

## Research Article

# Popularity Evaluation Model for Microbloggers Online Social Network

Xia Zhang,<sup>1,2</sup> Zhengyou Xia,<sup>1</sup> and Zhan Bu<sup>1</sup>

<sup>1</sup> College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

<sup>2</sup> Department of Computer & Software, Nanjing College of Information Technology, Nanjing 210016, China

Correspondence should be addressed to Zhengyou Xia; zhengyou\_xia@nuaa.edu.cn

Received 17 March 2014; Revised 15 May 2014; Accepted 16 May 2014; Published 3 June 2014

Academic Editor: Zhan Zhou

Copyright © 2014 Xia Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, microblogging is widely studied by the researchers in the domain of the online social network (OSN). How to evaluate the popularities of microblogging users is an important research field, which can be applied to commercial advertising, user behavior analysis and information dissemination, and so forth. Previous studies on the evaluation methods cannot effectively solve and accurately evaluate the popularities of the microbloggers. In this paper, we proposed an electromagnetic field theory based model to analyze the popularities of microbloggers. The concept of the source in microblogging field is first put forward, which is based on the concept of source in the electromagnetic field; then, one's microblogging flux is calculated according to his/her behaviors (send or receive feedbacks) on the microblogging platform; finally, we used three methods to calculate one's microblogging flux density, which can represent one's popularity on the microblogging platform. In the experimental work, we evaluated our model using real microblogging data and selected the best one from the three popularity measure methods. We also compared our model with the classic PageRank algorithm; and the results show that our model is more effective and accurate to evaluate the popularities of the microbloggers.

## 1. Introduction

Microblogging is a broadcast medium in the form of blogging. A microblog differs from a traditional blog in that its content is typically smaller in both actual and aggregate file size. Microblogging allows users to exchange small elements of content such as short sentences, individual images, or video links. Twitter and Weibo are all the famous microblogging and have more than hundreds of millions of users. Twitter and Weibo social networks have emerged as a critical factor in information dissemination, search, marketing, expertise, and influence discovery and potentially an important tool for mobilizing people [1–5]. Social media have made social networks ubiquitous, also given researchers access to massive quantities of data for empirical analysis. These data sets offer a rich source of evidence for studying dynamics of individual and group behavior, the structure of networks, and global patterns of the flow of information on them [6–8]. Popularity Evaluation Model for microbloggers is very important research field on social network such as

Sina Weibo. For example, companies choose popularities of microblogging users to run their commercials, by popularities of microblogging users to publish and forward, to achieve enterprise business advertising. Also in the study of online social networks, the network users need to study the role of other issues and also need a model or method of analysis of the popularities to the user. Therefore, Popularity Evaluation Model for microbloggers can be applied into commercial advertising, user behavior analysis and information dissemination, and so forth.

How to evaluate the popularities of microblogging users is an important research for online social network. Previous studies on the evaluation methods cannot effectively solve and accurately evaluate the popularities of the microbloggers [9, 10]. For example, the popularities of microbloggers are hard to evaluate based on transitional network structure models (PageRank algorithm [10]). It is well known that the more fans users has, which showed greater the popularities on Weibo social network [11]. According to the actual data statistics, we found that there were inactive users in Sina

Weibo. We referred to the users as “zombie.” The existence of “zombie” had no contribution to popularity of users, and this is why the relationships between the number of users’ fans and popularity were not close enough. Therefore, this method based on the fan list cannot truly reflect one’s connection strength or popularity.

In order to effectively and accurately evaluate the popularities of the microbloggers over time, we proposed an electromagnetic field theory based model to analyze the popularities of microbloggers in this paper. The concept of the source in microblogging field is first put forward, which is based on the concept of source in the electromagnetic field; then, one’s microblogging flux is calculated according to his/her behaviors (send or receive comments) in the microblogging platform; finally, we used three methods to calculate one’s microblogging flux density, which can represent one’s popularity in the microblogging platform. The remainder of this paper is organized as follows. We discuss the related work in Section 2. In Section 3, we proposed three kinds of user source in microblogging field as those in the electromagnetic field, which are positive, negative, and neutral sources. For every microblogger, the microblogging flux is calculated according to his/her behaviors (send or receive comments). Three methods are put forward to calculate one’s microblogging flux density in Section 4. In Section 5, we evaluated our model using real microblogging data and selected the best one from the three measure methods to evaluate the popularities of microbloggers. Finally, we conclude the paper in Section 6.

## 2. Related Work

Recently, online social networks [1] have gained significant popularity and are now among the most popular sites on the Web. Online social network researchers mainly focus on network-structure-model construction [2–5], user-behavior analysis [6], information dissemination [7], content recommendation [8], and so forth. Those research fields are closely associated. For example, information dissemination is mainly influenced by the user interest degree, number of users’ friends, and user’s behavior [12–15]. Mislove et al. [13] collected mass data from four social network sites; and they measured and analyzed the structures of online social networks. Zhao et al. [14] found weak connection in online social network, which has significant influence in evaluating the speed and the breath of the network information transmission. Kwak et al. [15] analyzed the topological structures of Twitter, and they found that the number of Twitter users’ followers is distributed according to a power law followed by an exponential cutoff. Letierce et al. [12] studied the label for the transmission of the information between users. At present, in the view of the network-structure-model researchers, user connection study and its weight measure are very popular. Yun et al. [16] analyzed five factors which influence mutual connections between users in Twitter. Chen et al. [17] analyzed the equivalence attributes of online users, which are mostly based on the user connection strength. In the field of content recommendation, some researchers focus on the category of the content, and then the content

will be recommended to users whose interest matching degrees are high. Content recommendation is widely used in the electronic commerce system, video sharing sites, and other fields [18]. Among these applications, collaborative filtering is a mostly used technology; for a given user, it is used to recommend those whose interests are similar to him/her as his/her potential friends. Saito et al. [19] and Tang et al. [20] conducted a series of experiments and found that, due to the different preferences, users with similar interests behave differently in spreading their topics. Yang and Scott [21] found that the mention rate from relevant users is an important factor which influences many aspects of information transmission, such as speed, scale, and scope. Lerman and Ghosh [9] analyzed the influence of network structure in information transmission based on Digg and Twitter data.

## 3. User Source in Microblog

In the microblogging social network, users release microblogs to share information, and they can also interact with each other by forwarding or commenting on microblogs. The behaviors of the microbloggers are described in Figure 1. As shown in Figure 1(a), the static circle represents a microblogger; lines with positive arrows mean that the microblogger is forwarded or commented on by others. As shown in Figure 1(b), lines with negative arrows mean that the microblogger posts microblogs or comments on other blogs. On microblogging platform, everyone can post microblogs or communicate with each other. However, their popularities are different. We often find on microblogging platform that even two microbloggers post similar number of blogs; their blogs receive different responses (the total numbers of followers and comments are far different). It agrees with the well-known Matthew effect, namely, “the rich get richer and the poor get poorer.” Those celebrities are more frequently followed or commented on by others. To study this Matthew effect in microblogging environment, we will first put forward the concept of the source in microblogging field and then give the criteria to calculate one’s microblogging flux (activity). Our model is based on the electromagnetic field theory, which is first in its kind.

In order to calculate one’s microblogging flux (activity), we first introduce the concept of the source in the electromagnetic field theory. An electromagnetic field (also EMF or EM field) is a physical field produced by electrically charged objects. It affects the behavior of charged objects in the vicinity of the field. The electromagnetic field extends indefinitely throughout space and describes the electromagnetic interaction. The field can be viewed as the combination of an electric field and a magnetic field. The electric field is produced by stationary charges and the magnetic field by moving charges (currents); these two are often described as the sources of the field. As the distribution of the charge or current in space is uneven, the charge density and current density are put forward to describe the distribution of the source. The electric field is produced by stationary charges; there are three kinds of point charges in nature, which are

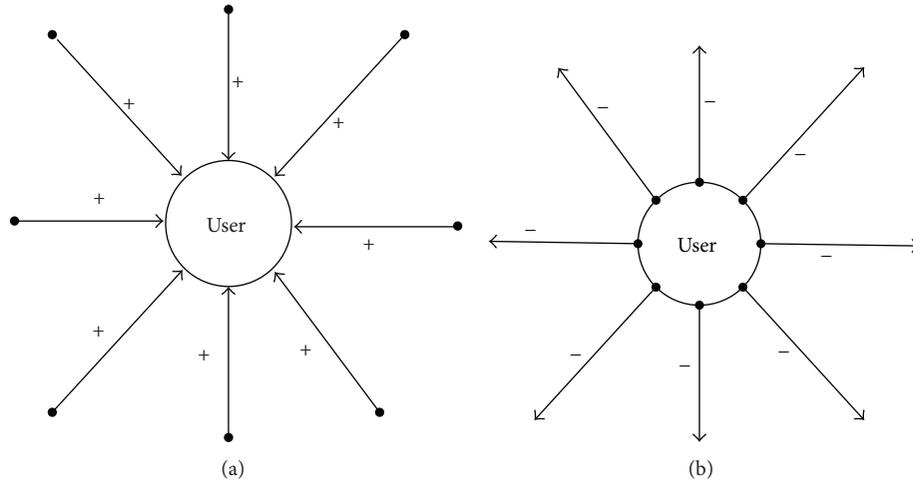


FIGURE 1: (a) The microblogger is followed or commented on by others; (b) the microblogger posts microblogs or comments on other users' microblogs.

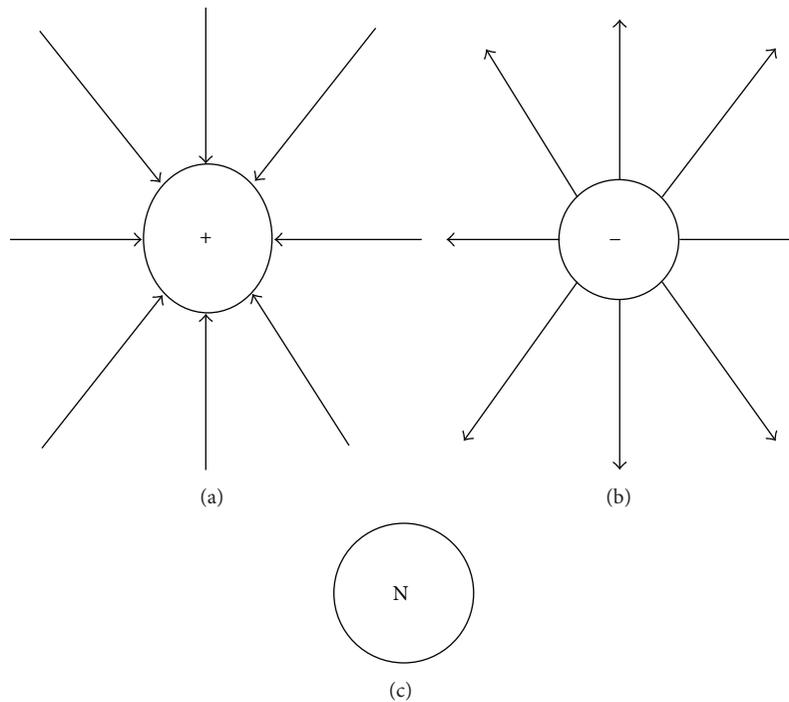


FIGURE 2: Three forms of charge in the nature.

(a) the positive charge, (b) the negative charge, and (c) the neutral charge as shown in Figure 2.

The point charge is a source of electric field which can produce an electric field. Similar with the source of electromagnetic field, in microblogging social network, the flux (activity) is produced by microbloggers. Therefore, we can take every microblogger as a source in microblogging field. In electromagnetism, the magnetic flux (often denoted as  $\Phi$ ) through a surface is the component of the magnetic  $B$  field passing through that surface. As shown in Figure 3,

the magnetic flux is properly defined as the component of the magnetic field passing through a surface, where  $B$  is the magnitude of the magnetic field (the magnetic flux density) having the unit of  $\theta/S$  (tesla),  $S$  is the area of the surface, and  $\theta$  is the angle between the magnetic field lines and the normal (perpendicular) to  $S$ .

According to the magnetic flux in a magnetic field, we also use this idea to study the microblogging flux. In the microblogging social network, the microblogging flux of a given blog is defined as the number of feedbacks (to be

TABLE 1: The parameters of microbloggers.

User_ID	Blog_Num	Forwarded_Num	Commented_Num	Flux	Source type
1440451110	1354	3	2	-1349	Negative
1038518860	4	30	0	26	Positive
1052073395	726	19	62	-645	Negative
1049364404	74	33	0	-41	Negative
1045865362	782	32	43	-707	Negative
1050862900	2058	143	260	-1655	Negative
1050881714	5	2	3	0	Neutral
1069689715	1	22	4	25	Positive
1085799815	361	14	52	-295	Negative
1088980567	182	35	5	-142	Negative
1094141601	91	18	0	-73	Negative
1146768700	231	22	55	-154	Negative
1117480841	239	37	83	-119	Negative
1159065410	2	6	7	11	Positive
1172610962	370	20	12	-338	Negative

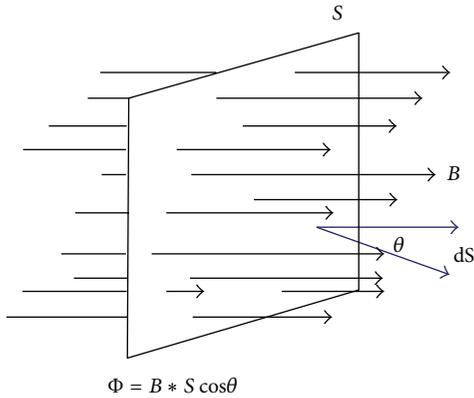


FIGURE 3: The magnetic flux in a magnetic field.

forwarded or commented on) subtracting one, as shown in the following formula:

$$\Phi_i = FNum_i + CNum_i - 1, \quad (1)$$

where  $FNum_i$  represents the forwarded number of this blog and  $CNum_i$  represents its commented number. If the blog receives no feedbacks, we will reduce one as the punishment. For each microblogger, we add the microblogging flux of all his/her blogs. Therefore, the microblogging flux of a given user is simplified as

$$\Phi = \sum_{i \in \text{Blogs}} \Phi_i = \sum_{i \in \text{Blogs}} (FNum_i + CNum_i - 1), \quad (2)$$

where Blogs represents the blog set of the given microblogger,  $i$  represents his/her  $i$ th blog,  $FNum_i$  represents the forwarded number of the  $i$ th blog, and  $CNum_i$  represents the commented number of the  $i$ th blog. According to formula (2), the microblogging flux of a given microblogger can be positive, negative, or zero. As shown in Figure 4(a), if a microblogger receives more feedbacks than the number of blogs she/he

posts, we define him/her as a positive microblogging source; if a microblogger receives less feedbacks than the number of blogs she/he posts, she/he will be defined as a negative microblogging source (see Figure 4(b)); otherwise, the left bloggers are defined as the neutral microblogging sources (see Figure 4(c)). In Table 1, we roughly identify several microbloggers from Sina microblogging platform as positive, negative, or neutral ones.

Thus, we can classify users in microblogging platform by considering their microblogging flux. For example, if one's microblogging flux is greater than 0, we see him/her as a positive microblogger; if one's microblogging flux is less than 0, we see him/her as a negative microblogger; otherwise, she/he is defined as a neutral microblogger. Figure 5 shows the distribution of three kinds of microbloggers randomly selected from the Sina Weibo sample, from which we can see that most microbloggers (59%) are positive; about 41% of microbloggers in the sample are negative; and only 1% of microbloggers are neutral.

#### 4. Popularity Evaluation Model for Microbloggers

In Section 2, we introduced the concept of microblogging source and classified microbloggers into three kinds according to their microblogging flux. We first propose Hypothesis 1 as follows.

*Hypothesis 1.* The popularity of a microblogger is decided by his/her microblogging flux.

In order to verify Hypothesis 1, we sort the microbloggers by their microblogging flux in descending order. Table 2 shows the microbloggers whose microblogging flux is arranged in top 16. We can manually compare their popularities by browsing their homepages in the microblogging platform. Take the user pair of ID1671526850 and

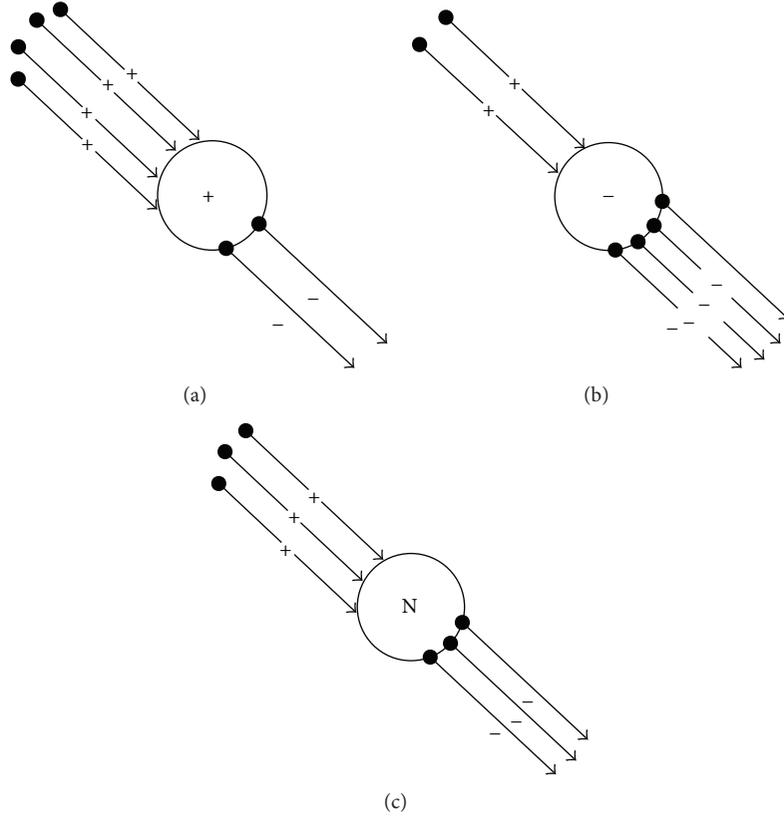


FIGURE 4: Three kinds of microblogging sources.

ID1660209951 as an example. The microblogging flux of the former is much greater than that of the latter as shown in Table 2. ID1671526850 has 158098 blogs and about 499 thousand fans. And ID1660209951 only has 57981 blogs and about 466 thousand fans. In this context, we believe that ID1660209951 is more popular than ID1671526850. That is because every blog of ID1660209951 attracts more feedbacks; even the number of his fans is less than that of ID1660209951. From this example, we can find that Hypothesis 1 may not be very reasonable which needs to be improved.

Microblogging flux may not accurately evaluate one's popularity; to address this problem, we propose a new evaluation metric which is called the microblogging flux density. We can learn from the electromagnetic concept that the microblogging flux is represented by the surface integral of a vector field, as shown in the following formula:

$$\Phi_M = \iint_A M \cdot dA, \quad (3)$$

where  $M$  is a vector field and  $dA$  is the vector area of the microblogging surface  $A$ , directed as the surface normal. Conversely, one can consider the microblogging flux the more fundamental quantity and call the vector field the microblogging flux density. Often a vector microblogging field is drawn by curves (field lines) following the "flow"; the magnitude of the vector microblogging field is then the line density, and the flux through a surface is the number of lines. The microblogging flux density (microblogging field)

is denoted by  $M$ . In the microblogging platform, we believe one's popularity is associated with the amount of feedbacks she/he receives, while the latter is also decided by the number of one's blogs and fans. Therefore, the microblogging surface  $A$  in formula (3) might be affected by these two factors, namely, the number of blogs (Blog\_Num) and the number of the fans (Fan\_Num). To verify our judgment, we propose three hypotheses as follows.

*Hypothesis 2.* One's microblogging surface  $A$  is mainly affected by the number of blogs.

*Hypothesis 3.* One's microblogging surface  $A$  is mainly affected by the number of fans.

*Hypothesis 4.* One's microblogging surface  $A$  is mainly affected by the number of blogs and fans.

Based on *Hypothesis 2*, the microblogging flux density can be calculated as the following formula, which is labeled as  $M_{\text{Blog}}$ :

$$M_{\text{Blog}} = \frac{\Phi_M}{w}, \quad (4)$$

where  $w$  is the number of one's microblogs.  $M_{\text{Blog}}$  reflects the average number of feedbacks of one's microblog, which is similar to divergence theorem in electromagnetic field theory. On the microblogging platform, some bloggers post

TABLE 2: Microbloggers whose microblogging flux is arranged in top 16.

User_ID	Blog_Num	Fan_Num	Microblogging flux	Microblogging flux per blog
1644395354	61,879	10,775,394	321,090,728	5,189
<b>1671526850</b>	<b>158,098</b>	<b>4,995,525</b>	<b>147,531,344</b>	<b>933</b>
<b>1660209951</b>	<b>57,981</b>	<b>4,663,119</b>	<b>101,086,971</b>	<b>1,743</b>
1567852087	88,358	3,805,427	92,930,548	1,051
1252373132	53,299	5,051,721	90,050,387	1,689
1657421782	46,420	4,369,588	70,721,633	1,523
1266286555	2,464	23,143,090	66,814,481	27,116
1657430300	50,027	3,373,765	63,745,465	1,274
1644572034	33,486	4,512,287	63,616,947	1,899
1134796120	53,801	1,288,228	50,956,066	947
1197161814	10,582	31,891,066	43,980,575	4,156
1195230310	5,060	28,644,548	43,755,799	8,647
1644574352	35,958	1,780,998	42,868,683	1,192
1182389073	51,302	13,662,475	41,258,257	804
1192329374	5,676	30,464,892	39,550,836	6,968

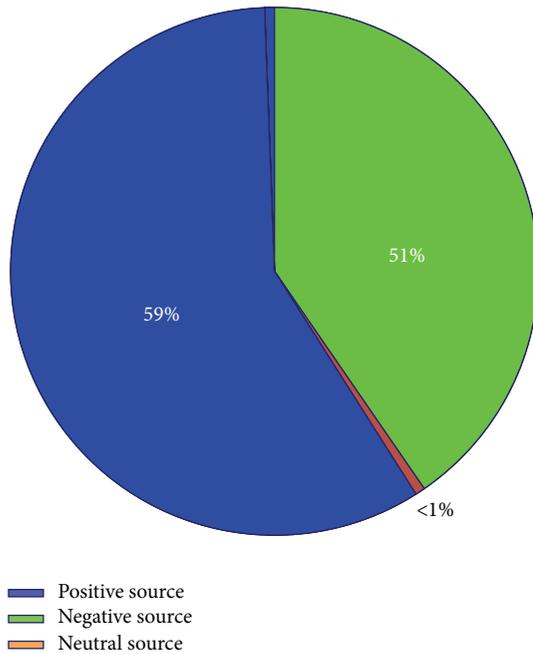


FIGURE 5: The distribution of three kinds of microbloggers.

a large number of blogs every day to attract the attention of others. Some of their blogs are interesting, and they can acquire a lot of feedbacks. However, most of their blogs are trite, which might not be forwarded or commented on. Microbloggers with this feature are called “twuilt.” Though the microblogging flux of those twuilts is high, the average microblogging flux from each blog is very low. Those twuilts are not really popular.  $M_{\text{Blog}}$  can reflect one’s popularity, which effectively reduces the influence from those twuilt IDs.

According to *Hypothesis 3*, the microblogging flux density can be calculated as in the following formula, which is labeled as  $M_{\text{Fan}}$ :

$$M_{\text{Fan}} = \frac{\Phi_M}{f}, \quad (5)$$

where  $f$  is the number of one’s fans. As the feedbacks of one’s blogs mainly come from his/her fans,  $M_{\text{Fan}}$  considers the average number of feedbacks from one of his/her fans. On the microblogging platform, some bloggers hire multiple IDs (usually called “zombie IDs”) to increase their popularity. Those zombie IDs are fake identities through which members of Internet community create the illusion of support for someone, pretending to be a different person. As those zombie IDs seldom give feedbacks to other bloggers, they cannot be seen as one’s true fans.  $M_{\text{Fan}}$  can reflect one’s popularity, which effectively reduces the influence from those zombie IDs.

*Hypotheses 2* and *3* consider the effects to one’s popularity from his/her number of blogs and fans. The former can reduce the influence from those twuilt IDs; the latter can reduce the influence from those zombie IDs, while *Hypothesis 4* combines these two effects and can lead to a new microblogging flux density ( $M_{\text{Blog.Fan}}$ ) as the following formula:

$$M_{\text{Blog.Fan}} = \frac{\Phi_M}{w \cdot f}. \quad (6)$$

$M_{\text{Blog.Fan}}$  fully considers all the factors that could affect one’s popularity. It reduces the influences from both twuilt IDs and zombie IDs. The detailed analysis of the three metrics will be discussed in the next section.

## 5. Experiments and Analysis

In order to validate the effectiveness of the above three popularity evaluation metrics, we will use real microblogging

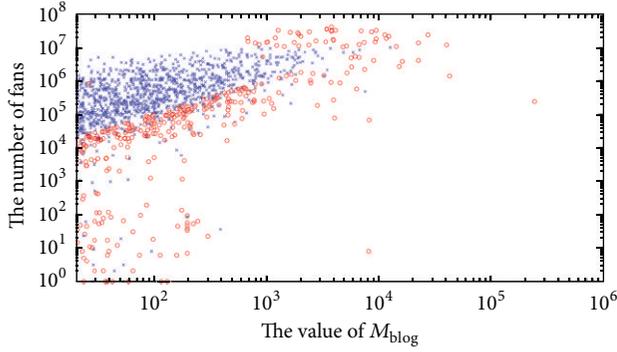


FIGURE 6: The microbloggers whose  $M_{\text{Blog}}$  are arranged in top 1500; the red nodes represent the VIP users.

data and conduct a series of experiments. The data used in this paper is crawled from Sina Weibo, which is a Chinese microblogging website. Akin to a hybrid of Twitter and Facebook, it is one of the most popular sites, in China, in use by well over 30% of Internet users, with a market penetration similar to what Twitter has established in the USA. It was launched by SINA Corporation on August 14, 2009, and has 503 million registered users as of December 2012. About 100 million messages are posted each day on Sina Weibo. In this paper, the collected data includes 8,945 users (include 901 VIP users), 20,147,746 blogs, and 925,669,059 comments.

Currently the performance of popularity evaluation methods is evaluated by manual inspection. For each microblogger an effort is made to interpret him/her as a “real” popular star by browsing their homepages in the microblogging platform. For example, we check how many original blogs she/he has posted, how often his/her fans give the feedbacks, and how do his/her fans like his/her posted blogs. However, such anecdotal evaluation procedures that require extensive manual effort are noncomprehensive and limited to small networks. In the Sina Weibo, to improve their popularities, some microbloggers might pursue the VIP privilege. Those VIP users are more easily accessed by other bloggers. We found in Sina Weibo that most of those popular microbloggers are VIP users. Although it cannot prove that all the VIP bloggers are very popular in Sina Weibo, we can use the number of VIP users as indirect indicators to verify the effectiveness of our proposed measure models.

**5.1. Comparison of Four Hypotheses.** To verify the effectiveness of the blog based flux density  $M_{\text{Blog}}$ , we select the microbloggers whose  $M_{\text{Blog}}$  are arranged in top 1500 (about 15% users in the data set). Figure 6 shows those microbloggers with their fan number and  $M_{\text{Blog}}$ , in which the red nodes represent VIP users. The number of VIP users in those 1500 microbloggers is 348, accounted for about 39% VIP users. Though some bloggers’  $M_{\text{Blog}}$  are very high, they are not VIP users, because the number of their fans is huge (see the blue crosses in Figure 6).

We take a similar experiment to verify the effectiveness of the fan based flux density  $M_{\text{Fan}}$ . Figure 7 shows microbloggers whose  $M_{\text{Fan}}$  are arranged in top 1500 with their blog number,

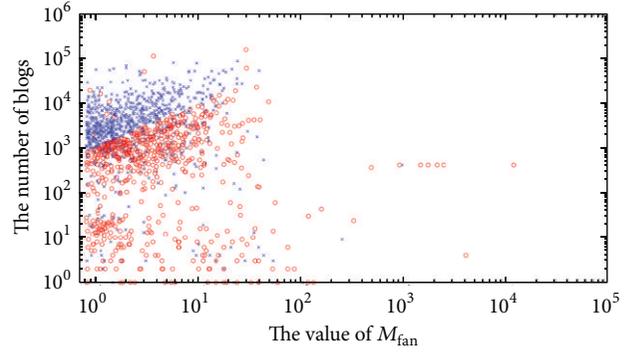


FIGURE 7: The microbloggers whose  $M_{\text{Fan}}$  are arranged in top 1500; the red nodes represent the VIP users.

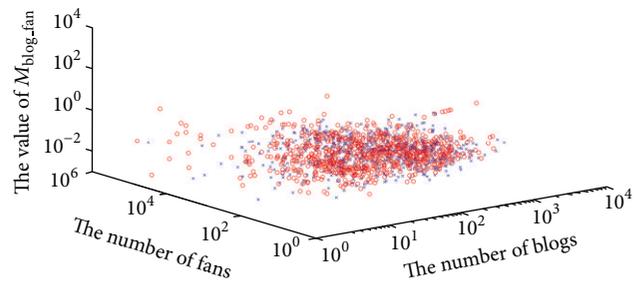


FIGURE 8: The microbloggers whose  $M_{\text{Blog-Fan}}$  are arranged in top 1500; the red nodes represent the VIP users.

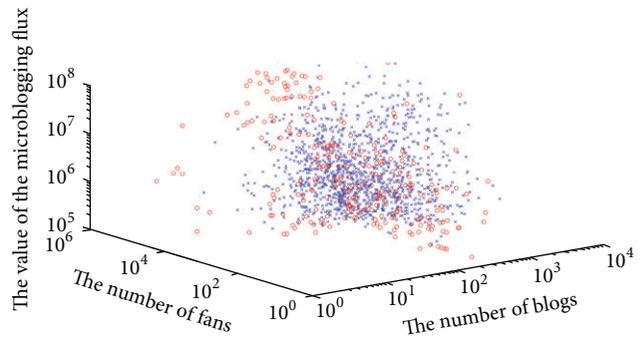


FIGURE 9: The microbloggers whose microblogging flux is arranged in top 1500; the red nodes represent the VIP users.

in which the red nodes represent VIP users. The number of VIP users in those blogger set is 656, accounted for about 73%. Some microbloggers are not VIP users even when their  $M_{\text{Fan}}$  are very high (see the blue crosses in Figure 7).

Then, we select the microbloggers whose  $M_{\text{Blog-Fan}}$  are arranged in top 1500. Figure 8 shows them with their blog number, fan number, and  $M_{\text{Blog-Fan}}$ . The number of VIP users (see the red nodes) is 862, which is accounted for about 95% VIP users. That means most of VIP users are included in those 1500 microbloggers.

To take the further validation of Hypothesis 1, we select the microbloggers whose microblogging flux  $\Phi_M$  is arranged in top 1500. Those users are shown in Figure 9, associated with their blog number, fan number, and the microblogging

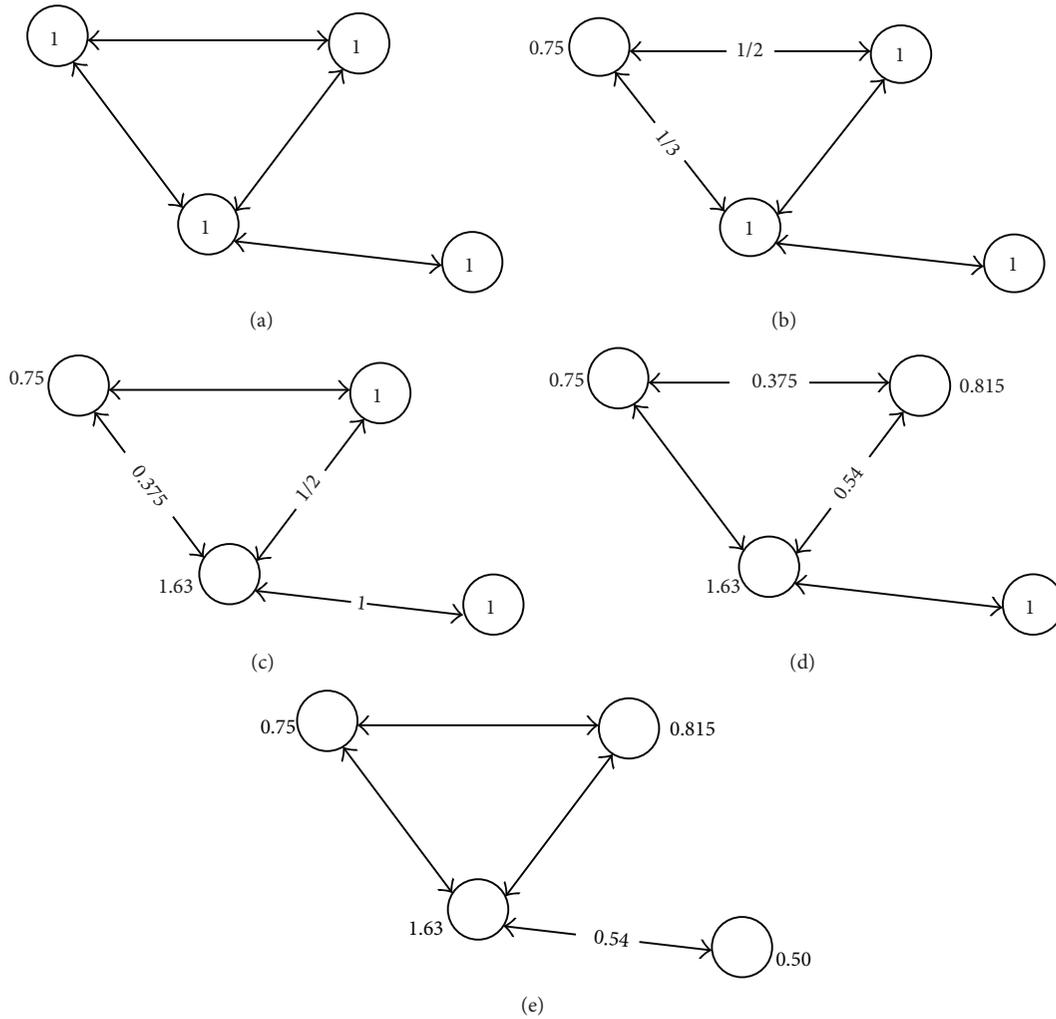


FIGURE 10: Visualization of the steps of PageRank algorithm.

flux. The number of VIP users in those 1500 microbloggers is only 246, accounted for about 18% VIP users. That means most of VIP users are not included in those 1500 microbloggers.

According to the above experiments, we can conclude that Hypothesis 1 < Hypothesis 2 < Hypothesis 3 < Hypothesis 4.  $M_{\text{Blog-Fan}}$  fully considers all the factors that could affect one's popularity, and it is the best metric to evaluate one's popularity in the microblogging platform.

**5.2. Comparison with PageRank Algorithm.** We also compare our evaluation model with classical PageRank algorithm. PageRank is a link analysis algorithm and first used by the Google web search engine, which assigns a numerical weighting to each element of a hyperlinked set of documents, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. The numerical weight that it assigns to any given element  $p_i$  is

referred to as the PageRank of  $p_i$  and denoted by  $\text{PR}(p_i)$  as shown in the following formula:

$$\text{PR}(p_i) = \frac{1-d}{N} + d \cdot \sum_{p_j \in M(p_i)} \frac{\text{PR}(p_j)}{L(p_j)}, \quad (7)$$

where  $p_1, p_2, \dots, p_N$  are the pages under consideration,  $M(p_i)$  is the set of pages that link to  $p_i$ ,  $L(p_j)$  is the number of outbound links on page  $p_j$ , and  $N$  is the total number of pages. When calculating PageRank, pages with no outbound links are assumed to link out to all other pages in the collection. Their PageRank scores are, therefore, divided evenly among all other pages. In other words, to be fair with pages that are not sinks, these random transitions are added to all nodes in the Web, with a residual probability usually set to  $d = 0.85$ , estimated from the frequency that an average surfer uses his or her browser's bookmark feature. The visualization of the steps of PageRank algorithm is shown in Figure 10. We initialize PR value of every node as 1; and then we update the PR value of each node according to formula (7) as shown in Figures 10(b)–10(e).

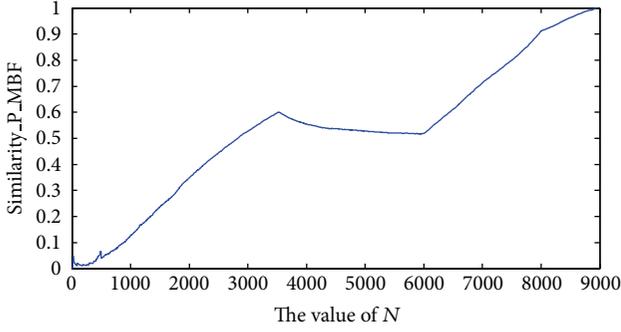
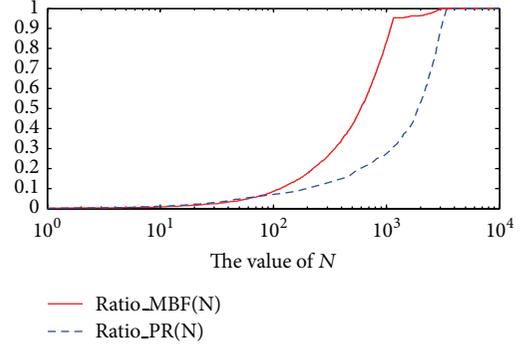


FIGURE 11: The similarity of the results of the two algorithms.

PageRank algorithm has been applied to traditional microblogger popularity evaluation. To compare our evaluation model and the classical PageRank algorithm, we first construct an interactive network according to the fan list and watch list of microbloggers. Then, PageRank algorithm is applied to calculate the PR value of each blogger. Blogger IDs with high PR value will seem as popular stars on the microblogging platform. We sort the microbloggers by their PR value in descending order; the top  $N$  bloggers are selected as Blogger  $SET\_A(N)$ . Similarly, we calculate the  $M_{\text{Blog.Fan}}$  of every microblogger and sort them by their  $M_{\text{Blog.Fan}}$  in descending order. The top  $N$  bloggers are selected as Blogger  $SET\_B(N)$ . The intersection of the two sets is denoted as  $SET\_A(N) \cap SET\_B(N)$ . In this context, the similarity of the results from the two algorithms can be calculated as follows:

$$\begin{aligned} \text{Similarity\_P\_MBF} &= \frac{\text{Card}(SET\_A(N) \cap SET\_B(N))}{\text{Card}(SET\_A(N))} \\ &= \frac{\text{Card}(SET\_A(N) \cap SET\_B(N))}{N}, \end{aligned} \quad (8)$$

where  $\text{Card}(\cdot)$  returns the number of elements in the set. To compare the results of two algorithms, we calculate  $\text{Similarity\_P\_MBF}$  with various  $N$  values as shown in Figure 11. In our data set, the total number of VIP users is only 901, which accounted for nearly 10% of all the microbloggers. The actual number of star bloggers might be lower than this value. As shown in Figure 11, as  $N$  grows from 1 to 1000 (accounted for nearly 10% bloggers in the data set),  $\text{Similarity\_P\_MBF}$  is very low.  $\text{Similarity\_P\_MBF}$  grows rapidly as  $N$  grows from 1000 to 3518.  $\text{Similarity\_P\_MBF}$  gets its local maximum value (0.6) when  $N$  is set as 3518. However, further increasing the value of  $N$  from 3518 to 5987 after this point is counter-productive, as indicated by the significant drop in  $\text{Similarity\_P\_MBF}$ . Then,  $\text{Similarity\_P\_MBF}$  grows with the growth of  $N$  and gets its global maximum value (1) when  $N$  is set as 8945. From this experiment, we can find that the results of the two algorithms are different to a great degree.

FIGURE 12: VIP ratios in top  $N$  bloggers sorted by the two algorithms.

VIP ratios in top  $N$  bloggers sorted by the two algorithms are denoted as  $\text{Ratio\_PR}(N)$  and  $\text{Ratio\_MBF}(N)$ , which are defined as follows:

$$\begin{aligned} \text{Ratio\_PR}(N) &= \frac{\text{Card\_VIP}(SET\_A(N))}{\text{Card}(SET\_A(N))} \\ &= \frac{\text{Card\_VIP}(SET\_A(N))}{N}, \\ \text{Ratio\_MBF}(N) &= \frac{\text{Card\_VIP}(SET\_B(N))}{\text{Card}(SET\_B(N))} \\ &= \frac{\text{Card\_VIP}(SET\_B(N))}{N}, \end{aligned} \quad (9)$$

where  $\text{Card\_VIP}(\cdot)$  returns the number of VIP users in the blogger set. As shown in Figure 12, as  $N$  grows from 1 to 1000 (accounted for nearly 10% bloggers in the data set),  $\text{Ratio\_PR}(N)$  grows slowly, while for  $\text{Ratio\_MBF}(N)$ , it grows slowly when  $N$  grows from 1 to 100; then it rises rapidly as  $N$  grows from 100 to 1000. When  $N$  is set as 1000,  $\text{Ratio\_PR}(1000) = 27.9\%$  and  $\text{Ratio\_MBF}(1000) = 83.2\%$ . We observe that our blog-fan based flux density,  $M_{\text{Blog.Fan}}$ , achieves the overall highest accuracy; apparently, it is better than the traditional PageRank algorithm in measuring one's popularity in the microblogging platform.

**5.3. The Current Popularities of Microbloggers.** The popularities of microbloggers are measured from overall point of view. In practice, the activities of microbloggers are changing over the time; therefore, their popularities also change dynamically. In this subsection, we analyze the current popularities of microbloggers. For every blog of a given microblogger, we keep track of its feedbacks. The feedback information of microblogs posted by ID1087770692 from August 4, 2011, to May 15, 2013, is shown in Table 3. The current popularity of a given microblogger is calculated as follows:

$$\begin{aligned} M_{\text{Blog.Fan}}^* &= \frac{\Phi_M^*}{w^* \cdot f} = \frac{\sum_{i \in \text{Blogs}^*} \Phi_i}{w^* \cdot f} \\ &= \frac{\sum_{i \in \text{Blogs}^*} (\text{FNum}_i + \text{CNum}_i - 1)}{w^* \cdot f}, \end{aligned} \quad (10)$$

TABLE 3: The feedback information of microblogs posted by user “1087770692” from August 4, 2011, to May 15, 2013.

Blog_ID	User_ID	Post date	Forwarded_Num	Commented_Num
3342638187376430	1087770692	2011-08-04	686	0
3342642545392029	1087770692	2011-08-04	1033	407
3342652108123960	1087770692	2011-08-04	2849	380
3342956334136552	1087770692	2011-08-05	359	146
3342969214160990	1087770692	2011-08-05	566	300
3343021802451966	1087770692	2011-08-05	7153	4589
3343090463253335	1087770692	2011-08-05	1146	420
3350669889615619	1087770692	2011-08-26	823	1263
3350672028496051	1087770692	2011-08-26	465	649
3350705175929327	1087770692	2011-08-26	801	262
3351097033072419	1087770692	2011-08-27	572	392
3577907916358736	1087770692	2011-08-27	387	610
⋮	⋮	⋮	⋮	⋮
3577963012832148	1087770692	2013-05-14	770	888
3578260099304910	1087770692	2013-05-14	1187	627
3578283050850408	1087770692	2013-05-15	9768	2969
3578306295561973	1087770692	2013-05-15	560	466
3578324054064691	1087770692	2013-05-15	916	1392

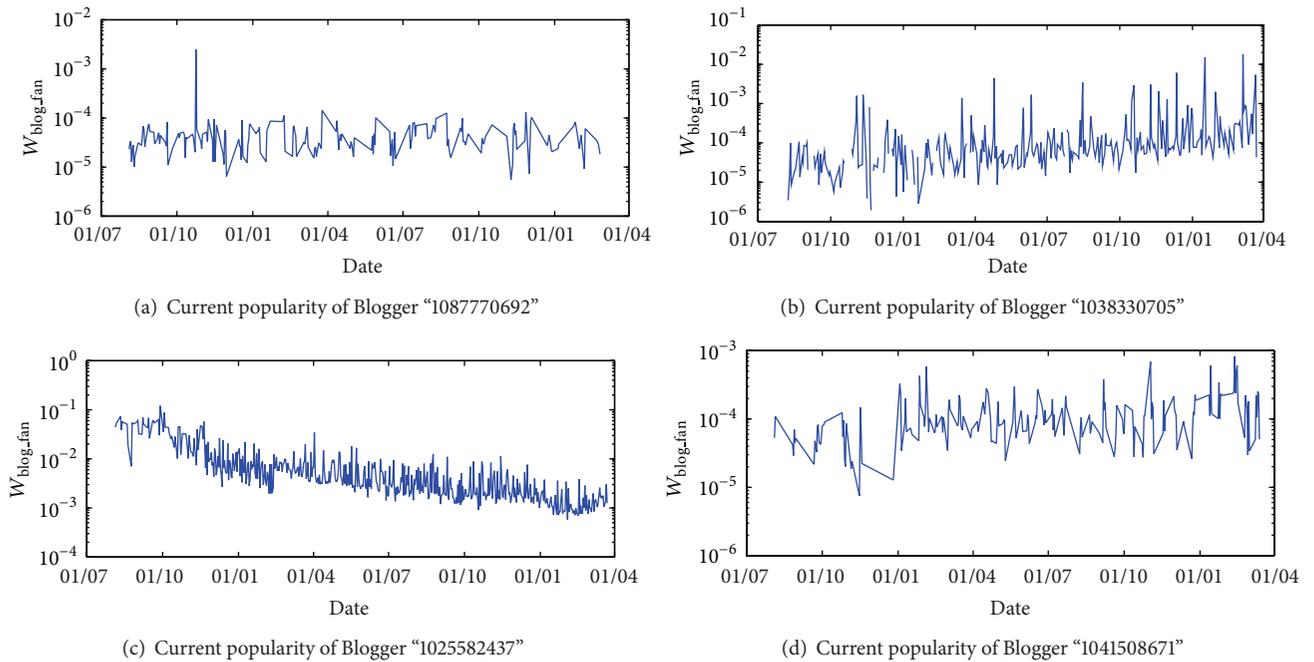


FIGURE 13: Current popularities of microbloggers from 2011 to 2013.

where  $Blogs^*$  represents the blog set posted by the given microblogger one day;  $w^*$  represents the number of blogs posted by him/her this day. As the number of one’s fans remains stable over the time, we use the total number of one’s fans  $f$  to represent the number of fans on the given day. For example, ID1087770692 posted three microblogs on August 4, 2011, and the number of his fans is 33,616,621. Therefore, the current popularity on August 4, 2011, can be

calculated according to formula (10), which is  $5.31 \times 10^{-5}$ . We can perform the analogous method to calculate the current popularities on the other days.

We chose four microbloggers to compare their current popularities and analyze their changing over the time. As shown in Figure 13, the current popularities of the four microbloggers may vary in a period, but the overall trends are relatively stable. The current popularities of ID1087770692

are mainly concentrated in the range between  $10^{-5}$  and  $10^{-4}$ ; the current popularities of ID1038330705 grow from  $10^{-5}$  to  $10^{-3}$ ; the current popularities of ID1025582437 drop from  $10^{-1}$  to  $10^{-3}$ ; the current popularities of ID1041508671 are mainly concentrated around  $10^{-4}$ . The microbloggers may sometimes post some interesting original blogs, which will attract many feedbacks from others; then his/her current popularity on that day may be a very high value. In general, the feedbacks of one's blogs mainly come from his/her close friends; therefore, the number of feedbacks of one's blog will remain stable over the time.

## 6. Conclusion

The popularities of microbloggers are hard to evaluate based on transitional network structure models (PageRank algorithm [14]). Some microbloggers may build a very long watch lists, the network built based on fan lists and watch lists cannot truly reflect one's connection strength, or popularity. In this paper, we proposed an electromagnetic field theory based model to analyze the popularities of microbloggers. The concept of the source in microblogging field is first put forward, which is based on the concept of source in the electromagnetic field; then, one's microblogging flux is calculated according to his/her behaviors (send or receive feedbacks) on the microblogging platform; finally, we used three methods ( $M_{\text{Blog}}$ ,  $M_{\text{Fan}}$ , and  $M_{\text{Blog\_Fan}}$ ) to calculate one's microblogging flux density, which can represent one's popularity on the microblogging platform. In the experimental work, we evaluated our model using real microblogging data and found that  $M_{\text{Blog\_Fan}}$  can best reflect one's popularity compared with other two metrics. We also compared our model with the classic PageRank algorithm; and the results show that our model is more effective and accurate to evaluate the popularities of the microbloggers. The contributions of this paper can be summarized as follows: (1) the proposed popularity evaluation metric- $M_{\text{Blog\_Fan}}$  is effective and reliable to evaluate the real influence of bloggers in Sina microblogging platform; (2) the popularities of microbloggers are different over the time; however, their overall trends are relatively stable. Some big oscillations may happen due to the contents of their released blogs.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This work is supported by the NUAA Fundamental Research Funds (NS2013090).

## References

- [1] K. Musiał and P. Kazienko, "Social networks on the Internet," *World Wide Web*, vol. 16, no. 1, pp. 31–72, 2013.

- [2] Z. Xia and Z. Bu, "Community detection based on a semantic network," *Knowledge-Based Systems*, vol. 26, pp. 30–39, 2012.
- [3] Z. Bu, C. Zhang, Z. Xia, and J. Wang, "A fast parallel modularity optimization algorithm (FPMQA) for community detection in online social network," *Knowledge-Based Systems*, vol. 50, pp. 246–259, 2013.
- [4] Z. Bu, Z. Xia, J. Wang, and C. Zhang, "A last updating evolution model for online social networks," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 9, pp. 2240–2247, 2013.
- [5] Z. Bu, Z. Xia, and J. Wang, "A sock puppet detection algorithm on virtual spaces," *Knowledge-Based Systems*, vol. 37, pp. 366–377, 2013.
- [6] Z. Xia, "Fighting criminals: adaptive inferring and choosing the next investigative objects in the criminal network," *Knowledge-Based Systems*, vol. 21, no. 5, pp. 434–442, 2008.
- [7] D. Scanzfeld, V. Scanzfeld, and E. L. Larson, "Dissemination of health information through social networks: twitter and antibiotics," *The American Journal of Infection Control*, vol. 38, no. 3, pp. 182–188, 2010.
- [8] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi, "Short and tweet: experiments on recommending content from information streams," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, pp. 1185–1194, April 2010.
- [9] K. Lerman and R. Ghosh, "Information contagion: an empirical study of the spread of news on digg and twitter social networks," in *Proceedings of the AAAI Conference on Weblogs and Social Media*, pp. 90–97, 2010.
- [10] P. Chen, H. Xie, S. Maslov, and S. Redner, "Finding scientific gems with Google's PageRank algorithm," *Journal of Informetrics*, vol. 1, no. 1, pp. 8–15, 2007.
- [11] Y. Xu, Y. Zhou, and K. Chen, "Observation on Spammers in Sina Weibo," in *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, Atlantis Press, 2013.
- [12] J. Letierce, A. Passant, J. Breslin, and S. Decker, "Understanding how twitter is used to spread scientific messages," in *Proceedings of the Web Science Conference*, pp. 1–8, 2010.
- [13] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM Internet Measurement Conference (IMC '07)*, pp. 29–42, October 2007.
- [14] J. Zhao, J. Wu, and K. Xu, "Weak ties: subtle role of information diffusion in online social networks," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 82, no. 1, Article ID 016105, 2010.
- [15] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, pp. 591–600, April 2010.
- [16] S. Yun, H. Do, and H. Kim, "Analysis of user interactions in online social networks," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 1–2, 2010.
- [17] C. C. Chen, T. H. Y. Chiu, Y. J. Joung, and S. Chen, "An examination of online social network properties with tie-strength," in *Proceedings of the Pacific Asia Conference on Information Systems (PACIS '11)*, pp. 41–53, July 2011.
- [18] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel, "Social media recommendation based on people and tags," in *Proceedings of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10)*, pp. 194–201, July 2010.

- [19] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Behavioral analyses of information diffusion models by observed data of social network," in *Advances in Social Computing*, vol. 6007 of *Lecture Notes in Computer Science*, pp. 149–158, Springer, Berlin, Germany, 2010.
- [20] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 807–816, July 2009.
- [21] J. Yang and C. Scott, "Predicting the speed, scale, and range of information diffusion in twitter," in *Proceedings of the AAAI Conference on Weblogs and Social Media*, pp. 355–358, 2010.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

