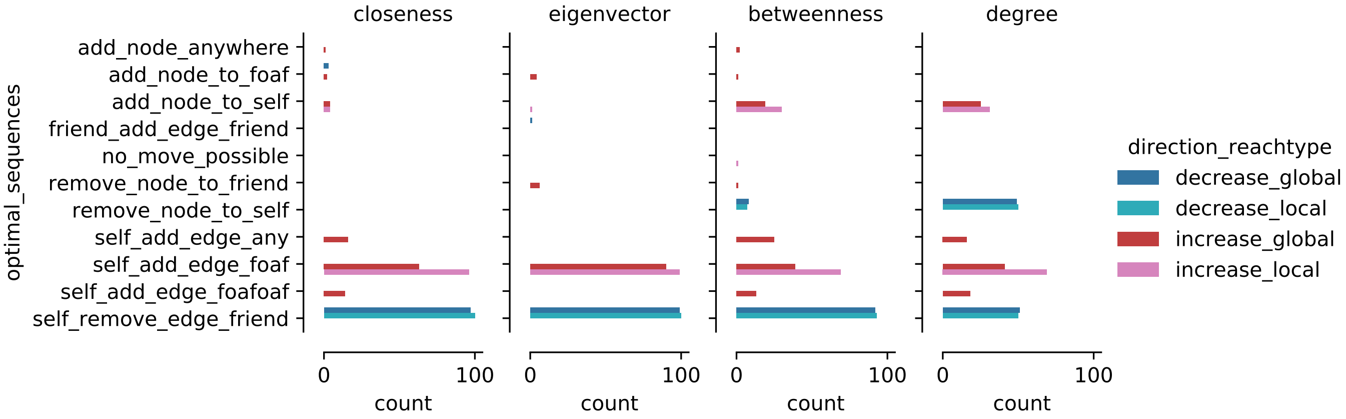

Fig. 4: Optimal move counts across 100 Scale free networks for each experiment (color and columns)

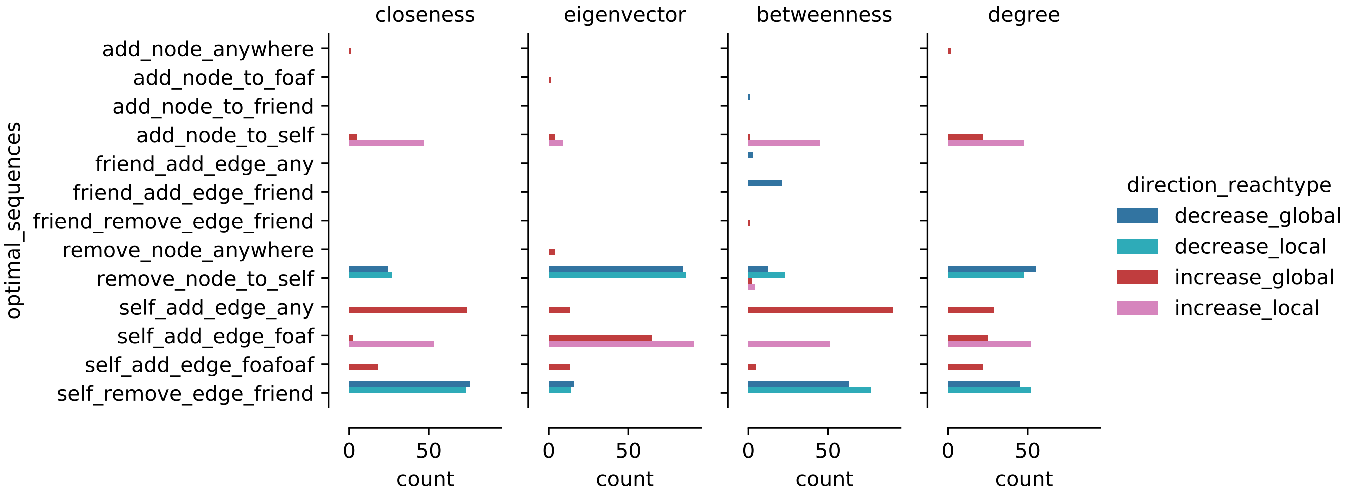

Fig. 5: Optimal move counts across 100 Small world networks for each experiment (color and columns).

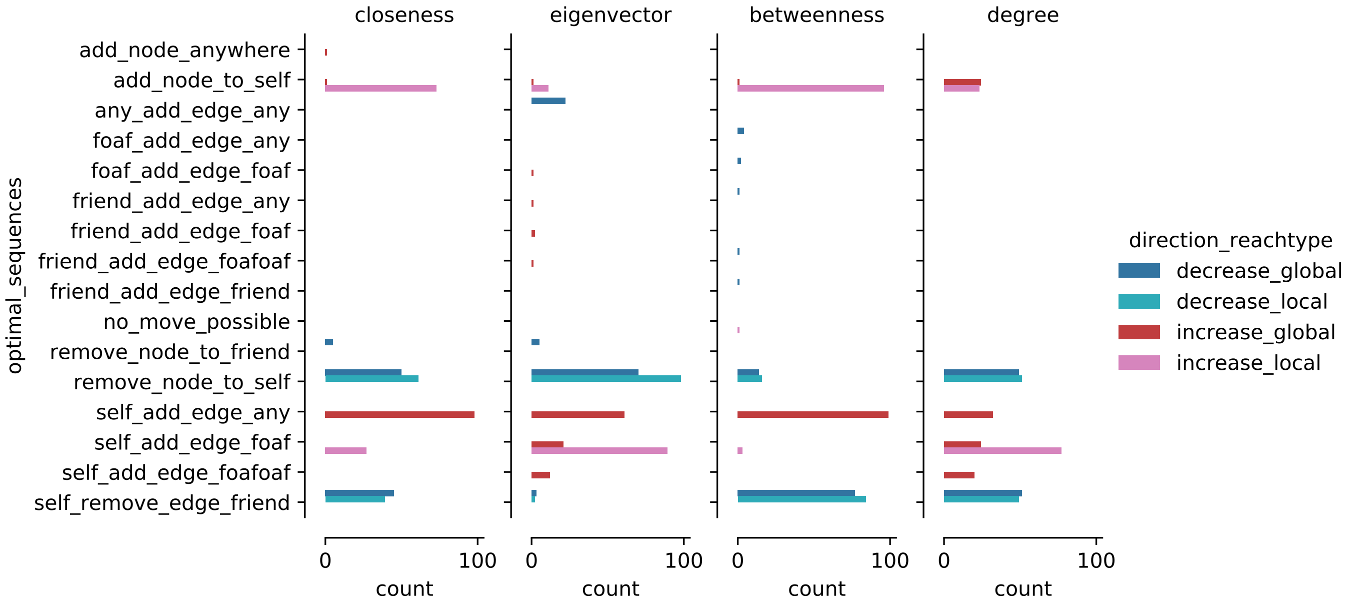

Fig. 6: Optimal move counts across 100 Cellular networks with 5 random graphs for each experiment (color and columns)

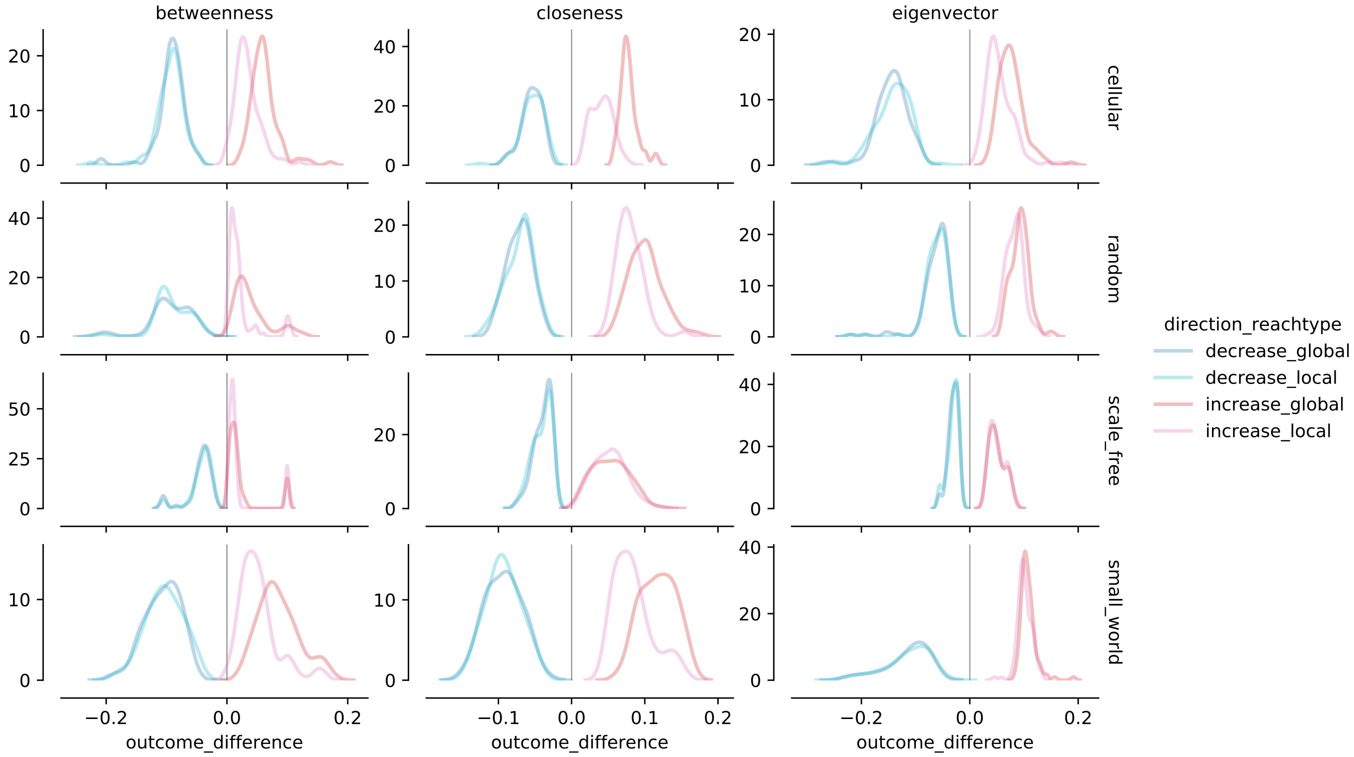

Fig. 7: Distribution of change in centrality score (outcome_difference) of the adversary before and after the adversarial moves across 100 randomly generated networks of the specified type (row) for each experiment (color and columns).

network types, possible reach types, and centrality measures. We also identify that the ego level moves are often sufficient to achieve the adversary's objective. Finally, we found that most of the tested network topologies are susceptible to the outlined attacks. Our findings also validate the approach taken by Waniek et al. [7] for decreasing node ranking.

Our work can be extended such that the role of an adversary is changed from a single node to a group or community of nodes. Adding the ability to handle directionality and larger graphs would also be needed for more comprehensive experiments. This could involve performing clever random walks, using heuristic shortcuts, or even machine learning and deep learning to better traverse the perturbation space without having to perform all possible exhaustive moves given a configuration. We plan to experiments with real-world data to study problems related to fake-news, marketing, supply chains, and resource allocation. Finally, the vulnerability of network metrics to these attacks can be used to inform the design of more robust metrics and network security strategies.

Acknowledgments

Research reported in this paper was sponsored in part by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] S. P. Borgatti, K. M. Carley, and D. Krackhardt, "On the robustness of centrality measures under conditions of imperfect data," *Social networks*, vol. 28, no. 2, pp. 124–136, 2006.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[3] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning visual classification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[4] N. Carlini and D. Wagner, "Audio adversarial examples: Targeted attacks on speech-to-text," in *2018 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2018, pp. 1–7.

[5] J. Ebrahimi, A. Rao, D. Lowd, and D. Dou, "HotFlip: White-box adversarial examples for text classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 31–36. [Online]. Available: https://www.aclweb.org/anthology/P18-2006

[6] J. Chen, Y. Wu, X. Xu, Y. Chen, H. Zheng, and Q. Xuan, "Fast gradient attack on network embedding," *arXiv preprint arXiv:1809.02797*, 2018.

[7] M. Waniek, T. P. Michalak, M. J. Wooldridge, and T. Rahwan, "Hiding individuals and communities in a social network," *Nature Human Behaviour*, vol. 2, no. 2, p. 139147, 2018.

[8] M. V. Avram, S. Mishra, N. N. Parulian, C.-L. Chin, and J. Diesner, "Social network adversarial perturbations," https://github.com/uiuc-ischool-scanr/social-network-adversarial-perturbations, Jul. 2019. [Online]. Available: https://github.com/uiuc-ischool-scanr/social-network-adversarial-perturbations

[9] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *International Conference on Learning Representations*, 2015. [Online]. Available: http://arxiv.org/abs/1412.6572

[10] S. Yu, Y. Vorobeychik, and S. Alfeld, "Adversarial classification on social networks," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, ser. AAMAS '18. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2018, pp. 211–219. [Online]. Available: http://dl.acm.org/citation.cfm?id=3237383.3237420

[11] H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, "GraphGAN: Graph representation learning with generative adversarial nets," 2018. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16611/15969

[12] H. Zhang, A. Goel, R. Govindan, K. Mason, and B. V. Roy, "Making eigenvector-based reputation systems robust to collusion," *Algorithms and Models for the Web-Graph Lecture Notes in Computer Science*, p. 92104, 2004.

[13] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 807–816.

[14] T. W. Valente, K. Coronges, C. Lakon, and E. Costenbader, "How Correlated Are Network Centrality Measures?" *Connect (Tor)*, vol. 28, no. 1, pp. 16–26, Jan. 2008.

[15] B. Karrer, E. Levina, and M. E. J. Newman, "Robustness of community structure in networks," *Physical Review E*, vol. 77, no. 4, 2008.

[16] J. Kim and J. Diesner, "Distortive effects of initial-based name disambiguation on measurements of large-scale coauthorship networks," *Journal of the Association for Information Science and Technology*, vol. 67, no. 6, pp. 1446–1461, Apr. 2015. [Online]. Available: https://doi.org/10.1002/asi.23489

[17] ——, "The effect of data pre-processing on understanding the evolution of collaboration networks," *Journal of Informetrics*, vol. 9, no. 1, pp. 226–236, Jan. 2015. [Online]. Available: https://doi.org/10.1016/j.joi.2015.01.002

[18] J. Diesner, C. Evans, and J. Kim, "Impact of entity disambiguation errors on social network properties," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2015. [Online]. Available: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10588

[19] S. Mishra, B. D. Fegley, J. Diesner, and V. I. Torvik, "Self-citation is the hallmark of productive authors, of any gender," *PLOS ONE*, vol. 13, no. 9, p. e0195773, Sep. 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0195773

[20] M. V. Avram, S. Mishra, N. N. Parulian, and J. Diesner, "Simulation data for adversarial perturbations to manipulate the perception of power and influence in networks. university of illinois at urbana-champaign," https://doi.org/10.13012/B2IDB-2134305_V1, 2019.

[21] A. Barrat and M. Weigt, "On the properties of small-world network models," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 13, no. 3, pp. 547–560, 2000.

[22] M. D. Humphries and K. Gurney, "Network small-world-ness: a quantitative method for determining canonical network equivalence," *PloS one*, vol. 3, no. 4, p. e0002051, 2008.

[23] A. Magner, S. Janson, G. Kollias, and W. Szpankowski, "On symmetry of uniform and preferential attachment graphs," *the electronic journal of combinatorics*, vol. 21, no. 3, pp. 3–32, 2014.

[24] P. D. Yates and N. D. Mukhopadhyay, "An inferential framework for biological network hypothesis tests," *BMC bioinformatics*, vol. 14, no. 1, p. 94, 2013.

[25] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.