Modeling Disinformation and the Effort to Counter It: A Cautionary Tale of When the Treatment Can Be Worse Than the Disease

Extended Abstract

Amirarsalan Rajabi
University of Central Florida, Complex Adaptive Systems Laboratory
Orlando, Florida
amirarsalan@knights.ucf.edu

Chathika Gunaratne
University of Central Florida, Complex Adaptive Systems Laboratory
Orlando, Florida
chathika.gunaratne@ucf.edu

Alexander V. Mantzaris
University of Central Florida, Complex Adaptive Systems Laboratory
Orlando, Florida
alexander.mantzaris@ucf.edu

Ivan Garibay
University of Central Florida, Complex Adaptive Systems Laboratory
Orlando, Florida
igaribay@ucf.edu

ABSTRACT

The problem of disinformation in online social networks has recently received a considerable amount of attention from the research community. It has been shown that online social networks are extensively getting exploited to alter public opinion and individuals' stance on a wide-range of topics. This study proposes an agent-based model that simulates a disinformation campaign by a group of organized users called *conspirators*, targeting a *susceptible* population, which are then opposed by a parallel organized group of users referred to as *inouclators* that try to act as a barrier to the spread of disinformation. The results of this study indicate that the process of inoculating a susceptible population against disinformation is mostly at the price of further polarizing the population.

KEYWORDS

disinformation; misinformation; polarization; agent-based modeling; social media

ACM Reference Format:
Amirarsalan Rajabi, Chathika Gunaratne, Alexander V. Mantzaris, and Ivan Garibay. 2020. Modeling Disinformation and the Effort to Counter It: A Cautionary Tale of When the Treatment Can Be Worse Than the Disease. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

While online social media is praised for its influence in democratizing conversations online [4], The increasing complexity and diversity of online social networks, and their ever-increasing adoption by a large portion of society, have made them an appropriate environment for the problem of disinformation. As many online platforms which do not charge their users for the usage directly, a business model based upon advertising for the users has been commonly adopted [6]. This model has allowed many new online services to be funded which the users are then exposed to paid content throughout their online experience. What is interesting in the evolution of the social network roles is that, social network ties also displayed an influence upon the economic choices made by its users [8]; namely that friends/acquaintances which may not have direct economic benefit can influence purchases. What is an interesting question is whether and to what extent has the increased amount of dedicated time to online social networks attributed to the users' formation of ideas beyond that of commercial interests.

The problem of disinformation has been observed and studied since 90s [7], although it has gained a recent growth of interest by the research community. Organized social media campaigns are shown to have been deployed in almost 70 countries [2]. Majority of these campaigns have been promoting disinformation and therefore can be considered to be *disinformation campaigns*. The malicious actors of these campaigns are deployed by cyber troops, gevernments, or political parties [2].

Models of opinion dynamics which stem from statistical physics, try to capture the process of social learning and formation of opinions in human populations [12]. In this work we propose a continuous model of opinion dynamics. This model contains users which disseminate disinformation throughout the network(*conspirators*), users which aim to inhibit the promotion of disinformation(*inoculators*), and users which are not directly engaged with content choice(*susceptibles*). The idea of introducing users with fixed opinions and studying their effect is discussed in literature([10] and [11]). Also, the introduction of inoculators is based on inoculation theory in social psychology. Originally developed by McGuire [9], the theory explains the protection of beliefs and opinions against external influence and manipulation.

2 MODEL DESCRIPTION

Our proposed model of opinion dynamics is inspired by the work of [1]. A population of N *susceptible* agents are connected to each

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

other through a network which is formed by Barabasi-Albert algorithm. The *conspirators* and *inoculators* get added to the network right after the formation of susceptibles' network. The number and position of the conspirators and inouclators is determined by model parameters.

A conceptual underlying state of the world Θ exists on which each agents has an opinion, and we assume that the true value of this variable is 1. $x_i^s(k)$, $x_i^c(k)$, and $x_i^i(k)$ show the opinion of susceptible agent i at time t, opinion of conspirator i at time t, and that of inoculator i at time t. The initial belief of susceptibles is a randomly generated real number between 0 and 2, is 0 for conspirators and 1 for inoculators:

$$x_i^s(0) \in [0,2] \qquad x_i^c(0) = 0 \qquad x_i^i(0) = 1 \qquad \frac{1}{N}\sum_i x_i^s(0) \approx 1 \tag{1}$$

Number of susceptible agents is represented by N, $\boldsymbol{\beta} \in [0,1]$ represents the ratio of susceptibles that are targeted, $\boldsymbol{\alpha} \in [0,1]$ represents the ratio of total number of inoculators to total number of targeted susceptibles. The number of conspirators and inoculators is then calculated by:

$$\begin{aligned} &\text{no. of conpirators} = N \cdot \beta \cdot (1-\alpha) \\ &\text{no. of inoculators} = N \cdot \beta \cdot \alpha \end{aligned} \tag{2}$$

$\boldsymbol{\rho}$ and $\boldsymbol{\rho'}$ are called conspiracy-target-log-rank and inoculation-target-log-rank and determine the accuracy with which conspirators and inoculators pick their targets. The target selection process in the model is based on eigenvector centrality of susceptible agents, and the idea is that a higher eigenvector centrality of a susceptible agent makes it a more *accurate* target.

The rules of interaction between different types of agents is as follows:

If i and j are both susceptibles:

$$\begin{cases} x_i(k+1) = x_j(k+1) = \frac{1}{2}[x_i(k) + x_j(k)] & \text{with probability p} \\ x_i(k+1) = x_i(k) \;\&\; x_j(k+1) = x_j(k) & \text{with probability } 1-\text{p} \end{cases}$$

If i is susceptible and j is conspirator:

$$\begin{cases} x_i(k+1) = \frac{x_i(k)+0}{2} & \text{with probability p} \\ x_i(k+1) = x_i(k) & \text{with probability } 1-\text{p} \\ x_j(k+1) = x_j(k) = 0 & \text{with probability 1} \end{cases}$$

If i is susceptible and j is inoculator:

$$\begin{cases} x_i(k+1) = \frac{x_i(k)+1}{2} & \text{with probability p} \\ x_i(k+1) = x_i(k) & \text{with probability } 1-\text{p} \\ x_j(k+1) = x_j(k) = 1 & \text{with probability 1} \end{cases}$$

3 EXPERIMENTS AND RESULTS

To understand and analyze the opinion state of susceptible population, the concept of ***collective-thought*** is introduced, which is the set of opinion of all members of susceptible population. Three quantities of collective-thought are measured in each experiment: $\overline{\Phi}$: collective-thought-mean, $\widetilde{\Phi}$: collective-thought-median, and $var(\Phi)$: collective-thought-variance.

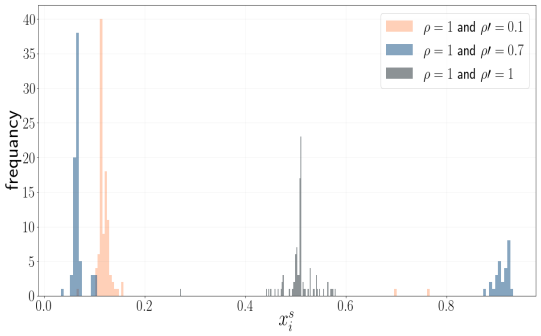

Figure 1: Effect of target selection on polarization.

In the first experiment, With varying probability of interaction $\in [0, 0.95]$, $N \in [50, 200]$, and $\alpha\&\beta \in [0.05, 1]$, and $\rho\&\rho\prime = 1$, the effect of α and β on collective thought was investigated. The results shows the expected effect that increasing α has on collective thought; as the ratio of inoculators to conspirators increases, $\widetilde{\Phi}$ gets closer to 1. The experiment shows another interesting criteria of the model: for α values of near 0, 0.5, and 1, $var(\Phi)$ minimizes, and tends to maximize near α values of 0.25 and 0.75. The effect of α is showing that unless the two campaigns are of similar number of actors ($\alpha = 0.5$), the variance of collective-thought tends to maximize in the absence of a total dominance by either side.

Second experiment investigates the effect of accuracy of target selection by conspirators and inoculators ($\rho\&\rho'$). The parameters selection for the experiment was as follows: $N = 100$, $\beta = 0.02$ and $\alpha = 0.5$, $\rho = 1$ and $\rho' \in [0, 1]$. Probability-of-interaction was also varied between [0,1]. The results of the experiment show a phase shift in the system: As the normalized eigenvector centrality of the targeted susceptible agent increases, the collective-thought quantities slightly increase, until reaching normalized-eigenvector of around 0.3 at which a phase shift seems to happen. After this point, $\widetilde{\Phi}$ tends to oscillate between extremes and there is a sudden increase in $var(\Phi)$. This phase shift denotes an unpredictable nature of the opinion state of susceptible population.

Figure 1 shows three realizations of the model with similar random seed, hence exact same network of susceptible, and similar parameter settings excluding $\rho\prime$ which takes the values of 0.1, 0.7, and 1. While for $\rho\prime = 0.1\&1$, opinion of majority of agents is close to 0 and 0.5 respectively(remarkably low polarization), for a $\rho\prime = 0.7$, opinion of susceptible agents is distributed across the two extremes (highly polarized population).

Result of the experiments indicate that the inoculation effort frequently produces a more polarized community. The negative effects of polarization in a social network are explained in the work of [5]. Future direction of this work includes designing a feature of the model to address the distinction between the dynamics of disinformation adoption and inoculation in susceptibles. The model should also be tested on other network formation algorithms.

REFERENCES

[1] Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. 2010. Spread of (mis) information in social networks. *Games and Economic Behavior* 70, 2 (2010), 194–227.

[2] Samantha Bradshaw and Philip Howard. 2019. The Global Disinformation Disorder: 2019 Global Inventory of Organised Social Media Manipulation. (2019).

[3] Michael J Brzozowski, Tad Hogg, and Gabor Szabo. 2008. Friends and foes: ideological social networking. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 817–820.

[4] Emilio Ferrara. 2017. Disinformation and social bot operations in the run up to the 2017 French presidential election. *arXiv preprint arXiv:1707.00086* (2017).

[5] Ivan Garibay, Alexander V Mantzaris, Amirarsalan Rajabi, and Cameron E Taylor. 2019. Polarization in social media assists influencers to become more influential: analysis and two inoculation strategies. *Scientific Reports* 9, 1 (2019), 1–9.

[6] Richard Hanna, Andrew Rohm, and Victoria L Crittenden. 2011. WeâĂŹre all connected: The power of the social media ecosystem. *Business horizons* 54, 3 (2011), 265–273.

[7] Peter Hernon. 1995. Disinformation and misinformation through the internet: Findings of an exploratory study. *Government information quarterly* 12, 2 (1995), 133–139.

[8] Raghuram Iyengar, Sangman Han, and Sunil Gupta. 2009. Do friends influence purchases in a social network? *Harvard Business School Marketing Unit Working Paper* 09-123 (2009).

[9] William J McGuire. 1961. The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. *Sociometry* 24, 2 (1961), 184–197.

[10] Mauro Mobilia. 2013. Commitment versus persuasion in the three-party constrained voter model. *Journal of Statistical Physics* 151, 1-2 (2013), 69–91.

[11] Mauro Mobilia, Anna Petersen, and Sidney Redner. 2007. On the role of zealotry in the voter model. *Journal of Statistical Mechanics: Theory and Experiment* 2007, 08 (2007), P08029.

[12] Alina Sîrbu, Vittorio Loreto, Vito DP Servedio, and Francesca Tria. 2017. Opinion dynamics: models, extensions and external effects. In *Participatory Sensing, Opinions and Collective Awareness*. Springer, 363–401.

[13] Cameron E Taylor, Alexander V Mantzaris, and Ivan Garibay. 2018. Exploring how homophily and accessibility can facilitate polarization in social networks. *Information* 9, 12 (2018), 325.