N24 Avram, M., Mishra, S., & Diesner, J. (2019). Adversarial perturbations for identifying strategies toward biasing the perceptions of power and influence in social networks.

Node impact in social networks is usually evaluated via centrality metrics. Calculating these metrics assumes reliable and accurate network data. In this work, we identify optimal adversarial network perturbations to such data that result in observed metrics to fulfill an adversary's objective. Here, the objective is to boost or diminish the true relevance of a node. This work is inspired by the application of adversarial attacks in machine learning and computer network security. Our technique can be applied to assess the robustness or vulnerability of network structures and metrics to adversarial attacks, to identify patterns used by malicious domains in altering search engine rankings in hyperlink networks, and to find network-level changes that social media designers can use to stymie the spreading of fake-news.

We aim to identify optimal perturbations in a network such that nodes appear in or disappear from the top k% nodes of a network (ranking defined in terms of centrality metrics). The set of adversarial moves include the addition or deletion of nodes or edges. Each move is associated with a cost (heuristic value), and each adversary has a cost-budget.

Our preliminary experiments on empirical and simulated network data reveal the following patterns: a) betweenness centrality can be optimally decreased by removing adjacent edges for all network topologies (random, scale-free, small-world, cellular), and b) closeness centrality can be optimally decreased by adding a node to a friend of a friend in scale-free networks. Additionally, the number of solutions and solution types to the adversarial perturbation problem vary based on network topology, allowed move set, and the adversary's budget. For example, it might be easier for a node to perform an adversarial move within instead of outside its ego-network, which is reflected in the costs of each move. Also, if global perturbations are allowed, then multiple solutions exist for an adversarial perturbation when the aim is to move the adversary node with low degree centrality to be among top degree centrality nodes. However, for other centrality measures, a fewer number of solutions exist.

In the future, we aim to extend the role of an adversary from a single node to a community of nodes. We also aim at finding more efficient search strategies and extending the framework to scale to larger graphs and search-space configurations. Our research is leading to the development of an open-source Python framework, which facilitates searching for adversarial moves. The framework uses a multi-threaded architecture to efficiently execute an exhaustive search through moves on user-provided network data. The experiments can be configured to constrain the search to an allowable set of perturbation moves; each of which is associated with a fixed cost, maximum allowable move budget, and target function to be optimized.

This work 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.