

InfluenceRank: An Improved Online Social Influence Model

Yun Bai*, Suling Jia, Meng Wu

School of Economics and Management, Beihang University, Beijing 100191, China

*Corresponding author. Email: baiyunbuaa@163.com

ABSTRACT

User influence is a popular research content in online social networks, and it plays an important role in marketing, public opinion management, and network relationships. Traditional research on user influence based on graph structure mainly considers whether users follow each other. However, "zombie fans" can make the influence analysis results inaccurate. Based on the PageRank algorithm, this study proposes a novel model for measuring user influence: InfluenceRank. User behavior and interaction information are introduced into the model through three indicators: activity, interaction and credibility. The experimental results, which are more comprehensive and persuasive, prove that the influence ranking of the InfluenceRank model on Microblog (Chinese Twitter) users is not limited to the number of users' fans.

Keywords: Influence, social media, PageRank

1. INTRODUCTION

Since the 21st century, online social networks have become an indispensable part of people's lives with the popularity of mobile smart terminals. In this trend, social applications with functions such as entertainment, making friends, and sharing information have already become ideal platforms for people to publish and obtain information. In the process of information exchange and dissemination, some users gradually show the higher influence and become opinion leaders in social networks. These opinion leaders are distributed in different fields and have huge fan scales. The content they post is often easier to believe and spread faster. Their speech and the behavior of their fan groups have a profound impact on national politics, economy, information security and social stability. If managers do not guide them, it is likely to cause public opinion disasters.

User influence analysis and modeling is a very important part of current social network research. Previous research on user influence can be traced back to the 1950s. The development of emerging technologies such as data mining has provided technical support and new directions for the research of online social networks. Such as precision marketing and advertising, emergency warning, and detection [1, 2]. After reviewing the previous literature on user influence, we find that these studies have two main aspects: measurement and propagation. The research of user influence measurement focuses on calculating the influence of users on social platforms with a certain model or ranking the influence of users. The study of influence propagation includes aspects such as the mode of communication, the path of transmission, and the evolution of influence.

In this paper, we study the measurement of user influence in online social networks. In the current research, the methods of measuring the influence of users on social networks can be roughly divided into three categories:

1.1. Measurement Based on Network Topology (MNT)

Users will form a certain topology in social networks. The topology regards users as nodes and the relationships as connections. The random walk model can measure the nodes, which mainly includes the eigenvector method [3], Katz method [4], and PageRank method [5]. When scholars measure connections, they consider the extent of interaction among users (nodes). For instance, Gomez-Rodriguez et al. [6] assume a certain distribution in advance to calculate the influence.

1.2. Measurement Based on Users' Behavior (MUB)

This method considers the behavior of users on social platforms, including posting, reposting, liking, and following. Yang and Leskovec [7] designed an influence function by reviewing text among users and studied the connections between users and those affected by them.

1.3. Measurement Based on Interactive Information (MII)

This method is based on the interaction of users. It mainly studies the transfer and diffusion of interactive information among users. The topic factor graph (TFG) was proposed

by Tang et al. [8] to calculate topic similarity and distribution among users for impact analysis.

Comparing the above three methods, we find: MNT is simple and extensible to implement, but ignores the behavioral characteristics and interests of users, which cannot describe the influence of users objectively. MUB is more effective in predicting users' behavior and more in line with the characteristics of social networks. Because the users' behavior data is required, which are often considered as private, MUB is not convenient in practical applications. User interaction information in online social networks includes text, images, and videos. Extracting useful information from these unstructured data and applying it to influence measurement models is a challenge for MII.

The remaining of this paper is organized as follows: Section 2 explains the proposed method InfluenceRank. In Section 3, the experimental results and analytical discussions are described. Finally, Section 4 concludes all the work.

2. PROPOSED METHOD

PageRank is one of Google's algorithms for webpage search. It calculates the ranking of various websites based

on the links among webpages. Scholars have applied the PageRank algorithm to the field of social influence [9]. This method only depends on the following relationship among users and ignores users' behavior and interaction information. This study establishes a user influence index system, and introduces a weighting factor to the PageRank algorithm and a new measurement model, InfluenceRank, is obtained finally.

2.1. Influence Evaluation Index System

There are many online social platforms at home and abroad. Microblog (Chinese Twitter) is a popular application in China, with 203 million daily active users in 2019. The users with great influence on Microblog are called "big V". They possess many followers and have a considerable voice on the Internet.

To measure and evaluate the influence of users in Microblog scientifically, we construct an influence evaluation system through the analytic hierarchy process, as shown in Table 1.

Table 1 Index evaluation system of microblog user influence

First-level index	Second-level index
Coverage	Number of followers Number of followings Sphere of influence
Activity	Number of posts Number of reposts Number of likes
Interaction	Number of comments Number of retweets Number of likes
Credibility	Official certification Self-media certification Big V or not
Blog quality	Benign content

2.2. PageRank Model

To implement the PageRank algorithm, an initial matrix $S = (s_{i,j})$ needs to be constructed to describe the basic webpage link situation. $s_{i,j}$ represents the probability of jumping from web page j to web page i , as follows:

$$s_{i,j} = \begin{cases} \frac{1}{L(j)}, & < j, i > \in A \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$L(j)$ represents the number of webpages linked from page j . However, there are some "selfish" webpages, which

means one does not link to others. The balance of access among webpages will be disrupted if the matrix S is constructed only according to formula (1). Because users will stay on these "selfish" pages all the time, which is not in line with actual conditions. To solve this problem, the user is assumed to randomly jump to any page with probability α , then $s_{i,j}$ can be modified to $m_{i,j}$:

$$m_{i,j} = \alpha s_{i,j} + \frac{1-\alpha}{N} \quad (2)$$

In practice, α is usually set to 0.85. N is the total number of all web pages. Next we calculate the PR value (to measure the ranking of webpage), and the formula is as

follows. The process of calculating the PR value is also a Markov process, and this process is convergent.

$$P_{n+1} = AP_n \tag{3}$$

where $A = \alpha S + \frac{1-\alpha}{N} ee^T$, and e is the unit vector.

2.3. InfluenceRank Model

Generally, when using the PageRank model to evaluate user influence in social platforms, only the user's following relationship in Microblog is considered. The Microblog user is regarded as a webpage, the followers are regarded as linked out pages, and the following are regarded as linked in pages. However, in this case, users who have a lot of "zombie fans" without actual influence will be enhanced. "Zombie fans" means fake fans, normally users who are generated automatically by the system or registered maliciously. This section constructs the InfluenceRank model by improving the PageRank based on the influence indicator evaluation system in Section 2.1.

2.3.1. Activity

After obtaining the relationship graph of users, this study selects the number of posts of a user to calculate his or her activity. Generally, the more posts, the higher the user's exposure, and the easier it is for users to disseminate their views. User activity can be calculated by the following formula:

$$Activity(i) = \mu \times \frac{wb_i}{WB} \tag{4}$$

wb_i is the number of posts of user i . WB is the total number of posts in the social network we study. μ is a normalization factor.

2.3.2. Interaction

Interaction reflects how often a certain user communicates with others. If a user has a lot of followers and posts often, but his posts are rarely reposted or commented, then this user's influence is obviously insufficient. This study

utilizes three indicators: reposts, comments, and likes to design an interaction calculation formula:

$$Interaction(i) = \theta \times \left(\frac{nz_i}{NZ} + \frac{np_i}{NP} + \frac{nd_i}{ND} \right) \tag{5}$$

nz_i is the number of reposts of user i , and NZ is the sum of reposts in the social network. Similarly, np_i and nd_i are the number of comments and likes. θ is a normalization factor.

2.3.3. Credibility

Officially authenticated users are more credible in social platforms, such as schools, businesses, celebrities, etc. We use a simple discrete function to describe the credibility of Microblog users.

$$Credibility(i) = \begin{cases} 1.5, & \text{Official certification} \\ 1, & \text{Selfmedia certification} \\ 1, & \text{Authentication} \\ 0.5, & \text{Interest certification} \\ 0, & \text{Non-certified} \end{cases} \tag{6}$$

Then we calculate the influence weight w of Microblog users, according to the above three indicators:

$$w(i) = Activity(i) + Interaction(i) + Credibility(i) \tag{7}$$

According to formula 2, we modify the transition probability among users according to the influence weight as in formula 8, and iteratively calculate according to formula 3 until the user's PR value converges.

$$z_{i,j} = \begin{cases} m_{i,j} \times w(i), & j, i \in A \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

3. EXPERIMENT AND ANALYSIS

We obtain the relevant data of 130 celebrities in the e-sports field on Microblog, calculate their influence through PageRank and InfluenceRank, and ranked them. The results (top5 listed) are shown in Table 2 and Table 3 respectively.

Table 2 Results of PageRank

Ranking	Username	Num of followers	Num of followings	Num of posts	PR-value
1	pdd_liumou	5326744	99	150	0.000491
2	Royal Esports Club	2645857	121	4444	0.000475
3	Commentary Miller	2092620	194	2682	0.000415
4	UziQAQ	3125439	109	333	0.000349
5	Dsm comentary	2113188	97	358	0.000296

Table 3 Results of InfluenceRank

Ranking	Username	Num of followers	Num of followings	Num of posts	PR-value
1	15WGaming Center	290865	281	14710	2.18E+303
2	Royal Esports Club	2645857	121	4444	2.14E+303
3	Commentary Miller	2092620	194	2682	1.99E+303
4	DOTAGame Alliance	1801197	212	15543	1.47E+303
5	DOTA2_HOHO	713081	273	4811	1.06E+303

To describe the difference between the two algorithms intuitively, we visualize Table 2 and Table 3 as Figure 1 and Figure 2.

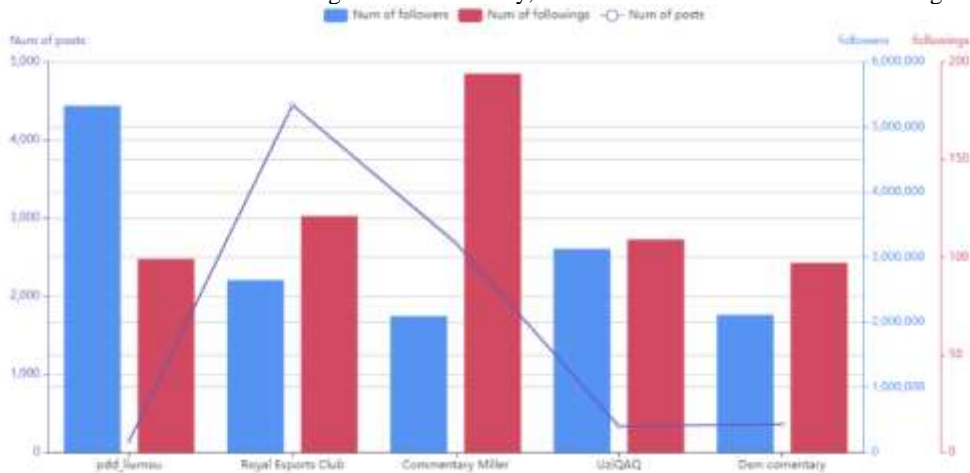


Figure 1 Results of PageRank

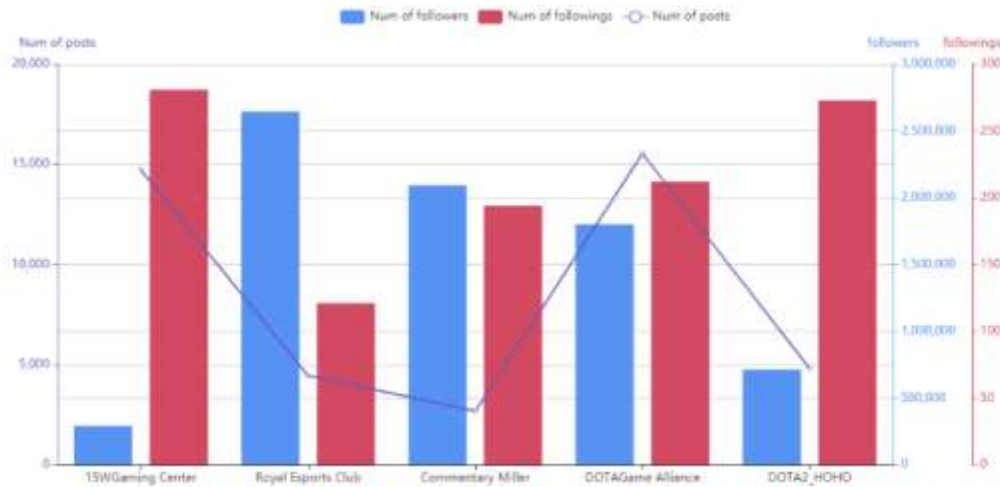


Figure 2 Results of InfluenceRank

In Figure 1 and Figure 2, we rank the top5 users of the two algorithms from left to right on the horizontal axis according to the influence. The blue bar, red bar, and purple polyline represent the number of fans, followers, and posts, respectively. Comparing the two figures, we draw the following conclusions:

1. The result of the PageRank algorithm screens out users with a huge fan base. The top five Big Vs have at least 2

million followers. According to Figure 1, the ranking has a clear correlation with the number of fans.

2. In Figure 1, although the number one user “pdd_liumou” has the most fans, he rarely posts. In actual data, he also rarely interacts with fans, indicating that the user is less active and interactive. It does not rule out that his followers are zombie fans. Therefore, he may not be the most influential in practice.

3. In the calculation of InfluenceRank, the influence ranking does not require that Big V has a considerable number of followers. Although the user "15WGaming Center" has a small group of fans, he posts frequently and has a good interaction with his fans, which has led him to rank first in overall influence.

4. CONCLUSION

This study introduces users' behavior and interaction information and improves the PageRank algorithm used to calculate user influence in online social platforms. Taking China's famous social platform "Microblog" as an example, we have established an influence evaluation index system for Microblog users. The three evaluation factors of activity, interaction, and credibility in this index system are designed and weighted. In combination with the PageRank model, an InfluenceRank measurement model is constructed. The experimental results show that the InfluenceRank model is more comprehensive and more in line with the actual situation in the calculation.

REFERENCES

- [1] Eesley, C. and Y. Wang, Social influence in career choice: Evidence from a randomized field experiment on entrepreneurial mentorship. *Research Policy*, 2017. 46(3): p. 636-650.
- [2] Sakaki, T., M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. in *Proceedings of the 19th international conference on World wide web*. 2010.
- [3] Bonacich, P., Some unique properties of eigenvector centrality. *Social networks*, 2007. 29(4): p. 555-564.
- [4] Katz, L., A new status index derived from sociometric analysis. *Psychometrika*, 1953. 18(1): p. 39-43.
- [5] Weng, J., et al. Twitterrank: finding topic-sensitive influential twitterers. in *Proceedings of the third ACM international conference on Web search and data mining*. 2010.
- [6] Gomez-Rodriguez, M., J. Leskovec, and A. Krause, Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2012. 5(4): p. 1-37.
- [7] Yang, J. and J. Leskovec. Modeling information diffusion in implicit networks. in *2010 IEEE International Conference on Data Mining*. 2010. IEEE.
- [8] Tang, J., et al. Social influence analysis in large-scale networks. in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009.
- [9] Luo, Z.-L., et al., A pagerank-based heuristic algorithm for influence maximization in the social network, in *Recent progress in data engineering and internet technology*. 2012, Springer. p. 485-490.