An Evolution Algorithm of Opinion Tendency in Internet

Yun Liu¹ Lei Deng¹ Michael R. Bartolacci² Qing-An Zeng³



¹ Key Laboratory of Communication & Information Systems Beijing Municipal Commission of Education, Beijing Jiaotong Universityy, Beijing, 100044, China liuyun@bjtu.edu.cn, 09111025@bjtu.edu.cn

² Information Sciences and Technology, Penn State University-Berks, Reading, PA 19610, USA mrb24@psu.edu

³ Department of Electronics, Computer and Information Technology, North Carolina Agricultural and Technical State University Greensboro, NC 27411, USA gzeng@ncat.edu

Received 1 April 2015; Revised 15 April 2015; Accepted 24 April 2015

Abstract. We empirically analyze a "hot topic" of online marketing to exam the predictive performance of an opinion evolution model. We compare our model with existing algorithms of online prediction and conduct an analysis using a test data set. Results show that our model and corresponding information influence rule have increased prediction performance.

Keywords: evolution algorithm, hot topic, Internet, opinion tendency

1 Introduction

The determination of attitudes and tastes is the ultimate goal of online opinion research. Prediction methods using online opinion data often integrate theory with practice when combined with promotion methods that utilize such data. The promotion methods, as it turns out, have many properties in common with the prediction methods: By understanding the opinion update rule and influence factors, the tendency of opinions can be predicted; and by understanding the opinion update rule and the technical changes of influence factors, the tendency of opinions can be exploited during promotion.

The research on online opinion prediction includes four main areas: Growth trends prediction, Prediction of diffusion depth and range, Prediction of emergency events and Automatic Post Machine. The number of research publications in these areas are limited.

In this paper, we summarize the methods of online opinion prediction and promotion and introduce a scenario which is then simulated, to test the performance of opinion evolution theory on opinion prediction. Finally, we compare the simulation results with empirical data, discuss the potential for improvement of the methodology and propose an evolution algorithm of opinion tendency based on several key factors.

2 Internet opinion promotion and prediction

In a social network, new products, new technology, new systems and ideas have to go through the process of information diffusion first before undergoing further development. Based on the RAS [1] model, the process of opinion diffusion over the Internet can be considered as consisting of three main steps: innovators learn of a specific "innovation", express their views based on the original information, and promote the process of diffusion over the Internet as accomplished by their interactions with others.

The existing promotional strategies utilizing online public opinions are mainly divided into two types: publicity promotion strategy for innovator groups or opinion leaders, and direct ways to promote public-

ity for ordinary users [2]. In publicity promotion for opinion leaders, corporate analysts carry out network data analysis, the sampling of market research and other means to get the user's consumption tendencies. Such methods are used to identify opinion leaders and core users of the product area. With the influence of innovators and opinion leaders, visibility and recognition of the product are promoted rapidly. In the implementation process, one should select different strategies depending on the circumstances. In publicity promotion strategy for ordinary users, corporations do not have to research before promotion, but can carry out marketing to normal users directly. Such a strategy cannot gain high efficiencies. For such reason, the publicity promotion strategy for ordinary users is the most popular way.

Watts and his coworkers have researched the two particular strategies. They found that the propaganda effect of traditional advertising and online mainstream media is reduced with the rising popularity of online applications. Online opinion leaders act in various fields in the Internet environment, which leads to two-stage diffusion in the network communication process. The research results show that publicity promotion strategy for opinion leaders has a higher rate of diffusion if one does not consider the cost factor, time and risk [3]. N. Meade and his coworkers described the process of information dissemination and promotion as a probability-based decision problem [4]. In this way, corporations calculate the expected return of each outreach strategy based on Game Theory, and select the maximum expected as final decision. S. Plous have studied information dissemination under decision theory. Results show that based on information parameters and evaluation criteria for a certain area, those promoting online can seek the optimal solution by a numerical method [5]. Zhineng Hu and his coworkers proposed some methods for promotion effect calculation based on numeric plan, decision trees, equal probability, regret value, and profit matrix [6]. Two models of intervention strategies-external pressure intervention strategy and embedded intervention strategy—were established based on the former opinion evolution models [7]. H. Cheng proposed a model to forecast the trend of online public opinion from the view of time and number, which based on social inertia theory and time continuity of online public opinion [8].

Research of online opinion prediction includes main areas: time trend forecasts of "hot topics", depth prediction of information dissemination, prediction of the information spread range, and trend prediction of opinion evolution. There are several definitions of online "hot topics", and research of this area is self-contained [9].

Tseng and his coworkers proposed a sorting algorithm of online topics based on information retrieval and trend index [10]. Zhou have captured the log contents, message number, the number of visitors, visits and comments proportion, scale and other data of a certain blog, analyzed the features of hot topic in directed network, and proposed a heat sorting algorithm based on parameter influence weights [11]. Chen, Luesukprasert and their coworker modeled the time characteristics and statements dimension of online "hot topics" to further define the features such as: broader spread range, clear continuity, and larger online communities [12]. Zhang counted the opinion interaction data in many Internet communities and studied the trend prediction of hot topic. Results show that "hot words" are a critical factor of "hot topics". To define and determine a "hot topic", researchers can observe the frequency and time curve of a certain word in texts, blogs, posts and messages by word frequency extraction [13]. Zheng and Li studied the Aging Theory, proposed a heat ranking algorithm and analyzed the definition of "hot topic" based on that algorithm [14]. Based on above discussions, we can see that most research achievements of online opinion prediction belong to "hot topic" detection and definition. However, the research on opinion trend prediction is limited.

3 Quantify of online hot topic

3.1 Quantification of individual behavior and character

3.1.1 Individual behavior in the initial stage

Researchers of statistical physics and sociology have created a good deal of statistics about the diffusion and evolution process of public opinion in social networks. Results show that when a beginner i starts to receive messages within a certain topic or area, his or her individual awareness is very low; therefore the awareness parameter takes on a small value. In the improved RAS model, a beginner's awareness is almost equal to 0 [15]. At this stage, individual *i* is influenced by the Spontaneous Symmetry Breaking.

The Spontaneous Symmetry Breaking plays a significant role in beginner's opinion decision. Hence it can also influence the opinion evolution of entire network in the early stage if the beginners have a cer-

tain proportion. Individuals with higher awareness, which means experts, can judge a message's credibility, opinion and other properties on sight. Although individuals with lower awareness cannot distinguish such messages, they will support a particular opinion and divided into groups in a short time based on Spontaneous Symmetry Breaking [16]. We can see from the above that a message's accept probability is depending on information properties, individual awareness and Spontaneous Symmetry Breaking effect.

3.1.2 The influence of opinion Leaders on environment preference

In a Majority Rule Model, probability p was used to represent the environment preference of network, which determines the opinion preference of the whole network or most group members. For example: if 80% of individuals in the network admired civilized language, then we have p = 0.8 in this range of topics. In the process of opinion evolution, if one opinion takes a position of absolute dominance, it can assimilate opponents and persuade them to follow its own opinion. However, if individuals of two sides are in a stalemate, environmental preference p plays a significant role in individual decision making. Research indicates that vulnerable groups could also win with high probability under the support of environmental preference. We can see that environmental preference can influence the opinion evolution process and help the network to reach consensus. The influence of an opinion leader is very similar to the influence of an environment preference. Opinion leaders can also affect the environment preference in many cases. Opinion leaders rarely change their opinions, which could break the impasse of opinion confrontation and lead to network consensus in two of the following ways [17]:

Firstly, the messages from opinion leaders tend to have high reliability and greater information dissmination intensity which guarantees the receipt and acceptance probability of the messages themselves. Secondly, the opinion leaders interact with each other more frequently than common individuals, but also have more extreme views, which means they are able to persuade more neighbors.

3.1.3 Opinion polarization in the final stage

The researchers that studied the statistical results of online public opinion have found that individuals always have a tendency of turning into "extremists" during the opinion interaction. The phenomena of being extreme always happen in the final stage of opinion evolution. Extreme opinions appear more in the applications where small-scale information dissemination and quick interaction such as microblogging and online social networks, occur. Microblogging information and personal status have very limited words. A microblogging message only has 200 words or and therefore has a limited scale of information that only allows the blogger to present conclusions without any discussions and proofs. Results of micro-Bo show that: because of the message scale restrictions, microblog users' posts are in lack of facts, which in fact reduce the opinion interaction and lead to a barrier effect.

If the word limit of the posts is deregulated and background links are introduced, the negative factors could be eliminated.

3.1.4 Parameter settings

Based on the algorithm in [18], when individual awareness grows to a certain value, its ability of receive messages will suddenly change. We could set the value of awareness parameters a_1 vary from 0 to 2 and follows a decaying exponential distribution. If this topic covers a wide range and is easy to understand, the parameter a_1 can be randomly set and mostly closed to 2.

Based on the survey and statistical results, we can see that individual awareness follows an approximately normal distribution in large datasets. Individual awareness follows the normal distribution and varies from—2 to 2 so the probability results can locate in [0, 1]. Individuals with higher awareness means it has a higher receipt probability and lower accept probability.

Parameter b_2 represents the individual's degree of steadfastness and its value varies from 0 to 5. Statistical results show that the "stubborn" degree of Internet users has a similar awareness parameter and follows a normal distribution. Individuals with higher b_2 tend to insist upon their own opinions and have lower probabilities to accept a new one.

3.2 Quantify of information and opinion update

3.2.1 Information influence

Although there are lots of posts which could cause network emergencies, dissemination of online media and the information push of online communities are the main reasons for the existence online hot topics, sensitive topics and emergencies. Based on the opinion evolution model with information influence [18], the early stage of opinion evolution can be seen as a neutral environment and the late stage, which equates to opinion consensus, opinion stalemate or opinion polarization, can be seen as an extreme environments. In a neutral environment, the opinion of promote information should not be too extreme. Information with extreme opinion will cause a huge difference between information opinion and individuals' opinion, the value of parameter P_i turns to a negative number, which could lower the individual's accept probability to this message.

When a new message reaches the Internet community, there are three basic reactions that will could happen among individuals: someone will receive the message successfully and accept its opinion, some may receive the message, but refuse to follow its idea, or some may miss the message or cannot understand it. Individuals with such three reactions will certainly have discussions and interactions among themselves. Such interaction could involve posting, reading, replying and reposting. From a theoretical model point of view, each individual which posts or replies to posts can be seen as an independent information source. Everyone commenting or posting in response to that individual can be seen as an opinion expression and persuasion to the others. Other individuals will certainly have a judgment or opinion update, and express their own opinions based on the received persuasion, which completes the circle of opinion interaction. In the process of opinion evolution, the Internet community may receive new messages continually and raise another round of opinion interaction among individuals. The message from Internet media obviously has a more powerful influence, which means it has a larger information credibility and intensity than other messages expressed by individuals. Such differences in credibility, intensity and opinion value can also influence the individuals' reactions which have a further impact on opinion formation. The above phenomenon can be described as Information Influence.

3.2.2 Information quantization

The quantification of the relevant parameters and incorporating them into model is the final stage of model initialization. The parameter a_0 which describes the intensity of message and diffusion determines the "hot" degree of the message's area. Messages of a hot topic will have a higher value of intensity. Note that it cannot be ensured that every individual in this community will receive the message. However, the messages of the "hot topic" or messages from the Internet media will have higher receipt probabilities. Based on the algorithm in [15], the value of parameter a_0 vary from—1 to 1. In the lifecycle of the "hot topic" or event, the value of the intensity parameter can be very high and kept constant by default. Meanwhile, the views or posts published by an individual do not have high values. The value of an individual's intensity and of information intensity can be set as 0.5 and -0.5, respectively. The floor parameter f equals to 0 under default conditions. The value of message creditability b_0 changes with information itself. In this topic, the parties kept blaming each other and defending themselves. Different from loyal supporters of the two sides, most of the Internet users do not have a firm stance and can hardly tell the right from wrong based on the speech from each side. Based on the above discussion, we can set most messages to have a common credibility. Such a setting cannot affect the essence of evolution. The value of parameter b_0 varies from 0 to 4. The credibility parameters of the given information equal to 2. The similarity between individual's opinion and message's opinion determines the value of parameter P_i .

4 Data set

Data sets for Internet opinion evolution research are difficult to obtain. There are many reasons for this difficulty including:

Most of the interactive topics of the Internet are strongly affected by the network environmental
preference: most topics have clear moral tendencies. For topics of this kind, users can tell right from
wrong and determine their positions easily by legal awareness and moral values without the opinion
interaction and discussion. One opinion overwhelming is always rendered during the process of dis-

cussion. From a theoretical point of view, the environmental preference is a strong and common interference to the process of opinion evolution.

- The lifecycle and dissemination range of most topics are very small. More than 80 percent of emergencies and explosive topics actually cannot be sustained for more than two weeks. And there are even less topics which have a longer lifecycle and high user participation.
- Every topic has some form of user interest attenuation and unstable user activity [18-19]. In the traditional sample survey, researchers can select a particular sample population and "keep up" to understand the latest focus of the crowd. However, in the research of Internet opinion evolution, the researchers sought to quantify the behavior of Internet users and the changing process of their opinion. Since a considerable proportion of users having erratic interests and are unable to keep abreast, the data acquisition is quite difficult.
- In order to restore some specific users' opinion evolution and development track, the researchers need to track the comments of specific users in different forums and different posts. However, this technology is still not perfect. If one wants the computer to achieve the above functions, one must integrate the behavior analysis algorithm, feature discovery and clustering, web crawler, semantic analysis by programming such features. Each of them have a high error rate and are difficult to merge with the others.

The ideal research subjects of public opinion, individual decision-making and opinion interacting in the Internet possess the following characteristics: long life cycle, high participation rate, weak environmental preference effect and clear evolutionary track of opinion. We can see from the above that it is very difficult to obtain such empirical data. After the above screening, we selected a "hot" networking event in first half of 2012 as our research focus for study. The information published through mainstream media is as follows: February 27, 2012, a Hong Kong newspaper published a message "A diva, 'actor A', has trouble with film crew". February 29, the film crew publicly announced they terminate the partnership with "actor A". On the same day, the message published by Hong Kong media spread on the Internet. Media Channel of People Website, "South China Sea" and Yahoo entertainment made reprints. March 1, "actor A" made microblogging that "diva speech is not real" and said he was maligned. The event started to attract the network media's attention. March 3, the film crew sends e-mails to the network media and explains salary, collaboration with production manager, arrangement of calendar and other affairs to them. They tried to prove that "A" had crossed the line. March 4, the "actor A" held a press conference, accusing the film crew of spreading rumors and that the crew kept modifying the script without his permission. On the same day, the crew rebuts the speech on a microblog. The accusations against "actor A" by online media had reached a peak. Up to March 5, related events are reproduced by the People's Website, China News, Yahoo, Sohu and many other mainstream media. Those websites then created news column and online comments. March 14, "actor A" accepted interviews with the media, accusing the film crew of modifying the script secretly and crowding collaborators, which lead to the end of the cooperation between A and B. March 14, "A" and "B" published statements accusing each other of wrongdoing in Sina microblog which triggered many film stars to declare their support for one or the other. The supports of two sides was also debated on Sina microblog and with corresponding attacks on each other. Meanwhile, the online discussion had peaked. Massive postings were posted in Tianya BBS, Phoenix, Tencent and other portals. A voting feature was opened in Phoenix and Tencent, which tried to investigate the support rate for both of the two sides with Internet voting. April 5, the former director of the film conducted an interview and said that during another actor beyond the film crew had modified the script as well as the staffing without being authorized to do so. These remarks were refuted by a number of the involved performers, but were ineffective in changing public attitudes. At this time online public opinion had become unified, that 'actor A' had the full support of Internet users who had an opinion on the matter. support rate.

The data on this topic and corresponding opinions were sampled from the following web portals: Tianya BBS, Phoenix, Tencent and Global Survey. The sampling ratio was 1%.

When we sampled on May 27, 2012, the number of newly generated message on this topic had stabilized and started to decay. Statistical results showed that support rates for 'actor A' reached 90%. The online public opinion had evolved into a stable opinion and then progressed into a decline stage. Up until May 27, 2012, the largest post on this topic in Tianyxa BBS gained 4,423,059 views and 285,357 replies. The discuss participation rate reached 6.45% [20]. The number of related posts grew to 2,094,113 and the number of total replies to 121,371,462. There were 2,392,718 users involved in the discussion up until that point in time. Up until March 1, 2013, the largest post on this topic in Tianyxa BBS garnered 5,792,918 views and 603,496 replies. There were 2,604,379 users involved in the discussion up until that point in time. It was obvious from examination of the posts that the number of new posts was less than the number of new replies, thus the growth rate of new users and new posts had slowed down dramatically in almost a year.

5 Model evolution and data contrast

From an opinion diffusion point of view, there is no essential difference between a single opinion leader and online media campaigns created by many entities. However, online media campaigns are much better with respect to information receipt probability. In order to reflect this feature, we set parameters such that the information from media can be received by every individual when we model the opinion interaction and evolution. This means that the floor parameter f and message intensity a_0 should be higher. According to F. Ding's research [20], even an individual with good memory can hardly change his opinion during a single "hot" topic's life circle. Fluctuations of individual preference can hardly influence the opinion evolution. So we reduced the weights of individual awareness W_i and individual preference b_2 . In order to study the evolution of the opinion and remove the influence of new posts' time characteristics and the users' interest attenuation, we sampled the discussion data for the support rate of "actor A" and "actor B" (other party involved in the dispute). First, we selected February 29 as the start of the opinion evolution, the carried out a simulation, and finally compared the simulation results with statistical results.

The statistical results shown in Fig.1 were sampled from the largest post of Tianya BBS and the sampling rate used was 1%. After 14 days of the topic accrued, the discussion had entered a stable period. So the 14th day was chosen as the start of the simulation. From Fig.1 it can be seen that the curve of the simulation results clearly show support growing in favor of "actor A". Based on the stage of model initialization, the simulation result can reflect the true tendency of opinion evolution.

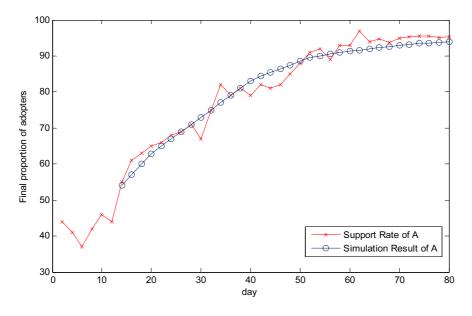


Fig. 1. The support rate of "actor A" as a function of time

Fig.2 shows the number of new posts in the first 20 days. When we started to sample, 340 days had passed after the topic emerged. In the first 5.88% of the total elapsed time, 25,331,010 posts had been made. In this period of time, the discussion had shown a significant growth, but public opinion had not yet reached a consensus yet. However, we can see from Figure1 that the there is a clear trend in terms of online support. Most replies were generated in the 6th and 14th day of our study. The number of replies had an obvious periodicity and attenuated over a relatively short time. The cyclical nature of the curve is related with information publication emanating from each side of the conflict. A decay function with a periodic component is an effective means for modeling a "hot topic" life circle' graphical representation.

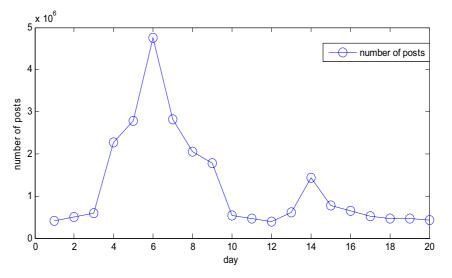


Fig. 2. The number of new posts as a function of time

The curve peaks appears when x = 6 and x = 14. We can see that the 6th and 14th day is also the peak of the information publication. This means that new information publication is an effective means to extend-the topic life cycle. Because topics in the early stage can easily attract public attention, the first peak takes a higher value then the second one.

Online "hot topics" always have a short term stochastic nature to them. However, from a statistical point of view, developments and changes in topic heat tend to show an obvious pattern. With the passage of time, user interest does rapidly attenuate.

6 Opinion prediction based on key factors

The process of opinion evolution and diffusion shows a two-stage property under the influence of opinion leaders. In our focus "hot topic", Internet users neglected the initial advantage of "actor B" advantage and grew support of "actor A". This phenomenon shows that the opinion leader can create a higher probability of acceptance for their opinion. In this event, the details of messages also influenced the process of opinion evolution. Sampling results show that "actor B's negative attitudes towards Internet users hindered the opinion interaction and created barrier effect. On the other hand, "actor A" published a large amount of messages with a positive attitude and won the support of Internet users. As two ultimate goals of opinion dynamic research, opinion promotion and prediction have many similarities. Once influencing factors have been determined, opinion prediction can be achieved. Once these influencing factors have been manipulated, the process of opinion evolution can take place and opinion promotion can be achieved successfully. Based on the above considerations, in can be seen that information from the media has an important effect on opinion diffusion. Messages with the right properties can win the trust of Internet users in the beginning and extend the life circle of a given topic. Successful promotion information should have properties such as a neutral opinion, high credibility, and the support of opinion leaders. It can then be easily utilized by online media and not views are promotional in nature.

7 Conclusions

We studied an opinion evolution approach for information spread over Internet-based media and carried out an empirical study of a real "hot topic". We the summarized the existing models and algorithms of Internet opinion promotion and prediction. Further, we discussed opinion interaction and opinion evolution under the notion of information influence. Finally, we discussed the promotion and prediction methods of Internet opinion. The results show that the process of online opinion evolution is determined primarily by opinion leaders. In the process of opinion evolution, the opinion leaders have more frequent discussion with other individuals which helps the diffusion of their own opinion. In the initial stage of opinion evolution, the frequent discussion among opinion leaders and other individuals helps the opinion leaders win the trust of the public. Opinion leaders can enlarge the spread range of their own opinion during this process. On the contrary, messages emanating from Internet news services have a disadvantage in this area: the authors of Internet news services rarely reply to comments generated by their work or participate in discussions. This serves to strengthen the barrier effect, lower the acceptance probability for their opinion, and enlarge the probability of success for a conflicting opinion.

Acknowledgement

This research is supported by National Natural Science Foundation under Grant 61371071, Beijing Natural Science Foundation under Grant 4132057, Beijing Science and Technology Program under Grant Z121100007612003, Academic Discipline and Postgraduate Education Project of Beijing Municipal Commission of Education.

References

- [1] Zaller, J. R. (2005). The nature and origins of mass opinion. New York, NY: Cambridge University Press.
- [2] Jamali, S., & Rangwala, H. (2009, November). Digging Digg: Comment mining, popularity prediction, and social network *analysis*. Paper presented at the International Conference on Web Information Systems and Mining, Shanghai, China.
- [3] Watts, D. J. & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34(4), 441-458.
- [4] Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation: A 25-year review. *International Journal of Forecasting*, 22(3), 519-545.
- [5] Plous, S. (1993). The psychology of judgment and decision making. New York, NY: McGraw-Hill.
- [6] Hu, Z. N., & Xu, J. P. (2003). Operations research. Sciences Press.
- [7] Zhang, L., & Liu, Y. (2008). Research on aggregation and intervention strategy in opinion evolution. *Journal of Internet Technology*, 9(5), 411-415.
- [8] Cheng, H., & Liu, Y. (2008). An online public opinion forecast model based on time series. *Journal of Internet Technology*, 9(5), 429-432.
- [9] Zhang, Y. C., Liu, Y., Ding, F., & Si, X. M. (2010). The research on stability of diffusion and competition between online topics. *International Journal of Modern Physics C*, 21(12), 1517-1529.
- [10] Tseng, Y. H., Lin, Y. I., Lee, Y. Y., Hung, W. C., & Lee, C. H. (2009). A comparison of methods for detecting hot topics. *Scientometrics*, 81(1), 73-90.
- [11] Zhou, E. Z., Zhong, N., & Li, Y. F. (2011). Hot topic detection in professional blogs. Active Media Technology, 6890, 141-152.
- [12] Chen, K. Y., Luesukprasert, L., & Chou, S. C. T. (2007). Hot topic extraction based on timeline analysis and multidimensional sentence modeling. *IEEE Transactions on Knowledge and Data Engineering*, 19(8), 1016-1025.
- [13] Zhang, Z. F., & Li, Q. D. (2011). Question holic: Hot topic discovery and trend analysis in community question answering systems. *Expert Systems with Applications*, 38(6), 6848-6855.
- [14] Zheng, D. H., & Li, F. (2009). Hot topic detection on BBS using aging theory. Web Information Systems and Mining, 5854, 129-139.
- [15] Deng, L., Liu, Y., & Zeng, Q.-A. (2012). How information influence an individual opinion evolution. *Physica A: Statistical Mechanics and its Applications*, 391(24), 6409-6417.
- [16] Hemsley, D. R., & Clark, D. M. (1980). Selective attention in reviews of the effectiveness of cognitive therapy. Behav-

ioural Psychotherapy, 8, 55-58.

- [17] Zimbardo, P., & Leippe, M. (1991). The psychology of attitude change and social influence. New York, NY: McGraw-Hill.
- [18] Liu, Y., & Xiong, F. (2013). External activation promoting consensus formation in the opinion model with interest decay. *Physics Letters A*, 377, 362-366.
- [19] Xiong, F., Liu, Y., & Zhang, Z. J. (2011). An opinion formation model with dissipative structure. *Physica A: Statistical Mechanics and its Applications*, 390(13), 2504-2510.
- [20] Ding, F. (2010). *Research on information interaction and diffusion in internet communities*. Beijing, Cnina: Beijing Jiaotong University.