2014 IEEE Joint Intelligence and Security Informatics Conference

Time Critical Disinformation Influence Minimization in Online Social Networks

Chuan Luo[1] Kainan Cui[2] Xiaolong Zheng[1] Daniel Zeng[1,3]
[1]State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]The School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China
[3]Department of Management Information Systems, University of Arizona, Tucson, USA
{chuan.luo, xiaolong.zheng}@ia.ac.cn, kainan.cui@live.cn, zeng@email.arizona.edu

***Abstract*—If a piece of disinformation released from a terrorist organization propagates on Twitter and this adversarial campaign is detected after a while, how emergence responders can wisely choose a set of source users to start the counter campaign to minimize the disruptive influence of disinformation in a short time? This practical problem is challenging and critical for authorities to make online social networks a more trustworthy source of information. In this work, we propose to study the time critical disinformation influence minimization problem in online social networks based on a continuous-time multiple campaign diffusion model. We show that the complexity of this optimization problem is NP-hard and provide a provable guaranteed approximation algorithm for this problem by proving several critical properties of the objective function. Experimental results on a sample of real online social network show that the proposed approximation algorithm outperforms various heuristics and the transmission temporal dynamics knowledge is vital for selecting the counter campaign source users, especially when the time window is small.**

Keywords—social networks; information cascades; competing campaigns; disinformation; submodular functions

I. INTRODUCTION

Recent years have witnessed an explosive growth of various social media sites such as online social networks, blogs, microblogs, social news websites and virtual social worlds. These different kinds of social media platforms have many benefits as a fast and widespread information propagation medium to report eye-witness accounts and share information on disasters, terrorist attacks and social crises as a collective effort [1]. In spite of these advantages, many warnings have been raised about the dark side of social media [2], [3]. For example, various social media platforms can be used by terrorist organizations and their supporters for a wide range of purposes, including recruitment, propaganda, incitement to commit acts of terrorism, and the dissemination of disinformation for terrorist purposes [4]. The disinformation can go viral and cause highly disruptive effects if emergency responders fail to take effective measures to contain the disinformation influence on social media.

In order for online social networks to serve as a more trustworthy source of information and a more reliable platform for disseminating critical messages, clearly it is urgent to have tools to limit the disruptive effects of disinformation soon after the disinformation cascade is detected by authorities. Borrowing the idea of viral marketing in business [5], one cost-effective strategy for emergence responders is to select the most influential source user set of a given size for counter information in online social networks. A counter information diffusion process that begins in such an influential set of users is expected to reach the greatest number of users in s short time, who would have been infected by the disinformation if emergence responders don't start the counter campaign. By making sure that most users hear about the correct information before the misleading one, the disruptive influence of disinformation can be minimized.

Although some recent work start to work on influence optimization problem in the context of competitive information diffusion on social networks [6]–[10], they build their work on discrete time models. However, it is common sense that social media users perform behaviors (i.e., retweet a post) asynchronously and artificially discretizing the time axis into bins may lead to not appropriate problem representation [11]. Besides, discrete time models fail to capture the important fact that different user pair may have different information transmission rate. For instance, if user A usually checks the status updates of her friend B, but pays a little attention on the other friend C, as a result, information transmission rate between user B and A can be quite different from that between user C and A. In this case, the solution to influence minimization problem based on discrete time models cannot be optimal since the rich temporal dynamics between users are neglected.

In this work, we propose to study the time critical disinformation influence minimization problem in online social networks based on the continuous time model of diffusion recently introduced by Gomez-Rodriguez et al. [12]–[14]. This model is able to account for different information transmission rates across different edges in the network. We will first build the continuous time multiple campaign diffusion model based on their model. Then we show that the optimization problem is a NP-hard problem. Fortunately, by utilizing several critical properties of the objective function, we are able to present a provable guaranteed approximation algorithm to find a set of source users for the counter campaign, which can minimize the disruptive influence of disinformation in a time window in online social networks.

978-1-4799-6364-5/14 $31.00 © 2014 IEEE
DOI 10.1109/JISIC.2014.20

The rest of the work is organized as follows. We introduce the related work in section II. The diffusion model and problem definition will be presented in section III. The problem definition will be given in section IV. The problem hardness and key properties of the objective function will be discussed in section V. The greedy algorithm and heuristics will be presented in section VI and VII respectively. Then we perform the experimental evaluation in section VIII. Finally, we give the conclusion in section IX.

II. Related Work

Richardson and Domingos [15], [16] were the first to study influence maximization as an algorithmic problem, motivated by marketing applications. In their work, they proposed heuristics for choosing a set of influential customers by modeling the social network as a markov random field. Kempe et al. [5] posed influence maximization in a social network as discrete optimization problem. They showed that solving the problem exactly is NP-hard for several models of influence such as Independent Cascade Model and Linear Threshold Model [17], and provide a simple greedy algorithm which can obtain an approximation guaranteed solution based on a natural diminishing property of the problem, submodularity. Since then, there have been substantial developments that build on their seminal work. Various efficient heuristics have been proposed to speed up the optimization problem [18]–[20]. Gomez-Rodriguez et al. [14] studied influence maximization in continuous time diffusion networks. Although our multiple campaign diffusion model is based on their work, we study the problem of time critical disinformation influence minimization as opposed to maximization.

Recently, influence maximization has also been studied on the context of competing cascades. Bharathi et al. [6] studied the game of innovation diffusion with multiple competing innovations such as when multiple companies market competing products using viral marketing. They presented a (1-1/e) approximation algorithm for computing the best response to an opponent's strategy. Goyal and Kearns [21] developed a game-theoretic framework for the study of competition between firms. This framework yields a rich class of competitive strategies. He et al. [7] studied influence blocking maximization in social networks under the competitive linear threshold model and proved the objective function is submodular under that model. Perhaps the work of Budak et al. [8] is the most similar to ours. They also proposed to study how to limit the spread of disinformation in social network. However, they built their work on the discrete time independent cascade model without considering information transmission time between people. In contrast to their work, we study how to minimize the disinformation influence in a timely manner by taking advantage of different pairwise transmission rate in the social network.

III. Diffusion Model

We will first introduce pairwise transmission function [11] to model the transmission time along directed edges in social networks. Then we will present the continuous-time multiple campaign diffusion model based on the transmission function defined on each edge.

A. Pairwise Transmission Function

To capture the observation that different pairwise interactions between nodes in the network occur at different rates, given a directed diffusion network $G = (V, E)$, this model associates each edge $j \rightarrow i$ with a transmission function, or the waiting time distribution, denoted as $f(t_i \mid t_j; \alpha_{ji})$. Formally, the transmission function $f(t_i \mid t_j; \alpha_{ji})$ for directed edge $j \rightarrow i$ is the conditional density of node i getting infected at time t_i given that node j was infected at time t_j, where α_{ji} is the information transmission rate between node j and node i, representing how fast the information spreads from node j to node i in the social network. Further, we assume this transmission function is time shift invariant. In other words, it depends on the time difference $\tau_{ij} = t_i - t_j$, and the pairwise transmission rate α_{ji}. Moreover, it takes positive values when $\tau_{ij} \geq 0$, and the value of zero otherwise, and as $\alpha_{ji} \rightarrow 0$, the expected transmission time between node j and node i becomes arbitrarily long. Specifically, we follow the previous work [11], [12] to consider the exponential distribution to model the information diffusion between nodes as shown in equation 1.

$$f(t_i|t_j;\, \alpha_{ji}) = \alpha_{ji} e^{-\alpha_{ji}(t_i - t_j)} \tag{1}$$

Essentially, the transmission function $f(t_i \mid t_j; \alpha_{ji})$ defines a probability distribution of the information transmission time from node j to node i in the social network. This is different from previous discrete time independent cascade model which associates each edge with a fixed diffusion probability. The probability distribution of information transmission time $f(\tau; \alpha_{ji})$ along a directed edge $j \rightarrow i$ is presented in equation 2, where α_{ji} is the information transmission rate of this edge, representing how fast the information spreads along the directed edge in the social network. We further assume that the transmission rate is information independent. In other words, the pairwise transmission time has no relationship with the kind of information propagating along the edge.

$$f(\tau; \alpha_{ji}) = \alpha_{ji} e^{-\alpha_{ji}\tau} \tag{2}$$

B. Continuous-Time Multiple Campaign Diffusion Model

Based on the pairwise transmission function described before, now we present the Continuous-Time Multiple Campaign Diffusion Model (*CTMCDM*) which models the diffusion of two campaigns propagating simultaneously in a social network. One of the two is the adversarial campaign, the purpose of which is to disseminate the disinformation or a rumor in a social network. The other one is the counter campaign which aims to limit the devastating effects of the adversarial campaign in a short time efficiently.

Let the initial set of active nodes for the adversarial and counter campaign be denoted by S_A and S_C, respectively. The diffusion process of adversarial campaign starts when the source node set S_A becomes infected (i.e., users adopt the disinformation) at time $t = 0$ by action of an external source

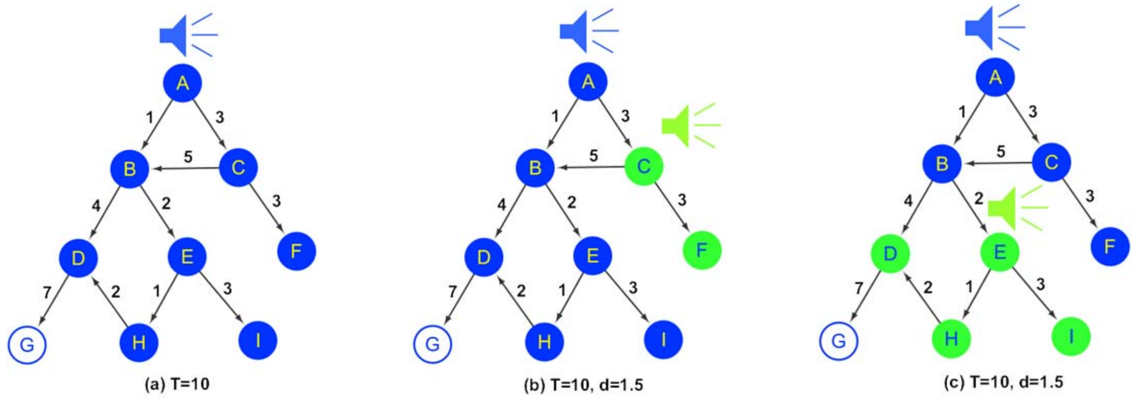

Figure 1. An example of continuous-time multiple campaign diffusion process. The number along each edge is the transmission time which is sampled from the probability distribution $f(\tau; \alpha_{ji})$. (a) The disinformation is released by A at time $t = 0$. If there is no counter information, 8 users (in blue color) will be infected up to time $t = 10$. (b) If the counter information is released by C at time $t = 1.5$, 6 users will be infected by disinformation and 2 users (in green color) will be "saved" by counter campaign. (c) If the counter information is released by E, only 4 users will be infected and 4 users will be "saved".

to the network. Then the disinformation is transmitted from the adversarial sources S_A along their out-going edges to their direct neighbors. Each transmission through an edge $j \rightarrow i$ entails a random spreading time, τ, drawn from the probability distribution $f(\tau; \alpha_{ji})$. We assume transmission times are independent and possibly distributed differently across edges since different edges are associated with different transmission rate α_{ji}. After some time delay d, the adversarial campaign is detected at time $t = d$ and the diffusion of counter information starts from the source node set S_C. When a node v first becomes active in adversarial or counter campaign at time t, the disinformation or counter information will be transmitted from node v along its out-going edges to its direct neighbors. The spreading time τ along an edge $j \rightarrow i$ is drawn from the distribution $f(\tau; \alpha_{ji})$. For some pairwise transmission time distribution, it may happen that $\tau \rightarrow \infty$ and node i is never infected. We assume that a node w becomes infected by the disinformation or counter information as soon as one of its parents (i.e., neighbors who are able to reach node w through an out-going edge) infects it, and later infections by other parents do not contribute anymore to the evolution of the diffusion process. In the case of the disinformation and counter information reaching a node w at the same time, we assume that the node w always adopts the counter information. Once a node becomes active in one campaign, it never becomes inactive or change campaigns. The diffusion process continues until there is no newly activated node in either campaign.

IV. Problem Definition

Suppose that an adversarial campaign starts propagating from a set of nodes S_A at time $t = 0$ and it is detected after a time delay d, at that point a counter campaign is initiated from a set of nodes S_C. The disinformation and counter information propagate in the social network as described with the Continuous-Time Multiple Campaign Diffusion Model.

We define $N(S_C; T, S_A, d)$ as the set of nodes infected by the adversarial campaign up to time $t = T$, and $M(S_C; T, S_A, d)$ as the set of nodes that the counter campaign prevents from adopting adversarial campaign up to time $t = T$. In other words, $M(S_C; T, S_A, d)$ is the set of "saved" nodes which would be infected by the disinformation if there is no counter information propagating in the social network. To limit the devastating influence of the adversarial campaign, the objective is to minimize the number of nodes that end up adopting the disinformation. Equivalently, we aim to maximize the number of "saved" nodes. Since the information propagation is a probabilistic process, we define the influence function $\sigma(S_C; T, S_A, d)$ as the average total number of nodes saved up to time $t = T$, i.e., $\sigma(S_C; T, S_A, d) = \mathbb{E}[|M(S_C; T, S_A, d)|]$.

Formally, given the social network $G = (V, E)$, the set of source nodes of adversarial campaign S_A, the detection time delay d, our goal is to find the set of source nodes of counter campaign S_C such that the diffusion of counter information in G saves the greatest number of nodes before a window cut off T, on average. Thus, we aim to solve:

$$S_C^* = \operatorname*{argmax}_{|S_C| \leq k} \sigma(S_C; T, S_A, d) \tag{3}$$

where the source set S_C is the variable to optimize and the time window T, the source set S_A for adversarial campaign, the detection time delay d, the source set cardinality k are given.

V. Key Problem Properties

We will first prove that the proposed problem is a NP-hard problem, and then present the monotone and submodular proof of the objective function of the optimization problem.

A. NP-hardness

Theorem 1. The proposed optimization problem is NP-hard for the Continuous-Time Multiple Campaign Diffusion Model.

Proof. If we let the time window $T \rightarrow +\infty$, the independent cascade model is a particular case of our Continuous-Time Multiple Campaign Diffusion Model. Then, we can follow the idea of Kempe et al. [5] to prove the NP-hardness of our optimization problem.

Consider an instance of the NP-complete Set Cover problem, defined by a collection of subsets $S_1, S_2, \ldots, S_m$ for a universe set $U = \{u_1, u_2, \ldots, u_n\}$; we will to know whether there exist k of the subsets whose union is equal to U. We show that this can be viewed as a special case of our optimization problem, in which the time window $T \rightarrow +\infty$, the disinformation and counter information propagate according to the independent cascade model, and the set of source nodes of adversarial campaign S_A contains only one node. Given an arbitrary instance of the Set Cover problem, we define a corresponding directed bi-partite graph with $m+n+1$ nodes: there is a node i corresponding to each set S_i, a node j corresponding to each element u_j, and a directed edge (i, j) with activation probability $p_{ij} = 1$ whenever $u_j \in S_i$. In addition, there is an adversary node w and a directed edge (w, j) for all u_j with activation probability $p_{wj} = 1$. The Set Cover problem is equivalent to deciding if there is a set of k nodes S_C (i.e., the source node set of the counter campaign) in this graph with $\sigma(S_C; T, S_A, d) \geq n+k$, when the time window $T \rightarrow +\infty$, the adversarial source set $S_A = \{w\}$, the detection time delay $d = 0$. Note that for the instance we have defined, the evolution of the diffusion process is a deterministic, as all probabilities are 0 or 1. Initially activating the k nodes corresponding to sets in a Set Cover solution results in saving all n nodes corresponding to the ground set U, and if any set of k nodes S_C has $\sigma(S_C; T, S_A, d) \geq n+k$, then the Set Cover problem must be solvable.

B. Monotone

By construction, the objective function $\sigma(S_C; T, S_A, d) \geq 0$ and if $S_C = \phi$, then $\sigma(S_C; T, S_A, d) = 0$. It also follows trivially that $\sigma(S_C; T, S_A, d)$ is monotonically non-decreasing in the adversarial source node set S_C, i.e., $\sigma(S_C; T, S_A, d) \leq \sigma(S_C'; T, S_A, d)$, whenever $S_C \subseteq S_C'$.

C. Submodular

A set function $F: 2^U \rightarrow \mathbb{R}$ mapping subsets of a finite set U to the real numbers is submodular if whenever $A \subseteq B \subseteq U$ and $s \in U \setminus B$, it holds that $F(A \cup \{s\}) - F(A) \geq F(B \cup \{s\}) - F(B)$, i.e., adding s to the set A provides a bigger marginal gain than adding s to the set B.

Theorem 2. Given the social network $G = (V, E)$, the set of source nodes of adversarial campaign S_A, the detection time delay d and a time window T, the objective function $\sigma(S_C; T, S_A, d)$ is a submodular function in the set of counter source node set S_C.

Proof. Since a node can only be infected by one campaign, a directed edge $j \rightarrow i$ will only be visited at most once. When the disinformation or counter information propagates from node j to node i, the diffusion time τ is sampled from the distribution $f(\tau; \alpha_{ji})$. In fact, it does not matter whether the diffusion time τ is sampled at the moment when node j starts transmitting the information, or if it was sampled before the whole propagation process and stored to be examined at the time when node j starts transmitting the information. So, although the whole information propagation is a probabilistic process, for a specific instance of information propagation, we can pre-sampled all the pairwise transmission time distribution $f(\tau; \alpha_{ji})$ to determine the time used for information propagates along this edge.

The key observation is that if we treat the pairwise transmission time as the stochastic edge length, then the state of a node is depended on the length of the stochastic shortest path from the adversarial source nodes to the node L_A and the length of the stochastic shortest path from the counter source nodes to the node L_C. Specifically, given the time window T and detection time delay d, if $L_C + d \leq L_A$ and $L_A \leq T$, this node will be saved by the counter information up to time $t = T$. And if $L_C + d > L_A$ and $L_A \leq T$, this node will be infected by the disinformation up to time $t = T$.

Consider the probability distribution of all possible transmission time along each directed edge in the social network. Thus, given a sample $\Delta\tau$ in the probability space, for $\forall a \in S_A$, we define $R(a; T)$ as the set of nodes that can be reached from node a in a time shorter than T if there is no counter campaign. Define I_A as the set of nodes that are infected by the disinformation by time $t = d$. Define $P(v, w)$ as the length of the shortest path from node v to node w. As a result, the proposed time critical disinformation influence minimization problem is equivalent to maximizing the number of nodes reachable from S_C in a new network $G' = (V', E')$, in which $V' = \{v \mid v \in V \land v \notin I_A\}$ and $E' = \{(v, w) \mid v \in V', w \in V' \land w \in W_v\}$ where W_v is defined as shown in equation 4.

$$W_v = \left\{ w \mid w \in \cup_{a \in S_A} R_{\Delta\tau}(a; T) \ \land \ P_{\Delta\tau}(v, w) + d \leq \min_{a \in S_A} P_{\Delta\tau}(a, w) \leq T \right\} \quad (4)$$

Since the reachability problem in a network is submodular and a non-negative linear combination of submodular functions is also submodular [5], the objective function of the proposed optimization here is submodular.

Algorithm 1 The greedy algorithm to select the counter source node set.

```
Initialize N_0 ← ϕ
for i = 1 to k do
    Set N_i ← N_{i-1} ∪ argmax_{u∈V} σ(N_{i-1} ∪ {u}; T, S_A, d)
end for
Set S_C ← N_k
```

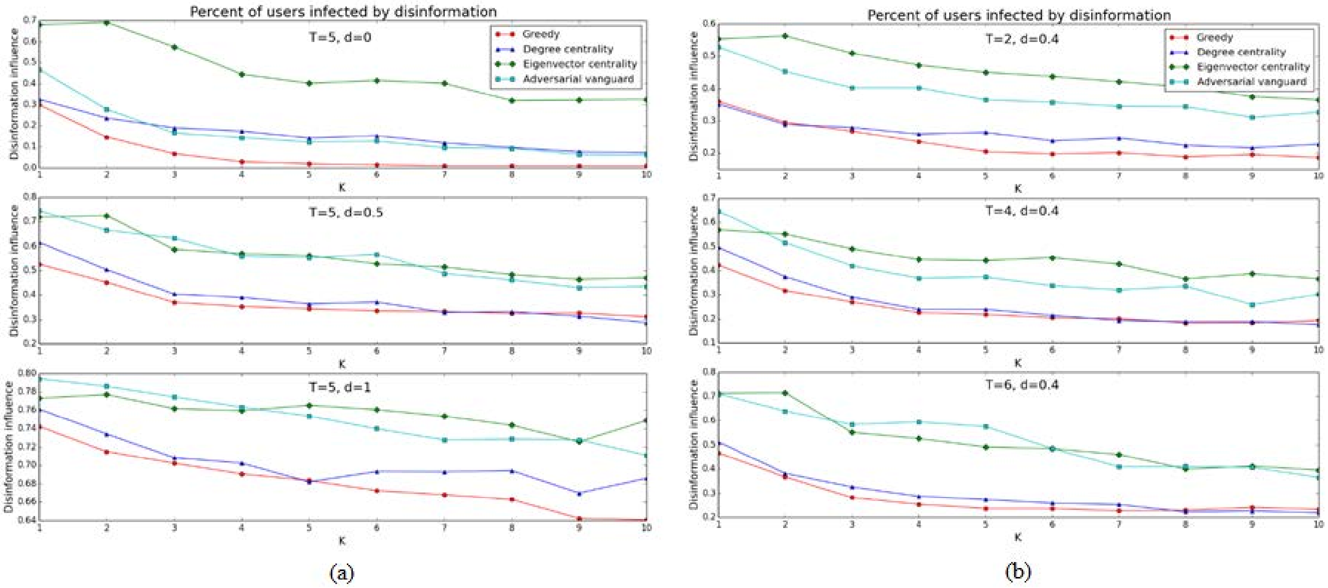

Figure 2. Evaluation of the greedy algorithm and the three heuristics in terms of the limited disinformation influence (i.e., percent of users infected by disinformation). X-axis denotes the number of source users for the counter campaign and Y-axis denotes the disinformation influence.

VI. Greedy Algorithm

By proving that the objective function of the time critical disinformation influence minimization problem is a monotonic submodular function, we are able to find a guaranteed near-optimal solution to the proposed problem. Specifically, we can apply the greedy algorithm which provides a (1-1/e) approximation of the optimal [22]. This algorithm builds the counter source node set S_C one node at a time, always greedily adding the node which results in the largest marginal gain in the objective function, i.e., the number of nodes that are saved by the counter campaign up to time T.

The value of function $\sigma(N_{i-1} \cup \{u\};\ T,\ S_A,\ d)$ can be estimated by simulating the random process. More specifically, we simulate the process multiple times for the set of $N_{i-1} \cup \{u\}$, resampling the transmission time along each directed edge in the social network from the distribution $f(\tau; \alpha_{ji})$ every time. Then a good estimate of $\sigma(N_{i-1} \cup \{u\};\ T,\ S_A,\ d)$ can be obtained by averaging all the sample values at each time.

VII. Heuristics

Because of the complexity of the proposed problem and the large scale of real world social network, even the polynomial time greedy algorithm may be too costly. We consider three different heuristics which can potentially perform well compared with the greedy algorithm.

The first two heuristics are structure-based including the "degree centrality" and "eigenvector centrality" [23], [24] which have been used to identify influential people in social network. The third heuristic is called "adversarial vanguard" which is the nodes that are expected to be infected by the disinformation just after the adversarial campaign is detected at time $t = d$. We will choose these adversarial vanguards as the source nodes for the counter information, since they are likely to drive the later propagation of disinformation. Specifically, we can implement this heuristic by running several rounds of simulation of disinformation propagation from S_A. For each round of simulation, we record the earliest infected k nodes after time $t = d$. The number of records of each node is its score. Then we select the top k nodes according to the score as the source node set for the counter campaign.

VIII. Experiment

We evaluate the proposed greedy algorithm and three heuristics including degree centrality, eigenvector centrality, and adversarial vanguard, on a sample of real online social network. We show that the greedy algorithm outperforms the three heuristics consistently and the rich transmission temporal dynamics information is critical in selecting the set of source users for counter campaign to limit disinformation influence in a time window.

A. Experimental setup

In order to evaluate the effectiveness of the greedy algorithm, we employ a Facebook-like social network originated from an online community for students at University of California, Irvine introduced in [25]. This online social network includes 1899 users that sent or received at least one message and 20296 directed edges among these users. To speed up the experiment process and focus on the solution quality evaluation, we further obtain a sample network from the original network by using Random

Jumping graph sampling algorithm [26]. The resulted social network G contains 500 users and 738 directed edges among these users.

For each directed edge $j \rightarrow i$ in the sample network, we associate it with a transmission rate α_{ji} drawn from a uniform distribution $\alpha_{ji} \sim U(0, 10)$. The transmission rate α_{ji} model how fast information spreads along the directed edge $j \rightarrow i$. We further suppose that the disinformation is released by three random users (i.e., $|S_A|=3$) at time $t = 0$. After a detection time delay d, the disinformation campaign is detected by authorities. Given the social network $G = (V, E)$, the set of source nodes of adversarial campaign S_A, the detection time delay d, our aim is to find the set of source nodes of counter campaign S_C such that the percent of users infected by the disinformation is minimized up to time $t = T$.

B. Experimental results

We set the time window $T = 5$. Fig 2(a) shows the percent of users infected by disinformation when the detection time delay $d = 0$ (the upper subgraph), $d = 0.5$ (the middle subgraph), $d = 1.0$ (the lower subgraph), respectively.

It is clear that the greedy algorithm outperforms the three heuristics consistently and degree centrality is the best among the three heuristics. We can also see that as the size K of the set of source users for counter campaign increases, the percent of users infected by disinformation decreases to a stable value generally. In addition, when the detection time delay d = 0, the stable percent of users infected by disinformation achieved by the greedy algorithm ($K = 10$) is nearly zero percent, while it can only be 64% when the delay $d = 1.0$. In other words, it is very vital for emergence responders to detect the disinformation campaign as quickly as possible.

Fixing the value of the detection time delay $d = 0.4$, Fig 2(b) shows the percent of users infected by disinformation when the time window $T = 2$ (the upper subgraph), $T = 4$ (the middle subgraph), $T = 6$ (the lower subgraph), respectively. Similar to the previous result, the greedy algorithm outperforms others and degree centrality is the best among the three heuristics generally. In addition, in terms of the disinformation influence limitation result, the difference between the greedy algorithm and the degree centrality heuristics is greater for smaller time window. In other words, the smaller the time window is, the more valuable the transmission temporal dynamics information become when selecting the set of source users of a given size for the counter campaign.

IX. CONCLUSION

In this paper, we have studied the problem of time critical disinformation influence minimization in online social networks. We mathematically formulize it as an optimization problem based on a continuous-time multiple campaign diffusion model. This model is able to account for different information transmission rates across different edges in the network. Even though this optimization problem turns out be a NP-hard problem, we present a provable guaranteed approximation algorithm by utilizing the monotone submodular property of the objective function. This approximation solution is able to find a set of source users for the counter campaign, which can minimize the disruptive influence of disinformation in online social networks in a time window. Experimental results show the presented approximation algorithm outperforms a set of heuristics for the optimization problem. In addition, the transmission temporal dynamics information is vital for selecting the set of source users for the counter campaign, especially when the time window is small.

Our future work lies in four aspects. First, we plan to consider more efficient heuristics to speed up the optimization problem and evaluate their effectiveness on real online social networks. Second, suppose every social network user has a damage cost representing how much loss she will suffer after being infected by the disinformation, then how to select a set of source users for counter campaign to minimize the total loss in a short time is an interesting question. Third, it would be interesting to take advantage of social balance theory [27] to address influence limitation problem in signed networks. Fourth, if the exact knowledge of who is the source of disinformation is not available to authorities, how to optimally limit the disinformation influence under the case of data missing is a promising direction.

We believe our work on time critical disinformation influence minimization in online social networks can help emergence responders effectively face the threats of increasing use of social media by terrorist organizations for adversarial purposes.

ACKNOWLEDGMENT

We thank Zhu Zhang and Saike He for invaluable discussions. This work was supported in part by the following grants: The National Natural Science Foundation of China under Grant No. 71025001, 71103180, 91124001, 61175040, and 71272236; The Beijing Natural Science Foundation under Grant No. 4132072; The Ministry of Health of China under Grant No. 2012ZX10004801; The Early Career Development Award of SKLMCCS; and by the Grant No. 2013A127.

REFERENCES

[1] O. Oh, M. Agrawal, and R. Rao, “Community Intelligence and Social Media Services: A Rumor Theoretic Analysis of Tweets During Social Crises,” *Manag. Inf. Syst. Q.*, vol. 37, no. 2, pp. 407–426, Jun. 2013.

[2] H. Chen, “Dark Web: Exploring and Mining the Dark Side of the Web,” in *Proceedings of the 2011 European Intelligence and Security Informatics Conference*, Washington, DC, USA, 2011, pp. 1–2.

[3] K. Lee, J. Caverlee, Z. Cheng, and D. Z. Sui, “Campaign Extraction from Social Media,” *ACM Trans Intell Syst Technol*, vol. 5, no. 1, pp. 9:1–9:28, Jan. 2014.

[4] United Nations Office on Drugs and Crime, “The Use of the Internet for Terrorist Purposes,” Oct. 2012.

[5] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2003, pp. 137–146.

[6] S. Bharathi, D. Kempe, and M. Salek, “Competitive influence maximization in social networks,” in *Proceedings of the 3rd international conference on Internet and network economics*, Berlin, Heidelberg, 2007, pp. 306–311.

[7] X. He, G. Song, W. Chen, and Q. Jiang, “Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model Technical Report,” in *Proceedings of the 12th SIAM International Conference on Data Mining*, Anaheim, CA, U.S.A., 2012.

[8] C. Budak, D. Agrawal, and A. El Abbadi, “Limiting the spread of misinformation in social networks,” in *Proceedings of the 20th international conference on World wide web*, New York, NY, USA, 2011, pp. 665–674.

[9] T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen, “Maximizing influence in a competitive social network: a follower’s perspective,” in *Proceedings of the ninth international conference on Electronic commerce*, New York, NY, USA, 2007, pp. 351–360.

[10] A. Borodin, Y. Filmus, and J. Oren, “Threshold Models for Competitive Influence in Social Networks,” in *Proceedings of the 6th International Conference on Internet and Network Economics*, Berlin, Heidelberg, 2010, pp. 539–550.

[11] N. Du, L. Song, M. G. Rodriguez, and H. Zha, “Scalable Influence Estimation in Continuous-Time Diffusion Networks,” in *Advances in Neural Information Processing Systems 26*, 2013, pp. 3147–3155.

[12] Manuel Gomez Rodriguez, D. Balduzzi, and Bernhard Schölkopf, “Uncovering the Temporal Dynamics of Diffusion Networks,” in *28th International Conference on Machine Learning*, Bellevue, Washington, 2011.

[13] M. G. Rodriguez, J. Leskovec, D. Balduzzi, and B. Schölkopf, “Uncovering the structure and temporal dynamics of information propagation,” *Netw. Sci.*, vol. 2, no. 01, pp. 26–65, 2014.

[14] Manuel Gomez Rodriguez and Bernhard Schölkopf, “Influence Maximization in Continuous Time Diffusion Networks,” in *29th International Conference on Machine Learning*, Edinburgh, UK.

[15] P. Domingos and M. Richardson, “Mining the network value of customers,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2001, pp. 57–66.

[16] M. Richardson and P. Domingos, “Mining knowledge-sharing sites for viral marketing,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2002, pp. 61–70.

[17] X. Zheng, Y. Zhong, D. Zeng, and F.-Y. Wang, “Social influence and spread dynamics in social networks,” *Front. Comput. Sci.*, vol. 6, no. 5, pp. 611–620, Oct. 2012.

[18] W. Chen, Y. Wang, and S. Yang, “Efficient influence maximization in social networks,” in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2009, pp. 199–208.

[19] W. Chen, C. Wang, and Y. Wang, “Scalable influence maximization for prevalent viral marketing in large-scale social networks,” in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 2010, pp. 1029–1038.

[20] W. Chen, Y. Yuan, and L. Zhang, “Scalable Influence Maximization in Social Networks under the Linear Threshold Model,” in *Proceedings of the 2010 IEEE International Conference on Data Mining*, Washington, DC, USA, 2010, pp. 88–97.

[21] S. Goyal and M. Kearns, “Competitive Contagion in Networks,” in *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, New York, NY, USA, 2012, pp. 759–774.

[22] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, “An analysis of approximations for maximizing submodular set functions—I,” *Math. Program.*, vol. 14, no. 1, pp. 265–294, Dec. 1978.

[23] M. S. A. L. Dipl. -Math. oec., D.-M. B. Friedl, and D. J. Heidemann, “A Critical Review of Centrality Measures in Social Networks,” *Bus. Inf. Syst. Eng.*, vol. 2, no. 6, pp. 371–385, Dec. 2010.

[24] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, “The Diffusion of Microfinance,” *Science*, vol. 341, no. 6144, p. 1236498, Jul. 2013.

[25] T. Opsahl and P. Panzarasa, “Clustering in weighted networks,” *Soc. Netw.*, vol. 31, no. 2, pp. 155–163, May 2009.

[26] J. Leskovec and C. Faloutsos, “Sampling from Large Graphs,” in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2006, pp. 631–636.

[27] X. Zheng, D. Zeng, and F.-Y. Wang, “Social balance in signed networks,” *Inf. Syst. Front.*, pp. 1–19, Jan. 2014.