IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 18, 2020, accepted March 2, 2020, date of publication March 5, 2020, date of current version March 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978522

Research on the Community Number Evolution Model of Public Opinion Based on Stochastic Competitive Learning

WENZHENG LI, YIJUN GU, DECHUN YIN, TIANYU XIA, AND JINGYA WANG
School of Information Technology and Cyber Security, People's Public Security University of China, Beijing 100038, China

Corresponding author: Yijun Gu (guyijun@ppsuc.edu.cn)

This work was supported in part by the National Social Science Fund of China under Grant 17CXW014.

ABSTRACT Human activities are usually collective, so clustering has become an important feature of human behavior. This paper studied the evolution of the community in the process of public opinion propagation so as to put forward a public opinion evolution model for the network community number. This study proposed the community number evolution model of public opinion based on stochastic competitive learning, and the proposed model consists of an increase in the number of communities and a decrease in the number of communities. The highlight of this model is that on the one hand, it realizes the research on the evolution of public opinions on the dynamic network; on the other hand, unlike other public opinion evolution models, this model pays attention to the community number increase and decrease rules in the evolution of public opinions. Then, as an extension of the community number evolution model of public opinion, the community number prediction model had been proposed. Based on Twitter data from the 2017 London Bridge attack, the proposed models were validated by experiments. In the verification section of this paper, two methods had been introduced as a comparison. The experimental results show that the community number evolution model of public opinion is correct.

INDEX TERMS Opinion propagation, stochastic competitive learning, community detection, community number, phase transition theory.

I. INTRODUCTION

The evolution of public opinion is a very important virtual social phenomenon with a profound impact on economy, politics, etc. [1], and has many applications in WeChat network [2], microblog network [3], signature network [4], e-commerce network [5] and other aspects.

In the research of public opinion evolution model, it can be traced back to the SIR (Susceptible Infected Recovered) model proposed by Kermack W O and McKendrick in 1932 [6]. As a classic model for studying the spread of infectious diseases, this model is often used to study public opinion propagation. Later, the model has undergone several improvements and evolutions, resulting in models such as SIS (Susceptible Infected Susceptible) model and SIRS (Susceptible Infected Refractory Susceptible) model [7]. In addition, Rogers proposed the innovation diffusion theory in 2003 [8], which was widely used by the public opinion propagation model. Later, sociological theory believed that opinion leaders often existed in public opinion propagation, and the role of important users in opinion propagation was noticed. Ellero and other experts used opinion leaders as special participants to conduct opinion propagation modeling analysis [9], [10]. Subsequently, experts such as Igor Kanovsky constructed the "0-1-2" model [11] based on the probability value of public opinion propagation. The model considered that the probability of whether public opinion propagated was positively correlated with the number of public opinion propagandist.

Nowadays, with the application of many technologies such as survival analysis techniques [12], matrix factorization [13], and mean-field approximation [14] to solve dynamic network analysis problems, the study of dynamic network evolution models has attracted more and more attention from scholars. Hence, a series of models related to the evolution of public opinion have been studied from various angles [15]–[25], and the evolution of public opinion has received extensive attention. In 2014, Tsang A proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

a model of opinion dynamics [15], which achieved the dynamics opinion model by building the trust function among members and based on the nature of homophilic network. In 2016, ChaoYu proposed a consensus formation model [16], which elaborated the process of consensus formation by constructing performance-driven and behavior-driven methods. The Gossiper-Media model [17] studies the interactive game between the public and the media in opinion propagation, confirming that competition among media will promote the formation of public opinion consensus. In combination with the influence network, Lejun Zhang designed a public opinion propagation influence network model and proposed a public opinion control point selection algorithm (POCDNSA) [18]. X Yin *et al.* proposed an agent-based online opinion formation model based on attitude change theory [19], group behavior theory and evolutionary game theory in the perspective of sociology and psychology. Similarly, based on sociology and psychology, Xi Chen *et al.* analyzes the evolution of public opinion based on PA (public authority) [20]. From the perspective of network members, A Mohammadinejad *et al.* studied the influence of leaders on the evolution of public opinion and proposed a framework to consensus opinion model within a networked social group [21]. Cordeiro M and Zhuang D proposed new dynamic community detection algorithms based on modularization from different aspects [22], [23], aiming to detect communities of dynamic networks as effective as repeatedly applying static algorithms but in a more efficient way. Han J, *et al* proposed ALPA (adaptive label propagation algorithm) [24], ALPA takes into account the information of historical communities and updates its solution according to the network modifications via a local label propagation process. Liu X, He D proposed a new information dissemination and opinion evolution IPNN (Information Propagation Neural Network) model [25] based on artificial neural network in 2019, which proposed new mathematical model reveals the relationship between the state of micro-network nodes and the evolution of macro-network public opinion.

However, there are still some problems with the existing public opinion propagation model. Social networks are often dynamically changing during the evolution of public opinion. W Yu found that the results of community detection are not isolated from time changes [13]. Deng J *et al.* also analyzed the dynamic changes of social networks [26]. However, the previous opinion evolution model ignores the fact that the communities of opinion network also change with the dynamics of opinion evolving. Moreover, the previous research only focused on the number of participants in the network and the changes of network heat, but did not analyze the community of the participants in the public opinion, and thus it is difficult to effectively analyze and control the public opinion propagation [27].

In recent years, algorithms based on competitive learning have also been applied to community detection and public opinion modeling. Competitive learning is one of the major achievements in unsupervised learning [28], having many applications in clustering [29], [30] and pattern recognition [31]. In 2008, an Isotropic multi-particle competition mechanism was proposed [32]. The basics of stochastic competitive learning is the particle competition mechanism. In 2012, the particle competition mechanism [33] was first applied to clustering and the clustering function was improved by combining nonlinear dynamics and mathematical models. Later, Silva, T.C *et al.* proposed a community detection algorithm [34] based on stochastic competitive learning algorithm, successfully combining dynamic models with community detection. As a dynamic process, stochastic competitive learning is essentially consistent with the evolution of public opinion networks. Therefore, this method had been adopted to solve the problem in the evolution of opinion network.

Moreover, in order to overcome the problems mentioned above, this study first puts forward the community number evolution model of public opinion (CNEM) based on the two processes consist of dynamic generation of network and communities merging based on stochastic competitive learning. The model consists of two parts: the addition of new communities and the communities merging. This composition is conducive to getting the real-time evolution rule of the network structure, judging the growth stage of the public opinion network and the addition and merging of the public opinion community at each moment, which is helpful for observer to make more timely decisions. Based on the stochastic competitive learning algorithm, this paper applies this real-time community detection algorithm to solve practical problems and analyzes the dataset of a real case and dataset of simulation cases. Based on situation prediction and model fitting, the validity of this model is verified.

The main contributions of the paper include three aspects. Firstly, according to the evolution of real social networks, the process of dynamic community members joining the network and community merger had been constructed, which is composed of time series data acquisition. Secondly, CNEM was constructed based on stochastic competitive learning and dynamic filling of community members. Because the model is dynamic, when community detection is carried out, corresponding relationships between communities in slices of different time can be found, therefore each community can trace its origin and discover its evolution process over time. This is of great significance to the analysis of public opinion dissemination. This feature of the model ensures that it has the function of predicting the evolution of public opinion community number. In order to predict the number of communities when the network structure is not available, this study proposed a community number prediction model. Thirdly, in order to verify the effectiveness of the model, the results obtained by CNEM, community number prediction model and the modularity-based algorithm [35], [36] and dynamic community detection algorithm (DCD algorithm) [22] has been fitted and analyzed respectively.

The following sections are organized as follows. Section II: This section showed the structure of the stochastic competitive learning algorithm and introduced the dynamic process of the algorithm. Section III: This section introduced the construction process of CNEM. Based on that, CNPM and its characteristics analysis of that model were also introduced. Section IV: This chapter uses 2017 London Bridge attack as a case for application analysis.

Finally, the results of the community number evolution model of public opinion (CNEM), the community number prediction model (CNPM), the modularity-based algorithm [35], [36] and DCD algorithm [22] were fitted together for analysis to prove the effectiveness of the model and the rationality of the prediction results of the prediction model.

II. REALATED WORK

Stochastic competitive learning is a competitive dynamics system composed of multiple particles. The algorithm implements dynamic community detection with unsupervised learning. Given a network $G = (V, E)$, V is the set of nodes, and E is the set of edge. In the stochastic competitive learning process, a set of particles $K = \{1, 2, 3 \ldots k\}$ is randomly placed in the nodes of the network. The goal of each particle is set to dominate new nodes while strengthening the degree of domination over the nodes that they have already dominated, and the set of nodes dominated by each particle is the community corresponding to the particles. When a particle visits any node, it would enhance its domination level on that node, and weaken the domination level of other particles on the same node [40]. The domination level can be expressed as domination elements in the domination level matrix as shown in equation (1). Since each time a particle visit a node, the particle's domination level on the node would be added by a parameter $\varepsilon = 1$, so the particle's domination level on the node is numerically equal to the times that the particle visits the node. [40], [33].

$$\overline{N}_i(t) \triangleq [\overline{N}_i^{(1)}(t), \overline{N}_i^{(2)}(t), \ldots, \overline{N}_i^{(k)}(t)]^T \tag{1}$$

In equation (1),(2), $Ni(t)$ records the total number of times that each particle in the network visits node i until time t, represents the domination level of each particle to node i at time t, and node i belongs to the community represented by the particle with the highest domination level value. The $\mathrm{Belong}^{(k)}(t)$ function defines the set of nodes belonging to particle k at time t, as shown in equation (2).

$$\mathrm{Belong}^{(k)}(t) = \{u | u \in V, \max(\overline{N}_u(t)) = \overline{N}_u^{(k)}(t)\} \tag{2}$$

In addition, in order to enhance the domination level of particles for nodes have been dominated, and further to promote the formation of distinct boundaries between communities, the particle in the network is guided to walk by a specific walking rule. This rule is the result of the combination of the random walk and the preferential walk. What's more, in order to prevent the particles from traveling for a long time in other particle-dominated communities, which would disturb the domination situation, stochastic competitive learning set the energy value for each particle [34].

III. THE COMMUNITY NUMBER EVOLUTION MODEL OF PUBLIC OPINION

A. DEFINITION AND HYPOTHESIS OF PUBLIC OPINION COMMUNITY

The participants of CNEM established by this study are users of Twitter, Weibo and other online media. The opinion community refers to the community structure in the public opinion propagation network composed of online media users. After the hot event occurs, online media users may take active follow-up reports, comments and other methods to promote the development of the topic. Since this paper focuses on the propagation of public opinion, the behavior "@" in Twitter and Weibo is also called the "mention" behavior, which is the most critical behavior in online public opinion propagation. Therefore, the network of public opinion propagation takes network media users and "mention" behaviors as nodes and edges. After a specific event outbreak, the relevant public opinion will be quickly propagated in the online network. The online media users will actively propagate their understanding of the event by the "mention" behavior, and the network often form communities in propagation network. The community is guided by one or several opinion leaders in the network. For such communities, this paper calls it the public opinion community.

This paper makes the following assumptions about the community number evolution model. The evolution model of public opinion community. Since this paper considers the change of the network structure in the process of generating the public opinion network, the directed edge formed by the original mentioned behavior in Twitter is regarded as the undirected edge in the merging operation. Meanwhile, since the key behavior of public opinion propagation is the active propagation of information, Twitter users who do not mention other people are not included in the public opinion propagation network. In the public opinion propagation network, it can be founded that the propagation speed of public opinions after the outbreak of events first surges and then gradually converges to zero as the public opinions fade [37]. In addition, users who have been following the event will continue to establish connections with other users, and that speed will gradually be faster than the speed of new users connecting to the public opinion network [38]. Therefore, the network density will gradually increase as time goes by.

B. THE CONSTRUCTION OF THE COMMUNITY NUMBER EVOLUTION MODEL OF PUBLIC OPINION

CNEM is composed of two processes: dynamic generation of network and community merging based on stochastic competitive learning.

In the public opinion propagation, the process of the public opinion network generating is often accompanied. Therefore, in order to simulate the time series opinion network generation process, the study uses a method of filling edges and nodes in time sequence in opinion network. In the domination level matrix introduced by equation (1), The domination elements represent the newly filled node will be initialized. When a new batch of nodes is filled into the network, if a newly filled node would not connect to another subgraph with particles, a new particle is added to the node. If the newly filled node has a connection with an existing node, the particle would not be added at that node. As the particle's walking, the newly filled nodes will be dominated by the nearby communities due to the network structure, and thus the dynamic community detection are completed. The change in the number of particles caused by the dynamic generation of the network is shown in equation (3). In equation (3), $K(t)$ represents the set of particles at time t, k_{new} represents the new particles, i represents the nodes in the network, and i_{new} represents the nodes added at the $t+1$ time.

$$K\,(t+1) = \begin{cases} K(t) + k_{new} & , \forall i \in V, (i, i_{new}) \notin E \\ K(t) & , \exists i \in V, (i, i_{new}) \in E \end{cases} \tag{3}$$

The number of communities can be increased by dynamic generation of network, and decreased by community merging based on stochastic competitive learning [34]. The original stochastic competitive learning algorithm determines the number of communities based on the static network. In a dynamic public opinion propagation network, the number of communities in the network should also be dynamically adjusted. The dynamics process in stochastic competitive learning will cause changes in the domination level matrix. Such changes can determine whether the communities dominated by the particles need to be merged. This study based on stochastic competitive learning proposed a method for judging whether public opinion communities need merging.

According to the definition of stochastic competitive learning, the element value in the domination matrix of the node determines which community the node belongs to. According to the random-preferential movements of the algorithm, if two communities gradually appear more connections between them in the process of opinion propagation, the two communities would tend to merge. In this situation, the particles corresponding to the two communities would walk more times in the other community due to the influence of the network structure. This situation reduces the difference in element values for the two particles in the domination level matrix of the nodes in the two communities. If the particle k_0 exists, the community dominated by the particle is named C_0. If the particle k_1 exists, the community dominated by k_1 is named C_1. In the domination level matrix of the node in the community C_0, when there is no significant difference between the element value of the corresponding particle k_1 and the element value of the corresponding particle k_0, the community C_1 could be considered to be merged into the community C_0. In the calculation, In the domination level matrix of all nodes in C_0, the element values in the domination level matrix corresponding to other particles are summed. If there is a certain particle k_1, the element values corresponding to the particle k_1 in the domination level matrix of all nodes in C_0 are summed. If the value is greater than

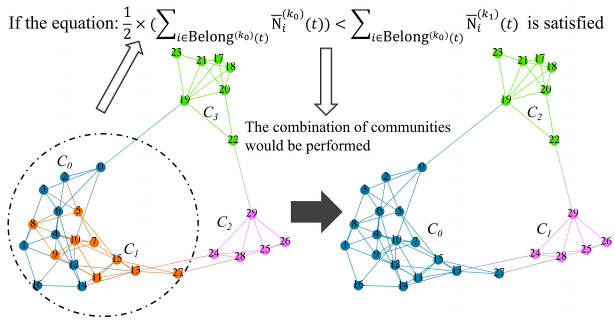

FIGURE 1. Community merging operation.

half of the sum of the elements corresponding to the particle k_0, the community C_1 controlled by the particle k_1 is merged with the current community C_0. The criterion for the merging operation is shown in equation (4). And ε is a parameter that controls the community merger speed which set as 2 in the experiment. If equation (4) is satisfied, the combination of k_0 and k_1 for the community would be performed.

$$\frac{1}{\varepsilon} \times (\sum_{i\in \mathrm{Belong}^{(k_0)}(t)} \overline{N}_i^{(k_0)}(t)) < \sum_{i\in \mathrm{Belong}^{(k_0)}(t)} \overline{N}_i^{(k_1)}(t) \tag{4}$$

After determining that the community are supposed to be merged, particle k_1 needs to be deleted, and the two particle-dominated communities need merging. When the merging operation of elements in domination level matrix is performed, it is necessary to add the element value of k_1 in the corresponding position in the domination level matrix to the corresponding position of k_0 and delete the element of k_1 in the domination level matrix. It should be noted that since the element value of all the positions of the domination level matrix after initialization is 1, it is necessary to eliminate the result of the initialization of the matrix when the merging operation is performed. The purpose of this operation is to prevent the overall increase of the corresponding element value of a particle in the domination level matrix due to multiple community merging. Therefore, the merging of elements in domination level matrix operation is shown in equation (5).

$$\overline{N}_i^{(k_0)}(t) = \overline{N}_i^{(k_0)}(t) + \overline{N}_i^{(k_1)}(t) - 1, i \in V \tag{5}$$

As shown in Figure 1, the communities C_0 and C_1 on the left side no longer have obvious boundaries, and if the equation (4) is satisfied in the calculation, the merging operation would be performed. The merged results are shown in the right side of Figure 1.

In summary, when new connectivity components are added to the public opinion network, the number of communities increases. When the two particle-dominated communities no longer have obvious boundaries, the community would merge, and after the merging of the community, the number of particles in the network will decrease. The above process builds a dynamic process of increasing and decreasing the number of particles. After the above criterion for increasing and decreasing the number of particles combined with the

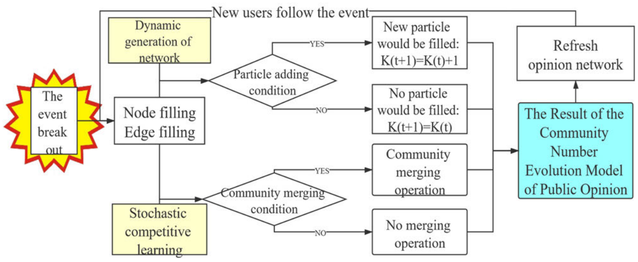

FIGURE 2. Schematic diagram of the community number evolution model of public opinion.

dynamic data set, the number of communities in the public opinion network can be simulated. The framework of community number evolution model is shown in Figure 2.

It can be got from Figure 2 that the model consists of two parts: the increase in the number of communities and the decrease in the number of communities. In the part of the increase in the number of communities, based on dynamic generation of network, the evolution rules of the increased number of communities could be constructed. In the part of the decrease of the number of communities, based on the community merging based on stochastic competitive learning, the evolution rules of the decreased number of communities could be constructed. CNEM is the combination of that two parts.

C. THE COMMUNITY NUMBER PREDICTION MODEL BASED ON THE COMMUNITY NUMBER EVOLUTION MODEL OF PUBLIC OPINION

1) THE COMMUNITY NUMBER PREDICTION MODEL

According to the community number evolution model proposed above, CNPM can be set up. This model can predict the number of communities when the network structure is not available. At time t, the total number of users in the public opinion network of a certain event is N, and the total number of edges is E. After the occurrence of the event, new Twitter users will continue to enter the public opinion network in time sequence, which is simulated by filling nodes and edges in the public opinion network. The number of nodes increased in unit time is S_n, and the number of edges increased is S_e. When any two public opinion communities in the public opinion network satisfy the equation (4), community merging is performed to reduce the number of particles, so the number of communities is also reduced.

In order to set up CNPM and calculate the generation and merging of the public opinion community separately, the speed of communities' number change can be defined as CNPM(t), CNPM(t) =CNPM_inc(t)-CNPM_dec(t). If CNPM(t) >0, the number of communities would increase. Among them, CNPM_inc(t) and CNPM_dec(t) are the speed of the change of the number of communities caused by the added particles and that caused by the merging of the communities respectively. According to the prediction model, the changes of the number of communities caused by the addition of particles would only lead to an increase in the number of communities, and the change of the number of communities caused by community merging would only result in a decrease in the number of communities. If the filled node does not have an edge with other nodes, the node becomes a new connected component. When a new connected component appears, a new particle needs to be added, and the number of communities increases. According to the random graph theory [41], the probability that any node value of degree centrality in the random network is d. The probability is shown in equation (6).

$$P(d) = C_{N-1}^{d} P^{d} (1-P)^{N-1-k} \approx \frac{<d>^{d}}{d!} e^{-<d>} \tag{6}$$

In equation (6), $<d>$ is the average value of degree centrality in the network, and the average value of degree centrality in the network can be expressed as $2E/N$. Therefore, the probability that a new node does not have an edge with other nodes in the network is the result of equation (6) when the d is 0. The result in this condition is shown in equation (7).

$$P(0) \approx \frac{<\frac{2E}{N}>^{0}}{0!} e^{-\frac{2E}{N}} = e^{-2E/N} \tag{7}$$

After getting the probability that the newly filled node is not connected to any other node, combined with the speed of the adding of nodes, the speed of the change of the number of communities caused by the newly added node can be obtained, as shown in equation (8). In equation (8), α is a positive parameter that controls the balance between CNPM_inc and CNPM_dec.

$$\text{CNPM\_inc}(t) = \alpha S_n e^{-2E/N} \tag{8}$$

According to equation (8), it can be found that the increase speed of the number of particles is directly related to the number of nodes in the current network and the number of edges. In the public opinion propagation network, in the case of only observing historical data, the network density variation rule of the public opinion propagation network shows an increasing trend [37]. Hence, the E/N value gradually increases. Therefore, the growth speed of particles would gradually slow down. If there are too many particles at a certain moment, it would automatically trigger the merging of the communities, and the communities would perform the merging operation. Since the purpose of the community merging step is to reduce the number of communities to the most appropriate value, the speed of community merging is positively correlated with the value K which is the total number of particles, which could be represented by CNPM_dec(t) $\propto K$. And positively correlated with $(\frac{E}{N} - \frac{E-e_i}{N-n_i})$ which is the increase speed of network intensity, which could be represented by CNPM_dec(t) $\propto (\frac{E}{N} - \frac{E-S_e}{N-S_n})$. Therefore, the change speed of the number of communities caused by community merging can be formed. The speed is shown in equation (9). In equation (9), β is a positive variable parameter.

$$\text{CNPM\_dec}(t) = \beta K(\frac{E}{N} - \frac{E-S_e}{N-S_n}) \tag{9}$$

In summary, based on the change speed equation of the number of communities caused by dynamic generation of network and equation caused by the community merging, community number prediction model could be got. The model is shown in equation (10).

$$\begin{aligned}\mathrm{CNPM}(t) &= \mathrm{CNPM\_inc}(t) - \mathrm{CNPM\_dec}(t)\\ &= \alpha S_n e^{-2E/N} - \beta K(\frac{E}{N} - \frac{E-S_e}{N-S_n})\end{aligned} \tag{10}$$

2) ANALYSIS ON THE RULE OF "PHASE-LIKE TRANSITION" IN PUBLIC OPINION PROPAGATION

This study found "Phase-like transition point" exists in the public opinion propagation. Phase transition phenomenon [39] refers to a situation similar to the phase transition phenomenon in a random graph model. Phase-like transition is a point like Phase transition point, which is an inflection point when the number of public opinion communities turns from increase to decrease in the process of model evolution. In other words, phase-like transition point is the point when the community number change speed turns from positive to negative. Hence, the change speed of the number of public opinion communities will increases first, then reaches the phase-like transition point, and finally decreases. What's more, in community number prediction model, the number of communities will eventually reach a convergence state, in other words, the number of public opinion communities tends to be stable.

Definition 1: phase-like transition point

Proof: In the public opinion propagation network, there is a peak in the public opinion propagation [38]. There will exist time t_0, which is the time that the node number's increasing speed S_n will reach a peak. Therefore, the value of S_n will be reduced after t_0.

$t \to 0$, CNPM_inc$= \alpha S_n > 0$.

Because CNPM_dec$(t) \propto K(\frac{E}{N} - \frac{E-S_e}{N-S_n})$, so when $t \to 0$, $K = 0$ and CNPM_dec=0.

$t \to 0$, CNPM(t) =CNPM_inc(t)-CNPM_dec$(t) > 0$

In the recession of heat: when $S_n \to 0$, According to the assumptions of the model.

CNPM_inc$(t) = \alpha S_n e^{-2E/N} = 0$

CNPM_dec$(t) = \beta K(\frac{E}{N} - \frac{E-S_e}{N-S_n}) = \beta K(\frac{S_e}{S_n}) > 0$

Hence, when $S_n \to 0$, CNPM(t) =CNPM_inc(t)-CNPM_dec$(t) < 0$

In summary, in the evolution of the model, there is a node with CNPM$(t) = 0$, so the phase-like transition point exists.

Definition 2: convergence state

Proof: Because the number of public opinion communities change at a speed: CNPM(t) = $CNPM\_inc(t) - CNPM\_dec(t)$

$$= \alpha S_n e^{-2E/N} - \beta K(\frac{E}{N} - \frac{E-S_e}{N-S_n})$$

According to the model assumptions: when $t \to \infty$, as the decreases of the event heat, the speed of nodes number change and the speed of network density change tend to stop. Hence, S_n, $S_e \to 0$, then CNPM_inc and CNPM_dec $\to 0$, then CNPM $= 0$. The number of public opinion communities has reached a convergence state.

The conclusion of Definition 1 is also consistent with people's common perception that after an event outbreak, the event would first be discussed by a small opinion community on online social network, and then the boundaries between communities will gradually become blurred due to the increasing network density. Later, different opinion communities began merging. The model would enter a decline phase when the number of communities reached the highest point. Definition 2 shows that in the later stage, the development of public opinion network will eventually make the formed public opinion community stop changing because of the decline of event heat, so it would reach the state of convergence.

In order to verify the universality of the "Phase-like transition" features in public opinion propagation, this chapter constructed a simulation network based on the known features of the public opinion network, and the "Phase-like transition" of the model were verified by experiments. Based on research on density changes in public opinion propagation networks [37], the number of nodes was set as variable N, the number of edges was set as variable E, and N was assumed to be a function varies with t, $N = t$, $E = 0.1(t/10)^{2.3}$. Based on the analysis of the topology of the Twitter network, most of the Twitter networks are star networks [44]. Therefore, when a simulation network is constructed based on the growth function of nodes and edges, a node with a large value of degree centrality will have more possibility to be a target node of a newly added edge, and a rule for newly added edge was set as random-preferential selection method with reference to the random-preferential walking rule [33]. When an edge was added into the network, the source node of the edge was randomly selected in the network, and the target node of the edge was selected by random-preferential selection rule. Random-preferential selection method is the combination of the random selection and the preferential selection. The random selection probability was shown in equation (11). In this equation, i was the nodes number, t was the time variable, and V_t was the set of nodes in the network at time t. The probability of preferential selection was as shown in equation (12). d_i^t and d_u^t represent the value of degree centrality of node i and u at time t, respectively. The probability value of selecting a node by using the preferential selection method was equal to the proportion of the value of degree centrality of the node in the total value of that of all nodes. When an edge was added to the constructed simulation network, as shown in equation (13), the probability that the node i was selected as the target node of the edge was the sum of the probability of random selection and preferential selection adjusted by the parameter γ.

$$P_{edge}^{rand}(i,t) = \frac{1}{\mathrm{len}(V_t)} \tag{11}$$

$$P_{edge}^{pref}(i,t) = \frac{d_i^t}{\sum_{u \in V_t} d_u^t} \tag{12}$$

$$P_{edge}(i,t) = P_{edge}^{rand}(i,t) + (1-\gamma) P_{edge}^{pref}(i,t) \tag{13}$$

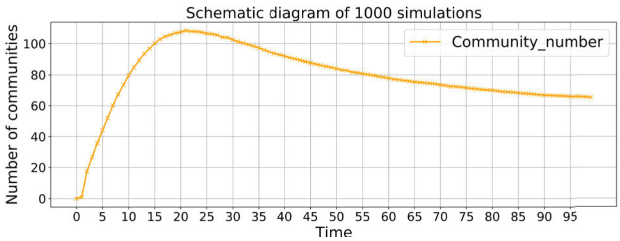

FIGURE 3. Schematic diagram of the simulation results.

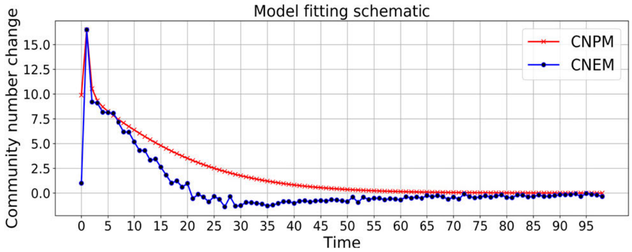

FIGURE 4. Schematic diagram of the model fitting results of the simulation graph model.

The constructed simulation network was analyzed by community number prediction model. In order to reduce the data fluctuation, the simulation structure was automatically constructed 1000 times and the average value of each experimental result was taken as the final result. Set γ as 0.5, the simulation results were shown in Figure 5. The abscissa ''Time'' was the time variable and the ordinate ''Number of communities'' was the number of communities calculated based on community number prediction model. It can be clearly found from Figure 5 that the simulation network built on the basis of the Twitter network structure has a clear phase-like transition point in the number of communities and the final number of communities shows a convergence trend.

The simulation network results obtained in Figure 3 can be fitted to community number prediction model with selected parameters. The fitting results were shown in Figure 4. ''CNEM'' was the change speed of community number calculated by CNEM, and ''CNPM'' was the result of CNPM.

After the model fitting was completed, the error analysis of model fitting result could be obtained. The Pearson correlation coefficient was 0.907 and the cosine similarity value was 0.863. Therefore, it can be concluded that the result of CNEM is basically consistent with that of community number prediction model based on the simulation diagram. Hence, it can be verified that CNPM conforms to the community number evolution model in the public opinion propagation network proposed in this paper. The establishment of the simulation network verifies the existence of characteristics in definition 1 and definition 2 in the propagation process of public opinion social network.

D. TIME COMPLEXITY OF THE COMMUNITY NUMBER EVOLUTION MODEL OF PUBLIC OPINION

Now the time complexity of the proposed model is analyzed. Assuming that there were n nodes and k particles in the network, and each particle corresponds to a community. And the number of steps each particle's walk every time was recorded as x.

Initialization step cost kn times of calculations to construct the domination matrix and energy matrix respectively. In the particle's walking step, once the particle walks, the node domination matrix would update xk times in the network. Since there are k elements in each domination matrix, the assignment operation occurs xk^2 times, as shown in equation (1). In community merger judgment step, firstly, the model would judge which communities the nodes belong to. This step cost nk times traverse. Next, the matching needs to be performed in all communities. This action is performed k^2 times. The judgment of the matching needs to traverse the value of the domination matrix of all nodes. Hence, this action costs nk times calculation. Hence, this step costs nk^3 times of calculations. In the community merger execution step, each merge operation is accompanied by an update of the domination matrix of all nodes, and the update cost nk times calculation. In addition, the number of the groups need merging is a constant.

In summary, the time complexity of the community number evolution model is $O(4nk + xk^2 + nk^3)$.Because k and x are constants, the time complexity can be regarded as $O(n)$.

The simulation network introduced above had been constructed once and it takes 93.0398 seconds to experiment the community number evolution model on it. Compared with the model, the algorithm based on modularity [35], [36] with the time complexity $O(n)$ would take 82.4375 seconds to complete this task, but communities cannot trace their increase and decrease of community number and discover their evolution process over time. The experimental environment of this article is 16GB of RAM, CPU model is i7-4790, and hard disk is 2TB.

IV. EXPERIMENTS

A. DATA ACQUISITION AND EVENT ANALYSIS

In terms of information acquisition, Twitter is the most commonly used media for people to socialize. As of the third quarter of 2018, monthly active users reached 326 million, and the average number of daily active users increased by 11% [42]. When a hot topic of online public opinion breaks out, most cyber citizens would choose to input keywords on Twitter to retrieve relevant information, and pay attention to the progress of the topic, mentioning the event to friends and other accounts with the mention behavior. This behavior is the main driving force for the spread of public opinion on the event.

This paper obtained the *2017 London Bridge attack*data and selected the information from the data to build a public opinion propagation network. The data were shown in Table 1 in which the ''Complete public opinion'' data represents all the public opinion information about the event, and the ''Public opinion network data'' was the network information of the public opinion propagation formed by the mention relationship.

TABLE 1. Data introduction to the 2017 London Bridge attack.

Complete public opinion data	Media user number	95415
	Tweets number	166853
Public opinion network data	Nodes number	13090
	Edges number	16019
	Clustering coefficient	0.049
	Average path length	4.894
	Average of degree centrality	2.448
	Diameter	15

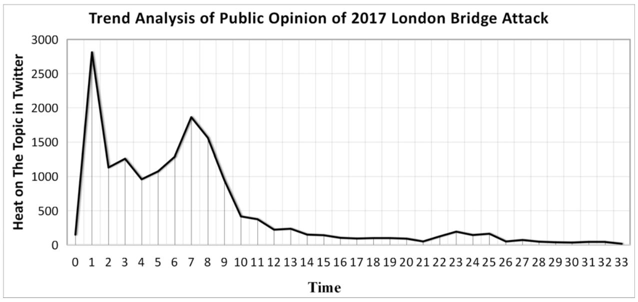

FIGURE 5. Schematic diagram of the heat of change in the 2017 London Bridge attack.

On June 3, 2017, an attack occurred in the London Bridge. The process of the event and the status of real-time public opinion were described as follows [30].

(1) Before June 3, a small number of users on Twitter expressed their concern about the security of the London Bridge and analyzed the possibility of terrorist attacks. However, because the event did not occur, the public opinion could not form a large scale.

(2) The period of June 3 to June 5 was the outbreak period of the topic. At this stage, net citizen's attention to this topic had rapidly increased. On the evening of June 3, 2017, a car rushed to the pedestrians on the London Bridge. Subsequently, three armed terrorists got off the vehicle and attacked pedestrians

(3) Later the terrorists were killed by the police. The London Police Department officially confirmed on the 7th that the number of deaths from terrorist attacks rose to eight. The number of comments forwarded by twitter reached 13,003.

(4) The period of June 6 to June 7, or even longer, was the plateau for the topic. The number of comments forwarded reached 1621.

(5) The period of June 8 to June 14 was the decline of the topic. The number of comments forwarded by twitter reached 1,265.

The public opinion development of the incident was visualized as Figure 5.

In figure 5, the abscissa "Time" represents the time change after the *2017 London Bridge attack*, the ordinate "Heat on The Topic in Twitter" represents the number of new users of the public opinion network, and the "Heat on The Topic in Twitter" corresponding to zero "Time" represents the number of tweets before June 3, after that time, statistics were taken every 6 hours. It can be found from the Figure that the number of tweets about the *2017 London Bridge attack* increased quickly at first, then increased slowly, and eventually converged.

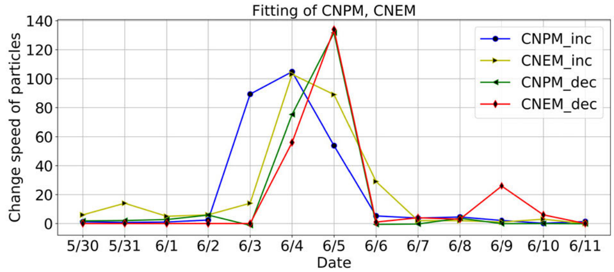

FIGURE 6. Model fitting diagram of the change speed of the number of communities.

B. VERIFICATION OF THE VALIDITY OF THE COMMUNITY NUMBER EVOLUTION MODEL OF PUBLIC OPINION

To verify the validity of the above-mentioned CNEM, this study needs to be carried out in two aspects: verifying whether CNPM can meet CNEM which is the basic of that model, and verifying whether CNEM meets the actual situation at each moment. Firstly, the increase and decrease speed of the number of communities obtained by CNEM was fitted to that obtained by CNPM. Then, the results of CNEM were fitted to the result of community number obtained by the modularity-based algorithm and DCD algorithm. According to the fitting results, if the results of CNPM were correlated to the results of CNEM, it can be proved that CNPM has a correlation with its basic model. If the result of CNEM and CNPM were related to the result of the modularity-based algorithm and DCD algorithm, it can be proved that the two methods of obtaining the community number proposed in the paper are effective.

This paper expands the model fitting experiment based on the acquired data, and the parameters $\alpha = 0.15$, $\beta = 0.26$. At the same time, the corresponding analysis and prediction were carried out. The change speed of the number of particles caused by the node filling was CNEM_inc(t), and the fitting result was shown in Figure 6. The change speed of the number of particles caused by the community merging was CNEM_dec(t), and the fitting result was also shown in Figure 6. Among them, CNPM_inc(t) represented the speed of the increase of the community number calculated by CNPM. CNPM_dec(t) represented the decrease speed calculated by that model. In Figure 6, the abscissa "Date" represented different dates and the ordinate "number of particles" represents the change speed of the number of communities. The speed value was the difference between the number of communities on a given day and that before the given day.

Cosine similarity was used to treat each set of corresponding data as a vector and get the correlation degree between the two sets of data by calculating the angle of the vector. When the cosine similarity closes to 1, the two sets of data were highly correlated. The Pearson correlation coefficient was used to judge the correlation of the change speed between the evolution model and prediction model. When the Pearson

TABLE 2. Evaluation table of the fitting effect on CNPM_inc and CNPM_dec.

Compared project	Cosine similarity	Pearson correlation coefficient
CNPM_inc-CNEM_inc	0.816	0.748
CNPM_dec--CNEM_dec	0.974	0.970

TABLE 3. Comparison of the changes of the number of communities obtained by the four methods.

Date	CNEM	Modularity-based	DCD	CNPM
5/30	15	14	15	0.157
5/31	5	5	6	-1.628
6/1	6	6	5	-1.114
6/2	12	14	15	-3.596
6/3	94	121	121	88.527
6/4	62	80	81	29.538
6/5	-109	-112	-111	-93.549
6/6	-1	1	6	14.306
6/7	-3	-3	5	5.833
6/8	-8	-24	4	4.145
6/9	-24	-1	-36	0.931
6/10	-9	-3	1	2.185
6/11	-4	-3	16	1.533

correlation coefficient was greater than 0.6, the two sets of data were considered to be correlated. When the value was greater than 0.8, the data between the two sets is considered to be strongly correlated. The result of model fitting of CNPM_inc and CNPM_dec was shown in Table 2.

It can be found from Table 2 that both the Cosine similarity value and the Pearson correlation coefficient value in the comparison of the two sets of data were close to 1. Therefore, it can be explained that the change of the public opinion community number caused by node filling and the change of that caused by community merging are strongly correlated with the estimated values based on the model. It can be explained that the increase of the number of communities caused by the filling of nodes and the decrease of the number of communities caused by community merging can be predicted by CNEM_inc and CNEM_dec. The advantage of this model is that the increasing and decreasing speed of the number of public opinion communities can be obtained respectively. And CNPM_inc(t) and CNPM_dec(t) can predict the condition of the rapid growth or rapid aggregation of public opinion networks. CNPM(t) can be get by combining CNPM_inc(t) and CNPM_dec(t). CNPM(t) is the speed change value of community number prediction model. At the same time, in order to verify that the CNPM(t) value of the model was valid for the estimated community number, modularity-based algorithm and DCD algorithm were introduced for comparison [36], [22]. The comparison of the changes of the number of communities obtained was shown in Table 3. The column named "Date" represents the date. "CNEM", "Modularity-based", "DCD" and "CNPM" respectively represented the change speed of the communities number obtained by CNEM, the change speed of the communities number obtained by the modularity-based algorithm, the change speed of the communities number based on DCD algorithm and that obtained by CNPM on the corresponding date.

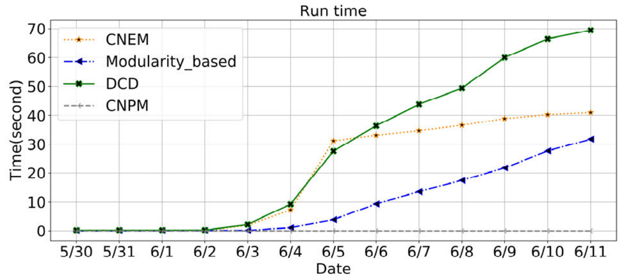

FIGURE 7. Time spent record.

TABLE 4. Correlation table of CNEM and CNPM.

	CNEM	Modularity-based	DCD	CNPM
CNEM	1.000	0.982	0.988	0.945
Modularity-based	0.982	1.000	0.965	0.943
DCD	0.988	0.965	1.000	0.939
CNPM	0.945	0.943	0.939	1.000

In the analysis of the *2017 London Bridge attack*, the CNEM spent 40.990 seconds, the module-based algorithm spent 31.747 seconds, and the DCD algorithm spent 69.516 seconds. The input data of CNPM is the network structure and the number of communities in the previous time slice, and the output can be completed by direct calculation, so CNPM spent 0.007 seconds only. The time consumption recorded at each time slice is shown in Figure 7.

Code run times would be influenced by many factors, such as coding habits and configuration of experimental equipment, so the listed code run times are for reference only.

The results of the changes in the number of communities obtained by the above four methods were analyzed according to the Pearson correlation coefficient, and the results of comparison were shown in Table 4.

Based on the comparison of four sets of data in Table 4 and the Pearson correlation coefficient, it can be concluded that the speed values based on CNPM, the modularity-based algorithm, DCD algorithm and the speed change values of the community based on CNEM have a correlation with each other. The model fitting can be completed, and the model fitting diagram was shown in Figure 8. Moreover, CNEM, CNPM and modularity-based, DCD algorithm can complete model fitting, which indicated that the community number evolution model and prediction model can get effective prediction results.

In Figure 8, "CNEM" represented the speed of change of the number of communities calculated by CNEM, "CNPM" represented the change speed of the number of communities predicted by community number prediction model, "DCD" represented that calculated by DCD algorithm, and "Modularity-based" represents that based on the modularity-based algorithm. It can be found from the results of Figure 8 that the model fitting curve and the result of modularity-based algorithm, DCD algorithm are basically consistent, and the model fitting effect is good. In addition,

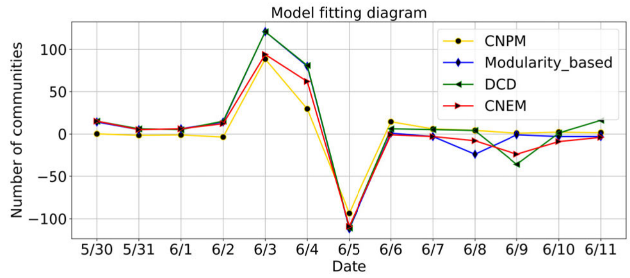

FIGURE 8. **Model fitting diagram of the prediction of the number of communities by the four methods.**

the results of the four methods in the fitting diagram are significantly higher than zero at the beginning and then less than zero, gradually approaching zero over time. This feature also confirmed the existence of the ''phase-like phase'' feature. Based on the above analysis, CNPM proposed in this paper can fit the changes in the number of communities, and the model can provide the speed of new community increase and the speed of community merging in the process of opinion propagation.

V. CONCLUSION

In the beginning, this paper analyzed and evaluated the existing public opinion evolution models and points out their two shortcomings. Firstly, the past public opinion evolution models were all carried out on static social networks, ignoring the changes in social networks with the public opinion evolving. Secondly, community is an important feature of human activity, which was ignored in the previous opinion evolution model. In order to solve the above problems, this study firstly used the nodes filling method in constructing the social network to simulate the dynamic generation process of the social network. Secondly, based on stochastic competitive learning, the community number evolution model of public opinion had been proposed. Then, this study built community number prediction model based on the above two points. On the one hand, the model realized the research of the evolution rule of the community number in the public opinion propagation and the analysis of the time series of the public opinion propagation network. On the other hand, the model can analyze the increase and the merger of the community in the evolution of public opinion. What's more, the model also found and proved the phenomenon of ''phase transition'' in the evolution of public opinion communities. Finally, this study validates the validity of the model with the combining of the data of the *2017 London Bridge attack*.

APPENDIX

The code and data involved in this paper are here: https://github.com/youguqiaomu/improved-stochastic-competition-learning

ACKNOWLEDGMENT

The authors sincerely acknowledge the reviewers for their suggestions which helped in improving the quality of the paper.

REFERENCES

[1] Y. Zhang, ''The evolution of public opinion in social simulation,'' in *Proc. 7th Int. Joint Conf. Comput. Sci. Optim.*, Jul. 2014, pp. 343–345.

[2] C. Ai, B. Chen, L. He, K. Lai, and X. Qiu, ''The national geographic characteristics of online public opinion propagation in China based on WeChat network,'' *GeoInformatica*, vol. 22, no. 2, pp. 311–334, Sep. 2017.

[3] X. Liu and C. Liu, ''Information diffusion and opinion leader mathematical modeling based on microblog,'' *IEEE Access*, vol. 6, pp. 34736–34745, 2018.

[4] X. Lin, Q. Jiao, and L. Wang, ''Opinion propagation over signed networks: Models and convergence analysis,'' *IEEE Trans. Autom. Control*, vol. 64, no. 8, pp. 3431–3438, Aug. 2019.

[5] Y. Wan, B. Ma, and Y. Pan, ''Opinion evolution of online consumer reviews in the e-commerce environment,'' *Electron. Commerce Res.*, vol. 18, no. 2, pp. 291–311, Apr. 2017.

[6] W. O. Kermack and A. G. McKendrick, ''Contributions to the mathematical theory of epidemics. II.—The problem of endemicity,'' *Proc. Roy. Soc. London. A, Math. Phys.*, vol. 138, no. 834, pp. 55–83, 1932.

[7] D. J. Daley and D. G. Kendall, ''Epidemics and rumours,'' *Nature*, vol. 204, no. 4963, p. 1118, Dec. 1964.

[8] E. M. Rogers, *Diffusion of innovations*. New York, NY, USA: Simon and Schuster, 2010.

[9] A. Ellero, A. Sorato, and G. Fasano, ''A new model for estimating the probability of information spreading with opinion leaders,'' Dept. Manage., Univ. Foscari, Venice, Italy, Tech. Rep. 13, 2011.

[10] E. Katz, P. F. Lazarsfeld, and E. Roper, *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Evanston, IL, USA: Routledge, 2017.

[11] I. Kanovsky and O. Yaary, ''Model of opinion spreading in social networks,'' 2011, *arXiv:1106.0872*. [Online]. Available: http://arxiv.org/abs/1106.0872

[12] E. G. Tajeuna, M. Bouguessa, and S. Wang, ''Modeling and predicting community structure changes in time-evolving social networks,'' *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 6, pp. 1166–1180, Jun. 2019.

[13] W. Yu, W. Wang, P. Jiao, H. Wu, Y. Sun, and M. Tang, ''Modeling the local and global evolution pattern of community structures for dynamic networks analysis,'' *IEEE Access*, vol. 7, pp. 71350–71360, 2019.

[14] A. Dadlani, M. S. Kumar, and S. Murugan, ''System dynamics of a refined epidemic model for infection propagation over complex networks,'' *IEEE Syst. J.*, vol. 10, no. 4, pp. 1316–1325, Jul. 2017.

[15] A. Tsang and K. Larson, ''Opinion dynamics of skeptical agents,'' in *Proc. Int. Conf. Autonomous Agents Multi-Agent Syst. Int. Found. Autonomous Agents Multiagent Syst.*, 2014, pp. 277–284.

[16] C. Yu, G. Tan, H. Lv, Z. Wang, J. Meng, J. Hao, and F. Ren, ''Modelling adaptive learning behaviours for consensus formation in human societies,'' *Sci. Rep.*, vol. 6, no. 1, Jun. 2016, Art. no. 27626.

[17] C. Zhang, X. Li, and J. Hao, ''The dynamics of opinion evolution in gossiper-media model with WoLS-CALA learning,'' in *Proc. 17th Int. Conf. Autonomous Agents MultiAgent Syst.*, 2018, pp. 2159–2161.

[18] L. Zhang, T. Wang, Z. Jin, N. Su, C. Zhao, and Y. He, ''The research on social networks public opinion propagation influence models and its controllability,'' *China Commun.*, vol. 15, no. 7, pp. 98–110, Jul. 2018.

[19] X. Yin, H. Wang, P. Yin, and H. Zhu, ''Agent-based opinion formation modeling in social network: A perspective of social psychology,'' 2018, *arXiv:1807.06182*. [Online]. Available: http://arxiv.org/abs/1807.06182

[20] X. Chen, X. Xiong, M. Zhang, and W. Li, ''Public authority control strategy for opinion evolution in social networks,'' *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 26, no. 8, Aug. 2016, Art. no. 083105.

[21] A. Mohammadinejad, R. Farahbakhsh, and N. Crespi, ''Consensus opinion model in online social networks based on influential users,'' *IEEE Access*, vol. 7, pp. 28436–28451, 2019.

[22] M. Cordeiro, R. P. Sarmento, and J. Gama, ''Dynamic community detection in evolving networks using locality modularity optimization,'' *Social Netw. Anal. Mining*, vol. 6, no. 1, p. 15, Mar. 2016.

[23] D. Zhuang, M. J. Chang, and M. Li, ''DynaMo: Dynamic community detection by incrementally maximizing modularity,'' *IEEE Trans. Knowl. Data Eng.*, to be published.

[24] J. Han, W. Li, L. Zhao, Z. Su, Y. Zou, and W. Deng, ''Community detection in dynamic networks via adaptive label propagation,'' *PLoS ONE*, vol. 12, no. 11, Nov. 2017, Art. no. e0188655.

[25] X. Liu, D. He, "Information propagation and public opinion evolution model based on artificial neural network in online social network," *Comput. J.*, Nov. 2019, Art. no. bxz104, doi: 10.1093/comjnl/bxz104.
[26] J.-W. Deng, K.-Y. Deng, Y.-S. Li, and Y.-X. Li, "Study on evolution model and simulation based on social networks," in *Proc. 8th Int. Conf. Natural Comput.*, May 2012, pp. 1238–1241.
[27] P. Xin, D. Gui-Shi, and T. Bin, "Construction and analysis of the opinion formation model based on the social network," *Oper. Res. Manage. Sci.*, vol. 20, no. 2, pp. 176–834, 2011.
[28] B. Kosko, "Stochastic competitive learning," *IEEE Trans. Neural Netw.*, vol. 2, no. 5, pp. 522–529, Sep. 1991.
[29] M. He, D. Zhang, and H. Wang, "Public opinion evolution model with the variable topology structure based on scale free network," *Acta Phys. Sinica Chin.*, vol. 59, no. 8, pp. 5175–5181, 2010.
[30] D. Liu, Z. Pang, and S. R. Lloyd, "A neural network method for detection of obstructive sleep apnea and narcolepsy based on pupil size and EEG," *IEEE Trans. Neural Netw.*, vol. 19, no. 2, pp. 308–318, Feb. 2008.
[31] D. Bacciu and A. Starita, "Competitive repetition suppression (CORE) clustering: A biologically inspired learning model with application to robust clustering," *IEEE Trans. Neural Netw.*, vol. 19, no. 11, pp. 1922–1941, Nov. 2008.
[32] M. G. Quiles, L. Zhao, R. L. Alonso, and R. A. F. Romero, "Particle competition for complex network community detection," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 18, no. 3, Sep. 2008, Art. no. 033107.
[33] T. C. Silva and L. Zhao, "Stochastic competitive learning in complex networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 3, pp. 385–398, Mar. 2012.
[34] T. C. Silva and L. Zhao, "Uncovering overlapping cluster structures via stochastic competitive learning," *Inf. Sci.*, vol. 247, pp. 40–61, Oct. 2013.
[35] V. D. Blondel, J. L. Guillaume, and R. Lambiotte, "Fast unfolding communities large networks," *J. Stat. Mech., Theory Express*, vol. 2008, no. 10, 2008, Art. no. P10008.
[36] J. Huang, H. Sun, and J. Han, "SHRINK: A structural clustering algorithm for detecting hierarchical communities in networks," in *Proc. 19th ACM Conf. Inf. Knowl. Manage. (CIKM)*, Toronto, ON, Canada, Oct. 2010, Art. no. 2.
[37] Y. Guo-Ping, X. U. Xiao-Bing, and B. School, "Research on the Internet public opinion after emergency occurrence based on system dynamics," *Inf. Sci.*, vol. 33, no. 10, pp. 52–56, Oct. 2015
[38] Z. Yan-Xin, W. Xiao-Ming, and L. I. Li, "Research on mobile cellular automata model for public sentiment dissemination in opportunistic networks," *Appl. Res. Comput.*, vol. 32, no. 2, pp. 543–546, Feb. 2015
[39] B. Bollobás and B. Béla, *Random graphs*. Cambridge, U.K.: Cambridge Univ. Press, 2001.
[40] T. C. Silva and L. Zhao, *Machine Learning in Complex Networks*. 2016, doi: 10.1007/978-3-319-17290-3.
[41] D. Lusher, J. Koskinen, and G. Robins, *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2013.
[42] *Q4 and Fiscal Year 2018 Letter to Shareholders [EB/OL]*. Accessed: Jul. 2, 2019. [Online]. Available: https://investor.twitterinc.com/events-and-presentations/event-details/2019/Twitter-First-Quarter-Earnings-Conference-Call/
[43] Stanford Network Analysis Project. *Stanford Large Network Dataset Collection [Eb/Ol]*. Accessed: May 15, 2019. [Online]. Available: http://snap.stanford.edu/data/index.html
[44] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.

WENZHENG LI received the B.S. degree from the People's Public Security University of China, in 2018, where he is currently pursuing the degree with the School of Information Technology and Cyber Security. His research interests include data mining and social network analysis.

YIJUN GU received the Ph.D. degree from the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. He is currently a Professor with the School of Information Technology and Cyber Security, People's Public Security University of China. His research interests include big data analysis, data mining, cybersecurity, and social network analysis.

DECHUN YIN received the Ph.D. degree from the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. He is currently an Associate Professor with the School of Information Technology and Cyber Security, People's Public Security University of China. His research interests include natural language processing, big data analysis, data mining, and social network analysis.

TIANYU XIA received the B.S. degree from the People's Public Security University of China, in 2019, where he is currently pursuing the degree with the School of Information Technology and Cyber Security. His research interests include machine learning and social network analysis.

JINGYA WANG received the master's degree in computer software and theory from the Department of Computer Science, Northwestern University, Xi'an, China. She is currently a Professor with the School of Information Technology and Cyber Security, People's Public Security University of China. Her research interests include network security and information technology.