

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Weiboerrank: Rank the Influence of Users in the Microblogging Sites

Xiao Yu, Ruzhi Xu and Peiyao Nie

School of Computer Science and Technology, Shandong University of Finance and Economics, 250000, Jinan, China

---

**Abstract:** Nowadays, microblogging has become an important way to share and propagate information. With rapid expansion of information in the microblogging sites, how to identify influential users is of important theoretical value and practical significance. In this study, an algorithm WeiboerRank based on Page Rank is proposed to rank the influence of users in the microblogging sites. It effectively combines user's influence on the followers, the influence on others by the followers and the influence brought by reposting. Compared with the existing work, WeiboerRank can measure the influence of users more effectively.

**Key words:** Microblog, page rank, influence, repost

---

### INTRODUCTION

The microblogging service is a platform to share, spread and acquire information based on the relationships of users. By using it, people can post 140-character messages and share their information instantly through their computers, tablets and even smart phones. There is no doubt that it has become a publishing platform for breaking news and events and an important channel to share and spread information as well. Currently, most research related to microblogging is based on twitter which is the most influential microblogging service in the world. While in China, Sina Weibo is the biggest microblogging service with over 500 million registered users and 100 million posts every day. The mechanism of Sina Weibo is similar to twitter, so the research on twitter can be applied to Sina Weibo and vice versa. In this study, the algorithm is validated in Sina Weibo datasets and it can be used in twitter as well.

The main activities of users in the microblogging sites are to talk about their daily routines, share and seek information (Java *et al.*, 2007). The users can be divided into elite users and ordinary users. (Wu *et al.*, 2011) Usually, elite users are more influential than ordinary users. Although ordinary users only follow a few elite users, the posts of elite users can spread to the ordinary users with a significant impact on society. Therefore, it is of important theoretical value and practical significance to find influential users in microblogging sites so as to identify, spread and share important information.

To find influential users, how to evaluate a user's influence becomes the key factor. The word influence is defined in Longman dictionary as "the power to affect the

way someone or something develops, behaves, or thinks without using direct force or orders." In the microblogging sites, it seems that the greater the number of followers is, the more power to affect people the user has, the more influential the user is. Therefore, comparing the number of followers becomes a way to measure the influence of users (Kwak *et al.*, 2010; Cha *et al.*, 2010). However, users in the microblogging sites not only can influence his/her followers but also can influence others by the followers. Some scholars studied this factor and used PageRank (Larry *et al.*, 1998) to rank the influence of users. (Kwak *et al.*, 2010; Cha *et al.*, 2010) Nevertheless, the reposted microblogs reflect the influence of the user as well. Though some studies considered the reposting factor (Leavitt *et al.*, 2009; Tinati *et al.*, 2012; Welch *et al.*, 2011), they failed to combine these factors together.

In this study, we propose an algorithm Weiboer Rank to rank the influence of users in the microblogging sites. It not only considers user's influence on the followers but also makes a comprehensive study of the influence on others by his/her followers and the impact brought by the reposted microblogs. We validate WeiboerRank using data in Sina Weibo and get reasonable results.

The rest of the study is organized as follows. The related work is discussed in Section 2. Section 3 describes the main ideas of WeiboerRank and the algorithm itself. The dataset and the experimental analysis are presented in Section 4. Finally, in Section 6, we discuss the conclusion and future work.

**Related work:** There has been a lot of research on the influence of users in the microblogging sites. Kwak *et al.* (2010) ranked users by the number of followers, Page

Rank and retweets respectively to find that the corresponding results were different. Cha *et al.* (2010) measured users' influence by the number of followers, retweets and mentions and they found users with the largest number of fans did not necessarily score high using other measures.

Tinati *et al.* (2012) classified users into five categories: Idea starter, Amplifier, Curator, Commentator and Viewer. They built a model to find the key users in a specific topic. The results showed that although sometimes the number of followers of Idea Starter was less than that of Amplifier, due to the fact that Amplifier retweeted the posts of Idea Starter, the influence of the Idea Starter is greater than that of Amplifier. It indicated that the retweeting mechanism had some impact on the influence of users.

Welch *et al.* (2011) studied the semantics of following and retweeting and ranked the influence of users in the follower graph and retweeting graph but they did them separately. TunkRank made an improvement on Page Rank. It took into account of the probability of reposting. However, it neglected the fact that the probability of reposting is different as to different users and even to the same user, the probability of reposting the messages from different users is not the same. We deal with the problem in this study.

Web Ecology Project (Leavitt *et al.*, 2009) defined influence as "the potential of an action of a user to initiate a further action by another user." And they studied the related factors such as retweets, mentions and replies. Bakshy *et al.* (2011) considered the influence as user's ability to post URLs which diffuse through the Twitter follower graph. They measured the impact of users by the attributes and the past behaviors of the users. Weng *et al.*, (2010) proposed TwitterRank based on PageRank and studied how to find topic-sensitive influential twitters. While the above studies measured users' influence from users' actions, the ability to diffuse URL and topics perspectives, we combine the user's influence on the followers and the influence on others by the followers and the impact brought by reposted microblogs effectively.

## COMPUTING THE INFLUENCE OF USERS

### Terminology:

- **Follow, followee and follower:** Following is a unidirectional relationship without the need of confirming by the user. You can follow anyone as long as you are willing to receive his or her update of status

- Followee is someone you follow
- Follower is someone who follows you
- **Repost or retweet:** When it is a verb, it means the behavior that people copy a valuable message someone posts as his own so that all his followers can see it. When used as a noun, it means that valuable message someone reposts. The repost is marked with "@user id"

**The main ideas of weiboer rank:** In this study, the influence of users is composed of two parts: Follower Influence and Repost Influence. Follower Influence is the influence on the followers and the influence on others brought by the followers, while Repost Influence is the influence when the user's microblogs are reposted.

### FOLLOWER INFLUENCE

**Follower Influence is related to two factors:** The number of the followers and the influence of the followers themselves. The greater the number of followers the user have, the more important the user is. Each follower of the user has different Follower Influence. The higher Follower Influence the followers have, the higher Follower Influence the user has. Since a user can follow many users, so the Follower Influence of a user should be shared between all the users he/she follows.

### REPOST INFLUENCE

If user A reposts a message from user B, then user A increases the Repost Influence of user B. Each one who reposts messages from the same user has different Repost Influence. Since a user can repost messages from many users, so the Repost Influence of the user should be shared between the users who he/she reposts messages from. The higher the frequency user A reposts messages from user B, the bigger the contribution user A makes to improve the Repost Influence of user B.

**WeiboerRank algorithm:** The influence of user *i* is composed of Follower Influence and Repost Influence. To any user *j* in the microblogging network, if *j* follows *i*, then *j* contributes some Follower Influence to *i*; if *j* reposts messages from *i*, then *j* contributes some Repost Influence to *i*.

The simple model of the user's influence in microblogging sites is defined as:

$$u(i) = \sum_{j=1}^n (FI_j * q + RI_j * (1 - q)) * u(j) \quad (1)$$

In the equation above,  $q$  is the impact factor to indicate the proportion of the Follower Influence, while  $1-q$  is the proportion of the Repost Influence. To each user  $j$ , the Follower Influence user  $j$  contributes to user  $i$  is  $FI_j * q * u(j)$  and the Repost Influence  $j$  contributes to  $i$  is:

$$RI_j * (1-q) * u(j)$$

$FI_j$  is the Follower Influence coefficient to indicate the ratio of the Follower Influence  $j$  contributes to  $i$  to the Follower Influence  $j$  has.

$$FI_j = \begin{cases} \frac{1}{F_j} & \text{if } F_j > 0 \text{ and } j \text{ follows } i \\ 0 & \text{else} \end{cases} \quad (2)$$

Where  $F_j$  is the number of followers user  $j$  has.  $RI_j$  is the Repost Influence coefficient to indicate the ratio of the Repost Influence  $j$  contributes to  $i$  to the Repost Influence  $j$  has:

$$RI_j = \begin{cases} \frac{r_{ji}}{\sum_{k=1}^n r_{jk}} & \text{if } \sum_{k=1}^n r_{jk} > 10 \\ 0 & \text{else} \end{cases} \quad (3)$$

Where  $r_{ji}$  is the number of messages that  $j$  reposts from  $i$ ,  $\sum_{k=1}^n r_{jk}$  is the total number of the messages  $j$  reposts. However, the simple model above fails to consider the fact that the user can follow someone randomly or repost a message randomly. In the same way as PageRank deals with the similar problem, we introduce the random walk model to modify the simple model.

The modified model of the user's influence in microblogging sites is defined as:

$$u(i) = (1-d) + d \sum_{j=1}^n (FI_j * q + RI_j * (1-q)) u(j) \quad (4)$$

**D is the dump factor:** We can use matrixes to represent the equations. Let  $U$  be an  $n$ -dimensional column vector of Weiboer Rank values, that is:

$$U = (u(1), u(2), \dots, u(n))^T$$

Let  $FI$  be the matrix of the Follower Influence coefficients, where  $FI_{ij}$  is defined in Eq. 2.

Let  $RI$  be the matrix of the Repost Influence coefficients, where  $RI_{ij}$  is defined in Eq. 3.

Let  $A$  be the matrix of influence coefficients, where  $A_{ij}$  is defined in Eq. 5.

$$A_{ij} = RI_{ij} * q + FI_{ij} * (1-q) \quad (5)$$

---

Algorithm weiboer rank

```

Compute the matrix fi
Compute the matrix RI
A ← FI*q + FI*(1-q)
U0:
m ← 1
Repeat :
Um ← (1-d)*e + dATUm-1
m ← m+1
Until:
||Um - Um-1|| < ε
Return Um
    
```

---

Fig. 1: Weiboer rank algorithm

Then WeiboerRank can be expressed as:

$$U = (1-d)e + dA^T U \quad (6)$$

The algorithm is shown in Fig. 1

## EXPERIMENTAL ANALYSIS

**Case study:** There are three users in the graph. The straight arrows indicate the following relationship of the users. If the arrow is pointed to user  $j$  from user  $i$ , it means that  $i$  follows  $j$ . The curved arrows indicate the reposting relationship. If the curved arrow is pointed to user  $j$  from user  $i$  and the number on the curved arrow is  $k$ , then it means  $i$  reposts messages from  $j$   $k$  times. For example, in Fig. 2, Wang, Li, Zhang follows each other. And Li reposted 5 posts from Wang and reposted 3 posts from Zhang.

Using the number of the followers, the influence of Wang, Zhang, Li are the same. Using PageRank, the influence of Wang, Zhang, Li are the same, too. Using WeiboerRank, in the order of Wang, Zhang, Li, the matrix of Follower Influence coefficients is:

The matrix of the Repost Influence coefficients is:

Let  $q = 0.5$ , the initial value of  $U$  is. Let the threshold be 0.0001 and the result of  $U$  is:

$$\begin{pmatrix} 0.3693 \\ 0.3183 \\ 0.3123 \end{pmatrix}$$

The ranking result (from high to low) is Wang, Zhang, Li.

In Fig. 2, the users follow each other and Wang's messages are reposted most frequently. Follow the intuition, we can conclude that Wang's influence should

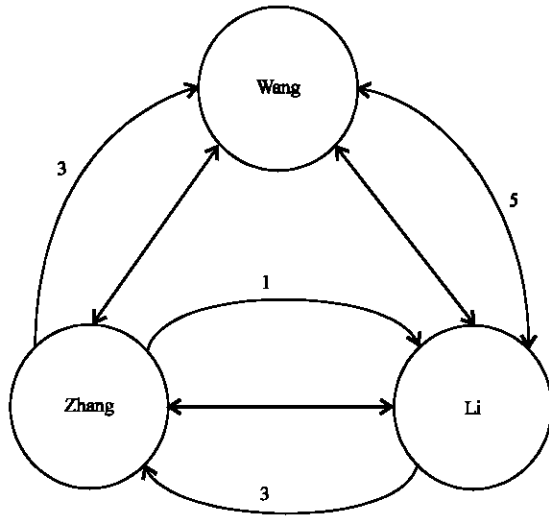


Fig. 2: Relationships between users

be the biggest, Zhang is second and Li is the last. The intuition is in line with the result of Weiboer Rank.

**Data set:** Using Sina Weibo as the data source, we crawled users’ profile, the microblogs users posted and the following relationship between users through Sina API.

Focusing on Kai-Fu Lee, we crawled all the users he followed in Sina Weibo and stored them in S.  $|S| = 450$ . To each user in S, we crawled all the posts he/she posted from January 1st 2013 to February 27th 2013 and saved them in T.  $|T| = 236, 920$ .

**Comparison of weiboer rank and other algorithms:** We rank the influence of the users in the dataset according to the number of followers they have in the dataset, PageRank and Weiboer Rank separately. The top 10 are shown in Table 1.

From Table 1, we can see that the results of the three algorithms are not the same. Take “People’s Daily” for example, it doesn’t show up in top ten when ranking by the number of followers or PageRank. However, it ranks second when ranking by WeiboerRank. The reason is that the three algorithms take different factors into account, Although People’s Daily ranks 56th out of 450 by the number of followers in the dataset, it ranks 21st by PageRank because the followers of People’s Daily are influential. Moreover, because its microblogs are reposted quite frequently, it ranks second by Weiboer Rank. The results show that Weiboer Rank is more reasonable. In the three ranking methods, ranking by the number of followers only considers the influence on the followers and ranking by Page Rank considers both the influence

Table 1: Ranking results of three different algorithms

	Number of followers	Page rank	Weiboer rank
1	Kai-fu lee	Kai-fu lee	Kai-fu lee
2	Chen lao	Chen lao	People’s daily
3	Manzi xue	Manzi xue	Top news
4	Xiaoping xu	Top news	Financial network
5	Hongwei zhou	Xiaoping xu	Sina technology
6	Ran wang	Ran wang	Micro-world
7	Jun lei	Shiyi pan	Sina video
8	Zhaohui wang	Sina technology	Manzi xue
9	Wensheng cai	Guowei cao	Excise book
10	Zhangying jingwei	Zhiqiang ren	Chenpeng li

on the followers and the influence on others by followers, while ranking by Weiboer Rank takes into account of the influence on the followers, the influence on others by followers and the influence brought by reposting. Therefore, compared with ranking by the number of followers and Page Rank, WeiboerRank is more comprehensive and reasonable.

### CONCLUSION AND FUTURE WORK

In this study, we propose an algorithm Weiboer Rank to rank the influence of users in the microblogging sites. Besides the influence on his/her followers, the algorithm also takes into account of the impact on others by the followers and the influence brought by reposting. Compared with the existing work, the experimental results of WeiboerRank are more reasonable.

However, the experiments still need improvements. The data set is small in the study and we will improve and validate the algorithm in the large-scale data sets in the future.

### ACKNOWLEDGMENT

This work was supported by The Ministry of education of Humanities and Social Science project(10JD710006) and Natural Science Foundation of Shandong Province (Project No. ZR2010FL008).

### REFERENCES

Bakshy, E., J.M. Hofman, W.A. Mason and D.J. Watts, 2011. Everyone’s an influencer: Quantifying influence on twitter. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, February 9-12, 2011, Kowloon, Hong Kong, pp: 65-74.

Cha, M., H. Haddadi, F. Benevenuto and P.K. Gummadi, 2010. Measuring user influence in twitter: The million follower fallacy. Proceedings of the 4th AAAI International Conference on Weblogs and Social Media, May 23-26, 2010, Washington, DC., USA., pp: 10-17.

- Java, A., X. Song, T. Finin and B. Tseng, 2007. Why we twitter: Understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD Workshop on Web Mining and Social Network Analysis, August 12-15, 2007, San Jose, CA., USA., pp: 56-65.
- Kwak, H., C. Lee, H. Park and S. Moon, 2010. What is Twitter, a social network or a news media? Proceedings of the 19th International Conference on World Wide Web, April 26-30, 2010, Raleigh, NC., USA., pp: 591-600.
- Larry, P., S. Brin, R. Motwani and T. Winograd, 1998. The pagerank citation ranking: Bring order to the web. Technical Report, Stanford Digital Library Technologies Project, USA.
- Leavitt, A., B. Evan, D. Fisher and S. Gilbert, 2009. The influentials: New approaches for analyzing influence on twitter. <http://www.webecologyproject.org/wp-content/uploads/2009/09/influence-report-final.pdf>.
- Tinati, R., L. Carr, W. Hall and J. Bentwood, 2012. Identifying communicator roles in twitter. Proceedings of the 21st International Conference Companion on World Wide Web, April 16-20, 2012, Lyon, France, pp: 1161-1168.
- Welch, M.J., U. Schonfeld, D. He and J. Cho, 2011. Topical semantics of twitter links. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, February 9-12, 2011, Kowloon, Hong Kong, pp: 327-336.
- Weng, J., E.P. Lim and J. Jiang, 2010. TwitterRank: Finding topic-sensitive influential twitterers. Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, February 3-6, 2010, New York, USA., pp: 261-270.
- Wu, S., J.M. Hofman, W.A. Mason and D.J. Watts, 2011. Who says what to whom on twitter. Proceedings of the 20th International Conference on World Wide Web, March 28-April 1, 2011, Hyderabad, India, pp: 705-714.