



The Three-Degree Calculation Model of Microblog Users' Influence (Short Paper)

Xueying Sun and Fu Xie(✉)

School of Information Science and Engineering, Shandong Normal University,
Jinan 250014, China
sunxueying@stu.sdnu.edu.cn, 1451850328@qq.com

Abstract. Highly influential social users can guide public opinion and influence their emotional venting. Therefore, it is of great significance to identify high-impact users effectively. This paper starts with the users' text content, users' emotions, and fans' behaviors. It combines the amount of information in the content and sentiment tendency with the fans' forwarding, commenting, and Liking actions. And based on the principle of the three-degree influence, the users' influence calculation model is constructed. Finally, the experimental results show that the three-degree force calculation model is more accurate and effective than other similar models.

Keywords: Three-degree · Microblog · User influence

1 Introduction

Taking the Sina Microblog social networking platform as an example, by September 2017, the number of monthly active users of Microblog was 376 million, increased 27% compared with the same period in 2016, of which China's mobile terminal accounted for 92% and the daily active users reached 165 million, an increase of 25% over the same period of last year. If we can find influential nodes in the network quickly and effectively, it will be better for the government to monitor public opinion and guide them in time. For the business community, it is possible to achieve targeted commercial promotion, advertising marketing, etc. according to the influence nodes in the network. Therefore, the research on the analysis, measurement, modeling, and dissemination of node influence in social networks has a very important theoretical and practical values.

According to the three-degree influence theory, this paper constructed a computing model based on users' blog content and fans' behavior. Firstly, calculate the content of the blog post and measure the amount of content information through the information entropy. Secondly, analyze and calculate the sentiment tendencies of the blog post. In addition, calculate the "forward-comment-like" influence of the user's three-tier. Finally, verify it by crawling data. Results showed that the three-degree influence calculation model proposed in this paper is more accurate and effective than other similar models.

2 Related Research at Home and Abroad

Research on influence attracted the attention of scholars from all walks of life in the early 20th century. When Triplett [1] studied social promotion theory, he discovered that when a person received attention from others, his/her performance would become more prominent. By the 1950s, Katz et al. [2] found that influence played a very important role in daily life or in political elections. In recent years, with the rise of large-scale social networks such as Microblog, various theories have shown that the distance between people has become shorter and shorter, and the links have become closer and more influential.

Zhang and Tang [3] pointed out three aspects of influence, the first is the individual's influence on one user only, means the ability of affecting other users and the probability of being affected by others; the second one was the influence between the two users of A and B, means A influences B and B influences A; the third is the group influence, which is the most complicated, refers to the influence of a group of multiple users on one user. Shi et al. [4] integrated user behaviors and blog post content and evaluated the predicted users and the influence of the issued microblog through the extraction of behavioral characteristics and content characteristics such as the users' blog posting time and "forward-like" behavior. Chen et al. [5] analyzed users' influence in terms of user behavior and community structure. First, it divided the network into groups. Then, the users' influence was measured from the number of fans of the user and the quality of the fans and the number of communities the user crosses. The authors in [6] evaluated users' influence from two aspects of user behavior and user relationship. First, analyze users' blog post content. Second, improved based on the PageRank algorithm, and calculates activeness for each node. Finally calculate the BRR (Behavior-Relationship Rank) of the node. and sort the nodes based on their value i.e. influence. Xu et al. [7] proposed a new social influence model, the PTIM model, which was an iterative combination of the characteristics of the user's fans and small-world attributes and has achieved relatively good results for all evaluation indicators. Fowler et al. [8] found that the happiness between people could reach a three-degree separation by studying the spread of happiness in social networks, that is the principle of the three-degree influence. The node affects the neighboring node, this is the first-degree effect; it can also affect the neighboring node's neighboring node, which is the second-degree effect; and it also affects the neighboring node's neighboring node of the neighboring node, which is the third-degree effect. The study found that the effect diminished with the separation of time and geography, and the influence beyond three degrees was negligible. In the literature [9,10], through real data experiments on Facebook and SNS, it pointed out that both the intensity of friendship between friends or the intimacy between users have a great impact on user behavior. Through the interaction intensity and intimacy analysis, a richer social graph can be obtained. The above studies on influence started from different aspects and can be roughly divided into three aspects: user relations, behavior, and content of blog post. The dissemination of information was interlocking and complex in the network. As above mentioned [5], attention was paid to the quality of the users'

fans. However, the dissemination of information is not merely a forwarding level and can be spread by multiple users.

3 The Three-Degree Calculation Model of Users' Influence

3.1 The Calculation of First-Degree Influence Based on the Content of Information

The more content a microblog user publishes, the greater the amount of information they contain and they are more likely to be forwarded. This article used information entropy to represent how much information the user content contained. For the analysis of the content of the blog post, this paper used the NLPIR semantic analysis system. NLPIR (Natural Language Processing and Information Retrieval Sharing Platform) is a word segmentation system developed by Dr. Zhang Huaping. The more information in the post content, the higher the complexity, the greater the entropy value, and the greater the probability of being followed, browsed, and forwarded. The keyword collection function of the NLPIR semantic analysis system can get the keyword set K , and the keywords' weight in the content also can be obtained.

$$P_i = \frac{W_i}{\sum_{i \in k}^n W_i} \quad (1)$$

where W_i refers to the weight of the extracted keywords in the content; and P_i refers to the probability that the i -th keyword appears in the content. The entropy [11] is calculated as follows:

$$E = - \sum_{i \in k} P_i \cdot \log P_i \quad (2)$$

While analyzing the content, we need to consider the topic degree of the blog post, whether the content fits the hot topic. Calculate the similarity of blog post content and topics. For text similarity, we use the cosine similarity to calculate the similarity of blog content and hot topics in this paper.

$$Sim = (content, hottopic) = \cos(K, H) = \frac{\sum_1^n (K_i \times H_i)}{\sqrt{\sum_1^n K_i^2} \times \sqrt{\sum_1^n H_i^2}} \quad (3)$$

In the formula, K_i , H_i refer to the weights of features extracted from the blog post content and feature weights of the current hot topics.

Finally, the influence of the post content can be expressed as:

$$TI = \begin{cases} E(1 + Sim(content, hottopic)), & Sim > 0 \\ E, & Sim = 0 \end{cases} \quad (4)$$

3.2 The Calculation of Second-Degree Influence Based on Bowen's Emotion

Microblog posts not only contain the views and opinions of users, but also embody their emotions. Therefore, the post's emotions also play a crucial role in the user influence. The study of text sentiment tendencies can be mainly divided into two aspects: the method based on dependency syntax analysis [12–14] and the method based on the term bag model [15–17]. Referring to the methods of these two aspects, the rules for calculating the sentiment orientation of sentences are given below.

Calculation rules 1. Emotional calculation of emotional words. The emotional word is recorded as “*word*”, emotional intensity is “*EIntensity*”, emotional positive and negative polarity is “*EPolarity*”. The calculation is defined as follows:

$$Value(word) = \Theta \times EIntensity(word) \times EPolarity(word), \Theta \in R \quad (5)$$

EIntensity, the emotional intensity, is the original part of speech, and refers to the emotional tendency of the word itself such as “praise”, “pampering” which has a clear emotional tendency; *EPolarity*, the positive and negative polarity of emotion is the positive and negative emotional tendency of the word. Θ refers to the harmonic coefficient which is the modified polarity of the word, and it is expressed as an adverb of degree such as “100%” and “compared”.

Calculation rule 2. Sentence calculation rules. V is the tendency value of emotional words, and its definition is as follows:

$$Value(sentence) = \sum_{i=1}^m V(word) \quad (6)$$

Among them, m refers to the number of emotional words in a sentence, and the emotional tendency of the entire sentence can be obtained by adding up the calculated emotional tendencies of each emotional word.

Calculation rule 3. The emotional value of blog post content is defined as follows:

$$Value(text) = \frac{1}{n-m} \sum_{i=1}^n V_i \times W_i \quad (7)$$

Among them, n is the total number of emotional sentences in the blog, and m is the number of sentences in the blog that do not contain emotions. V_i refers to the emotional value of the i -th sentence, and W_i refers to the weight of the i -th emotional sentence in the blog post.

The above calculation can divide the blog into three levels.

$$\begin{cases} Value > posThreshold, Positive Emotion \\ negThreshold < Value < posThreshold, Neutral Attitude \\ Value < negThreshold, Negative Emotion \end{cases} \quad (8)$$

PosThreshold and *negThreshold* are the propensity thresholds for positive emotions and negative emotions, respectively.

3.3 Three-Degree Influence Calculation Based on Fan Behavior

The behavior of fan users on a microblog mainly includes three types: forwarding, liking, and commenting. We use BI to represent behavior-influence, the expression is as follows:

$$BI = U(FI, CI, LI) \tag{9}$$

FI indicates the forwarding influence, which is how many people forwarded the blog post. CI indicates the influence of the comment, i.e., how many people have commented on the blog post. LI is the influence of praise, which is how many people like it.

User v is the fan of u , then v 's attention to u 's microblogs can be measured by the time difference between u publishing blog posts and v 's behavior. Assume that the blog post set posted by u is Mu and the time set is Tu , the blog set forwarded, commented and liked by v is Mv and the time set is Tv . Calculate the time difference of all the actions, and find the mean value. Where the smaller the mean value, the greater attention that v represents for u , the greater the likelihood of v 's forwarding. And calculate its variance S^2 . The smaller the fluctuation of the variance, the higher the reliability of the mean, and the more stable the attention of v to u .

The specific calculation is as follows:

$$M = \frac{\sum(T_v - T_u)}{N_{vu}} \tag{10}$$

where M is the mean of all time differences; $(T_v - T_u)$ is the time difference for a blog post; $\sum(T_v - T_u)$ is the sum of all time differences; and N_{vu} , which is how many time differences there are.

$$S^2 = \frac{\sum(T_v - M)^2}{N_{vu}} \tag{11}$$

S^2 is the variance of the time difference, and the smaller the variance, the more stable the attention degree. Finally, the variance is translated into the probability that the fan produces behavior. Based on the variance, the frequency of the behavior that generates the behavior should also be taken into consideration.

$$F = \begin{cases} \frac{n}{N}(1 - 0.01S^2), S^2 > 0 \\ 1, S^2 = 0 \end{cases} \tag{12}$$

where n is the total number of fans' actions. N is the total number of users' microblogs.

Therefore, the influences of forwarding, commenting, and liking are calculated as follows, respectively:

$$FI = \sum_{v \in Fan(u)} F_{vu} \tag{13}$$

$$CI = \sum_{v \in Fan(u)} F_{vu} \tag{14}$$

$$LI = \sum_{v \in Fan(u)} F_{vu} \quad (15)$$

We introduced α , β , γ , three parameters that represent the weights of the three types of behaviors, BI was calculated as follows:

$$BI = \alpha FI + \beta CI + \gamma LI \quad (16)$$

The principle of the three-degree influence showed that information was attenuated during propagation, so we introduced an attenuation factor ω , and as the distance increased, the value of the factor will decrease. For content information, emotions, and users' behavior, weighted fusion was based on their relationship. Calculated as follows:

$$SBI = a \cdot (TI \times Value) + b \cdot (\omega_1 BI_1 + \omega_2 BI_2 + \omega_3 BI_3) \quad (17)$$

SBI represents the sum of users' three-degree influence, which is users' influence; TI is the content influence; $Value$ refers to the emotional tendency value; BI_1 , BI_2 , BI_3 represent the influence of each layer, respectively; and a and b are the weights of the evaluation indicators.

4 Experiments

4.1 Data Collection

Through the Python crawler, the real data set obtained from the API interface provided by the Sina Weibo platform, was to verify the calculation model proposed in this paper. This article has crawled 946 microblogs from the official website of Shandong Normal University from November 2017 to the end of February 2018. There were 66,617 three-layer forwards, comments, and likes, including 5931 pieces of forwarding data, 12,065 pieces of review data, and 48,675 pieces of like data. In addition, we also got the relevant information from 500 hot topics including attributes as content, the number of fans, reading volume, and discussion volume. Before data is used, we has removed the inactive users.

4.2 Data Processing

For the analysis of the content of the blog, we calculated the content information of the blog based on the NLPPIR semantic analysis system of the Chinese Academy of Sciences, and calculated the emotional tendency value according to the sentiment calculation rules given above. When calculating fans' attention, we first counted the nicknames of the fans under Microblog, the time when they acted, and the number of microblogs each person interacted with, selected the appropriate time unit, and then got the fans' attention.

Figure 1 shows attention of some fans. The center point of the circle is 0, and it spreads out. The outermost layer is 1. We could see that most of the fans' attention was relatively low, and there were a few fans with high attention.

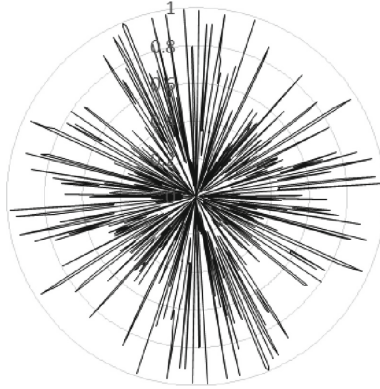


Fig. 1. Attention.

Three weights α, β, γ were introduced when we calculated each layer of users' influence. They represent the proportions of the three types of actions: forward, comment, and like. To calculate three weights, we used the analytic hierarchy process to get $\alpha = 0.539, \beta = 0.297, \gamma = 0.164$, as Table 1. For the introduced attenuation factor, we obtained that $\omega_1 = 0.136, \omega_2 = 0.098, \omega_3 = 0.024$ refers to literature [8, 18]. According to subjective evaluation method and reference [4, 5], $a = 0.3, b = 0.7$.

Table 1. The value of α, β, γ .

	Like (γ)	Comment (β)	Forward (α)
Like (γ)	1	2	3
Comment (β)	1/2	1	2
Forward (α)	1/3	1/2	1

4.3 Algorithm Comparison

Evaluation Index. The evaluation index adopted in this paper was the coverage rate [19], which was the ratio of the number of nodes affected by the selected user to the total number of nodes in the microblog. The greater the coverage rate, the wider the range of influence that the users spread, and the more people are affected, the greater the influence.

$$Coverage = \frac{\text{The number of affected nodes}}{\text{Total number of nodes}} \times 100\% \tag{18}$$

Compare. The purpose of the experiment was to prove that the algorithm proposed in this paper had a higher accuracy, so compared it with the PageRank algorithm and BRank algorithm [20], which only considered user behavior and the RBRank algorithm [21] which considered user relationship behavior. The result is shown in Table 2.

Table 2. Comparison between this algorithm and other algorithms.

Algorithm	PageRank	BRank	RBRank	This algorithm
Impact degree	20.9%	23.5%	26.1%	27.1%

From Table 2, we can see that the algorithm proposed in this paper improved the accuracy, but the difference of results was not very obvious. Analyzing the reasons, we believed that Shandong Normal University was the school portal, and the content and emotions conveyed were mainly positive. The fans were mainly students. The composition and structure were unitary, and the influence of fans was limited. Therefore, we needed to select other users with rich attributes to conduct the data acquisition and analysis again.

In the next experiment, we selected Fan Bingbing as a keyword and crawled its data for the past three months. The crawled attributes still included fans' nicknames, profiles, ID, followers, fans, microblogs, blog posts, and release dates. The number of forwarding, comments, number of likes, forwarders nickname, profile, ID, content of forwarded blog, forwarding time, totaled 280,460 items. According to the results of this algorithm, the comparison results are shown in Table 3.

Table 3. Comparison between this algorithm and other algorithms.

Algorithm	PageRank	BRank	RBRank	This algorithm
Impact degree	39.5%	47.6%	56.7%	63%

From Fig. 2, we can see that the RBRank algorithm and the algorithm of this paper had larger influence values than the PageRank and BRank algorithms. It explains that from multiple angles, selecting multiple eigenvalues in combination to evaluate users' influence is much more advantageous than simply calculating users' influence from one aspect. In addition, through the comparison, we found that the coverage rate of the proposed algorithm was the largest, indicating the best effect. Therefore, we reached a conclusion the proposed algorithm considered the users' sentiment content and emotional tendencies as well as the characteristics of the user behavior, expanding the depth of calculations, and calculating the three-degree influence was better.

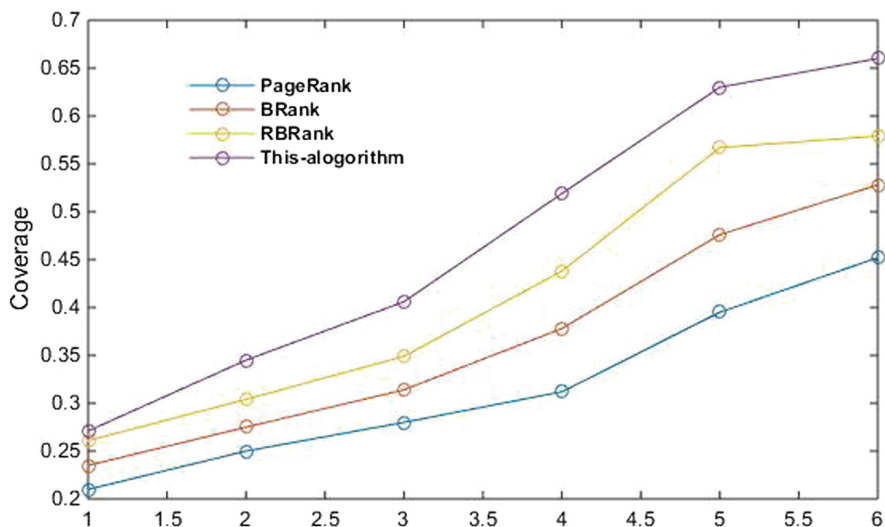


Fig. 2. Comparison between this algorithm and other algorithms.

5 Summary and Further Work

This paper proposed a calculation method based on the users' text content and users' behavior. Through the evaluation of content information and emotional tendency, the overall statistics on the users' forwarding, comments, likes, we constructed a users' influence computing model and improved accuracy. From the above experiment, we could see that selecting different experimental objects would have different effects, and the effect was more significant when the experimental objects were rich in attributes. Therefore, in future work, we will consider adding users' attributes to the calculation model to make the analysis of influence more perfect. Moreover, the cross relationship between fans is not considered in text, we should also be noted in the future work.

References

1. Triplett, N.: The dynamogenic factors in pacemaking and competition. *Am. J. Psychol.* **9**(4), 507–533 (1970)
2. Katz, E., Lazarsfeld, P.F.: Personal influence: the part played by people in the flow of mass communications. *Am. J. Sociol.* **21**(6), 1583–1583 (1955)
3. Zhang, J., Tang, J.: A review of social influence analysis. *Chin. Sci. Inf. Sci.* **47**(08), 967–979 (2017)
4. Shi, C., Tang, J., Hu, Y.: Predicting microblog user influence based on user behavior and blog content. *J. Chin. Comput. Syst.* **38**(07), 1495–1500 (2017)
5. Chen, Z., Liu, X., Li, B.: Analysis of influence of microblog users communication based on behavior and community. *J. Comput.* **35**(07), 1–6 (2018)

6. Kang, S., Zhang, C., Lin, Z., Shi, X., Ma, H.: Complexity research of massively microblogging based on human behaviors. In: Proceedings of the International Workshop on Database Technology and Applications, Wuhan, China, 27–28 November 2010, pp. 1–4. IEEE, New York (2010)
7. Xu, D., Liu, Y., Zhang, M., Ma, S.: Research on user influence based on online social network. *J. Chin. Inf. Process.* **30**(02), 83–89 (2016)
8. Fowler, J.H., Christakis, N.A.: Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham heart study. *BMJ* **337**, a2338 (2008)
9. Ilić, J., et al.: Proof of concept for comparison and classification of online social network friends based on tie strength calculation model. In: Zdravković, M., Trajnović, M., Konjović, Z. (eds.) Proceedings ICIST 2016 (2016)
10. Al-Ghaith, W.: Understanding social network usage: impact of co-presence, intimacy, and immediacy. *Int. J. Adv. Comput. Sci. Appl.* **6**(8), 99–111 (2015)
11. Li, R., Zhang, H., Zhao, Y., Shang, J.: Research on automatic abstracting technology of Chinese documents based on topic model and information entropy. *Comput. Sci.* **41**, 298–300 (2014)
12. Feng, S., Fu, Y., Yang, F., Wang, D., Zhang, Y.: Analysis of Bowen affective tendency based on dependency syntax. *J. Comput. Res.* **49**, 2395–2406 (2012)
13. Liu, F., Wang, L., Gao, L., et al.: A web service trust evaluation model based on small-world networks. *Knowl.-Based Syst.* **57**(2), 161–167 (2014)
14. Hu, M., Yao, T.: Chinese micro blog view sentence recognition and evaluation object extraction method. *J. Shandong Univ. (Sci. Edn.)* **51**(07), 81–89 (2016)
15. Liu, F., Wang, L., Johnson, H., Zhao, H.: Analysis of network trust dynamics based on evolutionary game. *Sci. Iranica Trans. E: Ind. Eng.* **22**(6), 2548–2557 (2015)
16. Paltoglou, G., Thelwall, M.: A study of information retrieval weighting schemes for sentiment analysis. In: Meeting of the Association for Computational Linguistics, 11–16 July, Uppsala, Sweden, 1386–1395. DBLP (2010)
17. Zhang, H., Li, H., Li, Q.: Research on automatic calculation algorithm of emotional word discovery and polarity weight. *J. Chin. Inf. Process.* **31**(03), 48–54 (2017)
18. Zhao, J., Liu, Y., Li, X., Wang, M., Mo, S.: The evaluation method of node influence based on the third degree theory. *J. Fuyang Teach. Coll.* **33**(04), 78–82 (2016)
19. Hou, W., Huang, Y., Zhang, K.: Research of micro-blog diffusion effect based on analysis of retweet behavior. In: International Conference on Cognitive Informatics and Cognitive Computing, pp. 255–261. IEEE (2015)
20. Li, X, Cheng, S., Chen, W., et al.: Novel user influence measurement based on user interaction in microblog. In: International Conference on Advances in Social Networks Analysis and Mining, pp. 615–619. IEEE (2013)
21. Huang, Y., Li, L.: Analysis of user influence in social network based on behavior and relationship. In: International Conference on Measurement, Information and Control, Harbin, pp. 682–686, August 2013