

## A Method to Calculate Influence Values of Information Dissemination in SNS Groups Based on Second-Level Contacts

Du Min

*School of Electronics and Information Engineering, Hunan University of Science and Engineering, 425199, People's Republic of China  
duminemail@126.com*

### **Abstract**

*It presents an analysis method based on second-level contacts in SNS groups to quantize influence values for information dissemination in this article. The method introduces “long tail effect” of weak relation on second-level contacts and entropy in the calculation process, quantizes influence values on inside or outside some node's group, and calculates action and reaction between adjacent groups. Simulation results show this method has more integrity and objectivity to evaluating values in messaging, prevents the exponential growth of strong views effectively, reduces formation and propagation effects of biased viewpoints. At the same time, it verifies the influence value between adjacent groups is the most critical factor for information dissemination, which is consistent with objective facts.*

**Keywords:** *social network services, long tail effect, entropy, information dissemination*

### **1. Introduction**

Social network services(SNS) [1] are important channels for people to keep in contact nowadays, such as Facebook, YouTube, Twitter, Google+ and et al, in which have millions of users. Users can add into different groups based on their personal interests, and can accept information from multiple newsgroups through social network services. There is strong relation or weak relation between users in information dissemination models [2]. Strong relation means users have personal strong homogeneity and closed relationship, and sustains strong emotional factors in one SNS group, just as user A connects user C directly. Weak relation implies stronger heterogeneity including extensive exchanges and not-closed interpersonal relationship, just as user A connects user B directly, and user B connects user C directly, and the relationship between user A and user C products by user B. However, weak relation between users can bring long tail effect and get huge economic benefits, which enhances the influence of information dissemination.

At the same time, it is very important to the spread of same information between different groups [3]. There are significantly different because of transmission capacity of users according to the general principles of sociology and hierarchical clustering in social systems [4-8].For example, when user A is a leader node in group G1, and meanwhile it is a general node in group G2, the influence ability of spreading messages maybe reduce through user A to communicate from group G1 to group G2. However, the influence ability of spreading messages maybe enhance through user A to communicate from group G2 to group G1.If user A is a general node in group G1 and group G2, the influence ability of spreading messages will be obvious.

## 2. Related Research

There are many models to reveal information dissemination such as linear threshold model (LT model) [24], independent cascade model (IC model) [11], infectious disease dynamics model [25-27], transmission dynamics model [28-30] and so on [9-13].

David and his team proposed linear threshold model in information dissemination mechanism [24], which was a way to portray the cumulative impact of characteristics. In the model, each node exists with active or inactive status, and only in one of the two. The activated weight  $b_{w,u}$  is set in directed edges, and  $(w,u) \in E$ .  $b_{w,u}$  represents the influence of node  $w$  to node  $u$ .  $\theta_u$  is the threshold and its value reveals the activated possibility of node  $u$  by his neighbors. When the sum of the weights from activated neighbors exceeds the threshold value,  $(\sum_{w \in N(u)} b_{w,u} \geq \theta_u)$ , that means node  $u$  is activated, otherwise activation failed. The smaller the threshold is, the more likely the node is activated.

Each node must be on only two states in independent cascade model [11], including active state or inactive state. At the same time, one node can only change from active state to inactive state, there is no other possible way to change.  $p_{v,w}$  Indicates the value of successful probability if node  $w$  is active and his neighbor node  $v$  is inactive, and node  $w$  transforms node  $v$  to be active in cumulative time  $t$ . When node  $v$  is surrounded by numerous activated neighbor nodes, the effects by neighbor nodes are in any order, but the impact probability affects by the value of  $p_{v,w}$ . The greater the probable value is, the huger the impact is, which means the value of  $p_{v,w}$  is larger, node  $v$  is activated more easily. And if node  $v$  becomes active, it will have actions to its neighbor nodes. Each node has same mechanism for communication until all of them are activated. Any node accepts some new diffusion affect independently in IC model, and has no relationship on diffusion effects of other nodes' history record.

There are SIS model and SIR model for classic infectious disease dynamics models [25-27]. In SIS model, individual is only in the two typical state, including susceptible status and infected status, abbreviated as status S and status I. A healthy individual may be converted to an infected individual on a certain probability. In contrast, an infected one can be recovered into a healthy state on a certain probability. However, each individual will be allowed only in one of three states in SIR model, consisted of susceptible status (state S), infected status (state I) and removed status (state R). Healthy individuals contacted with infected individuals can be infected, and infected individuals can be removed with a certain probability.

Many research works in transmission dynamics models [28-30] reveal that the process of information dissemination has clear directional characteristics in a small areas and reticular divergent characteristics in a wide range.

D.J. Watts and S.H. Strogate [29] simulated the simple spread of diseases in the small-world network. They found it was faster and easier in the small-world network than a rule network. R. Pastor-Satorras and A. Vespignani [30] studied the spread of diseases in scale-free networks of SIS model, and concluded that it does not need to set any threshold, which means that even the spread of diseases has a very small strength, but also can make disease spread quickly in the network. This conclusion has changed many conclusions in some traditional theories for spread of diseases fundamentally, which is, only when the propagation rate exceeds a threshold, large-scale propagation is possible.

But D. J. Watts and his team [28] put forward different views, and they believed a large-scale propagation is driven by the large number of influential vulnerable groups instead of influential individuals.

So we not only must consider independence activation behaviors of neighbor nodes for information dissemination in network, but also analyze a marked impact by neighbor nodes' groups or neighbor's neighbor nodes' groups together, which is called "the group effect" [14]. The effect of groups is a huge impact on node behavior and formation of

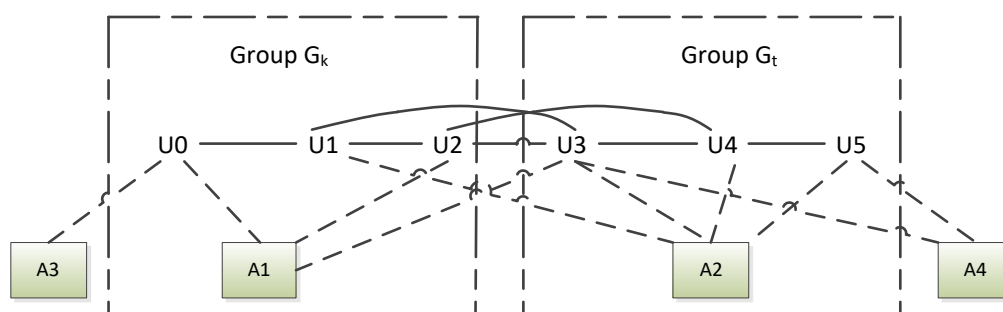
ideas. Yet many models cannot measure influence values of nodes and groups accurately.

It proposes an method to calculate the influence capability of nodes for information dissemination in SNS groups based on second-level contacts effectively, that uses “long tail effect” of weak relation in groups to quantify overall impact of groups and can avoid formation of biased viewpoints.

### 3. The Model of Groups on Second-Level Contacts in SNS

The model describes information of nodes, attributes, social connection, correspondence between attributes and correspondence between groups, shown in Figure 1. It is described as follows.

$$NG = \{U, A, EU, EA, G\}$$



**Figure 1. Description of Groups on Second-Level Contacts in SNS**

U is a set of nodes that means each node corresponds to a user in real social networks. A implies a set of attributes, which includes all values of properties in a real user, such as interest, active, time and events. EU shows a set of social connection that clears strong relation between two nodes through solid lines. EA indicates a set of correspondence between attributes, that draws weak relation between a user and its attributes through dotted lines. G is a set of correspondence between groups that shows many users in one group assemble together because of same properties.

### 4. Modeling

The paper creates different samples based on properties to measure influence values of nodes in the past. For example, the values of movie and constellation are different properties for users’ hobby, which makes to form two different groups. At the same time, nodes are divided into active state and inactive state [15-16]. The active state of one node reveals his joins in the information dissemination process after it receives messages from his neighbor and is activate by his neighbor. The inactive state indicates all his neighbors is activate by the node after it has joined in the process and is unaffected by his neighbors.

### 5. Quantitative Analysis

#### 5.1. Quantization for Nodes’ Opinions in One Group Based On Entropy

Entropy is an important parameter to measure information in information theory [17-20].When the value of entropy is smaller, which indicates it provides a greater quantity of information, has a large role and a greater weight. On the contrary, larger entropy means a small weight. The value of entropy on node  $U_i$  in his group  $G_k$  is defined to calculate the influence ability of information dissemination, shown in equation 1.

$$e_i = \frac{-p(U_i) \cdot \log_2 p(U_i)}{-\sum_{j=1}^{j=m} p(U_j) \cdot \log_2 p(U_j)}, \quad U_i \in G_k \quad (1)$$

There have  $m$  nodes in group  $G_k$ , marked  $U_1, U_2, \dots, U_m$ .  $p(U_j)$  means the communication ability of node  $U_j$  in group  $G_k$ , which is connected with node  $U_i$  directly.  $p(U_j)$  represents the value of influence, in where node  $U_j$  activates his neighbor node  $U_i$  successfully,  $p(U_j) \in [0,1]$ .

### 5.2. Calculate the Value of Influence about Node $U_i$ in its Group

The value of influence about node  $U_i$  in its group is affected by subjective information impacts and its strong relation in the group, which is calculated as follows.

$$v_i = c_i \cdot v'_i + (1 - c_i) \cdot e_i \quad (2)$$

$v_i$  measures influence of individual views at the current time.  $v'_i$  represents influence at the previous time.  $c_i$  is a degree of its confidence, defines the degree about accepting others' views and characterizes the role of subjective factor.  $e_i$  is a value of entropy, which describes the impact of environment on node  $U_i$  and this is an objective factor,  $c_i, e_i \in [0,1]$ .

### 5.3. Calculate Influence Values of Information Dissemination on Node $U_i$ and Group $G_k$

There will produce different degrees of influence in information dissemination because of the role of nodes in their group. So there must take into account the views of individual node  $U_i$  at the current time, thus it calculates the value of overall effect on the spread of the message based on group  $G_k$ . It is given below, as in equation 3 and equation 4.

$$w_i = \frac{v_i}{\sum_{j=1}^{j=m} v_j} \quad (\text{equation 3})$$

$$v_{G_k} = \sum_{i=1}^m w_i \cdot v_i / m \quad (\text{equation 4})$$

$w_i$  is the weight of node  $U_i$  in group  $G_k$ .  $v_{G_k}$  defines the value of overall effect on the spread of the message in group  $G_k$ .

### 5.4. Analyze the Influence of Information Dissemination Value between Groups

It is defined the correlation coefficient between groups, marked  $\rho_{k,t}$ , to represent the influence value between groups. Because it considers the overall impact between groups is more objective than part ones and can represent spread attitude and ability of majority nodes in the group.  $\rho_{k,t}$  is defined as follows.

$$\rho_{k,t} = \frac{\sum_{i=1}^m (v_{G_k} - \overline{v_{G_k}}) \cdot (v_{G_t} - \overline{v_{G_t}})}{\sqrt{\sum_{i=1}^m (v_{G_k} - \overline{v_{G_k}})^2} \cdot \sqrt{\sum_{i=1}^n (v_{G_t} - \overline{v_{G_t}})^2}} \quad (5)$$

The value of  $\overline{v_{G_k}}$  and  $\overline{v_{G_t}}$  indicates the expected value for information dissemination in group  $G_k$  and  $G_t$ .

### 5.5 Calculate the Value of Distribution Influence on Node $U_i$ in His Group $G_k$ to Adjacent Group $G_t$

Distribution influence function on node  $U_i$  in group  $G_k$  to adjacent group  $G_t$  reflects its relationship to its adjacent groups, which can determine the value of information influence based on the second-level contacts from the node as the center. It is shown as equation 6.

$$v(i, G_t) = \frac{v_i}{\rho_{i,j} * v_{G_t}} \quad (6)$$

When the larger value on  $v(i, G_t)$  is, that indicates the greater impact on node  $U_i$  to his adjacent group  $G_t$ , the greater likelihood of information dissemination and can increase the likelihood on information control dissemination through node  $U_i$ , which can get better results.

### 5.6. Structuring Analysis Equation

We construct an objective analysis function, marked as  $L(i, G_{k,t})$ , to analyze historical information in target groups. They are shown as equation 7 to equation 9.

$$L(i, G_{k,t}) = \rho_{k,t} \cdot \frac{\sigma(t)}{\sigma(k)} \cdot (i - \bar{v}_{G_k}) + \bar{v}_{G_t} \quad (7)$$

$$\sigma(t) = \sqrt{\sum_{i=1}^n (v_{G_t} - \bar{v}_{G_t})^2} \quad (8)$$

$$\sigma(k) = \sqrt{\sum_{i=1}^m (v_{G_k} - \bar{v}_{G_k})^2} \quad (9)$$

$L(i, G_{k,t})$  is the target function, the value of  $\sigma(t)$  and  $\sigma(k)$  represents respectively the standard deviation for differential transmission standards in group t and group d. When the standard deviation is minimized and the value of  $L(i, G_{k,t})$  is an analysis approximation according to the theory of linear regression analysis, information dissemination behaviors are most similar to real networks in group t, and the minimum is  $\sigma^2(k) \cdot [1 - \rho_{k,t}^2]$ .

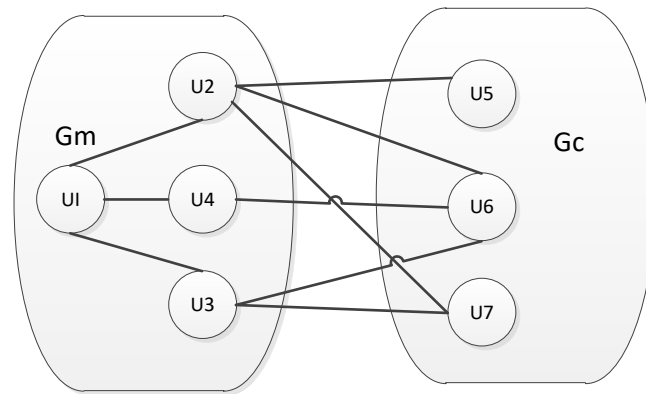
## 6. Simulation and Performance Analysis

### 6.1 Set Simulation Parameters and Relationship on Groups and Nodes

**Table 1. Simulation Parameters on Initial States of Nodes and Groups**

groups' name	nodes' name	nodes' out-degree	strength of node's view	nodes' attitude
$G_m$	U1	40	0.95	support
	U2	30	0.5	support
	U3	20	0.3	opposition
	U4	15	0.1	opposition
$G_c$	U5	35	0.95	opposition
	U6	25	0.5	support
	U7	15	0.95	opposition

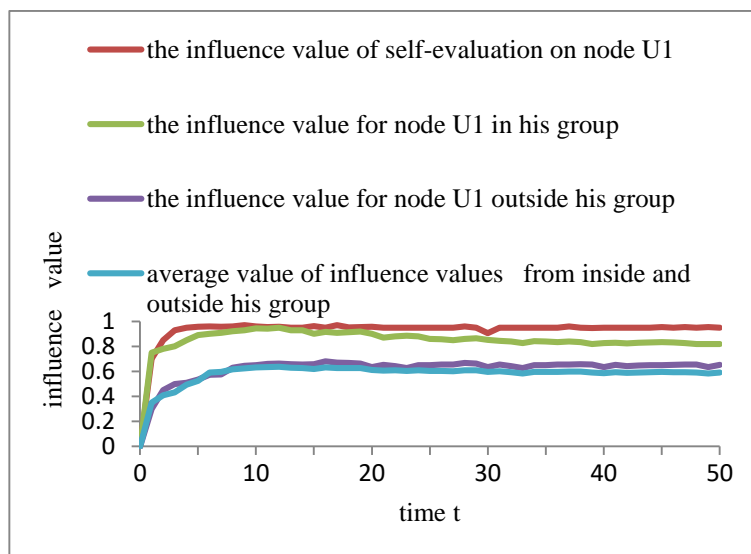
We set initial states of nodes and groups, and construct group  $G_m$  and group  $G_c$  according to two properties of movie and constellation. There are 500 nodes in every group, including leader nodes, general nodes and inactive nodes [21-23]. It covers 5 parameters in table 1. We set the specific relationship of nodes in group  $G_m$  and group  $G_c$ , shown in Figure 2.



**Figure 2. The Relationship Diagram of Nodes in their Groups**

### 6.2. Analysis of Simulation Results

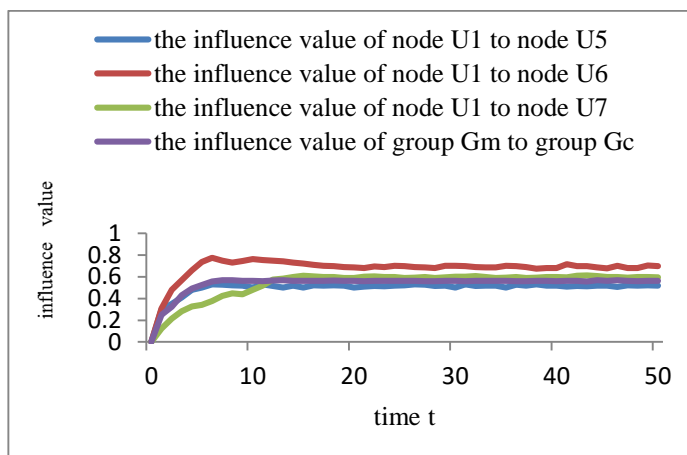
There are influence values from node U1 from inside or outside his group, shown in Figure 3. Node U1 is a leader node in group  $G_m$ , and its influence value of self-evaluation is 0.926, which is close to the strength of its view. The value of the strength of the view in group  $G_m$  is more than 0.846 because of the weak opposite view. But the value rapidly decreased to 0.617 when it is added more second-level nodes which are opposite to the view, that is close to the average value from group  $G_m$  and group  $G_c$ . This result shows that influence value for one node is more nearly closer to the overall evaluation in groups with increasing second-level nodes. Therefore, it concludes that the method has more integrity and objectivity of the evaluating values and avoids the exponential growth when numbers of strong views of information expand explosively.



**Figure 3. Comparison of Influence Values in Information Dissemination**

Node U1 has reaction influence to node U5, node U6 and node U7 after it is affected from group  $G_m$  and group  $G_c$ , shown in Figure 4. The reaction of node U1 to node U5 is stayed in a nice hobble because node U5 is a leader node in group  $G_c$  and has a strong attitude for messaging, so there is interacted between the two nodes and the value is 0.4966, which is closer to 0.5. The influence on node U1 to node U6 increases rapidly in a very short period because they have a same view to message, so the effect on node U1 to node U6 is far higher than other nodes from node U1 although node U6 is a general node. The reaction of node U1 to node U7 is a state of gradual penetration in a long period until

the result is close to the overall impression because node U7 has litter influence. The result shows that nodes with a strong view have a greater impact to second-level nodes with a same attitude. The rules of information dissemination obey by the overall factors unless the majority nodes in second-level groups have the same attitude to some nodes.

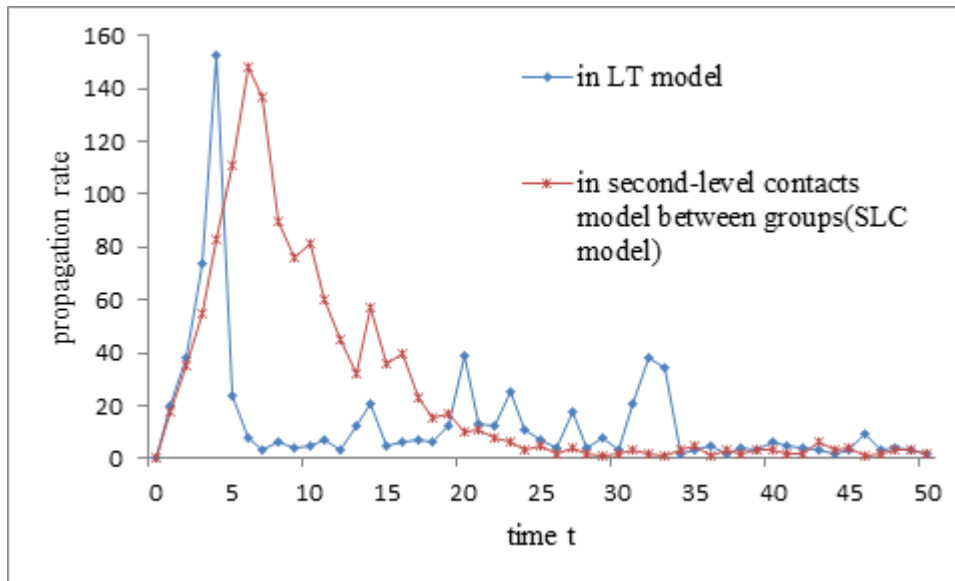


**Figure 4. Comparison of Reaction Influence from Node U1 and Interaction Force between Groups**

We set 160 nodes, observe information dissemination process before stabilizing during 50 time steps based on micro-blogging in the model.

We compare the relationship evolution on propagation rate for hot topics on time in figure 5, and give details of the hot topic in LT model and in second-level contacts model between groups(SLC model). We conclude there are several distinct phases of topic dissemination process, including a rapid growth phase, a decay phase and a decay phase until the minimum value gradually. The reason for this phenomenon is that the hot topic has intense attention by millions of people, is commented, forwarded, spread and erupted. But the spread of it is reduced with the weakening of people's interest until steady state. Therefore, the two models have similar velocity distribution characteristics from the speed of topics' distribution. They are subject to the law of the life cycle, have same experiences from birth, growth, aging to death. There is a clear peak, decays rapidly after it and follows a number of shocks.

At the same time, we observe there are different distributions. The topic reaches the peak at the 6th step in SLC model, decays rapidly and emerges a shock in the decay process. However, the growth and decay rate of the topic in LT model is significantly higher than in SLC model, and arise four shocks in middle and late stage. We consider there is more stable and holistic in SLC model than LT model, and avoids explosive expansion and dissemination of information on hot topics.



**Figure 5. Comparison of Relationship Evolution on Propagation Rate for Hot Topics and Time**

## 7. Conclusion

We propose a method to analyze the influence of information dissemination in SNS groups based on second-level contacts, which uses “long tail effect” of weak relation in groups and entropy effectively, leads weak relation of nodes into analysis of information dissemination, and quantifies information influence behavior. Simulation results show that the method has more integrity and objectivity of the evaluating values, and prevents the exponential growth effectively when numbers of strong views expand explosively. The interaction between groups is the most critical factor for messaging in information dissemination, which is consistent with objective factors and reduces the formation and propagation effects of biased viewpoints.

## Acknowledgment

This work was supported by key discipline for computer application and technology of Hunan University of Science and Engineering, supported by science and technology Planning project of Hunan Province (Grants NO. 2012FJ3051).

## References

- [1] D.M. Boyd and N.B. Ellison, “social network sites: definition, history, and scholarship”, *Journal of computer-mediated communication*, vol. 13, no. 1, (2007), pp. 210-230.
- [2] H.Q. Yu, W.J. Liu and Y.F. Qiu, “Second-level contacts recommendation of social network service based on subjects of users preference”, *Journal of computer applications*, vol. 32, no. 5, (2012), pp. 1366-1370.
- [3] M. Newman, “Modularity and community structure in network”, *The National Academy of Sciences*, vol. 103, no. 23, (2006), pp. 8857-8858.
- [4] Radicchi and C. Castellano, “Defining and identifying communities in Internet and other networks”, *Phys Rev E*, vol. 68, no. 2, (2003), pp. 26112-26119.
- [5] M.E.J. Newman, “Detecting community structure in network”, *Eur phys JB*, vol. 38, no. 2, (2004), pp. 321.
- [6] S. Boccaletti, V. Latora and Y. Moreno, “Complex networks: structure and dynamics”, *Phys Rep*, vol. 424, no.8, (2006), pp.175-308.
- [7] M.R. Sales-Pardo, A. Guimera and A. Moreira, “Extracting the hierarchical organization of complex systems”, *Proceeding National Academy Science USA*, vol. 104, no. 3, (2007), pp.15224-15229.
- [8] V.D. Blondel, J. Guillaume and R. Lambiotte, “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 10, no. 10, (2008), pp. 108-120.
- [9] M. Mathioudakis, F. Bonchi and C. Castillo, “Sparsification of Influence Network”, *Proceedings of the*



- 17th International Conference of KDD, (2011), Manchester Grand Hyatt, San Diego, CA
- [10] J. Guo, P. Zhang and B.X. Fang, "Personalized key propagating users mining based on LT model", Chinese Journal of Computers, vol. 37, no. 4, (2014), pp. 809-818.
  - [11] M. Kimura, K. Saito and R. Akano, "Extracting Influential Nodes on a Social Network for Information Diffusion", Proceedings of the 18th Data Mining and Knowledge Discovery, (2010), London, England.
  - [12] D. Kempe, J.M. Kleinberg and E. Tardos, "Maximizing the spread of influence through a social network", Proceedings of the 9th International Conference on Knowledge Discovery and Data Mining, (2003), Sydney, Australia.
  - [13] H. Ma, H. Yang and M.R. Lyu, "Mining Social Networks Using Heat Diffusion Processes for marketing Candidates Selection", Proceedings of CIKM, (2011), Glasgow, UK.
  - [14] M. Du, "Research on information dissemination model of social network services based on probabilistic hyper-graph", International journal of signal processing, image processing and pattern recognition, vol. 8, no. 6, (2015), pp. 267-274.
  - [15] L.X. Xiao, "Research on the optimization of enrollment data resources based on cloud computing platform", Review of computer engineering studies, vol. 2, no. 2, (2015), pp. 9-12.
  - [16] A. Josang, "A logic for uncertain probabilities", International journal of uncertainty, fuzziness and knowledge-based systems, vol. 9, no. 3, (2001), pp. 279-311.
  - [17] F.H. Huang, J. Peng and L.M. Ning, "Opinion evolution model of social network based on information entropy", Acta Phys. Sin, vol. 63, no. 16, (2014), pp. 16051.
  - [18] L. Xiong and L. Liu, "PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities", IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 7, (2004), pp.843-857.
  - [19] G. Zhi, F. Zeng, H. Liu and F. Kuang, "Realization of the communication between devices of bacnet/IP and real-time database", Review of computer engineering studies, vol. 1, no. 1, (2014), pp. 17-22.
  - [20] Z. Bo, X. Yang and H. Zhenhua, "Recommended trust computation method between individuals in social network site", Journal of Nanjing University of Aeronautics & Astronautics, vol. 45, no. 4, (2013), pp. 563-569.
  - [21] W.S. Chow and L.S. Chan, "Social network, social trust and shared goals in organizational knowledge sharing", Information & Management, vol. 45, no. 7, (2008), pp. 458-468.
  - [22] K.Y. Ae and P. Rasik, "A trust prediction framework in rating -based experience sharing social networks without a web of trust", Information Sciences, vol. 191, no. 5, (2012), pp. 128-145.
  - [23] A. Simpson, "On the need for user-defined fine-grained access control policies for social networking applications", Proceedings of the Workshop on Security in Opportunistic and Social Networks, (2008), New York, USA.
  - [24] K. David, K. Jon and T. Eva, "Maximizing the spread of influence through a social network", Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), (2003), Washington, USA
  - [25] N.T.J. Bailey, "The mathematical theory of infectious diseases and its applications", New York: Hafner Press, (1975).
  - [26] R.M. Anderson and R.M. May, "Infectious diseases of humans: dynamics and control", Oxford: Oxford University Press, (1992).
  - [27] O. Diekmann, J.A.P. Heesterbeek, "Mathematical epidemiology of infectious diseases: model building, analysis and interpretation", New York: John Wiley & Son publisher, (2000).
  - [28] D.J. Watts and P.S. Dodds, "Influentials, networks, and public opinion formation", Journal of Consumer Research, vol. 34, no. 4, (2007), pp. 441-458.
  - [29] D.J. Watts and S.H. Strogatz, "Collective dynamics of 'small-world' networks", Nature, vol. 393, no. 6684, (1998), pp. 440-442.
  - [30] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks", Phys. Rev. Lett, vol. 86, no. 4, (2001), pp. 3200-3203.

## Author



**Du Min**, she received her M.E. in Computer application (2008) from Central South University. Now she is associate professor of computer at information and engineering Department, Hunan University of Science and Engineering. Her current research interests include different aspects of information security and network security.

