Research on the Influence Maximization Problem of Microblog Network Based on Topological Potential

Qiwen Zhang, Qiaohong Bai, and Jun Li

Abstract—**Aiming at the problem of a single influence measure method in the microblog network, a comprehensive measurement method integrating microblog content topology potential and attributes is proposed. In order to better reflect the dissemination of information in the real network， in order to improve the convergence accuracy of particle swarm optimization, a follow-bee strategy is introduced, and an elite cloning strategy based on Logistic is proposed. At the same time, the simulated annealing algorithm is introduced to make the particle swarm algorithm better jump out of the local optimum. Finally, the improved algorithm is applied to the problem of maximizing the influence of Weibo network. Compared with other algorithms, the effectiveness of the integrated measurement method and the improved algorithm proposed in this paper is verified.**

Index Terms—**node topological potential, influence force, particle swarm, follower bee strategy, simulated annealing**

I. INTRODUCTION

In recent years, the era of Web3.0 has not only triggered an endless stream of new media [1], but also an explosion of information. For example, reference [2] analyzes the impact of COVID-19 vaccine on public health through emotion analysis and prediction based on social media. Alfred et al. [3] analyzes social media datasets to identify meaningful personal and social behaviors based on topic modeling by using the MOGA clustering method. Microblog, as a social network, has attracted extensive attention from researchers, especially in the exploration of users' influence [4-6], and mainly focuses on the study of user attributes and network structure. For instance, Kwak et al. [7] studied the data in Twitter and analyzed the influence of users by using PageRank algorithm of the number of followers and the number of retweets of microblog. Chen et al. [8] proposed the best effort algorithm, which increases the spread of influence by considering topics, estimates the upper limit of users' topic perceived influence, and used the upper limit to cut down a large number of users with low influence. In addition, an algorithm based on topic samples is proposed, which realizes the influence propagation of some topic distribution samples, and uses concrete information to avoid the actual influence of users with less influence. Experimental results also show that this method is obviously superior to baseline method. Lin et al. [9] comprehensively considered the value and topicality of users' comments, as well as the importance of users in the social network structure to calculate users' influence. Wang et al. [10] proposed Twitter user interest detection based on the LDA topic model, using other behavioral histories of the target user, such as tweets, retweets, favorites, and profile information. The topic list extracted by the LDA model is weighted, and then the topics are sorted according to the weight. The topic with the highest weight is the topic that the target user is most interested in. In terms of network structure, researchers have also applied topological potential theory [11] to social networks and achieved good results. Zhuang et al. [12] developed a multidimensional social influence measurement method based on structural, information-based and action-based factors to measure users' influence at the global level.Through sorting out relevant domestic and foreign literature, it is found that the factors considered by scholars in the research of influence are not comprehensive, and there is a lack of research on the combination of microblog attributes and network structure. Mao et al. [13] proposed a topological potential scheme for predicting influence nodes from large-scale online social networks. It solves the problems of influence attenuation and information dissemination in social networks, and evaluates the importance of nodes by calculating individual values of nodes and surrounding nodes using topological potential theory. Han et al. [14] proposed a topological potential method based on social network analysis, which measures the recommendation degree of recommendation trust nodes by considering both status and reputation. This method considers the credibility of each node and the trust measure between nodes, and can objectively evaluate the comprehensive recommendation degree of P2P recommendation trust nodes. In addition, there are many research algorithms [15-17] for the problem of maximization of influence in social networks, among which heuristic algorithm has been widely studied because of its fast running speed and its ability to obtain certain precision solutions.

Manuscript received March 14, 2022; revised October 14, 2022. This work is partially supported by the Research on HMS Scheduling Optimization and Control and Intelligence System in Manufacturing IoT Environment the National Natural Science Foundation of China (62063021), and National Natural Science Foundation of China (62162040) Research on Terrain Representation and Dissemination Adaptive Scheduling Strategy for Large-scale Social Network Influence Adaptability and Natural Science Foundation (20JRIORA573) Research on intelligent perception and behavior understanding method based on audio and video fusion.

Qiwen Zhang is an associate professor of Lanzhou University of Technology, Lanzhou 730050 China (e-mail: tdyy2010@126.com).

Qiaohong Bai is a postgraduate student of computer technology at Lanzhou University of Technology, Lanzhou 730050 China (corresponding author to provide phone: 18394598860; e-mail: 875008642@qq.com).

Jun Li is a director of Science and Technology Information Department at Lanzhou Modern Vocational College, Lanzhou 730050 China (e-mail: 8230270@qq.com).

Huang et al. [18], for example, proposed a community-based heuristic algorithm to mine influential nodes. The algorithm adopted a dial-and-conquer strategy to select influential nodes, taking into account topic perception and community correlation, so as to improve quality and efficiency. Jain et al. [19] proposed a new whale optimization algorithm based on social networks, which measures the influence of users and search for the top N opinion leaders. Aiming at the research in the application field of influence maximization, Ngamsa-Ard et al. [20] proposed a recommendation system based on location-based social network (LBSN) combined with points of interest (POI). The system is designed to improve the accuracy of POI recommender systems, supporting individual and group recommendations, considering friendships among users and their impact on LBSN and its POI ratings. Semangern et al. [21] proposes identifying bullying risk from social network messages, a method and accompanying tools that can help monitor the potential risk of cyberbullying to an individual, so that appropriate care can be sent in a timely manner. Since Particle Swarm Optimization has few requirements for optimization problems, as long as the quality of the solution can be evaluated, it can be used to solve, such as literature [22, 23]. Over the past 20 years, PSO has attracted a lot of attention in academia, and its effectiveness has been proven in the face of complex optimization problems. In recent years, the application of PSO algorithm to solve this problem in social networks has also achieved certain research results [24-28]. With the rise of weibo network, the literature [29] also proposed a method of maximizing the influence of microblogging based on PSO.

In this paper, a Simulated Annealing Discrete Particle Swarm Optimization algorithm (SA-DPSO) based on Simulated Annealing is proposed, which solves the problem that the influence maximization method is single and the node's own attributes are not considered. SA-DPSO is applied in the influence maximization of microblog network, which proves the effectiveness of this measurement method. The technical achievements of this paper are summarized as follows.

(1) By introducing topological potential theory and combining with the attribute characteristics of microblog network, a comprehensive microblog network influence measurement method is proposed, which solves the problem that influence measurement method is not comprehensive.

(2) In order to improve the problem of insufficient convergence accuracy of the PSO algorithm, the algorithm also introduces the follow-bee strategy, and proposes a Logistic chaotic sequence for elite cloning operations. In addition, the SA algorithm is also introduced to accept inferior solutions, so that PSO can better jump out of the local optimum.

The rest of this article is arranged as follows. Section 2 describes the calculation of propagation probability and the measurement factors of user influence. Then the details of the improved algorithm are shown in Section 3. Finally, the experimental results and conclusions are given in section 4 and 5 respectively.

II. Related work

A. Calculation method of propagation probability based on attributes of microblog

An independent cascading propagation model is adopted. In this model, any edge (u, v) has a propagation probability p_{uv}. The calculation formula is: $p_{uv} = \omega_{uv} * p$, where, p represents the basic propagation probability of the entire network, which is related to the information itself and other factors. If the relationship between users u and v is closer, v is more likely to be influenced by u. The effect of weight ω_{uv} is defined by content similarity and behavior similarity: $\omega_{uv} = (CS_{uv} + AI_{uv})/2$.

(1) Content Similarity: Weng et al. [30] pointed out, there was a certain connection between the concern relationship between weibo users and their topic similarity, that is, people with similar hobbies were more likely to influence each other. We express the keywords appearing in the microblog content of users u and v as vectors $k(u)$ and $k(v)$, respectively, and the cosine similarity between the two vectors is used to represent the content similarity. The calculation formula as follows.

$$CS_{uv} = \frac{|k(u) \cap k(v)|}{\sqrt{|k(u)| \cdot |k(v)|}} \tag{1}$$

(2) Behavioral Similarity: Akshay et al. [31] found that if a user v is interested in the information posted or forwarded by his follower u, v may interact with follower u by forwarding, replying, @, etc. In addition, if user v is more interested in users u, he is more likely to accept the u suggestions. The calculation is as follows.

$$AI_{uv} = \frac{|a(v,u)|}{\sum_{w \in N_v} |a(v,w)|} \tag{2}$$

Where, N_v represents the set of users followed by v, and $a(v,u)$ represents the behavior of v on u, such as comments,@, or retweets.

B. User Influence measurement factors in Microblog network

The influence of users in microblog network is related to many factors, such as the number of fans, activity, recent microblog quality, topics, following and being followed, forwarding and being forwarded, posting and being posted, commenting and being commented, mentioning and being mentioned, @ and by @, the closeness of connections between people in the network, the network distance between users, the effects of temporary events, the characteristics of the network and the individual itself, etc. At present, whether the method of measuring user influence is based on microblog attributes or on the improvement of PageRank algorithm, it only analyzes the different topological features of the microblog network or the interaction between nodes, and most of them directly use the network attributes of microblog network as the basis for analysis. In addition, some studies measure user influence only in terms of retweets, comments, mentions, etc. These measures take into account user interactions, however, only the aggregate result of interaction behavior on a single target is used, without considering the strength and interaction of user interaction behavior.

III. User Influence Strength Based on Topological Potential

A. Topological potential theory

The topological potential theory [32] was first proposed in physics, and has been applied to social networks in recent years to measure the importance of nodes, and achieved certain results. In the real microblog network, the interaction between nodes has local characteristics, and the influence of nodes decreases rapidly with the increase of distance, which is consistent with the characteristics of short-range field. Therefore, a gaussian function with short-range field characteristics and good mathematical properties is used to describe the interaction between nodes, and the corresponding field is the topological potential field. The concept of topological potential field was proposed under the influence of physical potential field, and was first applied to the research and application of complex network by academician Li [33], and analyzed the local characteristics of topological potential field: it is considered that the interaction between nodes is more obvious, which is in line with the close relationship within the community structure, so it has a large influence; on the contrary, in the low-potential area of the topological potential field, the nodes have little interaction, and there is almost no connection between the nodes, which is in line with the sparse connection between the nodes in the complex network, that is, the influence is small. In addition, Li et al. [34] proposed a new metric-topological potential distance (TPD) to mine the topological structure of the network, the attributes of nodes and the interaction between nodes for the node importance analysis and user role identification of directed social networks. The key nodes in the network, the experimental results show that the method can effectively identify the key nodes in the directed social network.

The microblogging network G can be regarded as a physical system containing N nodes and their interactions. There is an action field around each node, and any node in the action field will jointly acted by other nodes. The nodes in the microblogging network are not only affected by themselves, but also by the neighboring nodes in the network structure. Such joint influence generates the node potential, which is described by the topological potential of the node. All nodes of the microblog network are traversed to form ‘peak’, ‘valley’ and ‘slope’ nodes. ‘Peak’ refers to a node that has a greater influence on the connected nodes, ‘valley’ refers to the internal node with less influence, and ‘slope’ refers to a node at the relatively edge position, which will be related to the nodes of ‘peak’ and ‘valley’. Generate more interactions and connections, which are equivalent to the center of connection with the outside world. Fig. 1 shows the distribution of node potentials of a small karate club. It can be seen that this natural peak-valley structure can well reveal the hierarchy of complex network communities and the degree of node influence. In addition, the potential distribution of nodes reveals the structural characteristics of nodes in topological space.

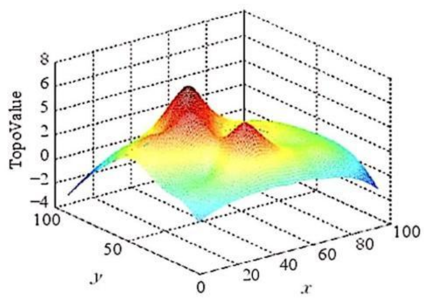

Fig. 1 Topological Potential Structure of Karate Network Nodes

Definition 1: Node topological potential. For a given microblogging network $G=(V,E,W)$, where V represents the set of nodes, E represents the set of edges between nodes, and W is the weight coefficient. According to the potential function definition of the data field, the topological potential of any node $v_i \in V$ can be expressed as:

$$\varphi(v_i)=\sum_{j=1}^{n}[m_j * e^{-(\frac{d_{ij}}{\sigma})^2}] \tag{3}$$

Where, the influence factor σ is used to control the influence range of each node, $\sigma \in (0,+\infty)$. It is found that the topological potential of a node is mainly determined by its second-degree neighbors, so $\sigma \approx 2$. $m_j \geq 0$ is represents the quality of a node v_i and describes the inherent properties of the node. d_{ij} is represents network distance between nodes v_i and v_j. For a given σ value, according to the mathematical properties of the gauss function, the influence range of the node is approximately $[3\sigma/\sqrt{2}]$ a jump local area, and the topological potential of the node is calculating as follow:

$$\varphi(v_i)=\frac{1}{n}\sum_{j=1}^{n}[n_j(v) * e^{-(\frac{j}{\sigma})^2}] \ \forall\, v \in V \tag{4}$$

Where, $n_j(v)$ represents the number of j hop-neighbor nodes of the node V. The normalized topological potential formula is obtained as:

$$\varphi(v_i)=\frac{1}{n}\sum_{j=1}^{n}[e^{-(\frac{d_{ij}}{\sigma})^2}] \tag{5}$$

B. Comprehensive impact measurement

Existing studies show that most of the definitions of influence use the social network attributes of nodes, and ignore the attributes of nodes themselves in the network structure, such as the topological potential of nodes. In the calculation process, topological potential not only measures the distance between nodes, but also describes the closeness of the connections between nodes' neighbors and their neighbors or neighbors with more hops. If all nodes are sorted according to the size of the topology potential value, and the sorting result is used as the evaluation standard for the importance of nodes, it may not reflect the importance of the content. Therefore, according to the previous summary and a large number of studies, user influence cannot be measured solely by a single structural feature. In addition to topological potential, we also consider the number of fans, user activity and recent microblog quality.

Definition 2: Number of followers F. In the microblogging network, there is a positive correlation between the number of followers and the popularity of users. The more famous a user is, the more followers he will have, just like the number of nodes in ordinary social networks, the greater the degree, the greater the influence of the nodes.

Definition 3: User activity Act. User activity is expressed in terms of the number of microblogs posted, reposted, and commented made by a user u within a certain period of time, expressed as $Act(u)$.

$$Act(u) = \frac{n_{post}(u) + n_{repost}(u) + n_{comments}(u)}{T} \tag{6}$$

Where, $n_{post}(u)$ 、 $n_{repost}(u)$ and $n_{comments}(u)$ are the number of microblogs posted by users, the number of reposts, and the number of comments in the time period T , respectively, and T is the length of the time period. The increase in user activity has a positive effect on increasing their influence. Through data analysis, it is found that there are a large number of inactive users in microblog, and these users post very few or even no microblog every day. At the same time, the network also has extremely active users, who post or retweet a large number of microblogs every day. Such users are generally advertising users. In order to eliminate the negative impact of the two types of users, only nodes that meet the activity threshold are reserved on the network.

Definition 4: Quality of recent tweets RTQ. An influential microblog will resonate with everyone, which will lead to a lot of forwarding, comments, likes, etc. A user who has little influence, by sending multiple tweets that people follow, becomes influential. Conversely, a user who has influence, if they haven't posted a popular tweet in a while, they're going to become less and less influential. Therefore, RTQ can be used as a factor to measure user influence, which is defined as follows.

$$RTQ(v) = \frac{\left(\sum_{i=1}^{RN} \left(FW_i(v) + PS_i(v) + CM_i(v)\right)\right)}{RN} \tag{7}$$

Where, $FW_i(v)$ 、 $PS_i(v)$ and $CM_i(v)$ are the average number of reposts, likes, and comments on the i Microblog posted by user v, respectively. RN is the total number of microblog posted recently.

Therefore, combined with analytic hierarchy process, the number of followers, activeness, recent microblog quality and topological potential are comprehensively considered, and the comprehensive influence measurement index is proposed as follows:

$$Influence(v) = \alpha_1 F(v) + \alpha_2 A(v) + \alpha_3 RTQ(v) + \alpha_4 \varphi(V_i) \tag{8}$$

Where, $F(v)$ 、 $A(v)$ 、 $RTQ(v)$ 、 $\varphi(V_i)$ are represent the number of followers, activity, the recent microblog quality and node topological potential respectively. α_i is the analytic hierarchy process coefficient.

C. SA-DPSO Algorithm

(1)Follow-Bee Strategy

In the artificial bee colony algorithm, the following bees search for new food sources near the food source, according to the information transmitted by the leading bees. When a good food source is found, the corresponding leading bees are notified to update their food sources. The following bees can improve the accuracy of the solution. Based on the follow-bee theory, this paper proposes an elite cloning operation.

The number of clones N_c is calculated from the following (9).

$$N_c = \sum_{i=1}^{n} round(S_i n/i + b) \tag{9}$$

Where, n the number of randomly selected elite particles, $S_i \in [0,1]$, in order to avoid the number of clones less than 0, so the addition b , b is an integer greater than or equal to 1, $round$ is the integer function.

Survivability S_i. The survivability S_i of a particle is the weight of the distance between any particle and the optimal particle and the particle densityρ_i. Where, ρ_i is the ratio of the number m of particles within the radius range r of the particle to the total number of particles, that is, $\rho_i = m/N$, so S_i is calculated as follows.

$$S_i = \mathrm{r1} * (1/d_{i,Pg}) + r2 * (1/\rho_i) \tag{10}$$

Where, $\mathrm{r1}, \mathrm{r2} \in [0,1]$, $d_{i,Pg} = |x_i - x_{Pg}|$, in the formula, $d_{i,Pg}$, ρ_i are between 0 and 1. $1/d_{i,Pg}$ the larger, that is, the closer the particle to the optimal value, the higher the particle fitness; ρ_i the larger, that is, the higher the number of particles clustered around the particle, the higher the density, continuing the search will result in local optima, so explore other areas. For the survivability S_i of the particle, the larger of S_i , the more adaptive it is, and vice versa, so the survivability of the cloning operation determines the number of clones.

The cloning method is as follows: select a few elite particles at random, and conduct random and regular cloning with Logistic sequence, such as follow.

$$X_i^{new} = X_i + (X_i^{max} - X_i^{min})/l * U_{r+1} \tag{11}$$

Where, X_i^{max} is the most influential particle in the area of elite particle density and X_i^{min} is the smallest particle in the area of elite particle density. l is a constant, depending on the specific problem. U_{r+1} is a chaotic sequential Logistic, where U_{r+1} the map of Logistics is a chaotic sequence is as follows .

$$U_{r+1} = \mu U_r(1 - U_r),\ r = 0,\ 1,\ 2,\ \ldots,\ 0 < U_0 < 1 \tag{12}$$

Where, $r = 0,1,2,\ldots,$ $0< U_0 < 1$, μ is the state control parameters of the system, it has been proved that when μ =4, the initial value $U_0 \notin (0.25,0.5,0.75,1)$, the system represented by the above formula is completely in a chaotic state, U_{r+1} traversing the (0,1) range.

Selection strategy: after the above operations on the particles in the population, the number of particles in the population has already exceeded the set value, so it is necessary to further select the particles in the population. The selection process of particles is as follows. After sorting the influence values of all particles, randomly select N particles from the top 50% of the particles as the initial population of the next algorithm, which effectively ensures that the selected individuals have advantages, but will not fall into local optimization because of being too good.

(2)Analysis of simulated annealing algorithm

The SA algorithm adopts Metropolis criterion as the probability of accepting a new solutions, as shown in (13). The Metropolis criterion accepts the optimal solution while accepting a deterioration value with a certain probability. In this process, the temperature is relatively high at the beginning, and the probability of accepting inferior solutions is relatively high, so that the particles have a high probability to conduct global search. With the gradual decrease of temperature and energy, the probability of accepting inferior solutions becomes smaller, however, the particles still have the chance to jump out of the local optimal solution, and finally converge to the global optimal solution.

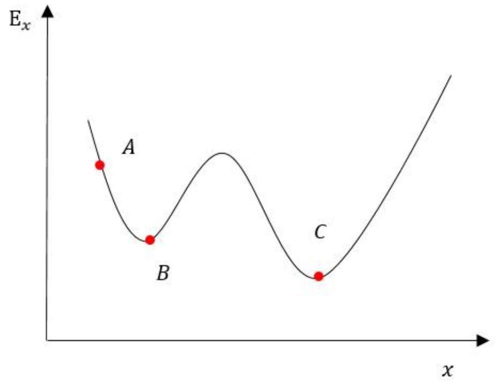

Fig. 2 Simulated Annealing Energy Diagram

$$P = \begin{cases} 1, & if\ f(x_k^{'}) < f(x_k) \\ \exp\left(\frac{f(x_k) - f(x_k^{'})}{T_k}\right), & if\ f(x_k^{'}) \geq f(x_k) \end{cases} \tag{13}$$

Where, $f(x_k)$ and $f(x_k^{'})$ are the energy of the solid in different states, and T_k is boltzmann's constant.

In Fig. 2, the abscissa represents the solution x obtained by the algorithm in the solution space, and the ordinate represents the energy E_x of the solution in the solution space. According to the figure, the search starts from A, and the energy corresponding to the function value decrease, the search process will continue. When point B is reached, only the local optimal solution is found. Since the simulated annealing algorithm is a random algorithm, it can accept a solution that is worse than the current solution with a certain probability. It will continue to search in the solution space randomly, continue to move to the right, and reach the peak point between B and C, so that jumped out of the local extreme value, and gradually converged to the global optimal solution C as the temperature dropped.

(3) Particle update method

In order to overcome the shortcoming of PSO, SA algorithm can be added to PSO to search better solution globally. The initial solution X is generated randomly and the range of influence is calculated. When a new solution is generated, a node in X is first selected, and then a node is selected from the candidate seed set for replacement to form a new solution $X^{'}$. Calculate the difference ΔT between the influence range of the new solution $X^{'}$ and the original solution X. If $\Delta T > 0$, accept $X^{'}$ as the new current solution, otherwise accept $X^{'}$ as the new current solution with a certain probability. After the annealing process, the one with the largest influence range from the 'optimal solution' of PSO and the 'optimal solution' of simulated annealing process is selected, and use them to replace part of the solution set of PSO to enter the next round of simulated annealing process.

The study found that the number of microblog users' fans, followers and followers of each other are the best indicators of the users' status [35]. Aiming at the characteristics of the microblog field, the standard PSO is improved to redefine the particle state update equation (14) (15):

$$v_i(n+1) = \omega v_i(n) + \mu_1 c_1 r_1 \left(p_{best,i} - x_i(n)\right) + c_2 r_2 \left(g_{best} - x_i(n)\right) \tag{14}$$

$$x_i(n+1) = x_i(n) + v_i(n) \tag{15}$$

Where, a disturbance μ_1 is added to the 'self-cognition' learning of particles, and μ_1 is defined as follow:

$$\mu_1 = \lg\left(\frac{F(u)+1}{N_g(u)+1}\right) \tag{16}$$

Where, F is represents the number of fans of user u, N_g is represents of the number of followers of user u.

The initial temperature and annealing temperature formula of SA algorithm are respectively:

$$t_{k+1} = \lambda t_k,\quad t_0 = f(p_g)/\ln 5 \tag{17}$$

In this equation, λ is the weight of the annealing constant. In addition, the internal energy of each particle (that is, the influence of the particle) needs to be determined. The probability of each influence value TF_{p_i} can be obtained by using the following formula:

$$TF_{p_i} = \frac{e^{-\frac{(f(p_i)-f(p_g))}{t}}}{\sum_{i=1}^{N} e^{-\frac{(f(p_i)-f(p_g))}{t}}} \tag{18}$$

(4) Algorithm flowchart

Fig. 3 is the overall flow chart of the algorithm.

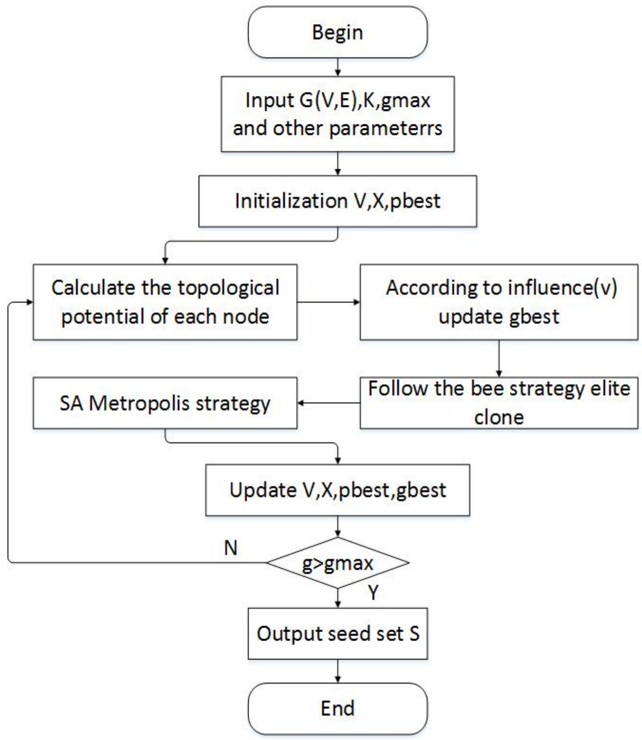

Fig. 3 Algorithm Flow Chart

(5) SA-DPSO Framework

The algorithm framework of SA-DPSO in microblog network influence maximization.

Algorithm SA-DPSO for Influence Maximization

1. Input：Graph $G=(V,E)$，the number of iterations g_{max}，the size of particle swarm N，the inertia weight w，the learn factors c_1，c_2，final temperature t_f，temperature difference ΔT，number of internal cycles q，the size of the seed set k.
2. Output：Output the $Gbest^*$ as the seed set S.
3. Step1：Initialization：
   a) Initialize iterator $g=0, t_g=t_0, count=0$；
   b) Initialize position vector：$X \leftarrow$Initialization（G,k,N）；
   c) Initialize velocity vector：$V \leftarrow 0$；
   d) Initialize $Pbest$ vector：$Pbest \leftarrow$Initialization（G,k,N）；
4. Step2：Select out the initial global best position vector $Gbest^*$ according to the $influence(v)$ value of each x_i；
5. Step3：Begin cycling
   a) Update the velocity vector V according（14）；
   b) Update the position vector X according（15）；
6. Step4：Use（17）to determine the initial temperature；determine the probability value of the fitness value of each particle at the current temperature according to formula（18），and compare it with a random probability to obtain the corresponding position as the global optimal position；
7. While $t_g < t_f$ do
   a) {calculate the range of influence of the set S
   b) generate a new seed set $S^{'}$
   c) count+=1
   d) $\Delta f = influence(S) - influence(S^{'})$
   e) if $\Delta f > 0$ then
      $S = S^{'}$
   f) else generate random numbers r
      if $\exp\left(\frac{\Delta f}{t_g}\right) > r$ then
      $S = S^{'}$
   g) if $count > q$ then
      $t_g = t_g - \Delta T$,count=0}
   h) return S.
8. Step5：Update $Pbest, Gbest$, and perform de-temperature operation；
   a) Compare and update the $Gbest^*$ ：$Gbest^* \leftarrow \max(Gbest^*, Gbest^{'})$；
   b) Update the $Pbest$ of the current generation；
9. Step6：Stop criteria：if $g = g_{max}$，stop the algorithm，otherwise，let $g \leftarrow g+1$ and go to Step 3；

Under the framework of this algorithm, the particle swarm algorithm mainly performs global search through iterative formula (14) and formula (15).And by introducing simulated annealing operation to optimize the optimal solution, because simulated annealing operation accepts the characteristic of inferior solution can optimize the local optimal space solution.

(6) Algorithm complexity analysis

The degree-based initialization method in step 1 requires a $O(N \cdot k)$ basic operation. The overhead of calculating the topological potential in step 2 is $O(n^2)$. Step 3(a) requires an $O(k \cdot logk \cdot N)$ action, and step 3(b) requires an $O(N \cdot k)$ action. Step 4 needs $O(g\overline{d}^2kM)$ action, which $\overline{d}^2k$ is to calculate $influence(S)$, d is the average degree of the network, M is the number of edges. Step 5(a) requires a $O(1)$ basic operation, and step 5(b) requires an $O(N \cdot k)$ operation. Therefore, the worst-case time complexity is $O(k \cdot logk \cdot N) + 3O(N \cdot k) + O(n^2) + O(g\overline{d}^2kM) + O(1)$. In addition, it takes a unit to set the operating time of the other steps, according to the O rules, in the worst case, SA-DPSO has a time complexity of $O(n^2 \cdot k^2 \cdot logk \cdot N \cdot \overline{d}^2 \cdot g_{max})$.

IV. Experimental results and analysis

A. Experimental environment and data collection and processing

All the code is written in python and the running computer is configured with Intel (R) Core (TM) i5-4590 CPU 3.30 GHZ, 8 GB of memory. The data in this study are all sourced from Sina Microblog. This data set is a Sina Microblog data set of 63,641 users, with 13,91718 user friend relationships. Due to the limitation of Sina Microblog, each user can only obtain the information of 200 followers at most, therefore, the friend relationship is not very complete. Data format: user uid, user nickname, user name, user location, user homepage url, user gender, user fans, user followers, user microblogs, user favorites, user creation time. For data processing, media microblogs with a large number of fans, a large number of microblogs, and frequent updates were selected. Finally, 200 micro-blog nodes and 10,781 sides were selected for research.

B. Analysis of Microblog Network Characteristics

Ucinet, a social network analysis tool, is used to analyze the network structure of the microblog users, including centrality and condensed subgroups. The study uses a directed network structure diagram to construct a binary directed network through the following relationship between microblog nodes. In the ‘following relationship matrix’ of the microblog social network, node A cares about node B, therefore, the connection between A and B is 1, that is, there is A directed line segment from A to B, otherwise, it is 0. However, in the microblog social network, the direction of information dissemination is opposite to the direction of ‘attention’[35], when A follows B, the information posted by B will be displayed on A’s homepage, that is, the information spreads from B to A. After data screening and processing, a 200×200 attention relationship matrix is finally sorted out, and a microblog network relationship diagram is generated. Since the drawn network attention map is too large, the attention relationship between lines cannot be seen in the figure, and will not be shown here.

Cohesive subgroup analysis: Cohesive subgroup is a subset of actors with strong, direct, close, frequent or positive relationships. Table I shows the condensed subgroups of the microblog network. From Table I, it can be concluded that the cohesive subgroup of the microblog network is divided into 8 subgroups. Among the 8 subgroups, 3 groups have large scale. Considering the number of fans and the number of followers, users with similar number of followers and followers are more likely to form subgroups. In addition, small groups are connected through a core user, however, no further connections are made, eventually forming the entire network. The combination of strong core users helps to cover more users and build a huge information transmission system. Meanwhile, the combination of core users in the network will make influential information spread more quickly. In addition, the densities of the eight subgroups are calculated as shown in the following matrix.

As can be seen from A matrix as Fig. 4, the density of subgroup 1 is 0.716, indicating that users in subgroup 1 communicate frequently and are closely connected with each other. The density of subgroup 5 is 0.048, indicating that the user communication of subgroup 5 is less and the relationship is loose. The density of subgroup 3 is 0, indicating that the users in the subgroup are not communicating.

TABLE I
CONDENSING SUBGROUP

subgroup	Nodes
1	1,2,3,29,5,131,7,8,34,10,61,37,163,14,15,66,17,143,19,70,21,47,23,24,25,51,27,78,179,30,55,32,169,46,171,42,167,50,200,65,41,178,119,194,103,180
2	26,6,144,16,44,40
3	4,190,188
4	104,60,72
5	152,12,9,33,99,145,126,38,111,162,57,58,68,86,123,74,87,75,77,148,18,142,93,20,134,135,94,161,174,112,151,22,91,129,155,43,83,71,153,122,117,130,100,102,165,166,141,67,140,108,127,90
6	95,113,92,150,109,98,114,116,31,59,121,53,11,94,48,197,79
7	189,139,13,149,157,185,110,164,136,137,120,88,184,176,182,181,168,133,96,160,173,199,138,125,177,56,156,118,80,195,183,159,172,186,187,193,175,28,54,105,192,196,170,146,128,198,115,69,154,191
8	124,52,39,132,101,158,107,64,36,62,76,89,35,85,106,81,82,45,84,147,73,49,63

$$A=\begin{bmatrix}0.716 & 0.413 & 0.406 & 0.225 & 0.054 & 0.045 & 0.150 & 0.147\\ 0.692 & 0.367 & 0.667 & 0.222 & 0.032 & 0.029 & 0.150 & 0.130\\ 0.775 & 0.444 & 0.000 & 0.111 & 0.045 & 0.039 & 0.147 & 0.188\\ 0.667 & 0.167 & 0.333 & 0.333 & 0.058 & 0.039 & 0.093 & 0.159\\ 0.708 & 0.490 & 0.346 & 0.327 & 0.048 & 0.045 & 0.140 & 0.186\\ 0.771 & 0.627 & 0.647 & 0.412 & 0.364 & 0.382 & 0.381 & 0.448\\ 0.730 & 0.383 & 0.220 & 0.367 & 0.036 & 0.027 & 0.108 & 0.119\\ 0.670 & 0.333 & 0.377 & 0.290 & 0.034 & 0.059 & 0.111 & 0.202\end{bmatrix}$$

Fig. 4 A Matrix

C. Comparison of influence ranking methods

In order to verify the validity of the measurement method in this paper, the following three commonly used influence measurement methods are used for comparison.

(1) UPR [36]: UPR is a measure of influence with only microblog attributes, that is, it only considers three factors: number of followers, activity, and quality of recent microblogs.

(2) PageRank (PR): The PageRank algorithm is a classic network algorithm, and the ranking of user influence in Weibo is carried out through this algorithm.

(3) Fans count: The FansC algorithm uses the fan count of the users in the micro bar to sort the impact. Twitter already uses this method, and it has been used as a comparison algorithm for many references in the literature.

In the comparison algorithm, the method in 3.2 is used, where the topological potential calculation formula for the node is (4), when $\sigma \approx 2$, the topological potential of the node is calculated as the number of nodes within the $6/\sqrt{2}$ hop range of the node , that is, the number of neighbor nodes set to 5 hops in the text. Table 1 shows the Top-20 users obtained on the basis of the three influence calculation algorithms of this paper, the number of fans and PageRank.

It can be seen from Table II, in the first 20 nodes, the method has 15 identical nodes with UPR, and the identical rate is 75%, indicating that the method has a high similarity with UPR, therefore, the method is reasonable. There are 9 identical nodes with the PR method, with an identical rate of 45%. For the 11 different nodes, the PR algorithm only considers the attributes of the node itself and does not consider the influence of other attributes on the node. There are 4 identical nodes with FansC method, and the same rate is only 20%. For 16 different nodes, FansC only considers the influence of the number of fans on nodes, which is too simple. Based on the above analysis, the difference between the proposed method and the UPR method is reflected in the consideration of topological potential value of nodes, which can become an indicator to measure the importance of nodes. In addition, the number of fans, microblog, forwarding and attention of nodes with differences is lower than that of other nodes on the whole. And node 3 with the largest number of likes ranks higher, reflecting the influence degree of likes on influence. To sum up, our algorithm not only considers the node's own attributes such as topological potential, but also the node's social attributes such as the number of fans, tweets, forwarding, comments, likes and followers, so the results are superior to PageRank and FansC methods.

TABLE II
TOP 20 USERS OF THE FOUR SORTING METHODS

Nodes	rank								
	ARTICLE	*UPR*	PR	FansC	Fans	Microblog	Forwarding	Comments	Likes	Follows
46	1	5	1	-	182728	15661	829	57	1380	2500
193	2	17	-	-	193	1380	1079	167	1061	129
8	3	15	-	-	399	3005	848	205	1542	649
56	4	3	-	4	781620	3502	1056	138	1060	198
19	5	20	-	-	31	233	1140	93	1288	171
78	6	7	6	-	9860	1409	207	13	82	1893
182	7	-	-	-	468	7990	165	84	207	363
169	8	2	17	2	900695	1464	1046	91	1070	1705
174	9	-	-	-	15	322	1566	660	1561	11
3	10	5	18	-	158394	30643	1195	20009	10095	989
50	11	19	11	-	161	297	701	46	1405	944
103	12	-	-	-	1512	4039	1054	115	1092	1388
163	13	-	10	-	3877	46633	1569	238	1668	1719
34	14	14	14	-	425	1801	300	9	63	381
24	15	4	-	-	570537	7327	335	19	1087	2722
200	16	6	16	17	16301	1829	113	23	88	1201
153	17	-	-	-	86	189	2625	704	1137	117
158	18	11	-	-	85475	9008	499	185	491	529
186	19	1	-	1	28186106	59323	524	92	1062	133
30	20	10	19	-	556	9747	950	94	1305	1177

D. Comparison of influence spread range

(1) Algorithm comparison

The experiment uses five existing algorithms as comparison tests, including the web based heuristic on network characteristics DPSO [24] and PageRank [37], CELF [38] greedy algorithm and SA based on objective function optimization [39], in addition, there is a hybrid heuristic and greedy algorithm based on the impact of topological potential maximization algorithm (TSH)[40]. Experimental parameter setting, PageRank jump factor $\varepsilon = 0.15$, in the SA algorithm $T_0 = 200{,}000$, $T_f = 10{,}000$, $\Delta T = 2{,}000$, $q = 200$. SA-DPSO population size and iteration times are 100 times, learning factor $c_1 = c_2 = 2$, inertia weight $\omega = 0.8$, other parameters are the same as in SA. The experiments all adopt IC model.

(2) Analysis of the scope of influence

The influence range of seed set is shown in Fig. 5, where abscissa is the number of seed set, and ordinate is the spread influence range.

Fig.5 shows the results of the experiment, where (b) and (d) are the magnifications of the parts in (a) and (c), respectively. In Fig. 5, SA-DPSO is compared with the five influence maximization algorithms respectively. It can be seen that SA-DPSO is superior to SA, DPSO, TSH and PageRank algorithms except CELF. DPSO algorithm is not as effective as SA-DPSO algorithm, because DPSO is more likely to fall into local optimization than SA-DPSO algorithm, resulting in premature convergence of the algorithm. In addition, the experimental results show that TSH algorithm is superior to SA, DPSO and PageRank algorithm, because TSH algorithm not only uses topological potential to heuristically select nodes with high weight as the seed set, but also uses CELF algorithm to determine the optimal subset. The PageRank algorithm in the figure has the worst performance and stability. When k<10, the influence performance of all algorithms is approximately the same. When k >10, even under different propagation probabilities, CELF, SA-DPSO and TSH algorithms show efficient performance. Especially in Fig 5 (c), it can be found that the solving performance of SA-DPSO and TSH algorithms are comparable to that of CELF algorithm. When k=30, p=0.01, SA-DPSO algorithm is better than SA, DPSO, TSH and PageRank algorithm by 39.72%, 18.04%, 8.67% and 45.56% respectively. When k=30, p=0.05, SA-DPSO algorithm is better than SA, DPSO, TSH and PageRank algorithm by 34.52%,18.01%,5.98% and 63.94% respectively.

Description:

The Fig.5 of (a) and (c) are very similar, because only the propagation probability of the algorithm parameters in the two figures is different, (a) is 0.01, and (c) is 0.05. The difference between the two figures lies in the size of the influence propagation range:

In figure (a), when the propagation probability is small and k=50, the maximum influence range of SA-DPSO is up to 2500; In (c), when the transmission probability is large and

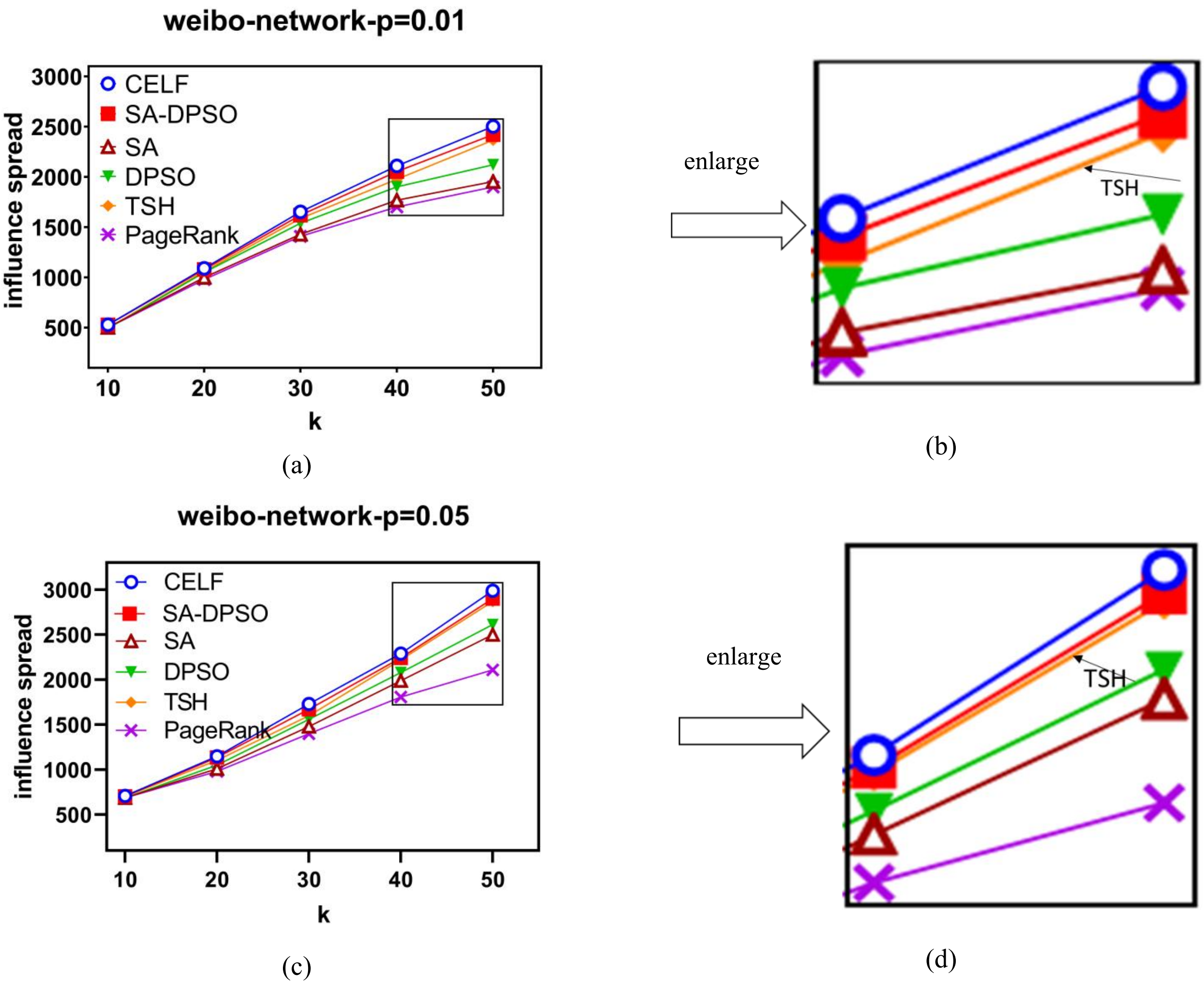

Fig. 5 Range of Influence

k=50, the influence range of SA-DPSO can reach nearly 3000.

In addition, in figure (b) and (d), the difference of figure (a) and (c) in the influence propagation range of the six algorithms in the node set 40-50 is amplified. From these two figures, it can be seen that the influence propagation range of each algorithm is larger, but the SA-DPSO algorithm can also catch up with the influence propagation range of CELF algorithm. And compared with figure (b), the SA-DPSO algorithm in figure (d) has better solving ability than TSH algorithm. According to the analysis of figures (b) and (d), the differences between algorithms are more obvious due to the improvement of propagation probability.

V. Summary

Influence measurement is still an issue worth discussing in social network analysis, and it is necessary to further study its influencing factors. Based on the theory of topological potential in physics, this paper integrates the method of topological potential and the attributes of the microblog network in the social network influence measurement method, so as to achieve the purpose of spreading the influence of microblog users in a wider range. In order to verify the effectiveness of the measurement method, PSO is used in this paper, and SA is introduced to solve the problem that PSO is easy to fall into local optimal. A real microblog network is used to test the effectiveness of the proposed algorithm. In addition, the experimental results verify the optimization ability of the SA-DPSO algorithm through comparison with other algorithms. The influence of SA-DPSO is similar to that of the CELF algorithm, but the running time of SA-DPSO algorithm is faster than CELF algorithm. In this study, the factors we considered for the measurement of influence are not very comprehensive. For example, we only investigated the characteristics of static Weibo network users, but did not consider the influence of time changes, so the next step can be to study the impact of time-dependent trends on the Changes in influence.

References

[1] Can U, and Alatas B, "A new direction in social network analysis: Online social network analysis problems and applications," Physica A: Statistical Mechanics and its Applications, 2019, 535, pp. 122372

[2] Syed AA, Gaol FL, Suparta W, Abdurachman E, Trisetyarso A, and M atsuo T, "Prediction of the Impact of Covid-19 Vaccine on Public Heal th Using Twitter," IAENG International Journal of Computer Science, vol. 49, no. 1, pp19-29 , 2022.

[3] Alfred R, Loo Yew J, Obit JH, Lim Y, Haviluddin H, and Azman A, "Social Media Mining: A Genetic Based Multiobjective Clustering Approach to Topic Modelling," IAENG International Journal of Computer Science, vol. 48, no. 1, pp32-42, 2021.

[4] Wang R, Rho S, Chen B-W, and Cai W, "Modeling of large-scale social network services based on mechanisms of information diffusion: Sina Weibo as a case study," Future Generation Computer Systems, vol. 74, pp. 291-301, 2017.

[5] Yu X, Wei XU, Zhaoxi S et al, "Analysis on Algorithms of Identifying Regional Influential Users in Micro-blogging," Computer Science, vol. 39, no. 9, pp. 38-42, 2012.

[6] Zareie A, Sheikhahmadi A, "A hierarchical approach for influential node ranking in complex social networks," Expert Systems with Applications, vol. 93, pp. 200-211, Mar 2018.

[7] Kwak H, Lee C, Park H, and Moon S, "What is Twitter, a social network or a news media?," in Proceedings of the 19th international conference on World wide web, 2010, pp. 591-600.

[8] Chen S, Fan J, Li G et al, "Online topic-aware influence maximization," Proceedings of the VLDB Endowment, vol. 8, no. 6, pp. 666-677, 2015.

[9] Qing L, Lixuan L et al, "Research on the Model of Social Network User Influence Quantification——Taking Sina Weibo as an Example," 2020 3rd International Seminar on Education Research and Social Science, no. 2018-8, pp. 202-207, 2021.

[10] Wang Y, Maeda A, "Twitter user's interest detection by using followee information based on lda topic model," in Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists, 2019, pp. 40-44.

[11] Wang Z, Chen Z, Zhao Y, and Niu Q, "A novel local maximum potential point search algorithm for topology potential field," International Journal of Hybrid Information Technology, vol. 7, no. 2, pp. 1-8, 2014.

[12] Zhuang Y-B, Li Z-H, Zhuang Y-J, "Identification of influencers in online social networks: measuring influence considering multidimensional factors exploration," Heliyon, vol. 7, no. 4, p. e06472, 2021.

[13] Mao YX, Zhou LJ, Xiong NX, "TPS: A Topological Potential Scheme to Predict Influential Network Nodes for Intelligent Communication in Social Networks," Ieee Transactions on Network Science and Engineering, vol. 8, no. 1, pp. 529-540, Jan 2021.

[14] Han QY, Yu L, Zhang R et al, "A P2P Recommended Trust Nodes Selection Algorithm Based on Topological Potential," in 1st IEEE International Conference on Communications and Network Security (CNS), Washington, DC, 2013, pp. 395-396.

[15] Li Y, Fan J, Wang Y, and Tan K-L, "Influence maximization on social graphs: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 30, no. 10, pp. 1852-1872, 2018.

[16] Peng S, Zhou Y, Cao L et al, "Influence analysis in social networks: A survey," Journal of Network and Computer Applications, vol. 106, pp. 17-32, 2018.

[17] Gong Y, Liu S, Bai Y, "Efficient parallel computing on the game theory-aware robust influence maximization problem," Knowledge-Based Systems, vol. 220, p. 106942, 2021.

[18] Huang H, Shen H, Meng Z et al, "Community-based influence maximization for viral marketing," Applied Intelligence, vol. 49, no. 6, pp. 2137-2150, 2019.

[19] Jain L, Katarya R, and Sachdeva S, "Opinion leader detection using whale optimization algorithm in online social network," Expert Systems with Applications, vol. 142, p. 113016, 2020.

[20] Ngamsa-Ard S, Razavi M, Prasad P, and Elchouemi A, "Point-of-interest (POI) Recommender Systems for Social Groups in Location Based Social Networks (LBSNs)–Proposition of an Improved Model," IAENG International Journal of Computer Science, vol. 47, no. 3, pp. 331-342, 2020.

[21] Semangern T, Chaisitsak W, Senivongse T, "Identification of risk of cyberbullying from social network messages," in Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science, San Francisco, USA, 2019, pp. 276-282.

[22] Chen G, Zhang A, Zhao C, Zhang Z, "Optimal Placement and Capacity of Combined DGs and SCs in Radial Distribution Networks Based on PSO-OS Algorithm," IAENG International Journal of Computer Science, vol. 48, no. 2, pp236-249, 2021.

[23] Li S, Liu H, and Du Y, "An Optimal Resource Allocation Scheme for Elastic Applications in Multipath Networks via Particle Swarm Optimization," IAENG International Journal of Computer Science, vol. 47, no. 2, pp278-283, 2020.

[24] Gong M, Yan J, Bo S et al, "Influence maximization in social networks based on discrete particle swarm optimization," Information Sciences, vol. s 367–368, no. C, pp. 600-614, 2016.

[25] Wang Q, Gong M, Chao S, and Wang S, "Discrete particle swarm optimization based influence maximization in complex networks," in 2017 IEEE Congress on Evolutionary Computation (CEC), 2017: IEEE, pp. 488-494.

[26] Yang J, Liu J, "Influence maximization-cost minimization in social networks based on a multiobjective discrete particle swarm optimization algorithm," IEEE Access, vol. 6, pp. 2320-2329, 2017.

[27] Zhou X, Ji F, Wang L et al, "Particle swarm optimization for trust relationship based social network group decision making under a probabilistic linguistic environment," Knowledge-Based Systems, vol. 200, no. 9, p. 105999, 2020.

[28] Singh SS, Kumar A, Singh K, and Biswas B, "LAPSO-IM: A learning-based influence maximization approach for social networks," Applied Soft Computing, vol. 82, p. 105554, 2019.

[29] Hao R, Yuman L, and Haimei C, "Analysis of Weibo Hot Topic Discovery Based on Particle Swarm Algorithm," no. 2016-12, pp. 51-54, 2021.

[30] Weng J, Lim E-P, Jiang J, and He Q, "Twitterrank: finding

topic-sensitive influential twitterers," in Proceedings of the third ACM international conference on Web search and data mining, 2010, pp. 261-270.
[31] Java A, Song X, Finin T, and Tseng B, "Why we twitter: understanding microblogging usage and communities," in Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, 2007, pp. 56-65.
[32] Domingos P and Richardson M, "Mining the network value of customers," in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, pp. 57-66.
[33] Wenyan G, Nan H, Deyi L, and Jiangming W, "A network community discovery method based on topological potential," Journal of Software, vol. 20, no. 008, pp. 2241-2254, 2009.
[34] H. Li, "Identifying Influential Nodes in Directed Social Network Based on Topological Potential," Journal of Chinese Computer Systems, vol. 42, no. 7, pp. 1492-1499, 2021.
[35] Jin Z, Yan L, Junyong L, and Yucheng D, "Identification method of microblog speculation accounts based on feature analysis," Computer Engineering, vol. 41, no. 4, pp, 48-54, 2015, DOI: 201504009
[36] Yuanying L, "Research on the method of social network influence maximization," PhD, Yanshan University, 2017.
[37] Brin S and Page L, "Reprint of: The anatomy of a large-scale hypertextual web search engine," Computer networks, vol. 56, no. 18, pp. 3825-3833, 2012.
[38] Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, and Glance N, "Cost-effective outbreak detection in networks," in Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 420-429.
[39] Jiang Q, Song G, Cong G, Wang Y, Si W, and Xie K, "Simulated annealing based influence maximization in social networks," in 25th AAAI Conference on Artificial Intelligence and the 23rd Innovative Applications of Artificial Intelligence Conference, AAAI-11 / IAAI-11, August 7, 2011 - August 11, 2011, San Francisco, CA, United states, 2011, vol. 1: AI Access Foundation, pp. 127-132.
[40] Xiufang W, Weinan N, Chengai S, and Liqing Q, "Influence Maximization Algorithm Based on Topological Potential," Computer Engineering and Design, vol. 041, no. 003, pp. 735-741, 2020.

Qiwen Zhang was born on February 23, 1975, master degree, master tutor. He graduated from Gansu University of Technology, majoring in computer and applied technology, with a bachelor's degree in engineering in 1999. He received the M.S. degree in computer and applied technology from Lanzhou University of Technology in 2005. He presided over or participated in the completion of a number of projects such as the National Science and Technology Support Program, the National Natural Science Foundation of China, and the Natural Science Foundation of Gansu Province, and more than 20 horizontal projects, and won the Gansu Province University Science and Technology Progress Award (first prize). Published more than 40 academic papers. Main research directions: intelligent information processing, knowledge discovery, computational intelligence.

Qiaohong Bai was born in1995. She is currently a master student in software engineering at Lanzhou University of Technology. Her main research areas are artificial intelligence and pattern recognition.She has written a small accepted paper.

Jun Li, male, born in September 1980, doctor of engineering, professor. In 2005, he graduated from Northwest Normal University, majoring in computer science and technology, with a bachelor's degree in engineering. In 2009, he graduated from Lanzhou University of Technology, majoring in computer application, with a master's degree in engineering. In 2015, he graduated from Lanzhou University of Technology, majoring in control theory and control engineering, with a doctorate in engineering.

He is currently the director of the Science and Technology Information Department of Lanzhou Modern Vocational College, and the leading talent (level 1) of Lanzhou in 2021. He has won a third prize of Gansu Province Science and Technology Progress Award, a third prize of Gansu Provincial Department of Education Social Science Achievement Award, a third prize of Gansu Province Science and Technology Information Award, and a second prize of Lanzhou Science and Technology Progress Award. He has presided over and completed 4 projects above the provincial level, 2 projects of education reform of the Department of Education, and 1 project of the provincial natural fund. Published more than 10 academic papers. The main research directions are artificial intelligence, information security, and big data.