# A Modified Method for User Authority Ranking in Chinese Microblog

JingXin

School of Economics and Management

Tongji University

Shanghai, China

**Keywords:** algorithms; social networking; web services; information management; internet; Chinese microblog; user authority ranking

**Abstract.** Nowadays, microblog has emerged as the most promising social networking service in the Web2.0 age with diverse features for information dissemination, interpersonal communication and many other aspects. Chinese microblog shows great potential owing to the rich marketing source and powerful media influence brought by the huge user population. As the microblog community is largely influenced by some authoritative users, identifying these users and understanding the information dissemination pattern are theoretically and practically significant for us to promote user services, develop further applications and seek commerce opportunities. In this paper, we proposed a modified method for user authority ranking. A user-message graph modified according to Chinese microblog was applied to our study, and an algorithm based on ObjectRank was used to calculate ueser authority scores. We analyzed Sina microblog, as it was the most widely adopted Chinese microblog and more resources for research was accessible.

## Introduction

In the past decade, microblog services have emerged as the most promising type of web services with considerable influence all over the world. Similar to traditional blogs and social network services, a microblog service works as a self-motivated media and an open communication platform. In microblog, the length of a message a user can send is limited to 140 characters, and the content of the information transmitted includes news, conversational, self-promotion, pointless babble, pass-along value, spam and so on [1]. The convenience of mobile Internet enables people share their opinions and experiences in real time. People addict to microblog as they receive the freshest news and keep in touch with family and friends.

In 2007, Fanfou website appeared online, and his founder, WangXing first introduced microblog to Chinese users [2]. Sina microblog, which has over 300 million users,is the leading microblog service in China. Though quite similar to Twitter in functions, Sina microblog has its own features for native users. Using a social network to propagate information can significantly affect customer decision. making [3,4]. In In addition, some public emergency was first spotted in microblog and useful information spread rapidly to rescue and avoid further losses.

In Sina microblog, users may play different roles, such as an information source, asocial relationship builder, an information seeker, or the one combining the intentions above [5]. Among all the users, some verified users such as celebrities, respected companies and trusted news papers come to the core of the social network as authorities and show great influence on their followers and in personal social community. As information dissemination is largely influenced by the core users and their behaviors, identifying these users and understanding the information dissemination pattern are important for us to collect useful information, promote user services, develop further applications and seek commerce opportunities in microblog.

## Research Foundation

**Overview of Sina Microblog.** Akin to a hybrid of Twitter, Sina microblog is the most popular Chinese microblog and take 90% of the market [6]. In Sina microblog, messages are called weibos, and the maximum length of them is limited at most 140 characters. Users can casually post weibos to

express them even if these weibos are not so meaningful or informative.A user can follow other users depending on his interest. He may follow his friends, a fashion magazine, even somebody he can't access to in reality, such as Kobe Bryant. Once user accounts are followed by a user, all weibos posted by the accounts are then displayed to the follower. Unlike traditional social network services, Sina microblog allows users to follow another account without any permission, which wins it an enormous number of users.

There are two important behaviors about the messages, "comment" and "repost". Users can comment and repost weibos , and both behaviors can reflect the influence of the original user to some degree. Compared to "comment","repost" contributes more to the dissemination of information and the prosperity of microblog. A reposted weibo chain can be regarded as a flow of information originated from a user to his followers and their descendant followers. In Sina microblog, another important behavior is "mention" A user can mention any other user account when they post, comment or repost. When a user is mentioned, he gets an"@", and a link directed to the content he is mentioned. The amount of mentioned reflects a user's popularity and influence. The more "@"a user gets the higher authority scores he gets.

**Related Work.** PageRank ,which is proposed by Larry Page and Sergey Brin in the late 1990s, is a cornerstone element of the Google search engine [7]. The principle of PageRank is ranking the authority of pages according to hyperlinks structures [8]. If the page v links to the page u, we consider v votes for u. We can calculate the authority score according to Eq.1.

$$R(u) = C \cdot \sum_{v \in B(u)} R(v)/N(v) \tag{1}$$

PR value is a page's score calculated according to PageRank. R(v) and R(u) respectively represents the PR value of page v and page u. B(u) represents the amount of all backward chaining ,while N(v) represents the amount of all forward chaining. Moreover, to make the PR value a constant, C play the role of scaling factors. We can approximate user authority in microblog applying PageRank algorithm by calculating the PR values of users and the amount of hyperlinks.

ObjectRank is an extension of PageRank. It performs link structure analysis over linked objects for measuring the importance of objects in the database. Unlike PageRank, it takes account of edge types and node types in order to deal with multiple kinds of edges and nodes. In ObjectRank, we differentiate each edge by setting an appropriate weight to control the flow of scores going through the edge [9]. In ObjectRank,, an authority transfer data graph is often used. Link structure analysis is applied to this graph in such a way that, for each node, the score is propagated along the outgoing edges, and the propagated scores are computed by the weighted average based on the edge weights [10]. To measure the user authority in microblog, PageRank brings damage in accuracy, while ObjectRank pays more attention to the score propagated along user behaviors.

## User Authority Ranking

**User-Weibo Graph.** To construct a user-weibo graph, we define a user-weibo schema graph, as illustrated in Fig.1. It corresponds to the authority transfer schema graph in ObjectRank. A user-weibo schema graph UWGS = (VS, ES) defines the structure and edge weights of a user-weibo graph [9]. Here, VS is the node set including user nodes and weibo nodes, and ES is the edge set including post, posted, follow, mention, comment and repost edges. Notice that the edge weights can sometimes be zero if the edge makes no difference on the target it directs to. In Fig.1, the zero-weighted edges are omitted.
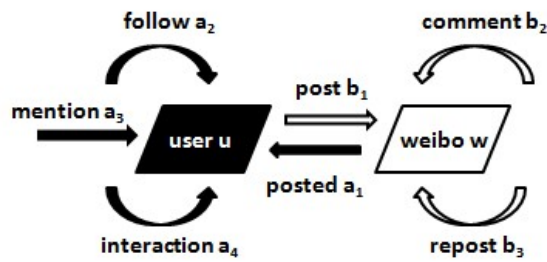
follow a$_2$    comment b$_2$

mention a$_3$    post b$_1$

user u    weibo w

posted a$_1$

interaction a$_4$    repost b$_3$

Figure1. A user-weibo schema graph

Here, $V$ is the node set which contains all weibo nodes and user nodes in obtained data. $E$ is the edge set which contains all existing edges in obtained data, namely, post, posted, follow, followed, and so on. The weight $w(e)$ set to the edge $e \in E$ from node $u \in V$ is calculated by Eq.2.

$$w(e) = \frac{w(e_s)}{Out\ Deg(u, e_s)} \qquad (2)$$

In the case of a user-weibo graph, we can capture the weibo chains appropriately. For example, as illustrated in Fig.2. When weibo W2 reposts W3, and W1 reposts W2, the score of W1 affects W3 through w2, both W3 and W2 gets the reposted score q delivered over this weibo. In case of the behavior of follow, user U3 follows U1 and U2, both the followed users get $p_3$ from U3. If U2 follows U1, U gets the accumulated score, $p_2+p_3$, as the authority of U2 is promoted by U3's following.   By a user-weibo graph, we can calculate scores of both users and weibos concurrently.



post

U1 ⟷ W1    q

posted

$p_2+p_3$

$p_3$

U2 ⟷ W2    repost

$p_3$

U3 ⟷ W3    q

follow
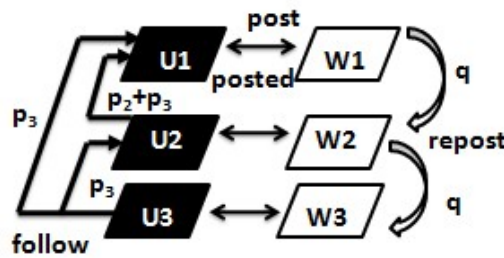
**Calculation.** Based on random surfer model, we calculate a user's authority score by applying Eq.3, which works in the same way of ObjectRank algorithm. Here, r is the ObjectRank score vector, d is the probability of random jump, and A is the transition matrix [2].

$$r = dAr + \frac{(1-d)}{|v|} e \qquad (3)$$

Having obtained a user-weibo graph, we calculate the scores of each user for subsequent ranking according to the detailed algorithm shown in Fig.3, the UARank (User Authority Rank). Here, the element $a_{ij}$ of the transition matrix A is the weight of the edge from node i to node j if it exists, otherwise, $a_{ij}$ is 0.

UARank
$r^0 \leftarrow [1, \cdots, 1]$
$\alpha \leftarrow 0$
Repeat
 $\alpha \leftarrow \alpha + 1$
 foreach $r_i^\alpha \in r^\alpha$
  $r_i^\alpha \leftarrow \sum_{e=(j,i)\in E} w(e)r_j^{\alpha-1} + (1-d)/|V|$
 end
 $r^\alpha \leftarrow r^\alpha / \|r^\alpha\|_1$
until $\|r^\alpha - r^{\alpha-1}\|_1 < \epsilon$
return $r^\alpha$
end

Figure3.   UARank algorithm

The score of node i in step α is calculated by summing up scores of all nodes which have the edge to i in step $\alpha - 1$ and scores by random jump. This calculation is iterated until all scores are converged, where convergence threshold ε is set to be sufficiently small.

## Conclusion

In this paper, we proposed a modified method to rank user authority in Chinese microblog. First, we modeled the interaction process of users and weibos by a user-weibo graph. Second, we took account for the influence of all kinds of user behaviors, including Chinese microblog's exclusive features. Finally, we created an integral scoring system on the basis of detailed definitions. To perfect the algorithm, observation of large amount of experiment data is necessary. Despite the limitations, our work contributes to the promotion of Chinese microblog and provides a foundation for further study in related issues.

## Reference

[1] YanChao-Zhang, YunLiu, HuiCheng, Fei Xiong, ChangLun-Zhang,A Method of Measuring User Influence in MicroBlog, Journal of Convergence Information Technology. Gyeongju, vol. 6, no.10, pp. 243–250, 2011.

[2] Li Jun, Chen Zhen, Huang Ji-wei, Micro-blog Impact Evaluation Study, Netinfo Security.Shanghai, vol. 3, no.10, pp. 10-13, 2012.

[3] Y.Kim, J.Srivastava, Impact of Social Influence in E-commerce Decision Making,in Proceedings of The Ninth International Conference on Electronic Commerce,ACM , pp. 233-242,2008.

[4] M.Fink, J.Spoerhase,Maximum betweenness Centrality:Approximability and Tractable Cases, WALCOM: Algorithms and Computation,pp. 9-20,2011.

[5] Yubao Zhang, Hui Wang, Zhihong Jiang, and Pei Li, Characterizing top user behavior of Sina microblog in the context of different events,in Proceedings of International Conference on Computing, Networking and Communications, pp. 451–455, 2012.

[6] K M Frahm, B Georgeot and D L Shepelyansky ,Universal emergence of PageRank, Journal of Physics A: Mathematical and Theoretical. UK, vol.44, 2011.

[7] Page.L., Brin.S., Motwani. R., Winograd.T, The PageRank Citation Ranking: Bringing Order to the Web, Technical report, Stanford Digital Library Technologies Project , 1998.

[8] Boyd.D, Golder.S, Lotan.G, Tweet, tweet, retweet: Conversational aspects of retweeting on twitter, in Proceedings of 43rd Annual Hawaii International Conference on System Sciences, 2010.

[9] Balmin.A, Hristidis.V, Papakonstantinou.Y, Objectrank:Authority-based key-word search in databases, in Proceeding of Very Large Data Base, 2004.

[10] Yamaguchi.Y, Takahashi.T, Amagasa.T, Kitagawa.H, TURank: Twitter User Ranking Based on User-Tweet Graph Analysis,in Proceedings of Web Information Systems Engineering, pp.240-253, 2010.