Heuristics Approach to Solve Stackelberg Game to Mitigate the Spread of Misinformation for Large Real World Networks

VAYAM AGARWAL, KATELINH JONES, SAHANA KARGI, DR. MARION SCHEEPERS, DR. FRANCESCA SPEZZANO

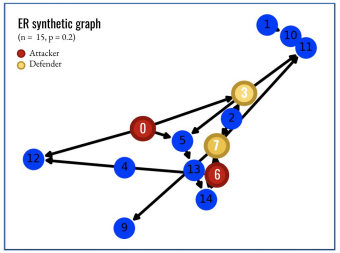
NSF Grant Number CCF-1950599

INTRODUCTION

With the increased presence of social media in our lives, the issue of combating fake news has become more prevalent than ever. Our project aims to find the users in a real-world social media network who when blocked, would optimally reduce the spread of fake news.

METHODOLOGY

We use principles of game theory and formulate a Stackelberg game; the attacker's action set and strategy is to choose a set of nodes that maximizes the influence in the graph while the defender chooses a set of nodes that minimizes the attacker's influence (I_A). We created a defender algorithm that is scalable for large networks while effectively reducing the attacker's influence. We have tested the algorithm on small Erdos-Renyi (ER) synthetic graphs and will implement it on a large network of Twitter users. Below is an ER 15 node synthetic graph where the defender and attacker both choose 2 nodes. The table compares the I_A when the defenders blocks 0, 1 and 2 nodes.



k_D	Defender Strategy	Attacker Strategy	I_A
0	[]	[7,0]	5.95
1	[7]	[3,4]	4.93
2	[7,3]	[0,6]	3.94

MATHEMATICAL FORMULATION FOR GAME

$$\min_{S_D} \quad \sigma(S_A^* | \mathcal{G}(S_D))$$

$$\text{s.t.} \quad |S_D| \leq k_D$$

$$S_A^* = \text{InfluMax}(\mathcal{G}(S_D))$$

$$\text{s.t.} \quad |S_A| \leq k_A,$$

k_A: Attacker's budget
k_D: Defender's budget
S_A: Nodes chosen by attacker
S_D: Nodes chosen by defender
$G(S_D)$: Graph when defender nodes are removed

MACHINE LEARNING APPROACH FOR LEARNING NODE WEIGHTS

We used a dataset comprised of a real-world Twitter network that includes users who follow one another as well as shared PolitiFact checked news articles (Shu et al.). There are 240 articles evenly split between being real and fake news. While there are over 23,000 users in the dataset, we used the largest connected network that had 16,114 users and 544,070 directed edges from followed to follower.

Parameter options for the attacker's and defender's algorithm that will be used include using weights on the nodes to determine which nodes to spread to. Each weight represents the likelihood of that node spreading fake news based on past behavior. These weights were found using a k-fold cross-validation Random Forest Classifier that was trained and validated using the users who shared more than 5 articles as the ground truth. Those used as ground truth were users who had shared more than 70% fake news were classified as a fake news spreader and those less than 30% were real news spreaders. The user features were provided in the dataset and represented unspecified features of the tweets. This classifier had an average of 87% AUROC and 92% Precision. After finding the model with the best AUROC, the featured model was used to predict the probability of either being a fake news spreader on the rest of the data. Then, these weights would be used in either the attacker's or defender's algorithm.

ATTACKER STRATEGY

The attacker wants to maximize their influence in the graph. Basically, the attacker is solving the Influence Maximization problem. There are two popular Influence Maximization algorithms: Greedy (Kempe et al., 2003) and CELF (Leskovec et al., 2007). These algorithms are not scalable in the network size we are dealing with.

For our purpose, we use def-CELF algorithm on top-ranked nodes based on degree and pagerank centrality. This makes our algorithm significantly faster, but may reduce accuracy.

DEFENDER STRATEGY

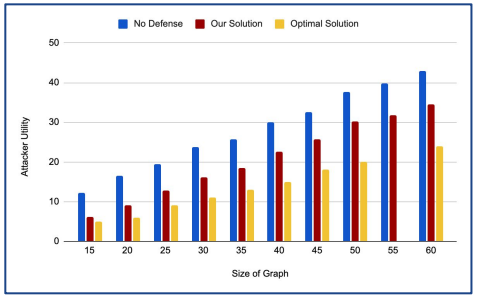
The defender's goal is to minimize the influence of the attacker. In our Stackelberg game, the defender makes the first move. The defender follows a greedy strategy to choose k_D nodes that will minimize the attacker's influence.

The algorithm works as follows: suppose the defender has to block 5 nodes. The defender first finds one node that minimizes the influence of the attacker. Then a second node which along with the first further minimizes the influence of the attacker and so on until it finds 5 nodes.

ONGOING WORK AND RESULT

The defender algorithm outputs a set of k_D nodes which, when blocked, minimizes the attacker's influence. We will run the defender algorithm on a large Twitter graph. We will compare our I_A results for the Twitter graph with the maximum influence the attacker would have had if the nodes blocked were chosen based on degree and pagerank centrality. Our assumption is that our algorithm will fare better than blocking based on centrality measures.

We also test our algorithm on synthetically generated Erdos-Renyi (ER) graphs. The defender chooses 5 nodes from graphs of varying seed sizes, and compare our Attacker's Influence with the optimal solution from Jia et al. and the influence when there were no nodes blocked.



The graph takes the average of 10 graphs of varying sizes (15-60) . We are comparing the average attacker influence with no defense strategy, our defense strategy and the most optimal solution. Our defense strategy performs slightly worse than the optimal solution as the number of nodes increases, but remains significantly better than the worst solution

REFERENCES

[1] Jia, F., Zhou, K., Kamhoua, C., & Vorobeychik, Y. (2020). Blocking Adversarial Influence in Social Networks. In Decision and Game Theory for Security (Lecture Notes in Computer Science, pp. 257-276). Cham: Springer International Publishing.

[3] Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 137-146.

[3] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., & Glance, N. (2007). Cost-effective outbreak detection in networks. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 420-429.

[4} Shu, K., Mahudeswaran, D., Wang, S., Lee, D.,& Liu, H. (2018). FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. arXiv.org.