Neurocomputing 333 (2019) 169–184

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Information propagation model based on hybrid social factors of opportunity, trust and motivation

Jihong Wan[a], Xiaoliang Chen[a,b,*], Yajun Du[a], Mengmeng Jia[a]

[a] School of Computer and Software Engineering, Xihua University, Chengdu 610039, China
[b] Department of Computer Science and Operations Research, University of Montreal, Montreal, QC H3C3J7, Canada

ARTICLE INFO

Article history:
Received 8 July 2018
Revised 1 November 2018
Accepted 7 December 2018
Available online 30 December 2018

Communicated by Prof. J. Cao

Keywords:
Social network
Information propagation
Social factors
Social trust
Game choice motivation

ABSTRACT

The propagation of information in the network is a complex dynamic process. Establishing an accurate information propagation model around common social factors benefits a lot from its evolution simulation for identifying the valuable information and regulating public opinions. In this paper, we propose Opportunity, Social Trust and Game Choice Motivation relying on triadic closure principle in social network and construct a novel information propagation model with respect to combine these three types of social factors. Firstly, the interest similarity between two users is convenient for measuring the opportunity to receive certain information. Secondly, the threshold of social trust is calculated by coupling users' network influence and content contribution. Thirdly, game choice with a rule to compute the best benefits has recognized as the motivation of users to spread a message. Finally, the game choice information propagation model based on page rank algorithm (GCIP-Page Rank) is proposed. The experimental results show that information propagation led by users experiences such stages as information contact, information trust and information propagation. At the same time, users' social trust can accelerate the spread of information in microblog social network by considering both the network structure and information content simultaneously.

1. Introduction

Propagation models have been used to explain and simulate how information is diffused over social networks. Large-scale online communities, such as Facebook, Twitter and Microblog, have become increasingly popular among people or information communication and social contacts. It is widely acknowledged that everybody has been making possible effort to post their opinions in their focused virtual communities. It seems that everybody gives top priority to enjoy the relaxed and non face-to-face online contacts. However, the very invention has also created some potential problems that almost everybody is exposed to. For example, an increasing number of people have become obsessed with transferring unverified gossip without doing anything meaningful in virtual communities.

Hence, developing a highly accurate information propagation model benefits a lot for studying potential information propagation processes in social media. Perfect examples can be found in such well-recognized information flow analysis as the prediction and the guide of public opinion in social networks. These studies have provided many reasonable technical means to avoid the propagation of public opinions that threaten social security and cause social unrest.

The model of information propagation in social networks has been widely focused. However, there has been an excessive abstraction in existing work during study the information propagation problem, ignoring some significant social, motivational and affective factors [1,2]. Examples can be found in traditional models that only expressed with the weightless undirected graphs. Such models would serve as an effective way to reduce the complexity. Nevertheless, constructing simplified models of a social network may not correctly reflect the real situation in the long run. Some of the most important characteristics with respect to social factors are lost. Consequently, we are likely to get a model that is contrary to the real situation.

This paper presents three kinds of social factors that affect the opportunity, trust and motivation of information propagation in social networks. Then, evolutionary game theory is introduced in this

* Corresponding author at: School of Computer and Software Engineering, Xihua University, Chengdu 610039, China.

E-mail addresses: jhwan03@aliyun.com (J. Wan), chexiaol@iro.umontreal.ca (X. Chen), dyjdoc2003@aliyun.com (Y. Du), mengjia163@aliyun.com (M. Jia).

URL: http://202.115.144.153/jsj/cxl/list.psp (X. Chen), http://202.115.144.153/jsj/dyj/list.psp (Y. Du)

https://doi.org/10.1016/j.neucom.2018.12.062
0925-2312/

paper to construct a information propagation model with hybrid social factors. This theory has provided finite rational hypothesis to individuals, training the disseminators in a social network how to adjust their game strategies by referring to the benefit of their neighbors [3–6].

A novel model with hybrid social factors is established according to the network topology, the content of microblog and the information of users. The contributions of this paper are summarized as follows:

- Relying on triadic closure principle in social network, three types of social factors are defined as the opportunity, social trust and game choice motivation.
- In the process of information propagation, propagation factors such as the social and emotional aspects of the propagator are introduced. From the two perspectives of network topology and content contribution, social trust can accelerate the dissemination of information in microblog social network.
- The characteristics of maximizing individual incomes in evolutionary game are used as the motive for users to choose information propagation.
- The experimental results show that integrating the three social factors of opportunity, trust, and motivation into the information propagation model by using evolutionary game as bridge can yield a more realistic model of information propagation.

The rest of this paper is organized as follows. Section 2 presents the relative work of microblog social network and information propagation models. In Section 3, three types of social factors of opportunity, trust and motivation, which are represented by the triadic closure principle in social networks, are described as the opportunity, social trust and game choice motivation in this paper. In Section 4, the game choice information propagation based on page rank (GCIP-Page Rank) algorithm is proposed, which is used to construct the information propagation model of hybrid social factors based on opportunity, trust and motivation. The experimental results are displayed and analyzed in Section 5. Section 6 makes conclusions and discusses the future works.

2. Related work

2.1. Microblog social network

Microblog social network has attracted many studies which an important platform for users to obtain information and social contacts in a timely manner.

Based on its topology features [7,8], the sharing mechanism [9], the prediction model [10], and the interaction between users [11], microblog social network is usually used as an effective and realistic medium for analyzing and verifying the laws of information propagation [7,12–14]. Han et al. [15] analyzed the structural properties of Sina microblog social network and compared it with Twitter. Guan et al. [16] analyzed the forwarding behavior of Sina microblog social network users in 21 hot events and found that males were more active.

Through the introduction of the relationship of human interaction in the real world, the social relationship among users has become an important factor affecting the effect of information propagation. What's more, some social factors that are specific to social people, such as motivation and trust, can also be reflected in microblog social network. Accordingly, we can effectively control and guide the network public opinion that threatens social security and causes social unrest by studying the process of information propagation.

2.2. Information propagation models

Social network analysis is the core of the social network theory research. It relies on the construction of information propagation model according to the actual communication mechanism, in which the communication mechanism and the sharing relationship mechanism in the social network complement each other. The information communication mechanism can accurately express the framework of a certain information from the publisher to the receiver (see Fig. 1).

Fig. 1 shows the forms, methods, and processes of information propagation, including such basic information as disseminators, channels of propagation, media and receivers. On the basis of the information carrier, encoding and decoding of symbolic information (e.g., languages, pictures, videos, etc.) enable information to be transmitted to the receivers. And the information has the meanings of information transfer, emotional fusion, and exchanging of ideas.

The research on information propagation models related to this article is mainly divided into the following three aspects:

2.2.1. Process-based propagation models

A propagation model based on propagation process describes the state of users receiving information and the changing process of the state of information propagation. The researchers proposed that the information propagation of online social networks has the same pattern as the epidemic spread of the real world [2,17,18]. Most of the information propagation models originate from the classical SI, SIS, SIR and SIRS epidemic models [19,20].

In the classical epidemic model, individuals in the population are mainly divided into three categories, each of which is at a typical stage in epidemic spreading. The basic stages include: S (susceptible) - susceptible period (also known as sensitive period); I (infectious) - infectious period (also known as contagious period); R (removed) - removed period (also known as immune or recovery period). In the model, an individual with a state of S will infect the disease with a probability of β and with a state of I will recover with a probability of γ.

In the SIS model, those individuals recovered will automatically become susceptible. In the SIR model, individuals will remain in the recovery state. However, individuals will become susceptible after recovery with probability α in the SIRS model. Differential equations are usually used to represent changes in the number of individuals in these states. For example, in the SIR model, the following equations are used to build the model [21]:

$$\frac{ds(t)}{dt} = -\beta i(t)s(t), \quad \frac{di(t)}{dt} = \beta i(t)s(t) - \gamma i(t), \quad \frac{dr(t)}{dt} = \gamma i(t) \tag{1}$$

where $s(t)$, $i(t)$ and $r(t)$ are the fractions of the population in each of the three states, and the last equation is redundant, since necessarily $s + i + r = 1$ at all times.

This model usually only constructs the information propagation rules, but does not consider the influence of the network topology on the spreading behavior. In view of the close relationship between the spread of information on the blog network and the characteristics of the blog network itself, Leskovec et al. [22] constructed the information cascade propagation model in the blog network based on the SIS model. Furthermore, Xiao et al. [23] obtained the hotspot propagation model based on user multidimensional attributes and evolutionary games, combined with the traditional SIR epidemic model, considering the external and internal driving factors that affect hot topic information diffusion.

These information propagation models mainly focus on the dynamic of the propagation process and the redistribution of individuals. But they do not take into account some key social factors

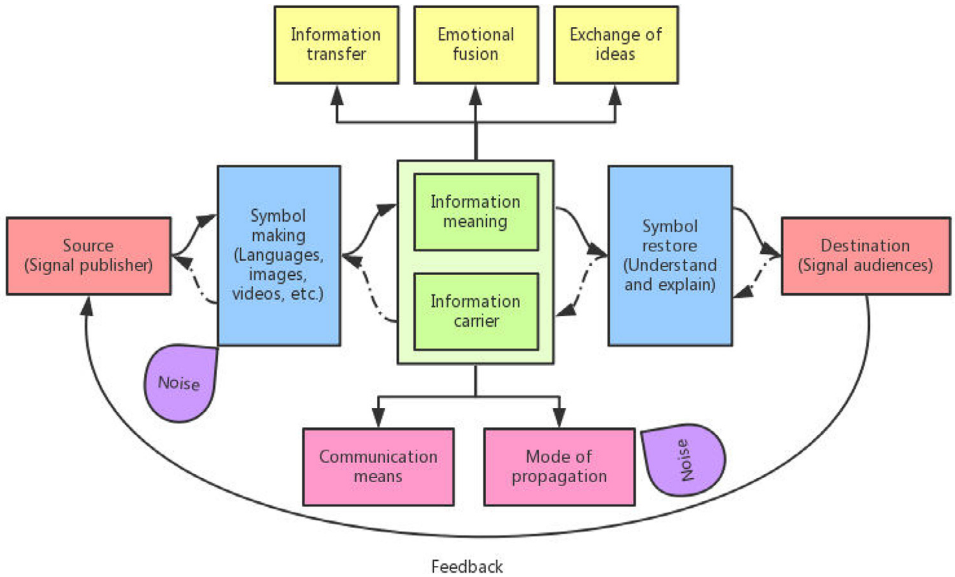


Fig. 1. Framework description of information communication mechanism.

(such as opportunity, trust and motivation) that users themselves have in the process of information propagation, and do not show the key propagation characteristics of the information propagation process.

2.2.2. Opinion leader-based propagation models

The comments or remarks of opinion leaders have certain influence in a particular field. Opinion leaders play an important role in information propagation, which they emerged in response to emergencies can influence and guide the decision-making choices of the public [24–27].

At present, most studies aim at accurately identifying the opinion leaders in the network [28–30]. For example, feature analysis and extraction of opinion leaders [31,32], cognition and personality testing of opinion leaders [33], and ranking of opinion leaders [7,34]. These efforts focus on the way of identification and accuracy of opinion leaders in the propagation of information.

Further works mainly consider the role and purpose of opinion leaders in the propagation of information. For example, in order to analyze the role of opinion leader in online social network (OSN) and the life cycle of Sina Microblog information diffusion, the public emergency information diffusion model and a new method grading opinion leader model were proposed [35].

Since then, Yong et al. [36] evaluated the influence of opinion leaders through trust indicators and distrust indicators between users, in which the indicators used knowledge scores (KS), match coefficients (MC), and Jaccard coefficients (JC) to reflect the strength of trust relationships between users by exploiting positive and negative link scores. Ortega and Vallejo [37] proposed a user polarization trust mechanism based on trust and reputation system to achieve trust and distrust propagation in social network.

Although these researchers adopt the user trust mechanism to determine the key role of opinion leaders in the process of information propagation, they do not pay attention to the relationship and importance of the trust mechanism between opinion leaders and users' information propagation.

2.2.3. Game theoretic-based propagation models

Some works focus on the information forwarding behavior from the perspective of game theory [38]. Especially, Jafari and Navidi [39] put forward a game-theoretic approach for modeling competitive diffusion over social networks, which shows that information content, topological structure of graph and the individual's initial tendency have significant influence on diffusion process.

Zinoviev and Duong [5] exploited the knowledge level, the degree of trust and influence of the actor to form the strategies of publishing and comment. And a game theory model of the one-way information forwarding and feedback mechanism of the star-type social network was proposed, which takes the participants' personality into consideration. Qiu et al. [6] used the utility function to combine the user's own features (such as knowledge, belief, persuasion, memory and reputation, etc) with the user's behavior characteristics (specially, propagation and reception) to obtain user's benefits in evolutionary propagation. Thus, a potential information propagation model in the form of comment and forwarding was constructed.

These studies only unilaterally regard trust as a measure of user behavior, and do not take into account the key bridge role that it plays in the whole process of information propagation from the network topology.

3. Three social factors and their algorithms

In this section, we first define the microblog social network. Then, three types of social factors of opportunity, trust and motivation, which are described by the triadic closure principle in social network, are defined as opportunity, social trust and game choice motivation, respectively.

Fig. 2 is a schematic diagram of the microblog social network. It means that after a microblog user issues a message, neighbor nodes with similar interests will have the opportunity to receive the information, which is represented by a dashed line. After the user's game revenue is measured, the neighbor node who maximize its benefit is selected to propagate. The dotted ellipse represents the series of information propagation.

A microblog social network is defined as a directed weighted graph $SN = (V, E, \Delta, \Phi)$, in which $V = \{u_1, u_2, \ldots, u_n\}$ and $E = V \times V = \{\varepsilon_{ij}\}$ are the set of users with size n and the association of similar interests between users, respectively; $\Delta = \{\delta^{\omega}_{u_i}\} \quad i = 1, 2, \ldots, n$ is the weight set of users in microblog social network, which refers to the social trust held by users in the process of information propagation; $\Phi = \{\varphi^{\omega}_{\varepsilon_{ij}}\} \quad i, j = 1, 2, \ldots, n$ is the weight set of associated edge between the microblog users, which mea-

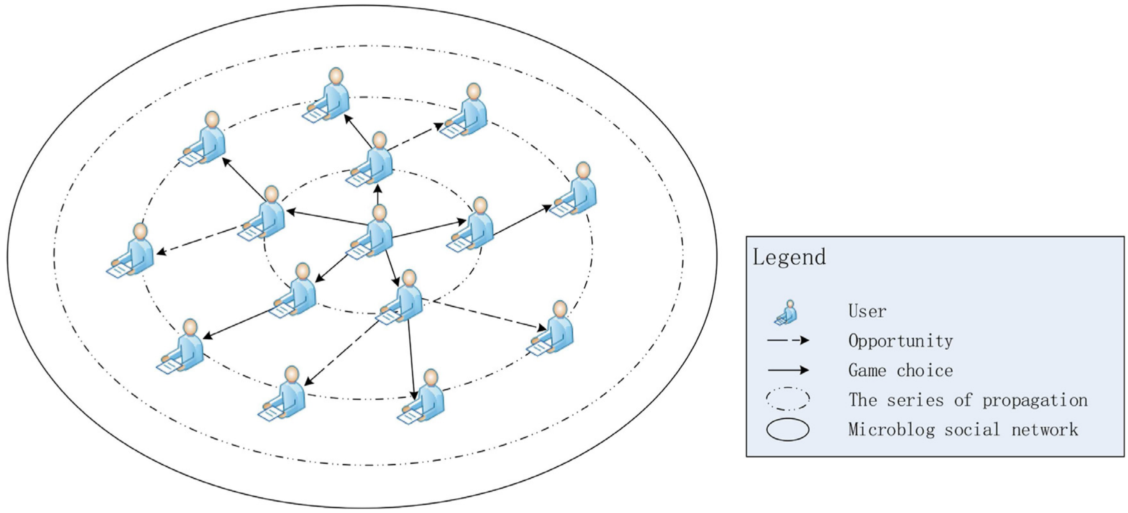


Fig. 2. Microblog social network schematic.

sures the chance of microblog users to receive a certain information.

3.1. Opportunity

Neighbor users with the similar interest of a user u_i may focus on a same topic and have a chance to receive the message m that user u_i posts on microblog. Consequently, the similarity of interest between two adjacent users is used as an opportunity metric for receiving certain information.

Focused interests for a user are usually presented as a Boolean vector, according to the interest characteristics of microblog users. Therefore, improved Tanimoto similarity coefficient can be used to calculate the similarity between individuals in Boolean metric.

Definition 1. ***Opportunity.*** The possibility for user u_j to receive information m transmitted by user u_i is determined by Eq. (2):

$$Opp_{\langle u_i,u_j,m\rangle,u_j\in N(u_i)} = \frac{\vec{u}_i{}' \bullet \vec{u}_j{}'}{|\vec{u}_i{}'|^2 + |\vec{u}_j{}'|^2 - \vec{u}_i{}' \bullet \vec{u}_j{}'} = \varphi^{\omega}_{\varepsilon_{ij}} \tag{2}$$

$$\vec{u}_i{}' = \{\vec{u}_{i1}{}', \vec{u}_{i2}{}', \ldots, \vec{u}_{ik}{}'\} \quad k = 1, 2, \ldots, Count(\vec{u}_i \cup \vec{u}_j)$$

$$\vec{u}_{ik}{}' = \begin{cases} 1 & \text{if } \chi_k \in \vec{u}_i, \chi_k \in \vec{u}_i \cup \vec{u}_j \\ 0 & \text{Otherwise} \end{cases}$$

Among them, m represents a certain message that is spreading on the microblog social network and $N(u_i)$ is the collection of all adjacent nodes of the user u_i. Notation $\vec{u_i}$ represents the interest vector of the user u_i and $\vec{u}'_i = \{\vec{u}'_{i1}, \vec{u}'_{i2}, \ldots, \vec{u}'_{ik}\}$ with size $k = Count(\vec{u}_i \cup \vec{u}_j)$ is its Boolean dimension representation. Operator $Count(vector\,\varepsilon)$ calculates the number of elements in vector ε. Let χ_k represents a interest in interest vector $\vec{u}_i \cup \vec{u}_j$. Then, an element $\vec{u}'_{ik}$ of $\vec{u}'_i$ equals 1 if $\chi_k \in \vec{u}_i$. Otherwise, $\vec{u}'_{ik}$ equals 0. $Opp_{\langle u_i,u_j,m\rangle,u_j\in N(u_i)}$ indicates that the user u_j receives the opportunity (possibility) of the information m transmitted by the user u_i. Eq. (2) finally obtains a opportunity weight value of the associated edge $\varphi^{\omega}_{\varepsilon_{ij}}$ between any pair of microblog users.

Example 1. Suppose that the interest vectors about two microblog users u_i and u_j are $\vec{u}_i = (photography, tourism, investment, cooking)$ and $\vec{u}_j = (tourism, automobile, music, investment)$, respectively. Their union is $\vec{u}_i \cup \vec{u}_j = \{photography, tourism, investment, cooking, automobile, music\}$. Then, according to Eq. (2), we can get the corresponding Boolean dimension interest vectors $\vec{u}'_i = \{1, 1, 1, 1, 0, 0\}$ and $\vec{u}'_j = \{0, 1, 1, 0, 1, 1\}$ and the similarity 1/3 between the two microblog users.

We regard the information received by users as a hot event in microblog social network. A set of opportunities for microblog users to receive the hot event is defined as:

$$sim\_opportunity = \{Opp_{\langle u_i,u_j,m\rangle,u_j\in N(u_i)} | i, j = 1, 2, \ldots, n\} \tag{3}$$

The output of the Algorithm 1 is the opportunity for microblog

Algorithm 1 An opportunity algorithm for microblog users to receive information.

Input:
A hot event in microblog.
Output:
sim_opportunity.
1: *uid_list* ← ∅, *micro_content* ← ∅, *sim_opportunity* ← ∅;
2: **for** each $i \in (1, n)$ **do**
3: Crawling the user ID (*uid*) under the microblog hot event; *uid_list* ⇐ *uid*;
4: **end for**
5: **for** each $j \in (1, length(uid\_list))$ **do**
6: **for** each $k \in (1, l))$ **do**
7: Crawling microblog content ($content_{jk}$) of the current user u_j; $micro\_content_j \Leftarrow content_{jk}$;
8: **end for** $micro\_content \Leftarrow content_j$;
9: **end for**
10: **for** each $g \in (1, length(uid\_list))$ **do**
11: *Jieba.analyse.set_stop_words*;
12: **end for**
13: **for** each $h \in (1, length(uid\_list))$ **do**
14: Calculate $opportunity_h$ for the user to receive the event information m; $sim\_opportunity \Leftarrow opportunity_h$;
15: **end for**
16: **return** *sim_opportunity*

users to receive the information in the process of propagation.

The main operations in Algorithm 1 consist of three parts, information crawling (Labels 2–9 in Algorithm 1), word segmentation (Labels 10–12 in Algorithm 1), and opportunity calculation

(Labels 13–15 in Algorithm 1). First, the time complexity of information crawling is known as $nlog(n)$ with n websites and $log(n)$ for set repeatability decision. However, set repeatability decision is not necessary here because of one-to-one correspondence between microblog posts and user ID. Hence, the time complexity of crawling shown in Labels 2–4 is (n) with n microblog posts. The crawling shown in Labels 5–9 is similarly $(n \times l)$ with $length(uid\_list) <= n$, which n and l represent microblog posts under the hot event and corresponding microblog posts of the current user u_k, respectively. Second, Jieba.analyse.set stop words is an outside standard library operation that is not generally consider as a time consumption in a new algorithm. Finally, the time complexity of opportunity calculation is (n) since mathematical formula calculation of known data is simply. Consequently, the comprehensive time complexity of Algorithm 1 is $O(n \times l)$ and the space complexity is $O(n)$ for a memory list to store crawling information.

Notation $'\leftarrow'$ and $'\Leftarrow'$ symbols in the algorithm represent assignment and adding values, respectively.

3.2. Social trust

Developing a reasonable measure of social trust devotes a lot to predict users actual feedback behavior. Determining the credibility of a message is necessary for a microblog user when he/she is likely to receive a message from their neighbors, which is usually depended on the social trust of microblog user who publish the message. Users thereby produce such feedback behaviors as reading, sharing, commenting, and praising. This paper comprehensively combines the two aspects of network structure and information content to measure user's social trust, that is, network structure and content contribution.

3.2.1. Users' social trust based network structure

As the influence of information content can not be measured according to network structure, the greater the influence of the network users, the greater the social trust. Hence, the network influence [40] of a user who posts a message is used in this work as one of standards of user's social trust.

Definition 2. ***Network opinion leader, NOL.*** The network influence of user u_i $NOL(u_i)$ is as follows:

$$NOL(u_i) = \frac{deg^{in}(u_i)}{\sum_{i=1}\sum_{j>1}\varepsilon_{ji}} \tag{4}$$

where $NOL(u_i)$ is influenced by the number of u_i's fans (that is, the in-degree of nodes in the network, $deg^{in}(u_i) = \sum_{j>1}\varepsilon_{ji}$).

3.2.2. Users' social trust based content contribution

The content contribution of the microblog user u_i $CC(u_i)$ in the information propagation process is mainly determined by the three parts: the frequency of publishing microblog posts within a specified period of time $MF(u_i)$, the amount of microblog posts $MN(u_i)$, and the number of user's feedbacks $MFB(u_i)$ (sharings $MS(u_i)$, coments $MC(u_i)$, and praisings $MP(u_i)$) from other microblog users.

User behavior relationships are listed as follows to explain the feedback.

(1) Microblog behavior "share", denoted by $ms(u_{j_{\langle i,j\rangle \in \varepsilon_{ij}}}) = \langle m_i, m_j, s\rangle$, is defined as the relationship between two microblog messages, where m_i represents the posted original microblog messages by user u_i, and $\langle m_i, m_j, s\rangle$ represents the sharing relationship between two users. The information propagation direction is $u_i \dashrightarrow u_j$ if user u_j shares user u_i's microblog information.

(2) Microblog behaviors "comment" and "praise" are represented by the relationship between users and messages. Different contributions of "comment" and "praise" are expressed as: a "comment" of the user u_j to message of the user u_i, $mc(u_{j_{\langle i,j\rangle \in \varepsilon_{ij}}}) = \langle u_j, m_i, c\rangle$, and a "praise" of the user u_j to message of the user u_i $mp(u_{j_{\langle i,j\rangle \in \varepsilon_{ij}}}) = \langle u_j, m_i, p\rangle$.

Definition 3. ***Behavior feedback.*** In microblog social network, microblog user u_i get the feedback from other microblog users $MFB(u_i)$ is organized as follows:

$$MFB(u_i) = \alpha_1 MS(u_i) + \alpha_2 MC(u_i) + \alpha_3 MP(u_i) \tag{5}$$

By Definition 3, user u_i gets the feedback from other microblog users $MFB(u_i)$, which is mainly composed of three parts:

- Microblog sharing numbers $MS(u_i)$, $MS(u_i) = \sum_j ms(u_{j_{\langle i,j\rangle \in \varepsilon_{ij}}})$.
- Microblog comment numbers $MC(u_i)$, $MC(u_i) = \sum_j mc(u_{j_{\langle i,j\rangle \in \varepsilon_{ij}}})$.
- Microblog praising numbers $MP(u_i)$, $MP(u_i) = \sum_j mp(u_{j_{\langle i,j\rangle \in \varepsilon_{ij}}})$.

Definition 4. ***Content contribution.*** The degree of content contribution of microblog user u_i in the process of information propagation $CC'(u_i)$ is specifically expressed as follows:

$$CC(u_i) = \alpha_4 MF(u_i) + \alpha_5 MN(u_i) + MFB(u_i) \tag{6}$$

$$CC'(u_i) = \frac{CC(u_i) - min[CC(u_i)]}{max[CC(u_i)] - min[CC(u_i)]} \tag{7}$$

where, $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 = 1$. Then, we normalized the content contribution of the user in the process of information propagation.

3.2.3. Computing users' social trust

The information will have a great impact in the process of propagation when the publisher of the information or the content of the information itself has certain authority, professionalism, and influence. Accordingly, adjacent users with similar interests will trust the information content and generate such feedback behaviors as reading, forwarding, commenting, or praising. Consequently, this paper uses the coupling value of the network influence of the microblog user and the content contribution of information as index to measure the social trust of microblog user.

Definition 5. ***Social trust.*** The social trust of microblog user u_i $ST(u_i)$ is defined as follows:

$$ST(u_i) = \eta_1 NOL(u_i) \times \eta_2 CC'(u_i) = \delta^{\omega}_{u_i} \tag{8}$$

Among them, $NOL(u_i)$ is the network influence of microblog user u_i who publish the information, $CC'(u_i)$ is the degree of content contribution of microblog user u_i in the process of information propagation, and $\eta_1 + \eta_2 = 1$. We take the social trust of the microblog user u_i as the node weight of the user in the network structure $\delta^{\omega}_{u_i}$.

The social trust set of users in microblog social network is defined as:

$$social\_trust = \{ST(u_i) | i = 1, 2, \ldots, n\} \tag{9}$$

The objective of Algorithm 2 is to output the social trust of users in microblog. The main operations in Algorithm 2 contain two phase: information crawling (Labels 2–7 in Algorithm 2) and microblog users' social trust calculation (Labels 8–12 inAlgorithm 2).

In the first phase, we crawl the microblog user ID (uid) under the microblog hot event by using the same way as the crawling process appeared in Algorithm 1. Then, the number of sharings $MS(u_i)$, comments $MC(u_i)$ and praisings $MP(u_i)$ are also crawled, respectively. Next, according to the uid in uid_info, this work continues to obtain the number of fans $deg^{in}(u_j)$ in the personal

Algorithm 2 An trust algorithm for users in microblog social network.

Input:
A hot event in microblog.
Output:
The social trust set *socail_trust* of microblog users.

1: $uid\_info \leftarrow \varnothing$, $social\_trust \leftarrow \varnothing$;
2: **for** each $i \in (1, n)$ **do**
3: $uid\_info \Leftarrow \{uid, MS(u_i), MC(u_i), MP(u_i)\}$;
4: **end for**
5: **for** each $j \in (1, length(uid\_info))$ **do**
6: $uid\_info \Leftarrow \{deg^{in}(u_j), MN(u_j), MF(u_j)\}$;
7: **end for**
8: **for** each $k \in (1, length(uid\_info))$ **do**
9: Calculate $NOL(u_k)$ and $CC(u_k)$ according to Eqs.(4) (5) and (6), respectively.
$uid\_info \Leftarrow NOL(u_k)$;
$uid\_info \Leftarrow CC(u_k)$;
10: The current microblog user's content contribution is normalized to $CC'(u_k)$ based on Eq.(7).
$uid\_info \Leftarrow CC'(u_k)$;
11: Finally, use Eq.(8) to calculate the social trust $ST(u_k)$ of the current microblog user u_k;
$social\_trust \Leftarrow ST(u_k)$;
12: **end for**
13: **return** *social_trust*

homepage about current microblog user u_j, the total number of microblog posts $MN(u_j)$, and the number of microblog posts in one month $MF(u_j)$ by crawling processes. Lastly, we add crawled information to list *uid_info*. Correspondingly, the time complexity of crawling shown in the first phase is (n).

Furthermore, we calculate the personal influence $NOL(u_k)$ and the content contribution $CC(u_k)$ of current microblog user u_k according to Eqs. (4), (5) and (7), respectively. Then, users' social trust based on the personal influence and the content contribution (Labels 8–12 in Algorithm 2) are calculated. Hence, the time complexity of users' social trust calculation is (n) by considering the basic equation calculation described in the second phase. Therefore, the comprehensive time complexity of Algorithm 2 is $O(n)$ and the space complexity is $O(n)$ for a memory list to store crawling information and calculation values.

In microblog social network, users generate the motivations of sharing, commenting, and praising based on social trust, then carry out the next step of information propagation. On the other hand, users always choose the messages with their best interest to share in the whole network. The motivation of choose can be described by choice strategies in game theory.

3.3. Game choice motivation

Every user in social network always chooses the information of the most beneficial to their own interests to spread and diffuse. Consequently, we use the game theory as the motivation of the user to choose the propagation of information. The motivation of participants in the process of information propagation is the key factor to ensure the propagation of information rules. In order to better illustrate the direction of information propagation in microblog social network, we adopt the sharing mechanism of information propagation characteristics (publishing, sharing, commenting and praising, etc.) to describe the process of information propagation. The Nash equilibrium theorem, the core of the game theory, is used to describe the motivation of participants in the information propagation process activities.

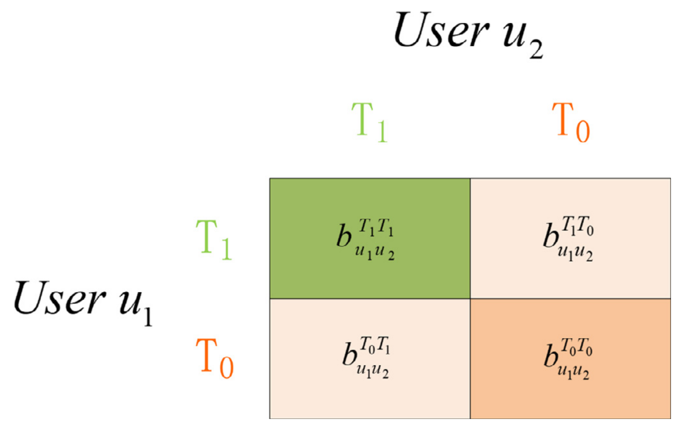


Fig. 3. Benefit matrix of microblog user u_1 in the process of information propagation.

The Nash equilibrium theorem assumes that the strategy chosen by users is the best response to each other, that is, users have consistency in their choose of strategies. Users achieve a balanced state in the process of interaction, in which no individual can increase its income by unilaterally changing the best strategy. Hence, the Nash equilibrium theorem can be used to judge the distribution of user's decisions in the stable state. A user in a social network can be regarded as a player in a evolutionary game. And each user's two possible actions, i.e., to share or not share, are corresponding with two strategies:

$$\begin{cases} T_1, & \text{share the information,} \\ T_0, & \text{not share the information.} \end{cases}$$

In the cooperative game model, behavior T_1 is considered as choice cooperation, and T_0 is regarded as abandonment of cooperation. The cooperation between the two parties needs to pay the cost. Hence, for microblog users u_1 and u_2 with similar interest, we have several possibilities for them to combine together. The profits of microblog user u_1 are defined as follows:

If both u_1 and u_2 make choice T_1, u_1 gets profit $b_{u_1u_2}^{T_1T_1}$;
If both u_1 and u_2 make choice T_0, u_1 gets profit $b_{u_1u_2}^{T_0T_0}$;
If u_1 makes choice T_1 while u_2 makes choice T_0, u_1 obtains profit $b_{u_1u_2}^{T_1T_0}$;
If u_1 makes choice T_0 while u_2 makes choice T_1, u_1 obtains profit $b_{u_1u_2}^{T_0T_1}$.

Consequently, a benefit matrix of microblog user u_1 in the process of information propagation is generated and shown in Fig. 3.

The propagation process of information in microblog social network is regarded as the process of users selecting the best decision through gaming and maximizing their own benefits in this paper. And we consider the game between the user and multiple neighbor nodes as a combination of multiple double games. The decisions made by users in microblog social network are highly autonomous and independent. Meanwhile, users' behavioral decisions are influenced by their neighbors. So we introduce the two concepts of neighbor propagation pattern and candidate propagation nodes.

Definition 6. ***Neighbor propagation pattern.*** Given microblog social network *SN*, vertex u in *SN*, and radius r, the neighbor propagation pattern p of node u that shows the neighbor signature by vertices within r hops away from u is denoted as $P_{SN}^r(u)$. A neighbor propagation path of *SN* is a path with r hops from the node u. A neighbor propagation tree of *SN* consists of all the longest paths starting from u with distances less than r. Hence, a neighbor prop-

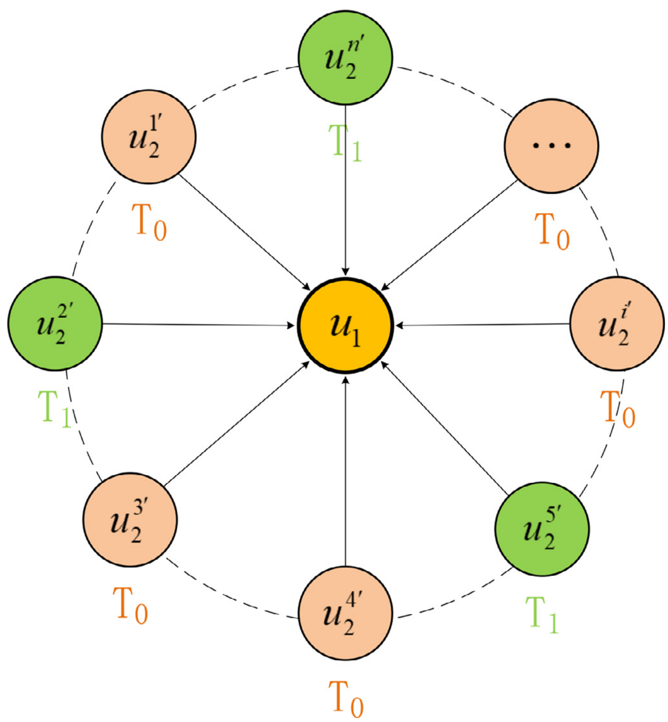


Fig. 4. Microblog user u_1 makes a game choice between actions T_0 and T_1 based on the behavior of neighboring users.

agation subgraph of SN is constructed by all vertices within r hops away from u.

Definition 7. ***Candidate propagation nodes.*** The candidate propagation nodes of microblog user u_i is the set of nodes in SN that satisfies the neighborhood pattern $P^r_{SN}(u_i)$:

$$C(u_i) = \{u_j | u_j \in V(SN), P^r_{SN}(u_i) = true\} \tag{10}$$

Fig. 3 illustrates the profit value when the user u_1 selects the specified user u_2 to disseminate or not to disseminate the information. In fact, when user u_1 conducts information dissemination, the set of candidate propagation nodes composed of the next propagation nodes ($r = 1$) of user u_1 is $\{u_2^{1'}, u_2^{2'}, \ldots, u_2^{i'}, u_2^{(n-1)'}, u_2^{n'}\}$, as shown in Fig. 4. The rule that the microblog user u_1 chooses which candidate propagation node to use as the next propagation node is based on the GCIP-Page Rank algorithm proposed in Section 4.

On the basis of opportunity and social trust, microblog users generate motives of comments, sharings and praisings by their own game revenue to spread the information. Therefore, the game motivation is defined as follows:

Definition 8. ***Game choice motivation.*** The behavioral motivation of different users at any moment is expressed as the greatest benefit of the game between different behaviors:

$$Max\{f | f : (b^{\Theta(\tau)}_{u_i u_j}) \rightarrow (b^{\Theta(\tau)}_{u_i u_j})'\} \tag{11}$$

$$\Theta(\tau) = \{T_x T_y | \forall x, y \in \tau\}, \quad \tau = \{0, 1\}$$

$$\left(b^{\Theta(\tau)}_{u_i u_j}\right) = \delta^{\omega}_{u_i} \times \varphi^{\omega}_{\varepsilon_{ij}}$$

where, the value 0 in the set $\tau = \{0, 1\}$ means that no messages are shared (resp. share messages). On the contrary, the value 1 in the set $\tau = \{0, 1\}$ means that microblog user u_i shares the message. No any benefit appears if the value equals 0, which will not contribute to the process of information propagation. Set $\Theta(\tau) = \{T_x T_y | \forall x, y \in \tau\}$ represents the choice of users to share or not to share the information.

In Definition 8, we define the benefit accepted by microblog users in the process of information propagation as the function mapping $f : (b^{\Theta(\tau)}_{u_i u_j}) \rightarrow (b^{\Theta(\tau)}_{u_i u_j})'$ from the initial benefit distribution $(b^{\Theta(\tau)}_{u_i u_j})$ to the steady state of benefit distribution $(b^{\Theta(\tau)}_{u_i u_j})'$ in GCIP-Page Rank. The initial benefit distribution vector is represented as $(b^{\Theta(\tau)}_{u_i u_j}) = \delta^{\omega}_{u_i} \times \varphi^{\omega}_{\varepsilon_{ij}}$, which is determined by the opportunity of microblog user u_i to receive this information and the user's own social trust. In other words, it is calculated by the weight of node $\delta^{\omega}_{u_i}$ and the weight of edge between two nodes with similar interest characteristics $\varphi^{\omega}_{\varepsilon_{ij}}$. Accordingly, $(b^{\Theta(\tau)}_{u_i u_j})'$ is the benefit distribution vector of steady state, the concrete calculation process can be seen Section 4.

In the process of information propagation, microblog user u_i will choose the most beneficial result from the candidate propagation nodes $C(u_i)$ corresponding to the user u_j as the next receiver of the information propagation. In fact, the process of user's game choice is the process of information propagation.

4. Information propagation model of hybrid social factors based on opportunity, trust and motivation

In this section, the game choice information propagation based on page rank algorithm (GCIP-Page Rank) is proposed and shown in Algorithm 3. We use the GCIP-Page Rank to calculate the steady state probability distribution vector value as the information benefit of each node. Compared with the benefits of different user, we choose the node with the largest benefits as the result of the game choice of the upper level microblog users, that is, the next information receiver in the process of information propagation. The GCIP-Page Rank algorithm is used to continuously iterate the user's choice process to build an information propagation model.

Algorithm 3 A game choice information propagation algorithm based page rank.

Input:
A hot event in microblog;
$\gamma = 0.5$; $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \eta_1, \eta_2$.
Output:
Information propagation process map $SN' = (V', E', \Delta', \Phi')$.
$SN' \leftarrow \varnothing$;
for each i and $j \in length[uid\_info]$ with $i != j$ **do**
  Obtain $Opp_{\langle u_i, u_j, m\rangle, u_j \in N(u_i)}$ (*that is*, $\varphi^{\omega}_{\varepsilon_{ij}}$);
  Get $ST(u_i)$ (*that is*, $\delta^{\omega}_{u_i}$);
  $\vec{x_0} = (b^{\Theta(\tau)}_{u_i u_j}) \leftarrow \varphi^{\omega}_{\varepsilon_{ij}} \times \delta^{\omega}_{u_i}$;
  The probability matrix **P** of the random walk process of the node users is calculated according to the Eq. (12);
  while $\vec{x}_{i+1} \neq \vec{x}_i\mathbf{P}$ **do**
    $\vec{x}_{i+1} = \vec{x}_i\mathbf{P}$;
  end while
  $(b^{\Theta(\tau)}_{u_i u_j})' \leftarrow \vec{x}_{i+1}$;
  Select the node user u_j corresponding to the maximum value in the steady-state probability distribution vector $Max\{(b^{\Theta(\tau)}_{u_i u_j})'\}$;
  $SN' \Leftarrow (u_j, \varepsilon_{ij}, \delta^{\omega}_{u_j}, \varphi^{\omega}_{\varepsilon_{ij}})$;
end for
return SN'

In web search engines, the Page Rank value of a node depends on the link structure of the web graph. We focus on the method of scoring and ranking based on link structure in the link analysis of Web search engine. The first technique of link analysis is to

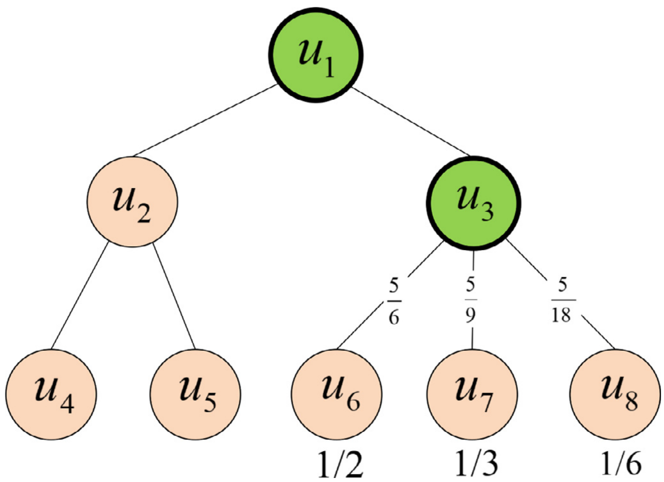


Fig. 5. Schematic diagram of initial state in information propagation.

add a score from 0 to 1 for each node in the Web graph, which is called Page Rank. The Page Rank algorithm of the web page is irrelevant to the query entered by the user, which means that the web page is ranked according to the Page Rank and has nothing to do with the query. Different from the Page Rank algorithm of web pages, the probability distribution of the initial state of our GCIP-Page Rank algorithm is the coupling vector value of the opportunity and the social trust. It reflects whether the user chooses to disseminate the information, which is determined by his opportunity (or possibility) to receive information from the upper layer microblog users and his own social trust.

We measure the opportunity as the weight of the edge between two adjacent nodes, and the social trust is the weight of the node. Suppose that the initial state of the node is 1, the corresponding initial state probability distribution vector is the coupling value of the opportunity and the social trust. As shown in Fig. 5, node u_1 is the publisher of a message, according to our rules, it spreads information to u_3. After u_3 receives the information, calculate the interest similarity of nodes between u_3 and its neighbor nodes u_6, u_7 and u_8 respectively. We take these values (5/6, 5/9, 5/18) as u_6, u_7 and u_8 to receive the opportunity measure of the information, respectively. The coupling value of the social trust of u_3 (3/5) and the chance measure with its neighbor nodes u_6, u_7 and u_8 is used as the initial state probability distribution vector $\vec{x_0} = (b^{\Theta(\tau)}_{u_3u_j}) = \delta^{\omega}_{u_3} \times \varphi^{\omega}_{\varepsilon_{3j}} = (1/2 \;\; 1/3 \;\; 1/6)$ in GCIP-Page Rank algorithm. We assume the opportunity relationship between u_3 and u_6, u_7, u_8 is bidirectional, and $\gamma = 0.5$.

The meaning of the GCIP-Page Rank (referred to as GCIP-PR) value proposed in this paper is to explain the probability of microblog users receiving information in the process of information propagation. And the GCIP-PR value of a user u_i is calculated as follows:

$$GCIP - PR(u_i) = \frac{1-\gamma}{N} + \gamma \sum_{u_j \in M_{u_i}} \frac{GCIP - PR(u_j)}{deg^{(out)}(u_j)} \tag{12}$$

where, M_{u_i} is a collection of all users that have outgoing edges to user u_i, $deg^{(out)}(u_j)$ is the number of outgoing edges of user u_j, and N is the total number of users.

We can calculate the GCIP-PR value of each user in the process of information propagation. According to the Eq. (12), the probability matrix $\mathbf{P}$ of the random walk process of the node users is calculated.

$$\mathbf{P} = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$

Table 1
A probability distribution vector sequence.

Probability vector	Probability distribution vector sequence		
$\vec{x}_0$	1/2	1/3	1/6
$\vec{x}_1$	1/4	1/2	1/4
$\vec{x}_2$	7/24	5/12	1/2
$\vec{x}_3$	13/48	11/24	13/48
$\vec{x}_4$	9/32	7/16	9/32
...	...	...	...
$\vec{x}$	5/18	4/9	5/18

According to the hypothesis, the initial state probability distribution vector is $\vec{x_0} = (1/2 \;\; 1/3 \;\; 1/6)$ in GCIP-Page Rank algorithm.

Then, the probability distribution after one step is $\vec{x}_1$:

$$\vec{x}_1 = (1/4 \;\; 1/2 \;\; 1/4) = \vec{x}_0\mathbf{P}$$

The probability distribution after the two step is $\vec{x}_2$:

$$\vec{x}_2 = \vec{x}_1\mathbf{P} = (1/4 \;\; 1/2 \;\; 1/4)\begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 5/12 & 1/6 & 5/12 \\ 1/6 & 2/3 & 1/6 \end{bmatrix}$$
$$= (7/24 \;\; 5/12 \;\; 7/24)$$

Repeat iteratively in the same way to obtain the probability distribution vector sequence shown in Table 1.

The case shown in Fig. 6 shows the information propagation process of microblog social network game choice. The final result $\vec{x} = (b^{\Theta(\tau)}_{u_1u_2})'$ is obtained when the GCIP-Page Rank mapping function is continuously iterated. In this paper, the maximum value of the steady state distribution vector is chosen as the result of user game choice. Therefore, the distribution converges to a steady state $\vec{x} = (5/18 \;\; 4/9 \;\; 5/18)$ after repeated iterations. In our example, we choose the maximum value of the steady state probability distribution vector 4/9 as the maximum benefit. Correspondingly, the next receiver of information that the node user u_3 should select in the information propagation process is the node user u_7.

The main contents of Algorithm 3 include three parts, the initial probability distribution calculation (Labels 4–6 in Algorithm 3), information iteration (Labels 7–11 in Algorithm 3) and selection of the propagation node (Label 12 in Algorithm 3). The preparation in the first part is to obtain two parameters: the Opportunity measure for users to receive information $Opp_{\langle u_i, u_j, m\rangle, u_j \in N(u_i)}$ (symbolized written as $\varphi^{\omega}_{\varepsilon_{ij}}$) and the social trust $ST(u_i)$ (symbolized denoted by $\delta^{\omega}_{u_i}$). The parameters are achieved according to Algorithm 1 and 2, respectively. Then, the initial probability distribution vector is captured by formula $\vec{x_0} = (b^{\Theta(\tau)}_{u_iu_j}) \leftarrow \varphi^{\omega}_{\varepsilon_{ij}} \times \delta^{\omega}_{u_i}$ in GCIP-Page Rank algorithm. Furthermore, the steady-state probability distribution vector $\vec{x}_{i+1} = (b^{\Theta(\tau)}_{u_iu_j})'$ is obtained by iteration. Finally, the node user u_j corresponding to the maximum value in $\vec{x}_{i+1}$ is chosen and added to the information propagation process map: $SN' \Leftarrow (u_j, \varepsilon_{ij}, \delta^{\omega}_{u_j}, \varphi^{\omega}_{\varepsilon_{ij}})$. A while loop is nested in the for loop in the whole process, so the time complexity of Algorithm 3 is $O(n^2)$ and the space complexity is $O(n)$ for a memory list to store the initial and steady-state values of current user.

5. Experiment

5.1. Experimental setup

5.1.1. Data collection and analysis

The experimental data in this paper is mainly obtained in two ways, as shown in Table 2. One is through the API interface provided by the Sina microblog platform, the other is through the web crawler. Since the interface provided by Sina microblog has a lot of access restrictions, it has great limitations in obtaining data. Therefore, the main way to obtain data is the second one.

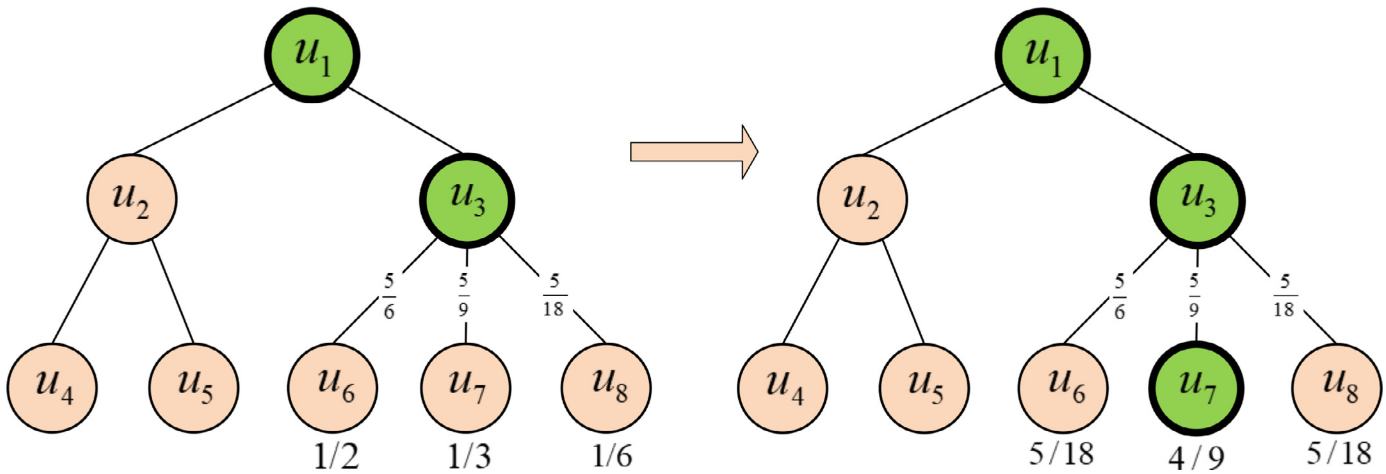


Fig. 6. Schematic process of propagation based on GCIP-Page Rank algorithm.

Table 2
Comparison of experimental data acquisition schemes (web crawler is mainly divided into "full collection" and "organized collection" according to the constraints of the acquired capability).

Program	Get timely	Complete acquisition	Identifying the heat	Distinguish true and false	Extensible	Stability
API	Keywords real-time acquisition	×	×	×	×	Very stable
Page crawler	Important priority	√	√	√	√	Relatively stable

KDD Cup 2012, Track 1[1] open dataset is used as Dataset 1. In addition, we crawl the data of three microblog hot events[2] from Sina microblog (http://m.sinna.com.cn), the largest Chinese online social network site, as samples of real data, which are Dataset 2 (Social event 1: Chengdu female driver was beaten), Dataset 3 (Entertainment event 2: Baihe Bai derailed) and Dataset 4 (International news event 3: THAAD incident), respectively. By setting a specific time interval, we get the number of related microblog posts of three hot events on the Sina microblog platform. Then, crawling such the details of the microblog hot event as the user ID, time of releasing, content of releasing, the number of sharings, the number of comments, and the number of praisings. Consequently, evaluate the results returned by the search and the percentage of valid articles in the amount of returned results are 95%, 96%, and 98%, respectively.

5.1.2. Experimental design

According to the information propagation model based on opportunity, trust and motivation proposed in this paper, the depth and breadth of information propagation on a hot event of microblog network are studied. The framework of the construction of information propagation model, as shown in Fig. 7, is specifically considered from the following steps:

(1) We obtain data from the microblog social platform, getting personal information and microblog posts from relevant users on a hot event. The natural language processing methods is used to analyze the text data, extract user's interest features, and form an interest feature set. According to the user's interest feature set, the improved Tanimoto coefficient is used to calculate the interest similarity between two adjacent users, which is defined as the opportunity measurement for the user to receive a certain information $Opp_{\langle u_i,u_j,m\rangle,u_j\in N(u_i)}$.

(2) A hot event of microblog is crawled to find the user ID in the information propagation. Firstly, the number of followees and followers of each node is crawled according to the user ID so that the individual influence of the user $NOL(u_i)$ can be calculated. Then, user's basic information is crawled according to the user ID, which includes the home page information, the amount of microblog posts $MN(u_i)$, the number of microblog posts within the specified time $MF(u_i)$, the number of sharings $MS(u_i)$, the number of comments $MC(u_i)$, and the number of praisings $MP(u_i)$. Therefore, the content contribution of the user $CC(u_i)$ is calculated according to Eqs. (5) and (6). Finally, the social trust measure of microblog user $ST(u_i)$ is obtained through the Eq. (8).

(3) Using the GCIP-Page Rank algorithm proposed in this paper, the probability of each node's steady state is calculated. The maximum probability is chosen as the result of the game choice, which the result is used as the object of the next information propagation. Among them, the method of combining opportunity measurement with social trust is used as the initial state probability distribution vector in GCIP-Page Rank algorithm. The model of information propagation process is well-established, using game evolution theory as the motivation for the user to choose disseminate information.

5.1.3. Experimental evaluation

In order to test the effect of the information propagation model proposed in this paper, we use three evaluation indexes of Precision, Recall and F1-Measure to evaluate the results of the information propagation model. The calculations are as follows:

$$Precision : P = tp/(tp + fp) \tag{13}$$

$$Recall : R = tp/(tp + fn) \tag{14}$$

$$F1 - Measure : F = \frac{2PR}{P + R} \tag{15}$$

5.2. Parameter analysis in the algorithm

Analytic Hierarchy Process (AHP) is a hierarchical weight decision analysis method developed by T. L. Saaty in the early

[1] https://www.kaggle.com/c/kddcup2012-track1/data.
[2] https://github.com/chinawan0230/InformationPropagation_Crawl_Data.

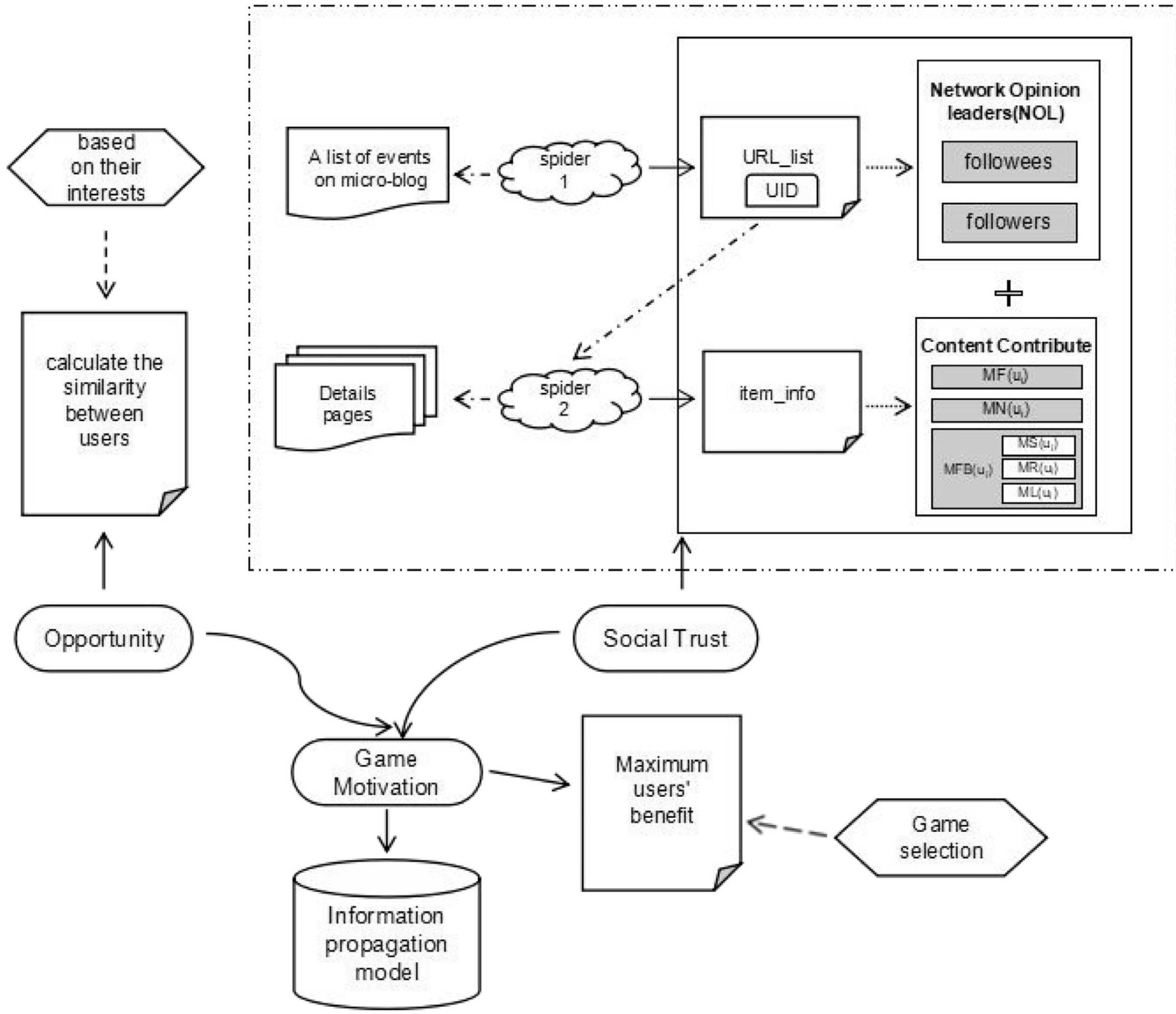


Fig. 7. Framework of the construction of information propagation model based on opportunity, trust and motivation.

Table 3
Content contribution interactive information judgment matrix.

Content contribute	$MF(u_i)$	$MN(u_i)$	$MP(u_i)$	$MC(u_i)$	$MS(u_i)$
$MF(u_i)$	1	1/3	1/5	1/7	1/9
$MN(u_i)$	3	1	1/3	1/5	1/7
$MP(u_i)$	5	3	1	1/3	1/5
$MC(u_i)$	7	5	3	1	1/3
$MS(u_i)$	9	7	5	3	1

Table 4
Social trust judgment matrix.

Social trust	$NOL(u_i)$	$CC'(u_i)$
$NOL(u_i)$	1	1/5
$CC'(u_i)$	5	1

1970s [41]. It is a qualitative and quantitative decision-making method. Because the AHP is more suitable for the target system with stratified and interlaced evaluation index, it can be used to solve the decision problem with multiple factors which is difficult to describe in the complex decision making problem. Therefore, we apply it to determine the weight value of each factor of the user's social trust. It can be divided into two steps: constructing judgement matrix and consistency checking.

1. Construction of judgment matrix

In the process of constructing the judgment matrix, the index level is determined by the quintile scale. According to the properties of the positive and negative matrices, the content contribution and social trust judgment matrix of microblog users in the information propagation process are constructed as shown in Tables 3 and 4, respectively.

The frequency of microblog posts $MF(u_i)$ is not as good as possible. On the contrary, it is better to publish high-quality posts in the peak period if the user has not many followers. Otherwise, it will easily result in relapsing, which affects the microblog experience of followers and causes followers to cancel the attention. Considering that microblog users may publish some posts (life diaries) that are not related to information propagation, these posts only show their value in important moments and important events. Therefore, the amount of microblog posts $MN(u_i)$ has a relatively small contribution to the propagation of certain information. And the contribution and influence of the information are relatively large in the process of information propagation. Among them, the number of shares of a certain microblog information is actually the number of forwards. It embodies the path and direction of in-

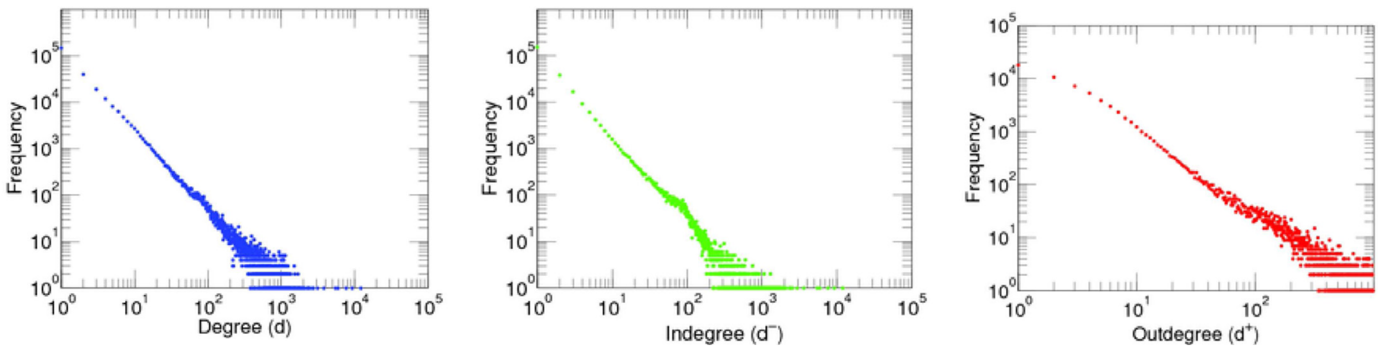


Fig. 8. Degree distributions of the Sina microblog.

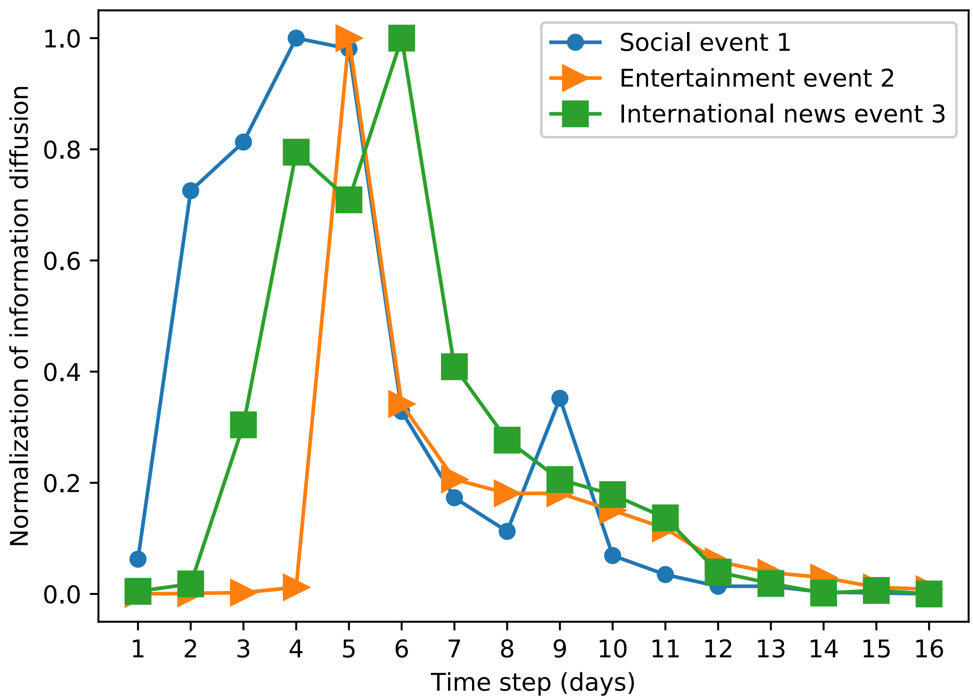


Fig. 9. Trends for the development of microblog hot events.

Table 5
The average random consistency index *RI* standard value (standard is different, *RI* value will have a slight difference).

Order of a matrix	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Table 6
User content contribution interactive information parameter setting.

Content contribution	Weight assignation				
	$MS(u_i)$	$MC(u_i)$	$MP(u_i)$	$MF(u_i)$	$MN(u_i)$
$CC'(u_i)$	$\alpha_1 = 0.568$	$\alpha_2 = 0.236$	$\alpha_3 = 0.096$	$\alpha_4 = 0.046$	$\alpha_5 = 0.054$

formation propagation in the network topology, and contributes greatly in the process of information propagation.

2. Consistency checking

In order to ensure the reasonableness of the judgment matrix, the consistency checking is needed. The consistency checking mainly includes two indexes: the consistency index (CI) and the consistency ratio (CR). The calculations are shown as follows:

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{16}$$

$$CR = \frac{CI}{RI} \tag{17}$$

In Eq. (16), $\lambda_{\max}$ is the maximum eigenvalue of the judgment matrix; n is the largest integer smaller than $\lambda_{\max}$; the matrix is consistent when $n = \lambda_{\max}$ and the smaller the *CI* the greater the consistency. Considering that the randomness leads to the deviation of the matrix consistency when checking the judgment matrix, the checking factor *CR* is introduced. The calculation is shown in Eq. (17), which the *CI* and the average random consistency *RI* are

Fig. 10. Propagation path of THAAD incident on microblog social network.

compared. Among them, the *RI* is related to the order of the judgment matrix. In general, the greater the order of the matrix, the greater the possibility of the consistent random deviation. The corresponding relationship is shown in Table 5. Hence, the microblog users' content contribution and social trust parameters calculated separately are shown in Tables 6 and 7.

Table 7
Social trust parameter setting.

Social trust	Weight assignation	
	$NOL(u_i)$	$CC'(u_i)$
$ST(u_i)$	$\eta_1 = 0.167$	$\eta_2 = 0.833$

5.3. Experimental results and analysis

We evaluate the information propagation process over Sina microblog social network. The Sina microblog dataset contains 293,437 users and 1,629,665 edges, where the edges means the connection between two users. We plot the degree distributions of the Sina microblog in Fig. 8, including the degree, indegree and outdegree of the network users. From the degree distribution of the Sina microblog, we can see that the users degrees vary from 1 to 10000 and 90 percents of the users are with less than 1000 degrees. And we can see that the Sina microblog network embodies the clustering phenomenon of users from the distributions of indegree and outdegree. It reflects the explosive propagation characteristics of information in microblog social network.

Three microblog hot events between 2015 and 2017 are used as datasets for the propagation of certain information on microblog social network, which include Dataset 2: Chengdu female driver was beaten, Dataset 3: Baihe Bai derailed and Dataset 4: THAAD incident. Trends for the development of microblog hot events as shown in Fig. 9.

Specially, we take the Dataset 4 (THAAD incident) as a key research object. And its propagation path is shown in Fig. 10 on microblog social network.

The circle and the size of the circle represent different microblog users and the social trust of microblog user in the Fig. 10, respectively. It is known by Definition 5 that the social trust of the microblog user is determined by the personal influence of the microblog user and its contribution to the content of the information. Different colored circles mean that they belong to different communities, which the community is composed of people with similar interests. The connections between the circles indicate the information propagation relationship between microblog users. And

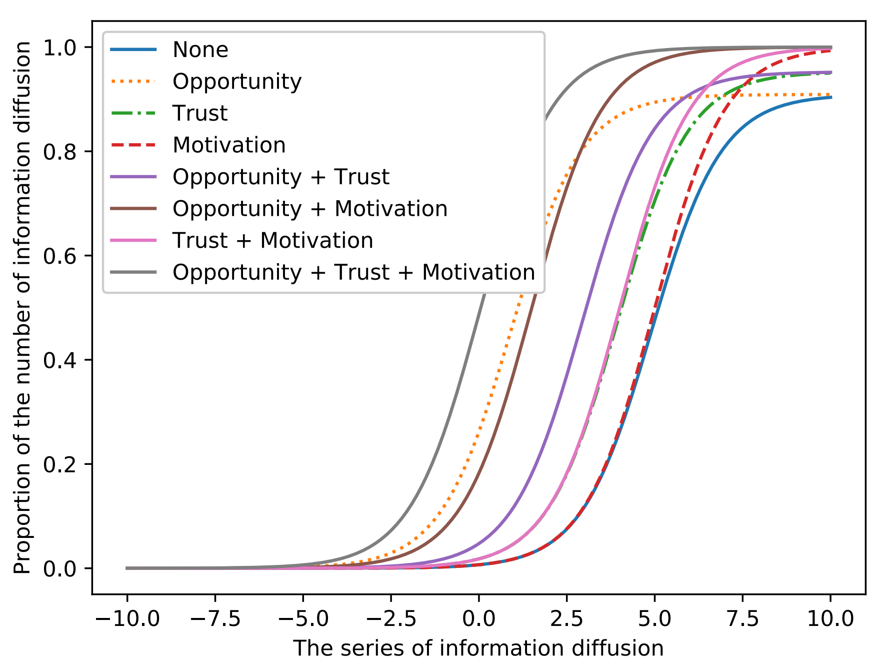
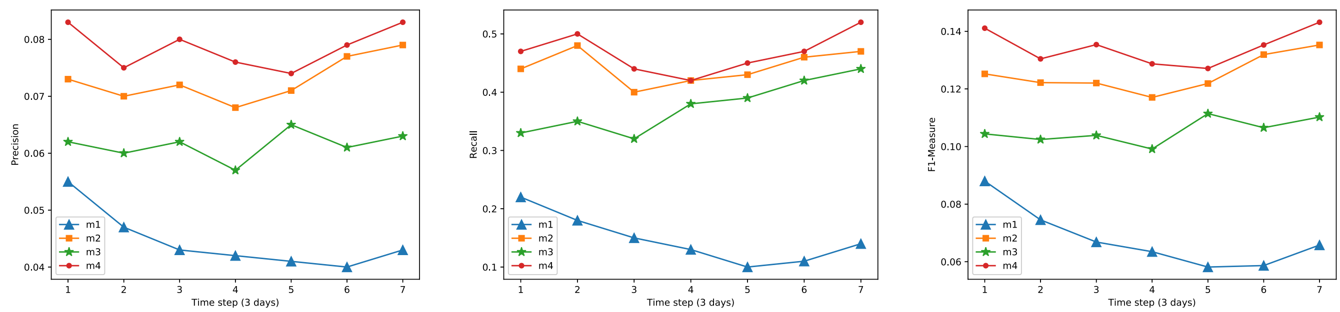
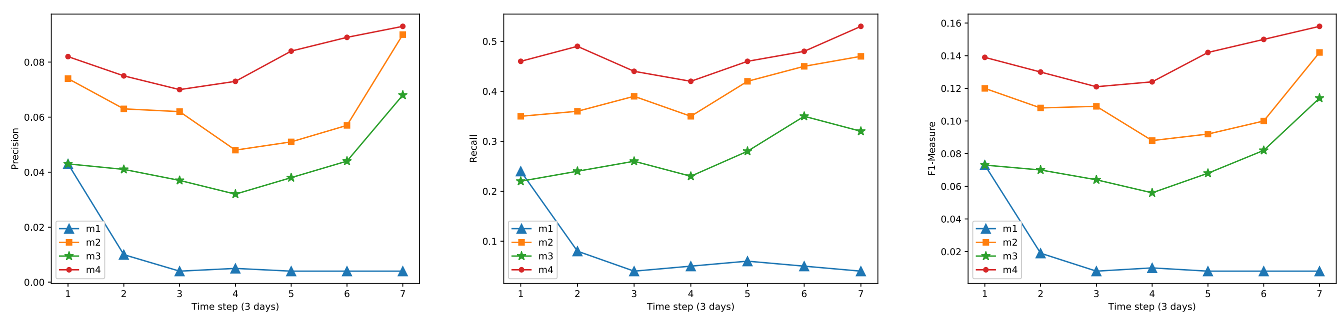


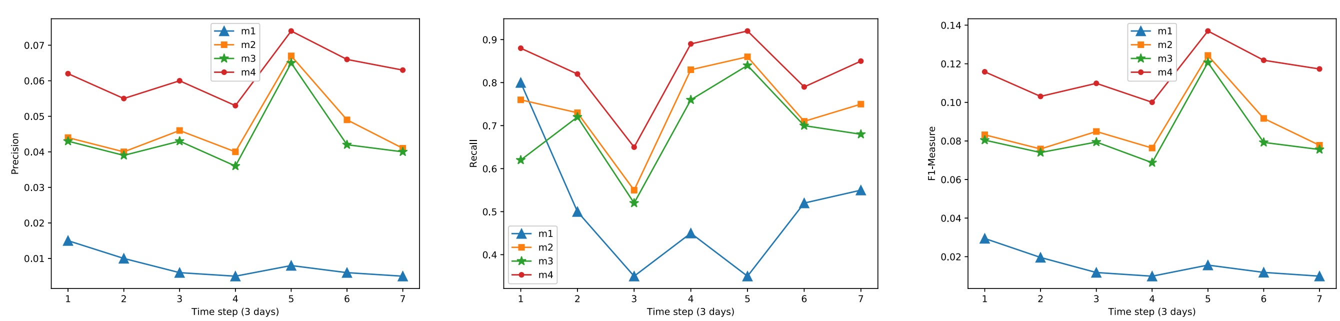
Fig. 11. Comparison of information propagation ratio based on hybrid social factors.



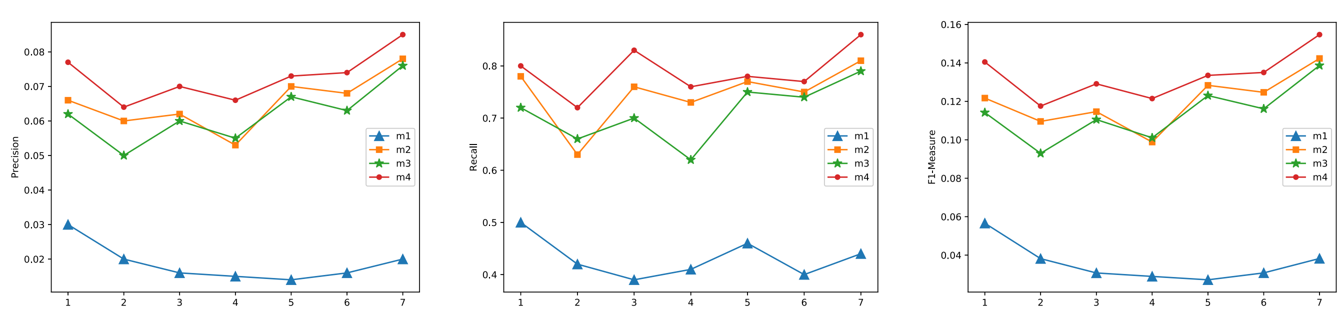
(a) Dataset 1



(b) Dataset 2



(c) Dataset 3



(d) Dataset 4

Fig. 12. Evaluative results of different datasets based on Precision, Recall and F1-measure in social network.

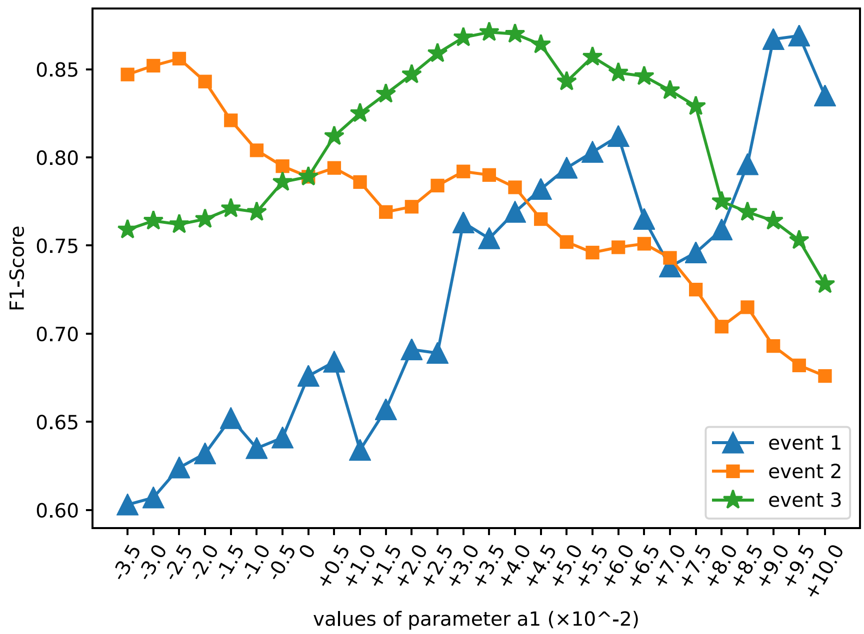


Fig. 13. Parameter α_1 affect the performance of the information propagation in F1-score.

the thickness of the connection between circles indicates the similarity of interest between microblog users. The greater the interest similarity, the thicker the connection is.

Next, the impact of social factors (opportunity, trust, and motivation) on the spread of information will be analyzed. Fig. 11 is a graph showing the percentage of event propagation in microblog social network with different social factors.

According to Fig. 11, it is found that the number of information propagation in microblog social network is similar to the S-curve. At the same time, the event propagation has experienced three stages: propagation contact (Series 0–1), information trust (Series 1–3) and information propagation (Series 3–6). In addition, we can see that different social factors have different perspectives and degrees of influence on information propagation. Further, we can observe three obvious phenomena: opportunity can promote communication behavior and shorten the time for propagation contact, social trust accelerates the propagation of specified information and shortens the life cycle of information propagation, and motivation influences the depth and breadth of propagation and makes important information as widely as possible for every individual in the microblog social network.

Goyal et al. [42] put forward the Continuous Time (CT) model and the Discrete Time (DT) model, and used these two models to combine the general threshold model to predict the information propagation process. Since the CT model has achieved better results than the DT model in the prediction results, we use the CT model in combination with the general threshold model as the comparison method. However, CT only calculates the probability of information propagation (social influence) and it does not consider the structural influence factors from the perspective of network topology. Therefore, Li et al. [43] combined two kinds of global influences: influence of network structure and influence of information propagation, and proposed a Game Theory (GT) information propagation model. Consequently, the methods evaluated and compared in our experiments are as follows:

- M(method)1: CT + General threshold model, baseline method combining the CT model [42] with the general threshold model.
- M(method)2: GT (Page rank), GT model [43] use Page rank to calculate user's influence and use the social payoff to diffuse.
- M(method)3: GT (Propagation cascades), GT model use propagation cascades to calculate user's influence and use the social payoff to diffuse.
- M(method)4: GCIP-Page Rank, the game choice information propagation based on Page Rank.

In Fig. 12, we compare the information propagation model based on opportunity, trust and motivation proposed in this paper with the general threshold model and the information propagation model based on the network game theory.

The experimental results show the effectiveness and superiority of the proposed method from three aspects of Precision, Recall and F1-Measure. By comparing experiments with these three methods, we can demonstrate the effect of the accuracy of parameters in the information propagation model based on opportunity, trust, and motivation on the propagation model. Because of many parameters involved in the model, we first evaluate the control parameters involved in the two algorithms in Section 3. By adjusting the parameters continuously in the experiment, we get the optimal value of these parameters finally (see Table 8).

We analyze the role of parameters in information diffusion from the whole. Here, we use the first parameter α_1 as an example to explain how this work optimizes the concerned parameters. Parameter α_1 affect the performance of the information propagation in F1-score as shown in Fig. 13. In addition, other parameters also use the same way to obtain.

Firstly, the estimated value of each parameter is determined by the analytic hierarchy process (AHP), and then the optimal parameter is obtained by adjusting the parameters continuously, which is measured by the F1-score. Furthermore, because the user's game revenue is the objective function, we use the selected sample values to conduct unsupervised learning, which are constrained by the extremum of the objective function. Finally, we calculate the optimal values of the parameters.

For α_1 in event 1, its value is continuously tuned from -0.040 to $+0.100$ (i.e. 0.528 to 0.668) on the basis of 0.568 (as shown in Table 6 in the manuscript paper). Finally, the parameter value that maximizes the F1-score (that is, the parameter value 0.658 corresponding to the F1-score of 0.869) is selected. Similarly, a1 in event2 and event3 is also obtained in the same way.

Table 8
The parameter values for microblog hot events 1–3.

Parameter	Description	Event 1	Event 2	Event3
α_1	Sharing rate of users	0.658	0.536	0.598
α_2	Review rate of users	0.297	0.277	0.214
α_3	Praising rate of users	0.015	0.023	0.037
α_4	Users release posts' frequency within selected time	0.007	0.073	0.069
α_5	The weight of users release microblog posts	0.023	0.091	0.082
η_1	The weight of personal influence in social trust	0.245	0.144	0.236
η_2	The weight of content contribution in social trust	0.755	0.856	0.764

From the Table 8, we can see that for a message, the dissemination of information is affected by two aspects: the person who disseminates information and the content itself. Among them, in a given certain field, the value, sometimes also called the significance of information (that is, the content contribution of information in the process of diffusion) plays a very important role in communication. This is reflected in the behaviors (such as sharing, review and praising) of users of the message in the social network. Furthermore, users sharing play an important role in social networks (corresponding to word-of-mouth transmission in the real world).

6. Conclusion and discussion

First of all, we take the similarity of interest between two adjacent users as an opportunity measure for users to receive certain information. Secondly, taking the characteristics of nodes in the process of information propagation that have certain social influence and discourse power in the professional field as the threshold criterion of the trust mechanism. Finally, considering the game characteristics that each user in the network always chooses the best interest information to spread and propagation, we use the game theory as the motive of the user to choose the propagation of information.

In view of the differences between Chinese culture and Western culture, there are many differences between Sina microblog and Twitter. However, the information propagation model proposed by us also applies to the Twitter platform because both Sina microblog and Twitter are broadcast social media platforms that share short and real-time information through a mechanism of concern and have the same user behavior. Therefore, both the calculation method of the user's personal influence and the user's contribution to the information propagation process are the same.Consequently, our GCIP-Page Rank algorithm is also applicable to Twitter users in the information propagation process to maximize the choice of benefits.

Acknowledgement

Fruitful discussion with Yue Wu, Xianyong Li and Yongquan Fan is gratefully acknowledged. The authors thank anonymous reviewers for many useful discussions and insightful suggestions. This work is supported by the National Natural Science Foundation (Grant Nos. 61872298, 61532009, 61602389, 61802316, and 61472329), the Innovation Fund of Postgraduate, Xihua University (No. ycjj2017176), the Students' Platform for Innovation and Entrepreneurship Training Program (No. 2018069), the Chunhui Plan Cooperation and Research Project, Ministry of Education of China (Nos. Z2015100 and Z2015109), Science and Technology Department of Sichuan Province (Nos. 2018026, 2017HH0083, 2016JY0244, and 2017RZ0009), Scientific Research Fund of Sichuan Provincial Education Department (Nos. 15ZA0130 and 15ZB0134), the civil aviation administration of China (No. PSDSA201802), Chengdu science and technology bureau (No. 2016-XT00-00015-GX), the Key Scientific Research Fund of Xihua University (Nos. z1412616 and z1422615), the Open Research Subject of Key Laboratory of Security Insurance of Cyberspace, Sichuan Province (No. szjj2015-057) and the University-sponsored Overseas Eduction Project of Xihua University.

References

[1] Z. Zhang, Z. Wang, The data-driven null models for information dissemination tree in social networks, Physica A: Stat. Mech. Appl. 484 (2017) 394–411, doi:10.1016/j.physa.2017.05.008.
[2] X.X. Zhan, C. Liu, G. Zhou, Z.K. Zhang, G.Q. Sun, J.J.H. Zhu, Z. Jin, T. Simos, Coupling dynamics of epidemic spreading and information diffusion on complex networks, Appl. Math. Comput. 332 (2018) 438–448, doi:10.1016/j.amc.2018.03.050.
[3] Z. Qin, F. Khawar, T. Wan, Collective game behavior learning with probabilistic graphical models, Neurocomputing 194 (2016) 74–86, doi:10.1016/j.neucom.2016.01.075.
[4] C. Jiang, Y. Chen, K.J.R. Liu, Graphical evolutionary game for information diffusion over social networks, IEEE J. Sel. Topics Signal Process. 8 (4) (2017) 524–536, doi:10.1109/jstsp.2014.2313024.
[5] D. Zinoviev, V. Duong, A game theoretical approach to broadcast information diffusion in social networks, in: Proceedings of the 44th Annual Simulation Symposium, ANSS'11, 2011, pp. 47–52.
[6] W. Qiu, Y. Wang, J. Yu, A game theoretical model of information dissemination in social network, in: Proceedings of the International Conference on Complex Systems, 2013, pp. 1–6, doi:10.1109/icocs.2012.6458551.
[7] H. Kwak, C. Lee, H. Park, S.B. Moon, What is twitter, a social network or news media? in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, 2010, pp. 591–600, doi:10.1145/1772690.1772751.
[8] S.A. Myers, J. Leskovec, The bursty dynamics of the twitter information network, in: Proceedings of the 23rd International Conference on World Wide Web, WWW '14, 2014, pp. 913–924, doi:10.1145/2566486.2568043.
[9] J. Li, N. Wu, Z. Feng, Model-based non-gaussian interest topic distribution for user retweeting in social networks, Neurocomputing 278 (2018) 87–98, doi:10.1016/j.neucom.2017.04.078.
[10] C.J. Hutto, S. Yardi, E. Gilbert, A longitudinal study of follow predictors on twitter, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, 2013, pp. 821–830, doi:10.1145/2470654.2470771.
[11] D. Antoniades, C. Dovrolis, Co-evolutionary dynamics in social networks: a case study of twitter, Comput. Soc. Netw. 2 (1) (2015) 14, doi:10.1186/s40649-015-0023-6.
[12] S. Wu, J.M. Hofman, W.A. Mason, D.J. Watts, Who says what to whom on twitter, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, 2011, pp. 705–714, doi:10.1145/1963405.1963504.
[13] Z. Tian, Q. Zhang, Empirical analysis of microblog information flow features based on complex network theory, Adv. Inf. Sci. Serv. Sci. 4 (7) (2012) 163–171, doi:10.4156/aiss.vol4.issue7.19.
[14] Y. Zhou, B. Zhang, X. Sun, Q. Zheng, T. Liu, Analyzing and modeling dynamics of information diffusion in microblogging social network, J. Netw. Comput. Appl. 86 (1) (2016) 92–102, doi:10.1016/j.jnca.2016.09.011.
[15] W. Han, X. Zhu, Z. Zhu, W. Chen, W. Zheng, J. Lu, A comparative analysis on weibo and twitter, Tsinghua Sci. Technol. 21 (1) (2016) 1–16, doi:10.1109/tst.2016.7399279.
[16] W. Guan, H. Gao, M. Yang, Y. Li, H. Ma, W. Qian, Z. Cao, X. Yang, Analyzing user behavior of the micro-blogging website sina weibo during hot social events, Physica A: Stat. Mech. Appl. 395 (4) (2014) 340–351, doi:10.1016/j.physa.2013.09.059.
[17] Z. Zhang, H. Wang, C. Wang, H. Fang, Modeling epidemics spreading on social contact networks, IEEE Trans. Emerg. Topics Comput. 3 (3) (2015) 410–419, doi:10.1109/tetc.2015.2398353.
[18] C. Li, G. Jiang, Y. Song, L. Xia, Y. Li, B. Song, Modeling and analysis of epidemic spreading on community networks with heterogeneity, J. Parallel Distrib. Comput. 119 (2018) 136–145, doi:10.1016/j.jpdc.2018.04.009.
[19] X. Zhang, S. Chang, Q. Shi, H. Huo, Qualitative study of a stochastic sis epidemic model with vertical transmission, Physica A: Stat. Mech. Appl. 505 (2018) 805–817, doi:10.1016/j.physa.2018.04.022.
[20] T. Kuniya, J. Wang, Global dynamics of an sir epidemic model with nonlocal diffusion, Nonlinear Anal. Real World Appl. 43 (2018) 262–282, doi:10.1016/j.nonrwa.2018.03.001.

[21] M.E. Newman, Spread of epidemic disease on networks, Phys. Rev. E: Stat. Nonlinear Soft Matter Phys. 66 (1 Pt 2) (2002) 016128, doi:10.1103/physreve.66.016128.
[22] J. Leskovec, M. Mcglohon, C. Faloutsos, N. Glance, M. Hurst, Patterns of cascading behavior in large blog graphs, in: Proceedings of the 2007 International Conference on Data Mining, 2007, pp. 551–556, doi:10.1137/1.9781611972771.60.
[23] Y. Xiao, C. Song, Y. Liu, Social hotspot propagation dynamics model based on multidimensional attributes and evolutionary games, Commun. Nonlinear Sci. Numer. Simul. 67 (2019) 13–25, doi:10.1016/j.cnsns.2018.06.017.
[24] Y. Zhao, G. Kou, Y. Peng, Y. Chen, Understanding influence power of opinion leaders in e-commerce networks: an opinion dynamics theory perspective, Inf. Sci. 426 (2018) 131–147, doi:10.1016/j.ins.2017.10.031.
[25] E. Bakshy, J.M. Hofman, W.A. Mason, D.J. Watts, Everyone's an influencer: quantifying influence on twitter, in: Proceedings of the 4th ACM international conference on Web search and data mining, WSDM '11, 2011, pp. 65–74, doi:10.1145/1935826.1935845.
[26] M. Cha, H. Haddadi, F. Benevenuto, P.K. Gummadi, Measuring user influence in twitter: The million follower fallacy, in: Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, ICWSM' 10, 2010.
[27] D.M. Romero, W. Galuba, S. Asur, B.A. Huberman, Influence and passivity in social media, in: Proceedings of the 20th International Conference Companion on World Wide Web, WWW '11, 2011, pp. 113–114, doi:10.1145/1963192.1963250.
[28] S.A. Ros, F. Aguilera, J.D. Nuez-Gonzalez, M. Graa, Semantically enhanced network analysis for influencer identification in online social networks, Neurocomputing (2017), doi:10.1016/j.neucom.2017.01.123.
[29] L. Yang, Y. Qiao, Z. Liu, J. Ma, X. Li, Identifying opinion leader nodes in online social networks with a new closeness evaluation algorithm, Soft Comput. 22 (2) (2018) 453–464, doi:10.1007/s00500-016-2335-3.
[30] Y. Zhao, S. Li, F. Jin, Identification of influential nodes in social networks with community structure based on label propagation, Neurocomputing 210 (2016) 34–44, doi:10.1016/j.neucom.2015.11.125.
[31] O. Biran, S. Rosenthal, J. Andreas, kathleen McKeown, O. Ranbow, Detecting influencers in written online conversations, in: Proceedings of the 20th International Conference Companion on World Wide Web, WWW '11, 2012, pp. 37–45.
[32] J. Goldenberg, S. Han, D.R. Lehmann, J.W. Hong, The role of hubs in the adoption process, J. Mark. 73 (2) (2009) 1–13, doi:10.1509/jmkg.73.2.1.
[33] E. Kelly, B. Davis, J. Nelson, J. Mendoza, Leader emergence in an internet environment, Comput. Hum.Behav. 24 (5) (2008) 2372–2383, doi:10.1016/j.chb.2008.02.013.
[34] F. Bodendorf, C. Kaiser, Detecting opinion leaders and trends in online social networks, in: Proceedings of the 2nd ACM Workshop on Social Web Search and Mining, SWSM '09, 2009, pp. 65–68, doi:10.1145/1651437.1651448.
[35] X. Liu, C. Liu, Information diffusion and opinion leader mathematical modeling based on microblog, IEEE Access 6 (2018) 34736–34745, doi:10.1109/ACCESS.2018.2849722.
[36] S.K. Yong, V.L. Tran, Assessing the ripple effects of online opinion leaders with trust and distrust metrics, Expert Syst. Appl. 40 (9) (2013) 3500–3511, doi:10.1016/j.eswa.2012.12.058.
[37] F.J. Ortega, C.G. Vallejo, Propagation of trust and distrust for the detection of trolls in a social network, Comput. Netw. 56 (12) (2012) 2884–2895, doi:10.1016/j.comnet.2012.05.002.
[38] J. Wan, X. Chen, Y. Du, M. Jia, Information diffusion model based on opportunity, trust and motivation, in: Proceedings of the China Conference on Information Retrieval, CCIR '18, 2018, pp. 185–197, doi:10.1007/978-3-030-01012-6_15.
[39] S. Jafari, H. Navidi, A game-theoretic approach for modeling competitive diffusion over social networks, Games 9 (1) (2018) 8–20, doi:10.3390/g9010008.
[40] G. Liu, F. Zhu, K. Zheng, A. Liu, Z. Li, L. Zhao, X. Zhou, Tosi: a trust-oriented social influence evaluation method in contextual social networks, Neurocomputing 210 (2016) 130–140, doi:10.1016/j.neucom.2015.11.129.
[41] B. Chandran, B. Golden, E. Wasil, Linear programming models for estimating weights in the analytic hierarchy process, Comput. Oper. Res. 32 (9) (2005) 2235–2254, doi:10.1016/j.cor.2004.02.010.
[42] A. Goyal, F. Bonchi, L.V.S. Lakshmanan, Learning influence probabilities in social networks, in: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, WSDM '10, 2010, pp. 241–250, doi:10.1145/1718487.1718518.
[43] D. Li, Z. Xu, Y. Luo, S. Li, A. Gupta, K. Sycara, S. Luo, L. Hu, H. Chen, Modeling information diffusion over social networks for temporal dynamic prediction, IEEE Trans. Knowl. Data Eng. 29 (9) (2017) 1985–1997, doi:10.1109/TKDE.2017.2702162.

Jihong Wan received her B.Sc. degree in Science and Technology of Computer from the Zhengzhou Normal University, Henan, China, in 2016. She is currently pursing the Masters degree in School of Computer and Software Engineering at Xihua University, Chengdu, China. She has published several papers in the fields of intelligence system and information retrieval. Her research interests include social networks and machine learning.

Xiaoliang Chen is an associate professor with the School of Computer and Software Engineering of Xihua University. He received the Ph.D. degree in school of Electron-Mechanical Engineering at Xidian University, Xian, China, in 2014. He is now a visiting professor of University of Montreal. He is a member of Network Information Service Committee in China. His research includes social networks and information retrieval.

Yajun Du received the Ph.D. degree in traffic information engineering and control from the School of Computer and Communicate, Southwest Jiao Tong University in 2005. He is presently a professor of computer science at XHU (Xihua University). He has published several papers in the fields of information retrieval, search engines, focused crawlers, and knowledge. He serves on the committees of Chinese information and is a PC member for some leading international conferences, WISE, ICIC, CCIR, and WMSE. His experience and research work focus on information retrieval, software engineering, search engines, Web mining, and computer networks.

Mengmeng Jia received his B.Sc. degree in Software Engineering from the Nanyang Normal University, Henan, China, in 2016. He is currently pursing the Masters degree in School of Computer and Software Engineering at Xihua University, Chengdu, China. He has published several papers in the fields of computer science and information engineering. His research interests include data mining and information retrieval.