

Association for Information Systems AIS Electronic Library (AISeL)

PACIS 2011 Proceedings

Pacific Asia Conference on Information Systems
(PACIS)

9 July 2011

Identifying Influential Users Of Micro-Blogging Services: A Dynamic Action-Based Network Approach

Meng Zhang

Renmin University of China, meng.zhang.ruc@gmail.com

Caihong Sun

Renmin University of China, chsun@ruc.edu.cn

Wenhui Liu

Renmin University of China, liu@163.com

ISBN: [978-1-86435-644-1]; Full paper

Recommended Citation

Zhang, Meng; Sun, Caihong; and Liu, Wenhui, "Identifying Influential Users Of Micro-Blogging Services: A Dynamic Action-Based Network Approach" (2011). *PACIS 2011 Proceedings*. 223.

<http://aisel.aisnet.org/pacis2011/223>

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2011 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

IDENTIFYING INFLUENTIAL USERS OF MICRO-BLOGGING SERVICES: A DYNAMIC ACTION-BASED NETWORK APPROACH

Meng Zhang, School of Information, Renmin University of China, Beijing, China,
meng.zhang.ruc@gmail.com

Caihong Sun, School of Information, Renmin University of China, Beijing, China,
chsun@ruc.edu.cn

Wenhui Liu, School of Information, Renmin University of China, Beijing, China,
wenhui-liu@163.com

Abstract

In this paper, we present a dynamic model to identify influential users of micro-blogging services. Micro-blogging services, such as Twitter, allow their users (twitterers) to publish tweets and choose to follow other users to receive tweets. Previous work on user influence on Twitter, concerns more on following link structure and the contents user published, seldom emphasizes the importance of interactions among users. We argue that, by emphasizing on user actions in micro-blogging platform, user influence could be measured more accurately. Since micro-blogging is a powerful social media and communication platform, identifying influential users according to user interactions has more practical meanings, e.g., advertisers may concern how many actions – buying, in this scenario – the influential users could initiate rather than how many advertisements they spread. By introducing the idea of PageRank algorithm, innovatively, we propose our model using action-based network which could capture the ability of influential users when they interacting with micro-blogging platform. Taking the evolving prosperity of micro-blogging into consideration, we extend our action-based user influence model into a dynamic one, which could distinguish influential users in different time periods. Simulation results demonstrate that our models could support and give reasonable explanations for the scenarios that we considered.

Keywords: User Influence, Micro-blogging platform, Action-based Network, Dynamic Model.

1 INTRODUCTION

Identifying influential persons in the society has been studied in the fields of sociology, communication, marketing and political science for a long time. Influential persons usually play a vital role in both political and economic world. For instance, advertisers could use these influential persons to apply virus marketing strategy for their products and politicians may want to get more supports through opinions of these influential persons. Identifying influential persons can help us better understand why certain opinions, trends or innovations are adopted and diffused faster than others and how we could help advertisers, politicians, marketers, and government design more effective campaigns and policies (Cha et al. 2010).

Sina micro-blog, a Twitter-like online service, is one of the most notable and widely used micro-blogging services in China. It allows twitterers (users) to publish tweets (with a limit of 140 characters) and build their social networks. Unlike most other social network services that require users to get permissions for befriending others, each twitterer of micro-blogging service is allowed to choose who he wants to follow without seeking any permission. We called this kind of social relationship "following". In fact, micro-blog service emerges as a social media more than a social network. Since people mainly publish their opinions other than making friends or joining social activities (Java et al. 2007). On January 25, 2011, Yu Zhengrong, a professor from Chinese Academy of Social Sciences, published a tweet to call for taking photos of those begging children in street, to help find abducted kids. This tweet was widely retweeted and spread quickly through Sina micro-blog. Then it was reported by newspapers and televisions, and aroused great social and governmental concerns on anti-abduction movements. Here, micro-blogging plays a significant role of general public in counter-abduction campaign. The 140-character micro-blog, channeled under the slogan of "taking, tweeting, and testifying", has become so powerful when so many citizens joined in, and it becomes a popular and important social medium.

In this paper, we present a model to measure user influence in a popular news media: micro-blogging services, from the view of information diffusion. For simplicity, we use the same terms of Twitter to describe micro-blogging services, and use Twitter as a representative micro-blogging service. The user who is following others is called "follower", while the user who is being followed is called "followee". Similarly, if one is following another person, and another person follows back, we call this kind of relationship "friend" relationship, and for each of them, another person is called "friend". By publishing tweets, a twitterer can broadcast his updates to all his followers.

A popular metric of user influence on Twitter is to measure the number of a user's followers (Leavitt et al. 2009). This approach has a basic assumption that the more followers that a user has, the more popular the user is. This seems reasonable intuitively. Nonetheless, it only considers one-step connections among users, ignores contents, link structure, and interactions among users (Leavitt et al. 2009). Another similar and popular user influence measure involves the ratio between the number of a user's followers and the number of other people the user follows. Although better than the method of counting followers only, the ratio approach is still imprecise, it ignores the ability of a user to interact with contents on the micro-blogging platform (Leavitt et al. 2009).

Cha et al. (2010) present three different types of user influence: indegree influence, i.e. counting the number of followers of a user; retweet influence, which measures the number of retweets containing one's name, indicating a user's ability to generate content with pass-along value; mention influence, which measures the number of mentions containing one's name, indicating a user's ability to engage others in a conversation (Cha et al. 2010). The latter two influence measures do consider the interactions of users from contents and conversation respectively. But they only count the total number of retweets or mentions a user has, not considering the information flow network among users, i.e. the link structure among users. For example, a user may have more influence if his tweet is retweeted by those influential users than by users who have less influential power.

PageRank algorithm is a widely used influence measure which can consider link structure. Weng et al. (2010) propose a topic-sensitive influence measure, in which they make an assumption that Twitter has high reciprocity in following relationships. TwitterRank, a PageRank alike algorithm, uses following link structure and weighted topic similarity to define transition probability matrix. This measure considers both the link structure and content, but ignores interactions among users. Besides that, the study of Kwak et al. (2010) shows that Twitter has a low level of reciprocity, which contradicts Weng's findings.

Considering Twitter both as a news media and a social network, the interactions among users and contents are two important factors to identify who are the most influential people. In this paper, we propose an action-based network approach to measure user influence. We distinguish user influence by the actions of micro-blogging users. We focus on two kinds of actions — retweets and replies. The purpose of retweets is to push content, while the main purpose of replies is for conversations or making comments. These actions could demonstrate values of influence in information diffusion. Since retweet and reply actions reflect different relationship among users, we build "action-based network" based on retweet or reply relationship respectively. For instance, in a retweet network, if user A retweets user B's tweets, then there is a link from A to B. Based on PageRank algorithm, an action-based user influence model is proposed, which considers both action network structure and the number of interactions among users. Since Twitter is an ecological system in which users' relationship and their environments are changing, in this paper, we also present a dynamic action-based user influence model.

In this paper, we have three main contributions. First, we propose an action-based network approach to identify influential users in micro-blogging platform. Considering micro-blogging as a news media, interactions between users and micro-blogging platform play a very important role in information diffusion. By emphasizing users' actions, our user influence model considers both user interactions and contents; by incorporating PageRank idea, our user influence model takes information flow path into account. Second, micro-blogging is an evolving system, user influence could change constantly. A dynamic user influence model is proposed in this paper to capture the ecology of user influence. Finally, we explore a potential framework to hybrid different user influence measures together.

The rest of this paper is organized as follows: Section 2 gives the definitions of user influence; Section 3 presents our proposed action-based user influence model; Section 4 describes a dynamic user influence model; Section 5 presents the design of experiments; Section 6 demonstrates the results of experiments and talks about related works; Section 7 draws the conclusions and Section 8 describes our future work.

2 DEFINING OUR USER INFLUENCE ON MICRO-BLOGGING SYSTEMS

Influence has been illustrated differently from different aspects because it is difficult to find a universal definition under all conditions. We have mentioned several types of user influence above. In this paper, we adopt the idea of user influence defined by Web Ecology Project (Leavitt et al. 2009):

“We define influence on Twitter as the potential of an action of a user to initiate a further action by another user. The term user is defined by Twitter's platform. The term action deserves further explanation.”

To identify user influence, we define two kinds of user actions intrinsic to micro-blogging: the retweet and the reply. Both actions are meant to pass content to other users in different ways. If a user responds to another user's tweet (reply action), it implies that the user is influenced by the tweet. If a user cited or paraphrase of another user's content (retweet action), it implies that the user is influenced to reproduce the content. But not all retweet actions (or reply actions) have equally importance. Since Twitter is not just a news media, but

a platform including content, user networking and interactions, actions could have networking effects on Twitter.

- **Definition 1** Content-based influence: The potential to make others initiate a content-based action. Specifically, content-based action refers to retweet action.
- **Definition 2** Conversation-based influence: The potential to make others initiate a conversation-based action. Specifically, conversation-based action refers to reply action.

How to measure the potential of an action? It should not just count the number of actions of each user, but it should consider the networking effects of interactions. The following section describes an action-based user influence model which demonstrates our definitions of influence.

3 AN ACTION-BASED USER INFLUENCE MODEL

Our definition of user influence in Section 2 is strongly user action related. Therefore, instead of using following relationship network, we construct action-based network to model the interactions of users. Each action, such as retweet or reply, represents a one-way link between users. For example, if user A retweet user B's tweet, then a directed link is formed from A to B. In this way, we can construct retweet network and reply network. Therefore content-based influence and conversation-based influence would be calculated via retweet network and reply network respectively. To the most of our knowledge, we are the first one who uses action-based network to measure user influence on micro-blogging services.

How to measure the potential of an action? Intuitively thinking, the potential of an action would be big if it interacts with a more influential user. In our model, we make an assumption that if a user is influential, then the potential of an action of this user is big. To quantify how much potential an action of a user, we introduce the idea of PageRank. In PageRank algorithm, the basic idea is that if one page is important, then another page that this page pointed to is also important. Through iterative computing, the importance scores diffuse among pages. Likewise, our model uses the same approach of PageRank algorithm but choose action-based network instead.

3.1 PageRank Algorithm

The basic idea of PageRank is as follows: If page u has a link to page v , then the author of u is implicitly conferring some importance to page v . For instance, *Yahoo!* is an important page, reflected by the fact that many pages point to it. Likewise, pages that *Yahoo!* pointed to are probably important (Haveliwala 2003). The following is a brief introduction of PageRank, the details of PageRank could refer to *The PageRank Citation Ranking: Bringing Order to the Web*.

Let N be the number of pages and, initially, every page has the same rank score $1/N$, denoted as $\overrightarrow{Rank}^{(1)}$ for page p . Now, imagine that there is a random surfer moves randomly from one node to another in the network. In each time of its movement, it brings the importance score from one node to another and the original score of node will not change. For a large number of times of iteration, the process can be expressed as the following calculation process:

Let M be the square, stochastic matrix corresponding to the directed web graph G . If there is a link from page j to page i , then let the matrix entry $m_{i,j}$ have the value $1/N_j$. Let all other entries have the value 0. Then, compute the rank vector repeatedly as:

$$\overrightarrow{Rank}^{(i+1)} = M \times \overrightarrow{Rank}^{(i)} \quad (1)$$

To guarantee the convergence of PageRank, the transition matrix M must be irreducible (Motwani & Raghavan 1995). We add a complete set of outgoing edges to all the nodes that

there is no outgoing edges before. This will not change the relative importance of other pages. Therefore, M is replaced by M' . The matrix entry d_{ij} of D is denoted as:

$$d_{ij} = \begin{cases} \frac{1}{n-1} & \text{if outdegree of } j \text{ is } 0. \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Then,

$$M' = M + D \quad (3)$$

It is possible that some users would "follow" one and another in a loop without "following" other users outside the loop and there are other users following some of them. Iteratively, other users' influence will "flow" into their loop and they will accumulate high influence without distribute their influence. This is called "rank sinks" (Page et al. 1998). To limit the effect of rank sinks as well as to guarantee convergence to a unique rank vector, a decay factor $1 - \theta$ is introduced. We construct M'' as:

$$M'' = (1 - \theta)(M + D) + \theta E \quad (4)$$

and $E = \begin{bmatrix} 1 \\ n \end{bmatrix}_{n \times n}$.

Then, the final PageRank algorithm could be expressed as:

$$\overrightarrow{Rank}^{(i+1)} = \{(1 - \theta)(M + D) + \theta E\} \times \overrightarrow{Rank}^{(i)} \quad (5)$$

3.2 An Action-based User Influence Model

The key to our action-based user influence model is to bias the transition probability matrix. Instead of following network, we use action-based network. We construct a weighted action-based network, for every link, its weight is the frequency of user's actions. Thus, our main idea of action-based user influence model is that if one user is influential, then, another users that this user retweet from (or reply to) is also influential; and the more retweet frequency (or reply frequency) is, the more influence that this user transfers.

In our model, we give out different transition probability according to different type of user influence and action-based network.

1) For content-based influence, the transition probability from user i to j is defined as:

$$P_{rt}(i, j) = \frac{|T_{rt}(i, j)|}{\sum_{k: s_i \text{ retweet from } s_k} |T_{rt}(i, k)|} \quad (6)$$

$|T_{rt}(i, j)|$ is the number of retweets published by i and retweet from j .

$\sum_{k: s_i \text{ retweet from } s_k} |T_{rt}(i, k)|$ calculates the total retweets that published by user i .

2) For conversation-based influence, the transition probability from user i to j is defined as:

$$P_{re}(i, j) = \frac{|T_{re}(i, j)|}{\sum_{k: s_i \text{ reply from } s_k} |T_{re}(i, k)|} \quad (7)$$

$|T_{re}(i, j)|$ is the number of replies published by i and reply from j .

$\sum_{k: s_i \text{ reply from } s_k} |T_{re}(i, k)|$ calculates the total reply that published by user i .

There is a scenario to explain Equation 6. If user a retweet 100 tweets totally and 40% of them is retweeted from user b , and 60% of them is retweeted from user c , then, for content-based influence, the transition probability from user a to user b is 0.4 and the transition

probability from user a to user c is 0.6. This example is presented in Figure 1. Likewise, for conversation-based influence defined in Equation 7, there exists the same principle except for counting reply action instead.

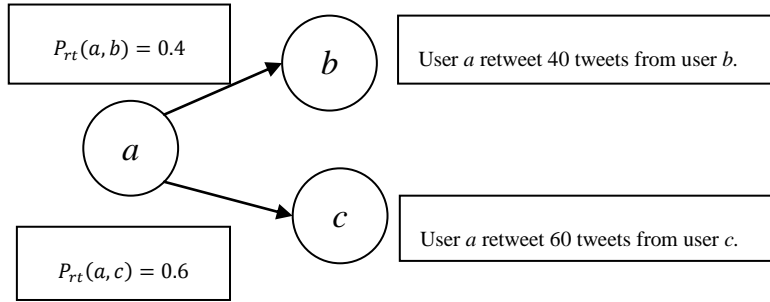


Figure 1. Example of Transition Probability

Except that the transition matrix M is computed according to Equation 6 and 7, our action-based user influence model follows PageRank algorithm described in Section 3.1.

4 A DYNAMIC ACTION-BASED USER INFLUENCE MODEL

Micro-blogging service is an ecological system; therefore, the user influence evolves too. How to capture time-frame characters of user influence becomes a very important question. The action-based network is dynamic and highly time-related since every action happens at a specific time, thus, there exist different action-based networks corresponding to different time fragment. Every action-based network within a time period is actually a snapshot of actions, it gives us a good opportunity to build dynamic model. Our original idea to build our dynamic model is based on the intuition that one user is influential at some time, could not be so influential in future. We call this phenomena "timeliness of user influence". Although timeliness of user influence has been widely exists, to the best of our knowledge, there is no similar influence model to entangle this problem in Twitter.

Moreover, computation efficiency is also a factor that promotes our work on building dynamic model. Incrementally measuring the change of user influence could be promising ways to measure user influence dynamics.

In our dynamic model, we divide time into small time fragments, each of time fragment covers actions initiated by users within this time fragment only. Then, we construct different action network corresponding to different time fragment. Finally, we calculate dynamic user influence by considering both the historical influence and new user influence arises in the current time fragments.

4.1 Dynamic Transition Probability

Since the action network is time-related, we could compute the transition probability for each time period easily. Similar to the basic model presented in Section 3, the transition probability for time period t is defined as follows:

- 1) For content-based influence, the transition probability from user i to user j is defined as:

$$P_{rt}(i, j, t) = \frac{|T_{rt}(i, j, t)|}{\sum_{k: s_i \text{ retweet from } s_k} |T_{rt}(i, k, t)|} \quad (8)$$

$|T_{rt}(i, j, t)|$ is the number of retweets published by i and retweet from j within time period t .

$\sum_{k: s_i \text{ retweet from } s_k} |T_{rt}(i, k, t)|$ calculates the total retweets that published by user i within time period t .

- 2) For conversation-based influence, the transition probability from user i to user j is defined as:

$$P_{re}(i, j, t) = \frac{|T_{re}(i, j, t)|}{\sum_{k: s_i \text{ reply from } s_k} |T_{re}(i, k, t)|} \quad (9)$$

$|T_{re}(i, j, t)|$ is the number of replys published by i and reply from j within time period t .

$\sum_{k: s_i \text{ reply from } s_k} |T_{re}(i, k, t)|$ calculates the total replys that published by user i within time period t .

4.2 Dynamic Rank Scores

In our model, we regard user's influence is an accumulation effect of all his influence on the previous time fragments, and the definitions are shown as follows:

- 1) For content-based influence of user i in the time period t is defined as:

$$s_{rt}(i, t) = \beta_{rt} \times \sum_{j=0}^{t-1} p_{rt}(i, t-j) \alpha_{rt}^j \quad (10)$$

$p_{rt}(i, t)$ is the content-based user influence ranking score within time period t , defined in Equation 5. α_{rt} is the decay factor for the effect of content-based user influence decline with time moving on and β_{rt} is a constant parameter.

- 2) For conversation-based influence of user i in the time t is defined as:

$$s_{re}(i, t) = \beta_{re} \times \sum_{j=0}^{t-1} p_{re}(i, t-j) \alpha_{re}^j \quad (11)$$

$p_{re}(i, t)$ is the conversation-based user influence rank score within time period t (see Equation 5). α_{re} is the decay factor for the effect of conversation-based user influence decline with time moving on and β_{re} is a constant parameter.

The decay factor represents the importance of user influence's timeliness. As time goes by, a user's influence may fade away, like an out-dated movie star.

The idea of time-related influence is quite significant especially in Twitter-like micro-blogging services. Compared with blog space, timeliness in micro-blog service is more important since a large number of micro-blog users use it to receive immediate information, even faster than TV news (Kwak et al. 2010), and some of its topics spread to the whole network within hours (Kwak et al. 2010). As a social media and social network platform, micro-blogging is kind of time sensitive. Measuring dynamic user influence of micro-blogging is quite necessary and practical.

4.3 A Dynamic Calculation Process

PageRank-like algorithm is kind of time-consuming, design an incremental algorithm is quite important. Our dynamic model could use incremental action-based network for efficient calculation. At the first time period, we calculate dynamic user influence with Equation 10 and 11. Then for each time period, we calculate a new part of action-based user influence rank score within the current time period. Furthermore, we combine the influence rank score

(see Equation 12 and 13). Equation 12 and 13 are another form of Equation 10 and 11 respectively. The purpose of Equation 12 and 13 is to demonstrate incrementally calculation.

1) For content-based influence of user i in the time t is defined as:

$$s_{rt}(i, t) = s_{rt}(i, t - 1) \times \alpha_{rt} + \beta_{rt} \times p_{rt}(i, t) \quad (12)$$

$p_{rt}(i, t)$ is the content-based user influence rank score within time period t . α_{rt} is the decay factor for the effect of content-based user influence decline with time moving on and β_{rt} is a constant parameter.

2) For conversation-based influence factor of user i in the time t is defined as:

$$s_{re}(i, t) = s_{re}(i, t - 1) \times \alpha_{re} + \beta_{re} \times p_{re}(i, t) \quad (13)$$

$p_{re}(i, t)$ is the conversation-based user influence rank score within time period t . α_{re} is the decay factor for the effect of conversation-based user influence decline with time moving on and β_{re} is a constant parameter.

5 SIMULATIONS AND EXPERIMENTS

5.1 Simulation Setup

To investigate the performance of our model, we implement experiments using data crawling from *Sina* micro-blog. We crawl 1000 users and their following links. In the following experiments, we assign each user a unique identify number ranging from 1 to 1000. Since our data only contains a small set of users, the retweet and reply networks are quite fragmented. Thus, in our experiments we use artificial retweet and reply networks generated. Similar to retweet and reply mechanisms in Twitter, one user can read all the tweets published by his followees and he can only retweet form his followees when generating artificial retweet and reply networks. Once a user publishes a tweet in a period, we assume that all the retweet actions of this tweet happen within the same time period. We generate tweets and their corresponding retweet actions based on the following two assumptions.

- **Assumption 1** Users do not publish their tweets evenly across all time periods.
- **Assumption 2** Different users' tweets have different retweet probability, and here we do not differentiate the content of tweets.

The reason for assumption 1 is that, intuitively, we assume that there exists a scenario that some user does not publish tweets evenly during his life, thus, our dynamic model could measure users who have temporal influence more accurately. For simplicity, we assume each user publishes the same number of tweets and each user could only publish all his tweets in one time period. The assumption 2 is trying to simulate that some tweets are more popular than others. Generally speaking, some users' tweets could have big chance to be retweeted.

5.2 Experimental Design

Experiment 1: Testing the effectiveness of our action-based user influence model.

In Experiment 1, we restrict all the tweets and actions into one time period. As we mentioned above, the entire user publish the same number of tweets, but their tweets have different retweet probability ranging from 0 to 0.5 randomly for all the users.

Experiment 2: Testing the effectiveness of our dynamic action-based user influence model.

In Experiment 2, we increase the number of time periods to 5, but each user could publish their tweets in only one period. For each user, the period that he publishes tweets is randomly selected. All the other conditions remain the same as Experiment 1.

6 RESULTS AND DISCUSSION

6.1 Results Analysis

In Experiment 1, we set every user will publish ten tweets and set the decay factor α_{rt} to 0.6, which is relatively high, since we want to observe the time effects soon. We set θ to 0.2. In our iteration process, we calculate 30 rounds for each rank score vector respectively.

Rank	No.	RS.	Indegree	Outdegree	Rank	No.	RS.	Indegree	Outdegree
1	65	0.0491	189	176	11	1	0.0074	43	73
2	68	0.0181	58	94	12	9	0.0063	72	71
3	45	0.0154	73	58	13	159	0.0061	23	68
4	28	0.0132	36	61	14	161	0.0057	30	80
5	51	0.0115	14	58	15	12	0.0054	40	97
6	60	0.0112	56	163	16	69	0.0054	51	68
7	50	0.0111	4	58	17	14	0.0052	43	82
8	11	0.0108	37	121	18	151	0.0050	27	56
9	47	0.0089	5	39	19	10	0.0050	28	41
10	43	0.0076	72	55	20	152	0.0048	27	88

Table 1. Top 20 Users Ranked by Rank Score in Action-based User Influence Model (denoted as RS.) & Indegree (number of followers), Outdegree (number of followees)

In Experiment 1, the simulation result demonstrates that rank score is correlated to the number of followers, as shown in Table 1. Generally, the top 20 users ranked by influence factor have relative larger indegree. This is because that, with large indegree, a user's tweet may have relative more retweets, even if he does not have high retweet probability. However, although one user has high indegree, it does not mean he has high influence too, since his published tweets may have very low retweet probability. Typically, user who has high influence also has high outdegree. The reason might be that if one user follows a lot of people, some of them may follow him back for respect. Then, he will have more followers and consequently, have high influence too. But this process might be completely reversed. The reason why users are influential remains further exploration.

Rank	RS.	Indegree	Outdegree	Rank	RS.	Indegree	Outdegree
1	65	65	65	11	1	180	1
2	68	176	60	12	9	68	9
3	45	178	11	13	159	588	31
4	28	179	57	14	161	60	69
5	51	175	12	15	12	570	159
6	60	177	54	16	69	69	132
7	50	31	68	17	14	4	28
8	11	45	152	18	151	32	24
9	47	43	14	19	10	58	45
10	43	9	161	20	152	1	51

Table 2. Top 20 Users Ranked by Rank Score (denoted as RS. in the table), Indegree & Outdegree respectively

Table 2 presents the top 20 users ranked by rank score, indegree and out degree respectively. As shown in Table 2, user 50 ranks 7 when ranked by rank score, while he does not rank in the top 20 either ranked by indegree or outdegree. Although ranks 7, user 50 only has 4

followers as shown in Table 1. That’s probably because among his 4 followers, some of them are very influential, once he publishes a tweet, his influential fan will retweet to a large number of users in his following network. There is, however, another possible explanation that he may have very high retweet probability. We think either explanation is possible or maybe both.

Rank	TP1			TP2			TP3			TP4			TP5		
	Cu	Ac	DRS	Cu	Ac	DRS	Cu	Ac	DRS	Cu	Ac	DRS	Cu	Ac	DRS
1	65	65	65	28	28	28	65	65	65	65	65	65	65	65	65
2	43 2	43 2	432	33 1	65	65	15 9	28	28	16 1	28	28	43 2	28	28
3	34 3	34 3	343	65	33 1	331	76 3	43	43	43 2	43	43	68	43	432
4	28	28	28	43	43	43	43	33 1	159	68	15 9	159	28	43 2	68
5	19 2	19 2	192	1	43 2	45	15 1	15 9	151	43	68	68	50	68	43
6	42 6	42 6	426	15 1	45	68	28	15 1	331	15 9	43 2	161	43	15 9	159
7	15 9	15 9	159	45	68	159	76 4	68	60	50 7	16 1	432	11	11	11
8	45	45	45	68	15 9	432	60	19 2	68	14 4	15 1	151	60	16 1	161
9	68	68	68	35	19 2	151	11	45	763	46	12	11	43 0	60	60
10	15 2	15 2	152	60	34 3	1	51	43 2	11	28	11	60	45	15 1	50

Table 3. Top 10 Users Ranked by “Current Time Period Rank Score” (Cu.), “Accumulated Rank Score” (Ac.) and “Dynamic Rank Score” in Dynamic Action-based User Influence Model (DRS.) across Time Period (TP1 ~ TP5)

In Experiment 2, as shown in Table 3, each user may have different influence in different time period. Our model could distinguish user influence from time dimension. Specifically, in Table 3, user 28 is most influential in time period 2 and user 65 is most influential in another 4 periods. Compared with accumulated user influence, our approach has a decay factor, which will emphasize the importance of timeliness. In time period 5, user 43 rank 3 by accumulated user influence, and user 432 and user 68 follows, while user 43 rank 5 by influence factor after user 432 and user 68. This is probably because user 43 has better performance in time period 2 and 3, but both user 432 and user 68 have better performance than user 43 in time period 4. Complying with the hypothesis that the closer to the current ranking time, the more important the influence score is, our model can “promote” user 432 and user 68 in rankings.

6.2 Related Work

Most micro-blog website including Twitter and Sina micro-blog use the number of followers as user influence indicator. This notion assumes that each follower has the same probability to see and to react to all of his followees. Another similar approach mentioned in Web Ecology Project (Leavitt et al. 2009) is to measure user influence by the ratio between the number of one’s followers and the number of one’s followees. As we claimed in the preliminary sections, both of these approaches ignore the user’s interactions with the content, which is denoted as actions in this paper.

Another two important user influence models related to our work, which adopted a PageRank-like algorithm, are TwitterRank and TunkRank. TwitterRank use topic similarity between two users to bias transition probability because they find strong presence of

homophily in user following network. The underneath idea is that the topic similarity between users and the following relationship are both strong indicators of influence (Weng et al. 2010). Again, this idea ignores the interactions between user and the content of micro-blogging platform. TunkRank uses a constant parameter to represent the probability that a user retweet another's tweet, which use PageRank-like algorithm to calculate user influence iteratively as:

$$Influence(X) = \sum_{Y \in Followers(X)} \frac{1 + p \times Influence(Y)}{|Following(Y)|} \quad (14)$$

In Equation 14, $Influence(X)$ represents the expected number of people who will read a tweet that X tweets, including all tweets of that user. $Following(Y)$ is the set of people that X follows. This notion assumes that the probability that one user retweet other users' tweet is the same among all users, and regards retweet network is same with following network (Tunkelang 2009).

The initial idea of the definition of user influence in our work comes from Web Ecology Project (Leavitt et al. 2009), which regard the influence as the ability of one user to make another user initiate another action. Following their research work, Cha et al. (2010) measure user influence (mention influence and retweet influence) by using the number of actions of each user. They compared two action user influences with indegree influence on empirical Twitter data. We agree that the user action is more representative than following relationship, and propose a user influence model to calculate the potential of an action of a user by a PageRank-like algorithm which involves an action-based network and user interactions.

7 CONCLUSIONS

This paper mainly focuses on proposing an action-based user influence model. Based on user definition of twitter given by Web Ecology Project (Leavitt et al. 2009), we give two kinds of user influence definition: content-based influence and conversation-based influence. Our user influence definition emphasizes on the interaction between Twitter user and the content of micro-blogging platform. To implement our user influence definitions, we introduce the idea of PageRank algorithm to build our action-based user influence model. The main contribution of our model is that we use action-based network to measure user influence. Many research on user influence measures on Twitter view user influence as the result of a serial of actions, not dynamic one, but in fact, micro-blogging service is an ecological system, user influence evolves as time goes by. In our paper, we present a possible tentative solution to explore this filed, which takes time dimension into considerations. By emphasizing on the importance of timeliness, we extend our model to a dynamic model. Besides that, we conduct simulation experiments to investigate the performance of our models. The experimental results show that our action-based user influence model performs well when micro-blogging service as a social media, and our dynamic model could distinguish the change of user influence in different time period.

8 FUTURE WORK

Our future work includes: 1) applying our model to empirical data sets, such as *Sina* micro-blog data set; 2) do more experiments to compare our models with other popular influence models; and 3) exploring the potential of hybrid user influence framework.

In this paper, we focus on measuring user influence in interactions between users and contents of micro-blogging platform. But to some extent, following relationship network does have impact on user influence. For example, although a user never retweet or reply to his followees, but he could get information or influence from his followees. In next step, we will explore the potential of combine action network and following network to capture both

following relationship and user interactions. To illustrate our idea, we consider four scenarios as follows:

- **Scenario 1:** There is no relevance between the probability, that one user's tweet is retweeted by others, and the number of one's followers.
- **Scenario 2:** There exists relevance between the probability, that one user's tweet is retweeted by others and the number of one's followers.
- **Scenario 3:** One user has many followers, but does not tweet a lot currently.
- **Scenario 4:** One user has many followers, and he always tweet a lot in all the time periods.

In Scenario 1 and 3, our proposed model could underestimate the potential of user influence since our model ignores the following relationship. In Scenario 2 and 4, our action-based network model works well. An intuitive idea is to propose a hybrid user influence framework which could hybrid different user influence together.

9 ACKNOWLEDGEMENT

This research is partly supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China under grant 10XN1029: Research on Financial Web Data Mining and Knowledge Management. The second author, Caihong Sun, is the corresponding author.

References

- Cha, M.; Haddadi, H.; Benevenuto, F. and Gummadi, K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. In *Association for the Advancement of Artificial Intelligence*.
- Haveliwala, T. H. (2003). Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 784-796.
- Java, A.; Song, X.; Finin, T. and Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. In *9th WEBKDD and 1st SNA-KDD Workshop'07*.
- Katz, E. and Lazarsfeld, P. F. (1955). Personal Influence: The Part Played by People in the Flow of Mass Communications. *New York: The Free Press*.
- Kwak, H.; Lee, C.; Park, H. and Moon S. (2010). What is Twitter, a Social Network or a News Media? In *WWW 2010*.
- Lazarsfeld, P. F.; Berelson, B. and Gaudet, H. (1968). The people's Choice: How the Voter Makes up His Mind in a Presidential Campaign. *New York: Columbia University Press*.
- Leavitt, A.; Burchard, E.; Fisher, D. and Gilbert, S. (2009). The Influentials: New Approaches for Analyzing Influence on Twitter. In *Web Ecology Project*, <http://webecologyproject.org>.
- Motwani, R. and Raghavan, P. (1995). Randomized Algorithm. *U.K.: Cambridge Univ. Press, 1995*
- Page, L.; Brin, S.; Motwani, R. and Winograd, T. (1998). "The PageRank Citation Ranking: Bringing Order to the Web". *Stanford Digital Libraries Working Paper*.
- Tunkelang, D. (2009). A Twitter Analog to PageRank. <http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank/>
- Watts, D. J. and Dodds, P. S. (2007). Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*.
- Weng, J.; Lim, E.-P. and Jiang, J. (2010). TwitterRank: Finding Topic-sensitive Influential Twitterers. In *ACM WSDM'10*.