Competitive Influence Minimization in Multi-Group Social Networks: An Opinion-Based Solution

Yuan Li, Jianming Zhu, Jianbin Jiao, and Qi Zhang

***Abstract*—Misinformation control has been a vibrant subject of research in online social networks (OSNs). With the diversity of OSNs, we observe that the emergence of group has also notably increased the exposure rate of misinformation. In the process of misinformation dissemination, the opinions adopted by an individual not only depend on the choices made by individuals' peers, but also highly depend on the individuals' own knowledge. However, we find that individuals' opinions are not consistent in the misinformation dissemination, and they are transferred dynamically. Motivated by these facts, we do some novel works on the problem of competitive influence minimization in multi-group OSNs. Firstly, we propose a novel dynamic competitive diffusion model. Secondly, a spontaneous mechanism and a contact mechanism are introduced to analyze users' opinion transfer processes, based on probabilistic discrete-time Markov chains. Thirdly, in order to make negative opinions be minimized, we use this model to work on a new Opinion Minimization (OM) problem. To quantitatively analyze this problem, a greedy algorithm of Equilibrium Opinion Minimization (EOM) is performed to select seed nodes. Furthermore, the experiments show that our proposed EOM algorithm outperforms the state of the arts, e.g., classical heuristic algorithms and local greedy algorithms of Influence Minimization in terms of minimizing opinion in steady-state opinion distribution.**

***Index Terms*—Competitive influence diffusion, markov chain, multi-group, online social networks, opinion transfer.**

I. Introduction

With the continuous development of network science and the rapid update of network technology, more and more new things have emerged on online social networks (OSNs) like multi-group OSNs, in which people acquire information not only from friends who have direct relations, but also from the groups they joined in. As a new module of OSNs, a group provides a convenient environment for the faster spread of information in large-scale networks. In multi-group social networks, users can be the audiences of information which they are interested in or follow hot topics at any time. Meanwhile, they can also be the producers and disseminators of their opinions. For example, many well-known doctors and scientists shared the correct anti-epidemic knowledge, in order to help people understand the new virus and protect themselves in the time of the COVID-19 suddenly broking out. Because of the convenience and virtuality, the multi-group OSNs become the primary site to access and share opinions/information for most people. The concerns are rising, which are the dissemination of authenticity and accuracy of the shared information, particularly the spread of some ideologies. Even so, misinformation is endless on the Internet. Research [1] has recognized that the spread of misinformation is faster and wider than good information. In multi-group OSNs, users may be influenced not only by their friends, but also by the members of group in whom they have no direct contact. It can be seen that the groups significantly increase the exposure of misinformation, and improve the frequency of misinformation interactions among people, and also expand the spread radius of misinformation [2]. For instance, on the evening of January 31, 2020, the two Chinese agencies jointly issued a piece of news that the study indicated that Shuanghuanglian oral liquid could inhibit the COVID-19 virus [3]. The news was spread widely in groups of OSNs in a very short time. Citizens in many places rushed to buy the oral liquids overnight, as a result of the drugs being sold out in all pharmacies and online shopping platforms. As Yang *et al.* [4] said, group influence usually will make individuals adopt the behaviors held by their neighbours, leading to the propagation of states throughout the network.

In recent years, the dissemination of misinformation in multi-group OSNs has become a major threat to public opinion, social stability, and economic development. Therefore, many researchers have begun to pay attention to the model of misinformation dissemination [5]–[8]. Lots of these models are based on general assumptions in order to fit more situations. However, these models do not take human knowledge into account. This is an important deficiency that these models ignore the differences in the acceptance of the authenticity of information among individuals, which consider only humans as network nodes, behaving homogeneously. In fact, it has been shown that different personality traits have direct influences on humans' interaction in OSNs [9], [10]. Otherwise, those models of misinformation dissemination only consider the

Manuscript received November 20, 2021; revised March 3, 2022; accepted April 13, 2022. Date of publication April 22, 2022; date of current version June 27, 2022. This work was supported by the National Natural Science Foundation of China under Grant 72074203. Recommended for acceptance by Dr. Shiwen Mao. *(Corresponding authors: Jianming Zhu; Jianbin Jiao.)*

Yuan Li is with the School of Electronics, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: liyuan110@mails.ucas.ac.cn).

Jianming Zhu is with the School of Emergency Management Science and Engineering, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: jmzhu@ucas.ac.cn).

Jianbin Jiao is with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: jiaojb@ucas.ac.cn).

Qi Zhang is with the School of Electronics, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: zhangqi203@mails.ucas.ac.cn).

Digital Object Identifier 10.1109/TNSE.2022.3168042

2327-4697 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

influence of users' friends who have direct relations in OSNs, and ignore the indirect influence of other users in their groups. Actually, information can be passed to others in groups, even though there is no direct ralations (network edges) among users in groups. And then, to get simplified models, the propagation algorithms adopted by these works, assume that users' opinions/states cannot be changed while the users are activated. That over-strong assumption ignored the fact the users' opinions will transfer from side to side according to their neighbors' opinions and also their own knowledge levels in OSNs. Abebe *et al.* [11] also commented that topic. Furthermore, most researchers limited their information dissemination models to a static network, and assumed that a user being activated in each round of dissemination has the same disseminated probability. So many researchers used the method named "flipping a coin" [12] to convert the network to a static propagation network when establishing propagation model. However, the probabilities of people spreading and receiving information will change over time in the real world. Therefore, when establishing an information dissemination model, it is necessary to take account into the impact of individual heterogeneity, the network topology structure and dynamic dissemination.

Additionally, in order to control of misinformation dissemination, dissemination models are often explored to study the problem of Influence Maximization [13] and Influence Minimization [12]. The goal of works on those problems is to find the sets of users to launch information diffusion process, which will have on a maximal or minimal impact on the numbers of activated users. As a new extension and application of Influence Maximization problem, Opinion Maximization problem [49] has also been studied. The goal of the opinion maximization problem is to maximize the overall opinions rather than the number of activated nodes. However, to the best of our knowledge, the Opinion Minimization problem has not been studied.

Those problems motivated us to work on the dynamic misinformation influence minimization in multi-group OSNs. We propose a novel diffusion model that incorporates both knowledge of users, their own personal opinions and multi-group network topology. For solving this problem, we propose a new extended method of Influence Minimization (IM), named Opinion Minimization (OM) problem, and designed an algorithm of Equilibrium Opinion Minimization (EOM) to select seed nodes. Our goal is to select the top-k nodes starting a truth campaign to minimize the sum of the equilibrium negative opinions for users affected by misinformation. The reason for formulating the Opinion Minimization problem is that in the previous IM problem, each view of misinformation can only be activated or inactivated, but in the real situation, the activated node may also have positive, neutral, and negative opinions. Therefore, in terms of misinformation control strategies, calculating and minimizing overall negative opinions shall play a more important role. In summary, the contributions of this paper are listed as follows:

1) The proposal of a novel dynamic competitive model for misinformation diffusion that considers the heterogeneity of individual knowledge, the impacts of individual's opinions on adopting misinformation, and also the realistic group topology in OSNs.
2) Based on the user's attitude towards misinformation, we propose a new way to classify user opinions.
3) We formulate an individual's opinion dynamics for the changers towards a misinformation based on a Markov chain model. Then we get the stationary distribution of the proposed process, and also find the steady-state solutions of three opinions.
4) Proposed algorithm of Equilibrium Opinion Minimization (EOM) effectively deals with the new problem of Opinion Minimization.

The rest of the paper is organized as follows. Section II includes the related work. We describe our network model in Section III. The dynamic competitive influence diffusion model, individual's opinion dynamics model and OM problem are shown in Section IV. Experimental results are given in Section V. Finally, Section VI concludes this paper and discusses our future work.

II. Related Work

A. Information Diffusion Model

Information includes true/positive information and fake/negative information. The diffusion of misinformation is similar to the spread of epidemics at the beginning, so many researchers use epidemic models to simulate the spread of misinformation in social networks [24]. As a continuous diffusion model, the two-state model (SIS) and the three-state model (SIR) are the two most used models. Domingos and Richardson [19] firstly put forward the problem of influence maximization (IM). Then Kempe *et al.* [13] transformed the IM problem into a discrete optimization problem, and proposed two diffusion models, the linear threshold model (LTM) and the independent cascade model (ICM). They also designed a greedy algorithm with $1-1/e-\varepsilon$approximation ratio since the function was submodular. Motivated by this work, a lot of works for IM (e.g., [25], [26], [27], [28], [29]) have been developed. But in the classic IM problem, the influence probability of node was assumed to be constant during each round of information dissemination, so the dynamics of the network couldn't be expressed. Therefore, Kempe *et al.* proposed a declining cascade model [30], which considered the attenuation effect of the influence between nodes. Also, many scholars have made an extension on the classic ICM and LTM. Hossein *et al.* [31] proposed an extended model of IC based on the distrust factors in real social networks. This model was called the perceptual cascade (SC) model. Feng *et al.* [32] proposed a multi-level attitude linear threshold model (LTMLA). Compared with the traditional LTM, this model proposed a method to model the positive and negative attitude towards an entity in the signed social network and the effect of interactions among users. In addition to the classic IC and LT model, some scholars have proposed other information dissemination models. In addition to the two classic models, some scholars have proposed other information dissemination models. Hosni *et al.* [5] considered the individual and social behaviors in OSNs, and proposed an individual behavioral statement

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

which simulated damped harmonic motion for rumors. Inspired by [6], Wang *et al.* [8] proposed an energy model to simulate rumor dissemination based on user experiences. Indu *et al.* [7] employed the concept of forest fire model to model the propagation of rumors in the Twitter dataset.

B. Misinformation Control Strategy

In the part of controlling misinformation dissemination, scholars have also proposed a new problem on influence minimization. It is the inverse problem of the influence maximization and its purpose is to design effective strategies to minimize the negative influence. The strategies can be divided into three categories: (i) Blocking a limited number of links [33], [34] in social networks, which usually removed a set of edges that plays a key role in information dissemination, so that the amount of dissemination is as low as possible. (ii) Blocking influential nodes [35], [5], [42], [43]. That method usually selected the most influential nodes in the network according to certain criteria, and removed them from the original network. Neither the blocking links nor the blocking nodes strategy considered the user experience in real social networks [8]. (iii) Choosing protector [12], [40], [41]. This method was to select a set of nodes and publish positive information which would be spreaded on the Internet. If users received positive information, then users would not spread misinformation. This was a way of competition between positive information and negative information. In summary, the first two strategies would destroy the network structure, and the latter strategy was to counteract the negative effect of misinformation by positive information. However, in real life, forcibly blocking nodes or links will cause users' bad emotions to a certain extent. Therefore, in our research, we adopted the competition method and choosed the top-k nodes to post positive information to guide user's opinions.

C. Opinion Maximization

The spreading process of misinformation can be understood as the dissemination of opinions. Some scholars have begun to pay attention to user's opinion. The Opinion Minimization problem is central problem studied in this paper, which is the inverse problem of Opinion Maximization. For a long time, the problem of Opinion Maximization has attracted extensive attentions of researchers. Gionis *et al.* [49] first studied Opinion Maximization problem which is defined as a process of choosing some set of k nodes so that positive opinion towards a given topic is maximized. Zhang *et al.* [50] studied the OM problem through a two-phase model called Opinion-based Cascade model based on the LT model. However, the opinion of each activated node only changed once. Shen *et al.* [51] proposed a LT-S diffusion model in signed social networks to study OM problem. The LT-S model was an extension of LT model incorporating both positive and negative opinions. However, the authors neglected the dynamics of user opinions. Liu *et al.* [52] considered a problem called AcTive Opinion Maximization (ATOM) to find a set of seed users to maximize the overall opinions spread toward a target product in a multi-round campaign. They assumed that a user opinion is derived from user preference data and the user opinion was unchanged after being determined. More recently, Nayak *et al.* [53] researched an efficient information diffusion using dynamic Bayesian networks with the goal of maximizing positive opinions of a chosen topic. He *et al.* [54] studied positive opinion maximization by using an Activated Opinion Maximization Framework (AOMF) in signed social networks. Hudson *et al.* [55] studied the OM problem by adopting the "Big Five" model from the social sciences and proposed a behavioral independent cascade (BIC) model that considers the personalities and opinions of user nodes when computing propagation probabilities for diffusion. The Opinion Maximization problem is suitable for solving the diffusion of positive opinions or innovations, but it is not applicable for negative opinions. Thus, in order to solve the problem of dissemination of misinformation, we innovatively propose the Opinion Minimization problem.

III. Network Model

In some works, an OSN is generally formulated as a directed or undirected graph $G(V,E)$, consisting of the set of nodes Vrepresenting the users, the set of edges Edenoting the relations among individuals (e.g., friendship, follows and cooperation). Let $|V| = N$ denotes the number of nodes and $e(u,v) \in E$ denotes the directed edge from node $u$to node $v(u,v \in V)$and that is to say node v is a child node of node u. In this kind of classical network, the nodes which have strong relations represented by edges just like friends, colleagues, or families, information may be propagated among users' existing edges. However, considering the diversity of OSNs, we observed there were different multiplex structures in OSNs. In the previous works Kuhnle *et al.* [38] presented community structures based on multiple social networks where information dissemination is also passed on edges between a pair of nodes. Ghoshal *et al.* [44] leveraged the community structure in the OSN to pick up the seed nodes statically to combat the spread of misinformation faster.

In our study, we focus on multi-group structures in an OSN. Different to classical network, in groups, information dissemination is by groups, no longer by edges. This indicates that information can be disseminated not only in strong relations, but also in weak relations. In fact, the multi-group structure is discovered in many broadly known and current social network, e.g., WeChat, Facebook, Youtube. For individuals, they could join several groups in an OSN simultaneously.

We model the multi-group OSN as a directed acyclic social network graph $G(V,E,g)$, where $V$is a node set (each node represents an individual), $E \subseteq V \times V$ denotes an edge set (each edge represents the connection relationship between two adjacent individuals), and $g$denotes a set of all groups in the OSN, $g^i(v_i, e_i) \in g$ is a small group formed spontaneously by users. $v_i$and e_i represent the node set and edge set in group g^i and $V(i) \subseteq V$, $E(i) \subseteq E$. As illustrated in Fig. 1. In Fig. 1, we can see that there are 10 nodes, and three groups, $g^1, g^2, g^3 \in g$. The user 1 and 5 belong to both $g^1$and g^2. The user 3 belongs to both g^2 and g^3. However, the user 1, 4 and 8 have no any relations in group g^1.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

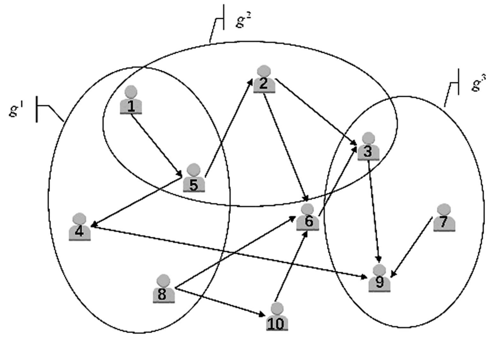


Fig. 1. An example to explain multi-group OSN $G(V, E, g)$ with 3 groups and 10 nodes.

The multi-group OSN is a novel network structure where groups are added to the ordinary social network. These groups are formed because of the same concern or topic of users, so that individuals with the same interests get together in a small cluster established by one user. There can be many groups in an OSN, and an individual can also belong to many groups. The relations between users in a group are uncertain. In other word, they may be friends, or followers, or they don't know each other. The dissemination of information in groups is different from the way in ordinary social network relying on direct relation of users. Even if a message is sent in a group where people don't know each other, it will be visible to all users in the group.

IV. Dynamic Competitive Influence Diffusion Model Formulation

In this section, the formulation and methods of our proposed model will be described in detail.

A. *Individual Heterogeneity Toward Diffusion Model*

In the formulated diffusion model, one important factor is that individual heterogeneity needs to be considered, although it is difficult to model humans' emotions, knowledge and strategies. Afassinou [56] had pointed out education significantly contributes to the rumor spreading cessation. It showed that the individual's background knowledge played a crucial role in misinformation diffusion process. Hence, we choose that quantifiable and important indicator as a parameter of heterogeneity. In total, an individual with much more knowledge has a more comprehensive understanding on an information or an object. For a misinformation, it is also easy to understand that people with high knowledge can analyze the credibility and authenticity of the information from more dimensions than people with low knowledge. Therefore, we defined knowledge as the individual's ability to evaluate the reliability and trustworthiness of misinformation. For each node $v \in V$, we give a parameter b representing the lack degree of knowledge of an individual. The value of b varies between [0, 1]. The larger the value, the less knowledge an individual owns.

When modeling the influence diffusion model, we also consider the misinformation attraction for different users. The success of misinformation diffusion between users indicates that an activated node is attracted by the misinformation and this node will choose to send this misinformaton to neighbors or group members. In order to better simulate the real situation, we define the value of attraction of misinformation for each individual. In the initial stage of the difussion process, a misinformation usually appears in an eye-catching way, so people are particularly attracted to it. At this stage, due to the attraction of the misinformation to the user is an easy dissemination process, the value of attraction shows a trend of rapid growth. But with the increase in the number of the same misinformation received, the value exhibits a gradual downtrend [6]. This phenomenon is easily understood from the experience of diminishing marginal effects. In addition, due to the individual heterogeneity of knowledge, the attraction of misinformation to different node will change. As being analyzed above, people with more knowledge are more likely to distinguish the authenticity of information. Therefore, misinformation will be less attractive to high-level knowledge users. Hence, we try to construct a function to simulate the change of attraction in different receiving points.

While analyzing the individual behavior in multi-group OSNs, we use the curve function in Planck's blackbody radiation law to fit our attraction problem because we find that the tendency of the attraction function accords with the curve appearance of the radiance of blackbody, wavelength and temperature in Planck's blackbody radiation law. In physics, Planck's blackbody radiation law describes that blackbody is an idealized object that completely absorbs external radiation of any wavelength without any reflection under any conditions. So, the absorption ratio of blackbody is 1. As the temperature is different, the color of the blackbody begins to change differently, showing a gradual process. When calculating the energy density of a blackbody, Planck's hypothesis gives the following theoretical expression for the power intensity [57]:

$$I(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \cdot \frac{1}{e^{hc/\lambda kT} - 1}, \tag{1}$$

where c is the speed of light in vacuum, h is called Planck's constant and k is Boltzmann's constant. In other words, the intensity $I(\lambda, T)$ of blackbody radiation depends on the wavelength λ of the emitted radiation and on the temperature T of the blackbody. Inspired by this function, we use an analogy that the attraction of misinformation to an individual is regarded as a function of individual knowledge background and individual receiving times. In order to conform to our model, we simplify the constant terms in (1) and reduce the power of the power of parameter to ensure that the value range is between [0, 1]. Therefore, we define the attraction of misinformation to an individual v as follow

$$A_v(t, b_v) = \frac{1}{t^2} \cdot \frac{1}{e^{1/tb_v} - 1}. \tag{2}$$

Here, $A_v(t, b_v)$is the attraction of misinformation to an individual v. The parameter $b_v$is the lack degree of knowledge of individual v. The parameter $t$is the times of receiving a

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

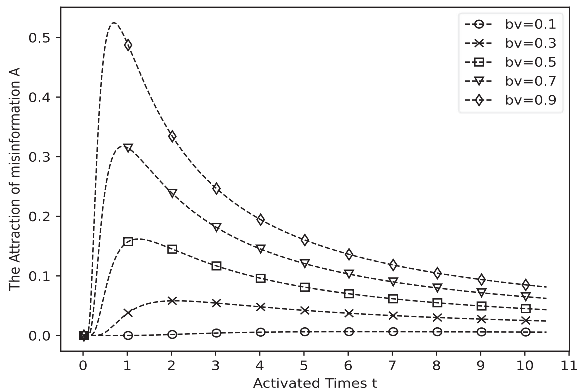


Fig. 2. The attraction of misinformation to users with different b_v. The horizontal axis indicates activate time t. The vertical axis represents the value of the attraction of misinformation to users $A_v(t, b_v)$.

misinformation for individual v. In (2), the individual's knowledge is analogy with the temperature, and the times of receiving misinformation is analogy with the wavelength in (1). That is to say, with the different knowledge, the attraction of misinformation begins to change differently, showing a gradual process similar to (1).

During the propagation process, the time when a node receives the misinformation for the first time is denoted as t_0. According to (2), we draw a line graph of $A_v(t, b_v)$ under different b_v showed as Fig. 2. In order to measure and visualize the user's b_v, we take 5 values to summarize the lack degree of knowledge for users. We can see that the greater the b_v is, the higher attraction of misinformation is. Highly educated people tend to recognize misinformation easily, therefore this kind of people is less likely to be attracted by misinformation along with the crowd. On the contrary, a person with a low level of knowledge is more likely to be attracted by misinformation, thus has a higher probability of spreading it. In addition, when the individual firstly receives a misinformation, the individual has a process of absorbing and understanding this message. Therefore, starting from t_0, the attractiveness of the misinformation to the individual has a rapidly increasing process. When the individual fully understands this message, the attractiveness of the misinformation will gradually decrease after been activated. The characteristic of rapid decay after the attraction reaches its peak is also a manifestation of the dynamic diffusion in the network. This is different from the traditional IC model that assumes that a user can only be activated once in the OSN. Based on the real situation, we believe that users are likely to receive the misinformation repeatly in a misinformation spreading cycle. If the user pays attention to it, he/she is activated by the message. It should be emphasized that the activated user does not necessarily spread the message. Next, we will introduce the activation process in detail.

B. *Propagation Process of Misinformation in Multi-Group OSN*

In a multi-group OSN G, each edge $e(u, v) \in E$contains a parameter p_{uv}, which represents the activated probability of node u sending the information to node v, and v accepting it. We call $p_{uv}$the successful probability of $u$activating v. Therefore, if the infected node $u$activates the node v successfully, then it indicates that v is activated. Otherwise, v is inactive. But in many previous studies, the successful activated probability p_{uv} was considered as a system parameter and is set artificially at the very beginning of the propagation process. In this way, the entire misinformation propagation process is equal to the process on a static network. As a result, it loses the dynamics in the real spreading process. However, we consider successful activated probability contains two aspects: the sending side and the accepting side. Therefore, based on the work in [8], the successful activated probability for u to v can be defined as follows

$$p_{uv}(t) = p_u{}^{send}(t) \cdot p_v{}^{acc}(t). \tag{3}$$

For the sending side, the node u is an activated node by a misinformation, and he/she supports this misinformation and will to send it to other nodes at time step t. The possibility $p_u{}^{send}$ represents the probability that an activated node will to spread the information. According to the above section, it is conceivable that the attraction of misinformation to user is an important factor in whether users are willing to forward the information after receiving the information. Thus, the possibility $p_u{}^{send}$ depends on the attraction of misinformation to an individual $A_u(t, b_u)$. In addition, the influence of the impression between individuals will also have an impact on the information diffusion. Generally, people are more willing to believe and follow people who are more educated or prestigious than themselves. Li *et al.* [10] researched social hot events in hot weibos (like tweet) and found that verified users tend to catch more attention. Most users are more willing to forward weibos posted by verified users. Here, the verified users are people who have an amount of knowledge and a high influence in a certain domain. Therefore, whether an individual is willing to spread misinformation also depends on who sent it to him/her. Kempe *et al.* [13] also claimed that nodes with higher in-degree have a greater authority to influence other nodes. Therefore, we use the ratio of the lack degree of knowledge b of spreading node s and accepting node u to evaluate the influence of the impression in the misinformation diffusion. Consequently, we define the sending probability of node u as follows

$$p_u{}^{send}(t) = A_u(t, b_u) \cdot \frac{b_u}{b_s + b_u} \cdot \alpha. \tag{4}$$

Here, α is an implicit parameter in the propagation process. b_u is the lack degree of knowledge of node u. b_s refers to the lack degree of knowledge of node s from which node u obtains the misinformation. That denotation can be interpreted as that the spreading nodes $u$with lower b_u(higher knowledge) have a greater authority to influence the spreading willing of node v, and this node u cannot be influenced by others easily. But it needs to be pointed out that in our model, we assume that the sending probability of the initial negative nodes and initial positive nodes are equal to 1.

For the receiving side, we need to discuss it separately, because the receiving method is different in the group and non-group of OSNs. This is the reason that we study the spread of

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

misinformation in multi-group networks. First, we talk about receiving misinformation in non-group. In the graph $G(V, E, g)$, if $e(u, v) \in E$and the activated node u sends the misinformation to the child node v through $e(u, v)$, acceptance probability of node v can evaluate the chance node v accepting a misinformation from their neighbor. The acceptance probability of node v is not only related to the degree of node v, but also related to the degree of the parent node u. Because the nodes with high-degree have a greater ability to influence their neighbors, in a meanwhile, they cannot be easily influenced [5]. Thus, in non-group OSNs, by considering the impact of both the sender u and the receiver v, the acceptance probability of node v can be formulated as

$$p_v{}^{acc}(t) = \frac{d(u)}{d(v) + d(u)}, \tag{5}$$

where $d(u)$refers to the out-degree of the node u, and $d(v)$ refers to the in-degree of the node v. During the propagation process, once a node v accepts a misinformation, it is activated. Thus, the attraction of misinformation to the node v begin to follow (2).

Second, in the group of OSN $G(V, E, g)$, due to the characteristics of the group, all the members of the group have an authorization to accept the information sent in the group. The dissemination of information in the group does not need to be based on direct ralations. However, the influence of the edge between users has little effect on the users in the group. But if an individual is willing to pay attention to the group, he/she has a greater chance of receiving the information. And we consider that the individual's willingness to pay attention to the group depends on the individual's stickiness to the group. To measure this stickiness, we use the number of friends the node followed in the group and the total number of nodes in the group as a description.

$$p_v{}^{group\_acc}(t) = \frac{|F(g^i(v))|}{|g^i(v)|}, \tag{6}$$

where $|F(g^i(v))|$represents the number of friends of node v in the group g^i, and $|g^i(v)|$represents the number of all nodes in the group g^i.

According to the above analysis, we derive (7) to calculate the probability that node v is activated by the spreading node u successfully. Here, we assume that in a time step, a node can only be activated by one parent node, or not be activated. However, even if an individual has accepted a misinformation and becomes an activated node, it does not mean that the node will pass it on. Because whether the activated node spreads the misinformation is determined by p^{send}.

$$\begin{aligned} p_{uv}(t) &= \begin{cases} p_u{}^{send}(t) \cdot p_v{}^{acc}(t), & e(u,v) \in E \\ p_u{}^{send}(t) \cdot p_v{}^{group\_acc}(t), & \{u,v\} \subseteq g(i) \cap e(u,v) \in \emptyset \end{cases} \\ &= \begin{cases} A_u(t, b_u) \cdot \frac{b_u}{b_s + b_u} \cdot \frac{d(u)}{d(v)+d(u)} \cdot \alpha \\ A_u(t, b_u) \cdot \frac{b_u}{b_s + b_u} \cdot \frac{|F(g^i(v))|}{|g^i(v)|} \cdot \alpha. \end{cases} \end{aligned} \tag{7}$$

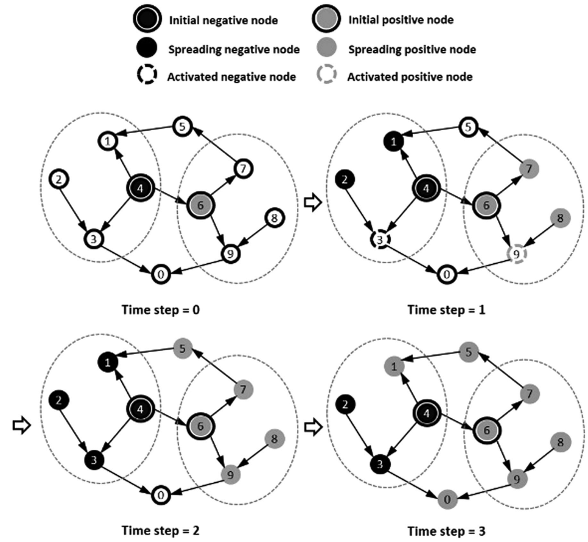


Fig. 3. An illustrative example of competitive spreading process in an OSN with two groups and 10 nodes.

In order to minimize the influence of misinformation, we select several nodes to spread positive information for competing with misinformation. It should be noted that nodes influenced by positive nodes will no longer be influenced by negative nodes in the dissemination of misinformation. For example, there is a network of 10 nodes with 2 groups in Fig. 3. It demonstrates the mechanism of competition in social networks. It shows a misinformation propagation process with a positive node in the social network. We assume that node 4 is the initial negative node and node 6 is the initial positive node at the beginning. In the time step 1, node 4 spreads the misinformation to node 1, 2, 3, 6 based on the links and group relationship. In this process, node 1, 2 and 3 are activated, and only node 1 and 2 are willing to spread it with probability $p_1{}^{send}$ and $p_2{}^{send}$. On the other hand, the initial positive node 6 sends positive information to neighbors and group members. Then, node 7, 8 and 9 are affected and node 7 and 8 are willing to spread positive information with probability $p_7{}^{send}$ and $p_8{}^{send}$. Although node 9 does not propagate, it is no longer affected by negative nodes. In time step 2, node 5 and 9 are influenced by node 7 and 8 to become spreading positive node. Meanwhile, node 3 is affected by spreading negative node 2 again and becomes a spreading negative node. In the time step 3, nodes 5 and 9 post positive information to nodes 1 and 0, respectively. Successfully, nodes 1 and 0 are both affected and they will continue to influence their neighbors and group members as they spread.

C. Individuals' Opinion Dynamics Formulation

1) Opinion Classification: Diffusion processes of misinformation in OSNs are based on spontaneous processes and contact processes [11]. As described above, due to the differences in knowledge levels of individuals in multi-group OSNs, different people have different opinions in a same misinformation. When users accept a misinformation, they may support, ignore, or refuse the misinformation based on their

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

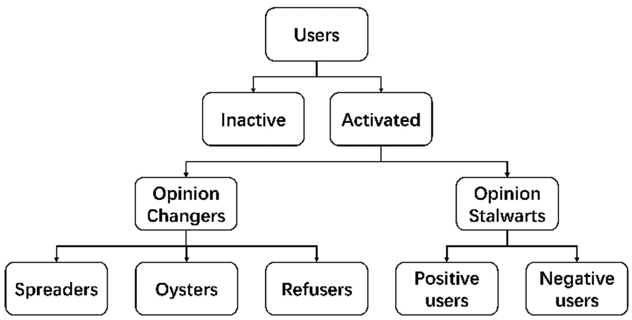


Fig. 4. The categories of user for opinion. Negative users and Spreaders all support and spread the misinformation, and Positive users and Refusers don't believe and deny the misinformation. The difference is that Stalwarts' opinion will not change, Changers' opinion may change over time.

subjective opinion. This is the reason why a node may not necessarily spread it after being activated by the misinformation. The process of an individual changing and revising the opinion based on their own knowledge levels without any external interference, is a spontaneous process. On the other hand, following a contact with neighbors who are highly educated or spread positive information, an individual's opinion might change. This process is a contact processe. Therefore, based on the obove two mechanisms we can firstly divide the user's opinions into two categories. One type is called opinion stalwarts. Once the opinions of this type of users are formed, they will not change. For example, those are the negative users who initially spread the misinformation, the selected positive users who post the anti-misinformation, and the users who are affected by positive information. The other type is called opinion changers. The users will make judgments and transfer their opinions on misinformation based on their own knowledge, experience and their neighbors' opinions at different stages. It can be seen that the opinion stalwarts' state is very stable, but opinion changers can lead to uncertainty in dissemination. Their opinions will mostly determine the speed and breadth of the spread of misinformation. Therefore, we mainly study the changers' opinion.

We further divided opinion changers into 3 states: 1) spreaders, believing and spreading the misinformation. 2) oysters, not spreading misinformation due to losing interest in it or questioning it, and 3) refuser, denying the misinformation based on own knowledge. In Fig. 4, we detail the categories of user's opinion.

2) A Markov Chain Approach for Opinion Dynamics: In order to simulate the real situation more accurately, we claim that the opinion of the user may change when the misinformation was received multiple times if the user is an opinion changer. And the change of the user's opinion only depends on the existing state, it has nothing to do with the past state. Because users need to make judgments based on their current state, and combine with the authority of the sender of misinformation to adjust their new state. But in the process of misinformation dissemination, once users receive the positive information send by positive nodes, they will become positive stalwarts. Since this change is inevitable, we only discuss the state transition of activated users who are not influenced by the positive node. According to this condition, we construct a discrete-time Markov chain to represent opinion changer's opinion dynamics model. Because discrete-time Markov chains is a great tool that enables the description of individual node dynamics as well as the determination of the macroscopic critical properties and the whole phase diagram [45].

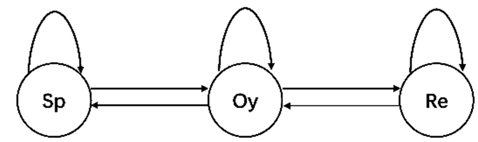


Fig. 5. A transition state of the proposed Markov chain process including 3 opinion states.

Before formulating the model for opinion transfer, we consider what conditions will change user's opinion. Inspired by the opinion susceptibility problem that was proposed by Abebe *et al.* [11], in which each agent had a fixed innate opinion that reflected the agent's intrinsic position on a topic, and a resistance parameter representing susceptibility to persuasion. In fact, whether it is believed in misinformation mostly depends on people's subjective judgments. Different knowledge backgrounds also determine different subjective judgments. Therefore, we use the lack degree of individual's knowledge b to judge the spontaneous process in our work. Moreover, mentioned in above sections, if an activated node by misinformation is not willing to spread the message with the sending probability p^{send}, the user will not become a spreader. To a certain extent, the sending probability also includes the influence of the parent node in the contact process, so the sending probability p^{send} should also be considered when building the model.

Subsequently, we let $\{X_t, t = 0, 1, 2, \cdots\}$ be a stochastic process which takes on a finite or countable number of possible values. If $X_n = i$, then the process is said to be in state i at time t. We define a state space $S = (Sp, Oy, Re)$. In graph $G(V, E, g)$, for any node $v \in OC$($OC$is opinion changers), $Sp_v(t)$, $Oy_v(t)$ and $Re_v(t)$are random variables indicating whether the node v has a supporting, questioning, refuting opinion about a misinformation at time step t, respectively. We assume that each time individual opinion will be updated and only related to the state at the previous moment. We can get $P\{X_{t+1} = j|X_t = i, X_{t-1} = i_{t-1}, \cdots, X_0 = i_0\} = P\{X_{t+1} = j|X_t = i\}$ where $i, j \in S$. Let P denote the matrix of one-step transition probabilities P_{ij}, so that the transition probability matrix is

$$\forall v \in OC, P_v = \begin{bmatrix} {p_v}^{send} & 1 - {p_v}^{send} & 0 \\ {p_v}^{send} & (1 - {p_v}^{send})b_v & (1 - {p_v}^{send})(1 - b_v) \\ 0 & {p_v}^{send} & 1 - {p_v}^{send} \end{bmatrix}. \tag{8}$$

In Fig. 5, we present the proposed Markov chain process diagram that illustrates the different transitions of individual opinion. And the spreaders and refusers are two poles in this process. Because we assume that the transition from one pole to another should at least pass-through oysters about a misinformation.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Dynamic Competitive Diffusion Model.

Input:
$G(V, E, g)$ including user's knowledge parameter b
Set of initial negative nodes NS who spread misinformation
Delayed detection time step tc
Set of the positive nodes PS
Output:
The diffusion of competitive influence
1: Initialize NS andPS, ensure $NS \cap PS = \emptyset$, and let $state_v =$ inactive, $state_{NS} =$ negative, $state_{PS} =$ positive, influence$_time_v = 0$
2: **while** $NS + PS \neq \emptyset$ do
3: **for** user $u \in NS + PS$ do
4: user $v \in changers$
5: **If** (1) user $v \in$child node of u or (2) user $v \notin$ child node of u and u, v are in the same group:
6: **If** user v does not influenced by u:
7: user v is in the state of inactive
8: **Else if** user v is activated with probability$p_{uv}(t)$:
9: influence$_time_v += 1$
10: **If** node u is inPS:
11: $state_v = \ positive$
12: $v \rightarrow$ positive nodes
13: $changers = changers/\{v\}$
14: **Else:**
15: **If** $state_v == \ inactive$:
16: $state_v = \ activated$
17: **If** user v spread the misinformation with probability $p_v^{sent}(t)$:
18: **If** $state_v == \ refusers$:
19: $state_v = \ oysters$
20: **Else** if $state_v == \ positive$:
21: $v \rightarrow$ spreading positive nodes
22: **Else:**
23: $state_v = \ spreaders$
24: $v \rightarrow$ spreading negative nodes
25: **Else:**
26: **If** $state_v == \ spreaders$:
27: $state_v = \ oysters$
28: **Else if** $state_v == \ oysters$:
29: **If** user v refuse it with probability $(1 - p_v^{sent}(t))(1 - b_v)$:
30: $state_v = \ refusers$
31: **Return** the diffusion of competitive influence

According to the previous analysis, Algorithm 1 presents the details of the dynamic competitive influence diffusion model. Starting from line 5, the influence diffusion in the group and non-groups are discussed separately. Lines 17 to 30 update the states of the nodes.

D. *Proposed Opinion Minimization Problem*

Now our goal is to minimize the influence of a misinformation as much as possible. Different from the previous Influence Minimazation problem to minimize the amount of activated negative nodes at the end of propagation process, in order to make negative opinion towards a misinformation be minimized in multi-group OSNs, we formulate a novel Opinion Minimization problem which is defined to be the process of choosing a set of k positive nodes and minimizing the negative influence. In this formulation, we take advantage of the steady-state convergence theorem of the Markov chain to analyze the stationary distribution of states of opinions. Then, we select the targeting positive nodes by constructing the negative Equilibrium Opinion Minimization framework. At last, targeting positive nodes are discovered and located precisely.

1) Steady State Analysis: In this section, we perform a steady-state analysis for the Markov chain of opinion transfer.

According to the steady-state convergence theorem of the Markov chain, it is required to satisfy irreducible and aperiodic features. Then the stat j and steady-state probability π_j have the following characteristics:

$$\lim_{n\to\infty} r_{ij}(n) = \pi_j, \forall i, j \in S. \tag{9}$$

Theorem 1. The proposed Markov chain is irreducible.

Proof. Theorem 1 can be proved rather directly. It is easy to observe that $\forall i, j \in S$, $\exists n$, let $p_{ij}^n > 0$ through the transition probability matrix. That is to say, the state j is accessible from i through n steps. In the same way, the state i is also accessible from j through n steps. All states are mutually reachable. Therefore, the proposed Markov chain is irreducible. ∎

Theorem 2. The proposed Markov chain is aperiodic.

Proof. In the transition probability matrix, we set $d(i)$as a period of state i. $d(i) = \gcd |\{n > 0 : p(X_n = i|X_1 = i) > 0\}$, here $\gcd |\{\cdot\}$represents to take the greatest common divisor in the set. For the three states, $d(Sp) = \gcd |\{1, 2, 3, 4 \cdots\}$, $d(Oy) = \gcd |\{1, 2, 3\}$, $d(Re) = \gcd |\{1, 2, 3, 4 \cdots\}$. Their periods are equel to 1. Therefore, the proposed Markov chain is aperiodic. ∎

Through the above analysis, the Markov chain of opinion transfer is irreducible and aperiodic. We can get its balance (11). Then the stationary distribution of the state π_j satisfies the following conditions:

$$\sum_{j\in S} \pi_j = 1, \tag{10}$$

$$\pi_j = \sum_{j\in S} \pi_i P_{ij}, (i, j \in S). \tag{11}$$

The steady-state solution through the above two equations can be acquired. For simplicity, we use abbreviations as p and b to represent $p_v{}^{send}$ and $b_v$in transition probability matrix P. The solution of this system is presented as follows

$$\begin{cases} \pi_{Sp} = \frac{p^2}{p-(p-1)(-p+b(p-1)+1)} \\ \pi_{Oy} = \frac{-p(p-1)}{p-(p-1)(-p+b(p-1)+1)} \\ \pi_{Re} = \frac{-(p-1)(-p+b(p-1)+1)}{p-(p-1)(-p+b(p-1)+1)}. \end{cases}$$

2) Opinion Minimization Problem Definition: Based on the above steady-state solution, we propose an Opinion Minimization (OM) problem which aims to minimize the total negative opinion objective function $\xi(\cdot)$, defined as

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

$$\min \xi(TS)= \mathrm{E}\left[\vec{1}^{\mathrm{T}} \sum_{u \in OC}^{N} \pi_{Sp}^{u}(T)\right]. \quad (12)$$

Here, TS denotes the selected set of target positive seed. $\pi_{Sp}^{u}(T)$represents the steady-state probability of the state Sp which is referred to user $u$support and spread misinformation by the end of a time interval T. $\mathrm{E}[\vec{1}^{T} \sum_{u \in OC}^{N} \pi_{Sp}^{u}(T)]$ denotes the expected number of overall steady-state probabilities of nodes owned the state Sp. T is a time horizon for the diffusion process.

Our proposed OM problem is completely different to Influence Minimization (IM) problem. Though those two problems both want to find k target nodes to launch a dissemination process, the core factors are how opinion and activation are represented and formulated. In our problem, opinion is represented as a real value obtained from the sum of steady-state probability, however, activation in IM is a binary state (active or inactive). Therefore, IM aims to minimize overall influenced nodes. Our proposed OM problem is to select k positive nodes $\{v^*\}$and minimize the sum of the probability of negative equilibrium opinions when the state of user $u$affected by misinformation is spreader at the end of propagation process.

Theorem 3. Our Opinion Minimization problem is NP-hard under the Dynamic Competitive Diffusion Model.

Proof. It is known that if a problem is any generalization of an NP-hard problem, then the problem is also NP-hard. Since the IM problem [12] has been proved NP-hard, then we prove Theorem 3 by reducing an instance of the NP-hard IM problem to an instance of the OM problem under our proposed model. In our model, when all nodes in $G(V, E, g)$ have the same knowledge, and the sending probabilities of all nodes are equal to 1. IM problem is a special case of our problem. Essentially, the same knowledge makes individuals homogeneous, so that knowledge influence does not impact dissemination probabilities. $\forall u \in V$, ${p_u}^{send} = 1$, means that the inactive node will become a spreader after activation, i.e., weakening the step of opinion transfer. Therefore, solving this instance of the OM problem is equal to solving the IM problem. Thus, our proposed OM problem is also NP-hard. ■

3) Greedy Strategy for Discovering Positive Seeds: Here, given a multi-group OSN $G(V, E, g)$, a greedy algorithm named Algorithm of Equilibrium Opinion Minimization (EOM) is designed based on (12) to find a set of positive seeds. This algorithm is based on the dynamic competitive influence diffusion model and the steady-state solution of opinions. The detailed process is given in Algorithm 2.

Based on the above EOM algorithm, we present a time complexity analysis. Assume that we perform the EOM algorithm in a multi-group OSN graph $G(V, E, g)$ of $|V|$ users, $|E|$ connections and $|g|$ groups. Most of the time cost lies in the loop of calculating $\pi_{Sp}^{u}(T)$. In order to calculate the steady-state probability, our algorithm has to visit each user, connections and groups at most once, which has the time complexity of $O(|V| + |E| + |g||V^g|) = O(|E|)$. Then, for discovering the top-k target positive seed, we need k iteration to repeat calculating the steady-state probabilities. By considering the extremely awful situations where all the nodes are opinion changers, the time complexity for each k iteration is $O(|V|)$ times. Synthetically, we conclude the total time complexity for our proposed EOM algorithm is $O(k|V||E|)$.

Algorithm 2: An Algorithm of EOM.

Input:
$G(V, E, g)$ including user's knowledge parameter b
Set of initial negative nodes NS who spread misinformation
Delayed detection time step tc budget k (Select top-k positive nodes)
Output:
The set of target positive seeds TS
1: Initialize $TS \leftarrow \emptyset$
2: **while** the seed size $|TS| <= k$ **do**
3: Select user $v^* \leftarrow \arg\min \vec{1}^{T} \sum_{u \in OC}^{N} \pi_{Sp}^{u}(T)$
4: $TS \leftarrow TS \cup \{v^*\}$
5: $changers = changers - v^*$
5: **Return** TS.

V. Experiments

A. Data Sets

In this part, we ran our two algorithms on a PC with i5-7300U CPU (2.6 GHz) and 8GB RAM. Experiments are performed on three real data sets using Python 3.8 to test and evaluate the performance of the algorithms. The details of data sets are mentioned in Table I.

The first data set is an email network [46] that was generated using email data from a large European research institution. The data set also contains "ground-truth" community memberships of the nodes. Each node belongs to exactly one of 42 departments at the research institute. The second data set is the network including social circles from Facebook [47]. We extract one subgraph with 24 circles from this data set. The third data set is from Youtube which is video-sharing web site [48]. In this network, users can create groups where other users can join. The first two are small social networks, and the third set is a large data set. They are all from Standard Large Network Dataset Collection in Standford open datasets (http://snap.stanford.edu/data/index.html).

B. Experiment Setup

In these three networks, we removed isolated nodes which can't interact with other nodes. In all experiments, we randomly choose 1% of the total nodes of each data set as the initial negative nodes. These initial negative nodes launch a dissemination of misinformation. Since individual knowledge characteristics are not available in these three data sets, we have adopted a method of assigning values based on prior knowledge. We randomly initialized the lack of knowledge of each node b according to the educational structure of netizens announced in the 46th "Statistical Report on Internet Development in China". According to the 5 values of b in Fig. 2, we add knowledge values to users in the three networks proportionally. That is, the proportion of users with $b = 0.1$ is 8.8%, $b = 0.3$ is 10%, $b = 0.5$ is 21.5%, $b = 0.7$ is 40.5%, and $b = 0.9$ is 19.2%.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

TABLE I
DATA STATISTICS

Data Sets	#Groups	#Nodes	#Edges
Email-Eu-Core	42	1005	25571
Ego-Facebook	24	333	5038
YouTube	8385	1134890	2987624

We evaluate the performance of our EOM algorithm and the four baseline algorithms. One of algorithms is Degree Centrality (DC) which refers to high-degree nodes may outperform other centrality-based heuristics in terms of influential identification. The second algorithm is PageRank (PR) which is used by Google to identify the rank/importance of web pages. The PR score is a well-known measure index. The larger the PR score, the more important the node is. The third one is based on Reverse Influence Set (RIS) sampling method that captures the influence landscape of the whole network through generating a set of random Reverse Reachable (RR) sets [40], [58]. The last one is Influence Maximization via Martingales (IMM) algorithm that is based on RIS, but it utilizes a series of martingale-based estimation techniques to improve its computational efficiency [59]. The four baselines are currently popular and classical optimization approaches, which have been widely used in dealing with the IM problem. The first two algorithms are classical heuristics. We selected the nodes with the top k nodes of highest DC and PR score as positive nodes to initiate the truth movement respectively in the three data sets. The last two algorithms are based on local greedy algorithms. For RIS algorithm, we use Reverse Sampling to randomly generate enough random RR sets. Then we use the greedy algorithm to find the node set which covers the most RR set as the target positive seed set. For IMM algorithm, it added the martingale approach in the phase of node selection. It should be pointed out that in the RIS and IMM algorithms, the influence probability between two nodes needs to be presetted, so we use the probability of the first influence between two nodes as the initial weight of the edges.

For each algorithm in each data set, we repeat the propagation process for 100 times and take the average value as the general feature to ensure the reliability of the results.

C. Experimental Results

In our experiments, we first considered the delayed effects of misinformation. In the real scene, misinformation is usually spread for a period of time before being discovered. In other words, the misinformation has diffused through the social network for some times. And then at a certain time instant, it is detected by the system. When the misinformation is detected, we started the presented competitive strategies to publish positive message to counteract negative influence from further diffusion. Therefore, we performed a comparative experiment with different delay time steps of a misinformation. In this experiment, we set the number of searched positive target nodes $k = 1$in each data set. To verify the delay effect, we let the delay times equal to 1, 2, and 3, respectively. Fig. 6 shows the experimental results of how the delay time influenced the final number of negative nodes (not including the initial negative nodes) in different algorithms. The vertical axis represents the final number of negative nodes at the end of the spread. The starting point on the horizontal axis is sarting time. It represents the number of negative nodes after delaying the corresponding time step, that is, the number of negaive nodes before the positive node is added. Other points on the horizontal axis represent different comparing algorithms and our proposed EOM algorithm. In the three data sets, we can see that the earlier we start to add positive nodes, the lower the final number of negative nodes will be. Because if we detect the misinformation at an early stage, the number of negative nodes of the entire network will probably be relatively lower. Thus, if we start to join competing nodes, the misinformation can be prevented immediately. And then, the final negative nodes will be constrained to a lower level. Obviously, in this process, the starting time to join positive nodes has an enormous impact on the final misinformation infection in the entire social network. In addition, the proposed EOM algorithm is also better than other algorithms on three data sets.

On the contrary, if we don't add any interventions, let negative information spread on the network, can negative nodes infect the entire network? We conducted experiments on three different networks, and made the misinformation to spread for 30-time steps. Fig. 7(1) shows that the diffusion results of the three states of users under the initial level of knowledge (the most people with low knowledge). It can be seen that without any interventions, the three states of users will eventually reach a balance which can prove that users with different opinions will reach a steady state as mentioned in the above chapter. Therefore, on the basis of the transfer of individual opinions, negative nodes can not infect the entire network. And the three networks all exhibit the function of opinion dynamic balance. Moreover, the number of spreaders grows the fastest in short time, then in the two small data sets, the number of spreaders slowly decays to a stable value. However, in the large data set YouTube, the number of spreaders has been increasing until it stabilizes. Contrarily, the number of refusers gradually increased in all data sets.

In order to compare the self-balancing characteristics of different networks, we try to change the characteristics of users in the network. We increased the proportion of high-knowledge people in three networks. The proportion of people with $b = 0.1$ is 50%, $b = 0.3$ is 10%, $b = 0.5$ is 30%, $b = 0.7$ is 5%, and $b = 0.9$ is 5% (i.e., the majority of users on the two networks are people with high levels of knowledge). The experimental results are shown in Fig. 7(2). The steady-state of opinion is also revealed. But the difference is that the number of spreaders grows slowly. The fastest growing is on the number of refusers. This reflects that the more users with high knowledge in the network, the less likely it is for misinformation to be spread on a large scale. The experiment shows that knowledge level has great impact on the final size of misinformation, that is to say, the more there are educated individuals within the population, the weaker is the misinformation influence. This is consistent with the conclusion of [56].

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

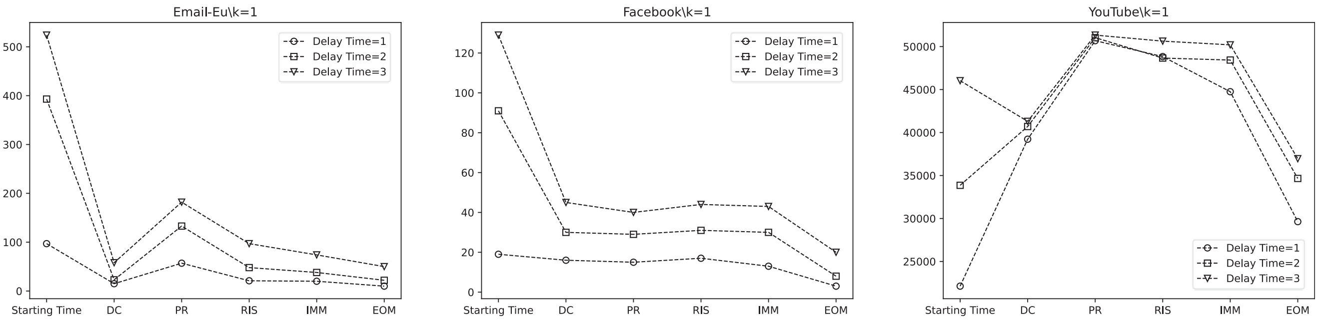
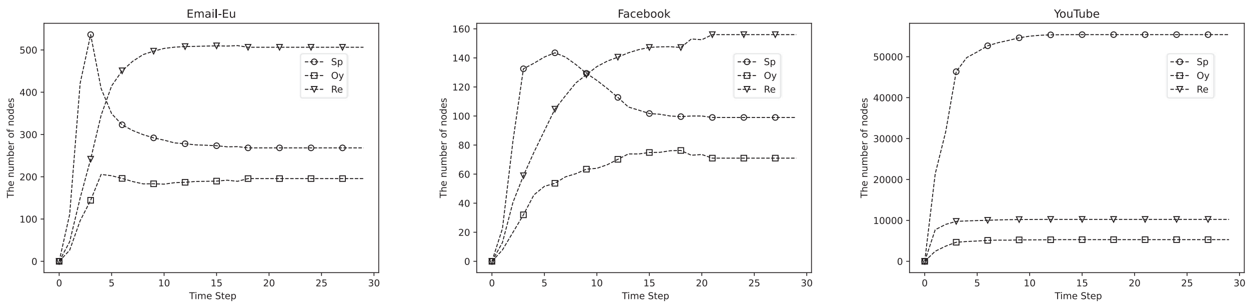
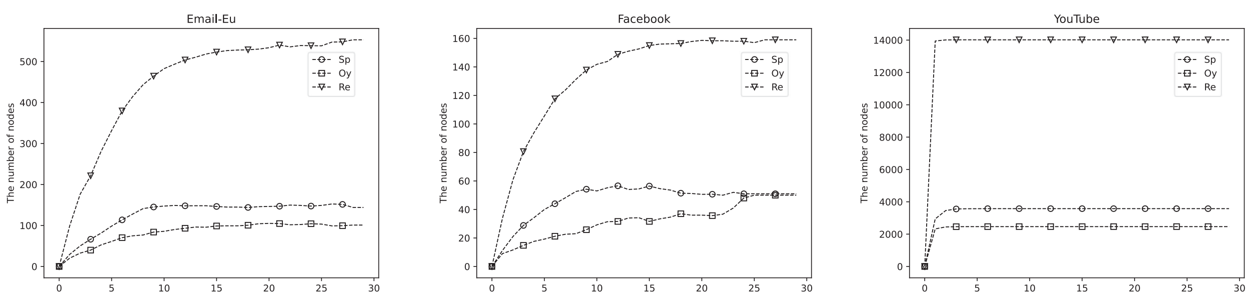


Fig. 6. The final number of negative nodes under different algorithms with different delay times (1,2,3) on the three data sets, respectively. The number of initial positive node is one. The earlier we start to add positive nodes, the lower the final number of negative nodes will be.



(1) The proportion of people with b = 0.1 is 8.8%, b = 0.3 is 10%, b = 0.5 is 21.5%, b = 0.7 is 40.5%, and b = 0.9 is 19.2% (A large proportion of people with low knowledge).



(2) The proportion of people with b = 0.1 is 50%, b = 0.3 is 10%, b = 0.5 is 30%, b = 0.7 is 5%, and b = 0.9 is 5% (A high proportion of people with high knowledge).

Fig. 7. Network self-propagation graph without any initial positive nodes. The horizontal axis indicates time steps. The vertical axis represents the number of users with different states. (1) represents the original network. (2) represents the network with a high knowledge. The more users with high knowledge in the network, the less likely it is for misinformation to spread on a large scale.

For evaluating the performance of the proposed approach, we discuss the number of negative nodes that are finally affected under different algorithms in different cases. We set 3 different delay times to launch the truth campaign. In each delay time, we set 11 different numbers of positive nodes (1, 2, 3, 4, 5, 6, 10, 15, 20, 25, 30) for comparison. As shown in Fig. 8, the diamond-shaped mark curve stands for our proposed EOM greedy algorithm, circle mark for DC algorithm, square mark for PR algorithm, trigonal mark for RIS algorithm and X-shaped mark for IMM algorithm. Obviously, from the Email-Eu and Facebook figures, we can see that the final numbers of spreaders for a misinformation are decreased to different numbers after the introduction of competitive strategies for the three network data sets. But for the large data set of YouTube, when only a small number of positive nodes are added, the final negative nodes will tend to increase. As the number of positive nodes increases, the growth of negative nodes will be suppressed. It is worth being noted that when only one positive node is added to the two small networks (Email-Eu and Facebook), the decline rate of negative nodes is the largest. And as the number of competing positive nodes increases, the number of people who ultimately believe the misinformation decreases. But in the YouTube network, when a small number of positive nodes are added even at a very early time (delay time $= 1$), it is not enough to effectively control the spread of misinformation. This is because the amount of data is too large. However, although the more positive nodes show better control effects, selecting the appropriate number of positive nodes can reduce

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

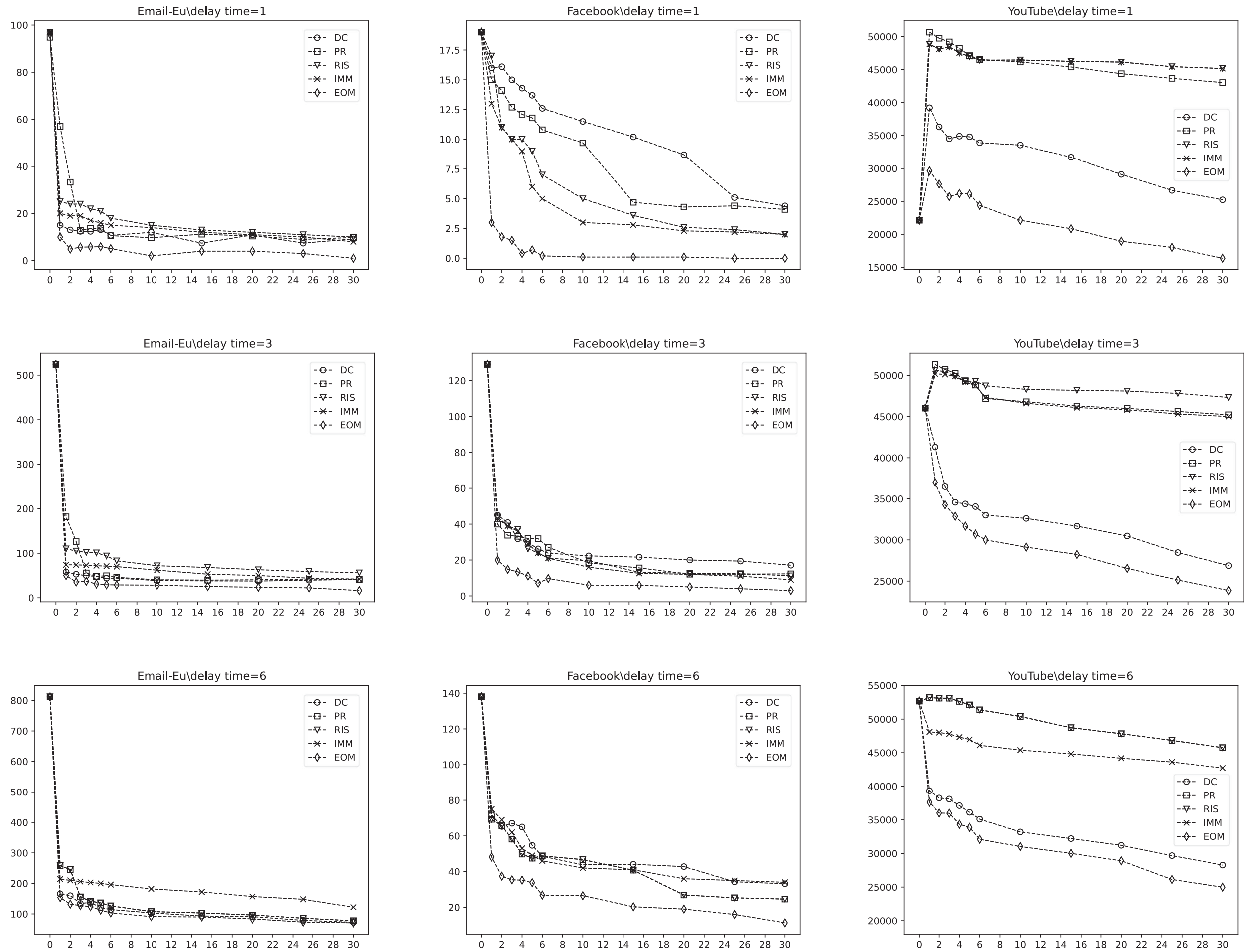


Fig. 8. The performances of different algorithms for different numbers of positive nodes and delay times in three networks. The horizontal axis indicates the number of initial positive nodes. The vertical axis indicates the final numbers of the spreaders.

the control cost as much as possible. For example, in the two small dada sets, when the delay time is 1 and 3, selecting 3-5 positive nodes can effectively control misinformation in the whole network. In YouTube, it needs more positive nodes to control misinformation. When the delay time is 6, no matter which network needs more positive nodes. Therefore, the constraints of cost can also become our main future work. In addition, we can also intuitively see that the proposed EOM algorithm outperforms all the other algorithms, since the final number of spreaders infected by the misinformation is minimal at the end of propagation under each intervention strategies in the three data sets.

Though the proposed EOM algorithm shows the best performance in controlling the misinformation, the computational complexity is a bottleneck. Thus, the further exploration of control strategies for misinformation may enlighten us on designing better strategies with less cost.

VI. Conclusion

In this paper, we investigate the dynamic competitive influence minimization problem for controlling the spread of misinformation. We first propose a novel dynamic competitive diffusion model incorporating with group feature, individual knowledge, and opinion transfer. Different from the previous classic propagation model which mainly concentrate on the activated nodes and the static opinion formation process, the proposed diffusion model combines the dynamic activation process with the opinion transfer process in multi-group social networks. In addition, we divide users into opinion changers and opinion stalwarts, and further divide the opinions of changers into three states in the dynamic opinion formation process. We formulate the changers' opinion transfer processes based on a Markov chain model and use the probability of equilibrium opinions to analyze the likelihoods of nodes as spreaders. At last, in the part of discovering positive seed set, we propose a new OM problem which is different from IM problem, and design an EOM algorithm to find a set of positive seeds greedily. Experiments implemented on three real world social networks show the efficiency of our method. Although the greedy algorithm usually well obtains superior influence propagation, it consumes a large amount of running time. Therefore, this method is more suitable for some small data sets to guarantee optimal result. For future research, we will plan to devise a more time-efficient algorithm that minimizes competitive influence on large-scale data sets. And we will further explore the problem of misinformation control under multiple strategies.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *J. Sci.*, vol. 359, no. 6380, pp. 1146–1151, 2018, doi: 10.1126/science.aap9559.

[2] J. Zhu, P. Ni, G. Wang, and Y. Li, "Misinformation influence minimization problem based on group disbanded in social networks," *J. Inf. Sci.*, vol. 572, no. 3, pp. 1–15, 2021, doi: 10.1016/j.ins.2021.04.086.

[3] People's Daily Online–Technology Channel, "*Chinese Academy of Sciences found that Shuanghuanglian oral liquid can inhibit the new coronavirus,*" 2020. [Online]. Available: https://scitech-people-com-cn.translate.goog/n1/2020/0131/c1007-31566098.html

[4] G. L. Yang, T. P. Benko, M. Cavaliere, J. C. Huang, and P. Matjaz, "Identification of influential invaders in evolutionary populations," *J. Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-43853-9.

[5] A. I. E. Hosni and K. Li, "Minimizing the influence of rumors during breaking news events in online social networks," *J. Knowl.-Based Syst.*, vol. 193, pp. 105452, 2020, doi: 10.1016/j.knosys.2019.105452.

[6] S. Han, F. Z. Zhuang, Q. He, Z. Z. Shi, and A. Xiang, "Energy model for rumor propagation on social networks," *J. Phys. Statist. Mech. Appl.*, vol. 394, pp. 99–109, 2014, doi: 10.1016/j.physa.2013.10.003.

[7] V. Indu and S. M. Thampi, "A nature - inspired approach based on forest fire model for modeling rumor propagation in social networks," *J. Netw. Comput. Appl.*, vol. 125, pp. 28–41, 2019, doi: 10.1016/j.jnca.2018.10.003.

[8] B. Wang, G. Chen, L. Fu, L. Song, and X. Wang, "Drimux: Dynamic rumor influence minimization with user experience in social networks," *IEEE Trans., Knowl. Data Eng.*, vol. 29, no. 10, pp. 2168–2181, Oct. 2017, doi: 10.1109/TKDE.2017.2728064.

[9] G. Seidman, "Self-presentation and belonging on facebook: How personality influences social media use and motivations," *J. Pers. Individual Differences*, vol. 54, no. 3, pp. 402–407, 2013, doi: 10.1016/j.paid.2012.10.003.

[10] Y. Li *et al.*, "What are Chinese talking about in hot weibos?," *J. Phys., A, Stat. Mech. Appl.*, vol. 419, pp. 546–557, 2015, doi: 10.1016/j.physa.2014.10.043.

[11] R. Abebe, J. Kleinberg, D. Parkes, and C. E. Tsourakakis, "Opinion dynamics with varying susceptibility to persuasion," in *Proc. 24th. ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2018, pp. 1089–1098, doi: 10.1145/3219819.3219983.

[12] C. Budak, D. Agrawal, and A. E. Abbadi, "Limiting the spread of misinformation in social networks," in *Proc. 20th. Int. Conf. World Wide Web*, 2011, pp. 665–674, doi: 10.1145.1963405.1963499.

[13] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th. ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2003 pp. 137–146, doi: 10.1145/956750.956769.

[14] W. Chen, C. Wang, and Y. J. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. 16th. ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2010, pp. 1029–1038, doi: 10.1145/1835804.1835934.

[15] H. Nguyen and R. Zheng, "On budgeted influence maximization in social networks," *J. IEEE J. Sel. Areas Commun.*, vol. 31, no. 6, pp. 1084–1094, Jun. 2013, doi: 10.1109/JSAC.2013.130610.

[16] A. Saxena, W. Hsu, M. L. Lee, H.L. Chieu, L. Ng, and L. N. Teow, "Mitigating misinformation in online social network with top-k debunkers and evolving user opinions," in *Proc. Web Conf.*, 2020, pp. 363–370, doi: 10.1145/3366424.3383297.

[17] G. F. de Guilherme, F. A. Rodrigues, P. M. Rodríguez, E. Cozzo, and Y. Moreno, "A general markov chain approach for disease and rumour spreading in complex networks," *J. Complex Netw.*, vol. 6, no. 2, pp. 215–242, 2018, doi: 10.1093/comnet/cnx024.

[18] M. L. Fransen, E. G. Smit, and P. W. Verlegh, "Strategies and motives for resistance to persuasion: An integrative framework," *J. Front. Psychol.*, vol. 6, pp. 1–12, 2015, doi: 10.3389/fpsy.g.2015.01201.

[19] P. M. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. 7th. ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2001, pp. 57–66.

[20] M. Newman, "The spread of epidemic disease on networks," *J. Phys. Rev. E Stat. Nonlinear Soft Matter Phys.*, vol. 66, no. 1, 2002, Art. no. 016128, doi: 10.1103/PhysRevE.66.016128.

[21] F. D. Sahneh and C. Scoglio, "Epidemic spread in human networks," in *Proc. 50th. IEEE Conf. Decis. Control Eur. Control Conf.*, 2011, pp. 3008–3013, doi: 10.1109/CDC.2011.6161529.

[22] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *J. IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 1–14, Feb. 2009, doi: 10.1109/TNET.2008.925623.

[23] L. Yang, X. Yang, and Y. Wu, "The impact of patch forwarding on the prevalence of computer virus: A theoretical assessment approach," *J. Appl. Math. Modelling*, vol. 43, pp. 110–125, 2017, doi: 10.1016/j.apm.2016.10.028.

[24] J. Daley and G. Kendall, "Epidemics and rumors," *Nature*, vol. 204, p. 1118, 1964.

[25] C. Aslay, L. Lakshmanan, L. Wei, and X. Xiao, "Influence maximization in online social networks," in *Proc. 11th. ACM Int. Conf. Web Search Data Mining*, 2018, pp. 775–776, doi: 10.1145/3159652.3162007.

[26] D. Nan, Y. Liang, M. F. Balcan, M. G. Rodrigueza, and S. Le, "Scalable influence maximization for multiple products in continuous-time diffusion networks," *J. Mach. Learn. Res.*, vol. 18, no. 2, pp. 1–45, 2017.

[27] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. 13th. ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2007, pp. 420–429, doi: 10.1145/1281192.1281239.

[28] N. Ohsaka, T. Akiba, Y. Yoshida, and K. I. Kawarabayashial, "Fast and accurate influence maximization on large networks with pruned monte-carlo simulations," *Proc. 28th. AAAI Conf. Artif. Intell.*, 2014, pp. 138–144.

[29] Y. Tang, X. Xiao, and Y. Shi, "Influence maximization: Near-optimal time complexity meets practical efficiency," *Proc. ACM SIGMOD Manage. Data (SIGMOD '14)*, 2014, pp. 75–86, doi: 10.1145/2588555.2593670.

[30] D. Kempe, J. Kleinberg, and É. Tardos, "Influential nodes in a diffusion model for social networks," in *Proc. 32nd. Int. Colloq. Automata, Languages Program.*, vol. 3580, 2005, pp. 1127–1138.

[31] H. P. Maryam, Z. Kamran, R. N. Ahmad, and D. Peter, "Maximizing the spread of positive influence in signed social networks," *J. Intell. Data Anal.*, vol. 20, no. 1, pp. 199–218, 2016, doi: 10.3233/IDA-150801.

[32] W. Feng, G. Wang, and D. Xie, "Maximizing the spread of positive influence under LT-MLA model," in *Proc. Asia-Pacific Serv. Comput. Conf.*, vol. 10065, 2016, pp. 450–463, doi: 10.1007/978-3-319-49178-3_34.

[33] K. Masahiro, S. Kazumi, and M. Hiroshi, "Blocking links to minimize contamination spread in a social network," *J. ACM Trans. Knowl. Discov. Data*, vol. 3, no. 2, pp. 1–23, 2009, doi: 10.1145/1514888.1514892.

[34] A. K. Nandi and H. R. Medal, "Methods for removing links in a network to minimize the spread of infections," *J. Comput. Oper. Res.*, vol. 69, pp. 10–24, 2016, doi: 10.1016/j.cor.2015.11.001.

[35] S. Wang, X. Zhao, Y. Chen, Z. Li, K. Zhang, and J. Xia, "Negative influence minimizing by blocking nodes in social networks," in *Proc. 17th. AAAI Conf. Late-Breaking Develop. Field Artif. Intell.*, 2013, pp. 134–136.

[36] P. Crucitti, V. Latora, and M. Marchiori, "Model for cascading failures in complex networks," *J. Phys. Rev. E Stat. Nonlinear Soft Matter Phys.*, vol. 69, no. 4, 2004, Art. no. 045104, doi: 10.1103/PhysRevE.69.045104.

[37] G. Golnari and Z. L. Zhang, "The effect of different couplings on mitigating failure cascades in interdependent networks," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2015, pp. 677–682.

[38] A. Kuhnle, M. A. Alim, X. Li, H. Zhang, and M. T. Thai, "Multiplex influence maximization in online social networks with heterogeneous diffusion models," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 2, pp. 418–429, Jun. 2018, doi: 10.1109/TCSS.2018.2813262.

[39] C. Song, W. Hsu, and M. L. Lee, "Temporal influence blocking: Minimizing the effect of misinformation in social networks," in *Proc. IEEE 33rd Int. Conf. Data Eng.*, 2017, pp. 847–858.

[40] J. Zhu, S. Ghosh, and W. Wu, "Robust rumor blocking problem with uncertain rumor sources in social networks," *J. World Wide Web*, vol. 24, no. 1, pp. 229–247, 2020.

[41] A. I. E. Hosni, K. Li, and S. Ahmad, "Darim: Dynamic approach for rumor influence minimization in online social networks," in *Proc. Int. Conf. Neural Inf. Process.*, vol. 11954, 2019, pp. 619–630, doi: 10.1007/978-3-030-36711-4_52.

[42] G. Tong *et al.*, "An efficient randomized algorithm for rumor blocking in online social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 845–854, Apr.–Jun. 2020, doi: 10.1109/TNSE.2017.2783190.

[43] J. Zhu, P. Ni, and G. Wang, "Activity minimization of misinformation influence in online social networks," *IEEE Trans. Comput. Social Syst.*, vol. 7, no. 4, pp. 897–906, Aug. 2020, doi: 10.1109/TCSS.2020.2997188.

[44] A. K. Ghoshal, N. Das, and S. Das, "Influence of community structure on misinformation containment in online social networks," *J. Knowl.-Based Syst.*, vol. 213, no. 1, 2020, Art. no. 106693, doi: 10.1016/j.knosys.2020.106693.

[45] S. Gomez, A. Arenas, J. B. Holthoefer, S. Meloni, and Y. Moreno, "Discrete-time markov chain approach to contact-based disease spreading in complex networks," *J. EPL (Euro-Phys. Lett.)*, vol. 89, no. 3, 2010, Art. no. 38009, doi: 10.1209/0295-5075/89/38009.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.

[46] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *J. ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, pp. 2–es, 2006, doi: 10.1145/1217299.1217301.
[47] J. McAuley and J. Leskovec, "Learning to discover social circles in ego networks," in *Proc. 25th. Int. Neural Inf. Process. Syst. Conf.*, vol. 1, 2012, pp. 539–547.
[48] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *J. Knowl. Inf. Syst.*, vol. 42, no. 1, pp. 181–213, 2015.
[49] A. Gionis, E. Terzi, and P. Tsaparas, "Opinion maximization in social networks," in *Proc. SIAM Conf. Data Mining*, 2013, pp. 387–395.
[50] H. Zhang, T. N. Dinh, and M. T. Thai, "Maximizing the spread of positive influence in online social networks," in *Proc. IEEE 33rd Conf. Distrib. Comput Syst.*, 2013, pp. 317–326, doi: 10.1109/ICDCS.2013.37.
[51] C. Shen, R. Nishide, I. Piumarta, H. Takada, and W. Liang, "Influence maximization in signed social networks," in *Proc. Web Inf. Syst. Eng. Conf.*, vol. 9418, 2015, pp. 399–414, doi: 10.1007/978-3-319-26190-4_27.
[52] X. Liu, X. Kong, and P. S. Yu, "Active opinion maximization in social networks," in *Proc. 24th. ACM SIGKDD Conf. Knowl. Discov. Data Mining (KDD '18)*, 2018, pp. 1840–1849, doi: 10.1145/3219819.3220061.
[53] A. Nayak, S. Hosseinalipour, and H. Dai, "Smart information spreading for opinion maximization in social networks," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 2251–2259, doi: 10.1109/INFOCOM.2019.8737538.
[54] Q. He *et al.*, "Positive opinion maximization in signed social networks," *J. Inf. Sci.*, vol. 558, no. 2, pp. 34–49, 2021, doi: 10.1016/j.ins.2020.12.091.
[55] N. Hudson and H. Khamfroush, "Behavioral information diffusion for opinion maximization in online social networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1259–1268, Apr.–Jun. 2021, doi: 10.1109/TNSE.2020.3034094.
[56] K. Afassinou, "Analysis of the impact of education rate on the rumor spreading mechanism," *Phys. A, Stat. Mechanics Appl.*, vol. 414, pp. 43–52, 2014, doi: 10.1016/j. physa.2014.07.041.
[57] S. J. Ling, J. Sanny, and W. Moebs, "CHAPTER 6 photons and matter waves 6.1 blackbody radiation," in *University Physics, Volume 3*. Houston, TX, USA: Rice Univ., 2019. [Online]. Available: https://openstax.org/details/books/university-physics-volume-3
[58] C. Borgs, M. Brautbar, J. Chayes, and B. Lucier, "Maximizing social influence in nearly optimal time," in *Proc. 25 Annu. ACM-SIAM Symp. Discrete Algorithms*, 2014 pp. 946–957, doi: 10.1137/1.9781611973402.70.
[59] Y. Tang, Y. Shi, and X. Xiao, "Influence maximization in near-linear time: A martingale approach," in *Proc. ACM SIGMOD Int. Conf. Manage. Data May*, 2015, pp. 1539–1554. [Online]. Available: https://doi.org/10.1145/2723372.2723734

Yuan Li received the B.S. degree in information management and information system from Dalian Maritime University, Dalian, China, in 2010, and the M.S. degree in logistics engineering in 2013 from the University of Chinese Academy of Sciences, Beijing, China, where she is currently working toward the Ph.D. degree with the School of Electronics, Electrical and Communication Engineering. Her research interests include social networks and machine learning.

Jianming Zhu received the Ph.D. degree in operations research from the Academy of Mathematics and System Science, Chinese Academy of Sciences, Beijing, China. He is currently a Professor with the School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China. From December 2017 to December 2018, he was a Visiting Scientist with the Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA. His research interests include algorithm design and analysis for optimization problems in data science, wireless networks, social networks, and management science.

Jianbin Jiao was an Associate Professor with the Harbin Institute of Technology, Harbin, China, from 1997 to 2005. Since 2006, he has been a Professor with the University of the Chinese Academy of Sciences, Beijing, China. His research interests include image processing and pattern recognition.

Qi Zhang received the B.S. degree in digital media technology from Huaqiao University, Quanzhou, China, in 2016, and the M.S. degree in software engineering from the University of Electronic Science and Technology, Chengdu, China, in 2020. He is currently working toward the Ph.D. degree with the School of Electronics, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China. His research focuses on machine learning.

Authorized licensed use limited to: Universidad Nacional de Colombia (UNAL). Downloaded on February 25,2023 at 20:02:41 UTC from IEEE Xplore. Restrictions apply.