Analyzing the Influential People in Sina Weibo Dataset

Qing Liao[†], Wei Wang[†], Yi Han[‡], Qian Zhang[†]
[†]Department of Computer Science and Engineering
Hong Kong University of Science and Technology, Hong Kong
{qnature, gswwang, qianzh}@ust.hk
[‡]Department of Electronic Engineering and Computer Science
Peking University, Bei Jing
yihan@pku.edu.cn

***Abstract*—With the increasingly rapid growth of micro-blogging services, influence analysis is becoming a very important topic in this area. Sina Weibo, one of the largest mirco-blogging services in China, has provided a new operation *comment-only*, which allows users to give feedback on a post without forwarding. However, most existing works focus on Twitter, which fail to consider this new operation, therefore no previous works are suitable for the Sina Weibo dataset. In paper, we propose a new influence measurement method called *WeiboRank* on Sina Weibo which considers comment operation. Furthermore we analyze why a particular user is influential based on tracing the source of influence to find out which aspects contribute to influence. The experiment show that we can accurately find the most influential individuals among entire social networks, while the running time of the algorithm increases linearly with any increase in data size, which is suitable for large scale networks.**

I. Introduction

Nowadays, micro-blogging services, such as Twitter, Google+ and Sina Weibo, have become an important way to keep up with the news and exchange information. It not only provides a new way to collect huge data set that reflects behaviour patterns of users in large scale networks, but also affects people in the real world. Twitter, as the original mirco-blogging service, has been studied for the past several years. Besides Twitter, a similar platform, called Sina Weibo, one of the Chinas largest micro-blogging services, gained 200 million registered users from August 2009 to August 2011. Studies on the properties of micro-blogging formed by user relationships and interactions have attracted much interest. Analyzing the social influence of celebrities in online social networks (OSNs) has attracted many researchers. This is mainly due to the many benefits, such as helping us to better understand what the latest trends are and how advertising can be more effective in the online commercial field.

There are several methods to measure influence based on different online platforms, in particular, Twitter [1][2][4][5][6][7][8][9][10]. However, there are several limitations to these previous works. First, Sina Weibo, which is one of the largest micro-blogging services in China, exclusively provides a new operation called *comment-only* so that existing works are not suitable for this platform. Second, previous works have mainly focused on the accuracy of identifying influential users and the efficiency of the algorithm. However, they do not theoretically analyze why a particular user is important.

In order to overcome the above two shortcomings, we propose a novel influence measurement called *WeiboRank* which considers a new operation comment-only. Comment-only can reflect the information feeback ability which Twitter cannot. Furthermore, the comment-only operation can avoid "follower-buying" and "retweet-buying" where two business cases can purchase some accounts to follow a particular user or retweet others tweet to build a fake social influence. WeiboRank consists of two parts; a single-dimension influential measurement and identifying influence people via Skyline. In the first part, we adopt PageRank-like to evaluate the influence under a single dimension and in the second part, we adopt Skyline to extend the influence result into follow, repost and comment dimensions to help identify influential people in the Sina Weibo dataset. Then we propose a new concept called "dependence" to trace the source of influence. Tracing the source of influence helps us to understand why a particular user is influential as we discover which factors and users contribute to the influence.

To the best of our knowledge, this is the first work that analyzes the social influence in Sina Weibo. The contributions of this paper are threefold:

- We compare the differences between Twitter and Sina Weibo and study the relationship between comment, repost and follow through qualitative and quantitative aspects. We show that existing methods for Twitter are not suitable for Sina Weibo.
- We propose a novel method called WeiboRank which considers the new operation: comment, to evaluate the influence score based on Sina Weibo.
- We theoretically analyze why a particular user is important based on tracing the source of influence in the rather large Sina Weibo dataset.

The rest of this paper is organized as follows. Section II presents observations on Sina Weibo vs. Twitter and states

(a) Retweet in Twitter

(b) Reply in Twitter

(c) Three operations in Sina Weibo

Fig. 1. Operations on Twitter vs. Sina Weibo

the problem. Details about the novel influence measurement method called WeiboRank discussed in Section III. Section IV analyzes why a particular user is influential based on tracing its source of influence. In Section V the experimental results show that we can find most of the influential individuals in the Sina Weibo. Some related work is listed in Section VI. Finally, we summarize the paper in Section VII and discuss future work.

II. Preliminaries

A. Sina Weibo vs. Twitter

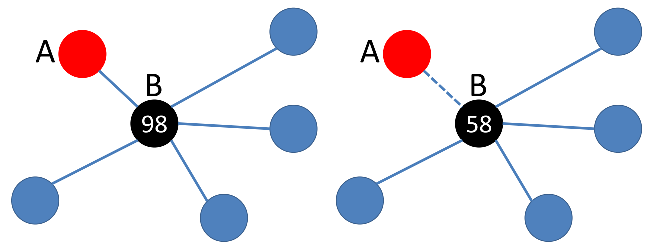
In Twitter, there are only two types of operations provided:

- **retweet:** retransmit the original tweet to their followers;
- **reply:** respond to or comment to others' tweets, as long as retweeting.

Users can retweet or reply a tweet as shown in Figure 1(a) and 1(b). Both retweet and reply operations retransmit the original tweet, which means if a user retweets or replies to a tweet, his or her followers can receive the original tweet.

In Sina Weibo, three types of operations are provided:

- **repost-only**: similar to retweet in Twitter;
- **repost-and-comment**: similar to reply in Twitter;
- **comment-only**: similar to reply in Twitter but without retweeting.



(a) Influence Score With Node A (b) Influence Score Without Node A

Fig. 2. Influence Score of Node B

As shown in Figure 1(c), Sina Weibo allows users to comment and repost at the same time. If a user does not choose the repost-and-comment operation, the user will do repost-only or comment-only operations. Both Repost and Repost-and-comment retransmit the original post, which operates similarly to Twitter. However, Sina Weibo has comment-only. If the user adopts this new operation, the comment will only show under the original post as in the bottom of Figure 1(c) and that user's followers are unable to access the comment and original post.

The comment-only operation is the key difference between Sina Weibo and Twitter, because Twitter does not distinguish between the effects of retransmission and the comment operation. Repost and comment are two different ways to transmit information in Sina Weibo, however, the repost operation is mainly focused on the retransmission of information. It also has a stronger ability to diffuse information. However, the comment operation mainly focused on the feedback ability and communication ability of information.

B. Problem Definition

There are two definitions which need to be declared before we state the problem.

1) Evaluating Social Influence: In our work, we consider the new operation, comment-only and adopt Page Rank to evaluate influence in Sina Weibo.

2) Tracing the Source of Influence: Some scholars exploit entities social influence by analyzing the topological structure. Such influence can be considered as the social standing of entities, is usually measured by scoring functions. Different vertices in a social network may have different scoring function values. However, what makes such difference in values?

Example 1: Figure 2(b) represents the induced subgraph of Figure 2(a) by voiding the links connecting to B. The labeled number on vertices represents the a link-based authority measure. In order to analyze the dependency from A to B, we separate the vertex A from the network by disconnecting its links. If A can effect the influence score of B, we say that A is one of the sources of influence on B. If edge $\{A,B\}$ was disconnected, the influential score B decrease dramatically. We say there has a strong interdependency between A and B. Otherwise, we say that is has a weak relationship. We consider that A is one of the sources of the B's influence.

Such situation is quite common in Sina Weibo. For instance, a famous user A may not public posts by himself, but repost a specified user B's posts frequently. A's broadcast amplifies

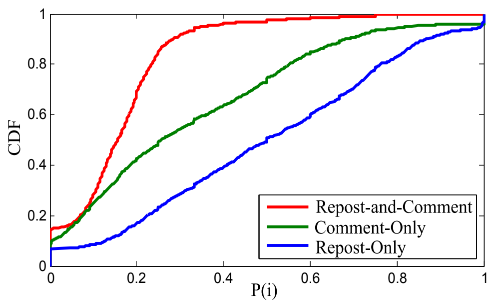


Fig. 3. CDF of Three operations

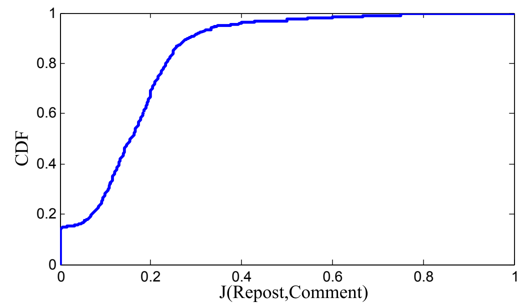


Fig. 4. The CDF of Jaccard Similarity Coefficient Between Repost and Comment

the propagation range of B's information. B's influence score can also be boosted by A's support. Therefore, we consider A is one of the sources of user B's influence. There also exists some artificially constructed linkage structure which can be used for boosting some specified vertices' influence ranking score. Study of these issues will help to detect the linkage web spam sites, or improve the ranking algorithms or search results.

3) Problem Statement: We consider follow, repost and comment to identify the most k influential users in Sina Weibo dataset and trace the sources of influence to analyze why a particular influential user is important.

III. Measuring Social Influence

A. Correlation among Operations

As mentioned before, since most existing works are based on Twitter, the comment-only operation has not been considered before. Two questions need to be answered first. Question 1: Is the comment-only operation important to measure influence on Sina Weibo? Question 2: Can the comment operation be replaced by existing operations(follow and repost)?

Question 1 considers the importance of the comment-only operation in Sina Weibo. If only a few users use the comment-only operation, it may only have a small impact on Sina Weibo, which means there are no major differences between Sina Weibo and Twitter. Question 2 tries to understand the effect of the comment-only operation on Sina Weibo. If there is a strong correlation between comment and repost/follow, it would indicate that the effect of the comment operation is similar to the repost or follow operations under the influence measurement so the comment operation is no longer needed.

First, we answer Question 1. We randomly collect 1132 posts on Sina Weibo and calculate how many people use these three operations relate to each post. Figure 3 shows the Cumulative Distribution Function (CDF) of the three operations on Sina Weibo. $P(i)$ is the percentage of the special behaviours *i* on a message. The definition of $P(i)$ is as follows:

$$P(i) = \frac{Number(i)}{Number(R \cap C) + Number(C) + Number(R)}, \quad (1)$$

where i is the operations, and $i \in \{R \cap C, C, R\}$, C the comment-only operation, R the repost-only operation, and $R \cap C$ the repost-and-comment operation. According to Figure 3, we find that repost-only and comment-only operations are common in Sina Weibo, and repost-only operations are more popular than comment-only operations. However, most of the users do not use repost and comment operations at the same time. In conclusion, the comment-only operation is quite common in Sina Weibo. Since Twitter does not have the comment-only operation, there is a big difference between Sina Weibo and Twitter.

Then, we answer Question 2. We analyze the correlation between repost and comment, because Twitter does not distinguish between these two operations. Its "Reply" function is a combination of the "Repost" and "Comment" functions.

When user u publishes a post d, and user v has commented on the post d, $c(v,d) = 1$, otherwise, $c(v,d) = 0$. When user u publishes a post d, and the user v has reposted the post d, $r(v,d) = 1$, otherwise, $r(v,d) = 0$.

We adopt the Jaccard Similarity Coefficient to calculate the correlation between repost and comment. The formulation is shown below:

$$J(d) = \sum_{v \in V-u} \frac{c(v,d)r(v,d)}{c(v,d) + r(v,d) - c(v,d)r(v,d)}, \quad (2)$$

where $\sum_{v \in V-u} c(v,d)$ is the number of users who commented on the message d, $\sum_{v \in V-u} r(v,d)$ is the number of users who reposted on the message d and $\sum_{v \in V-u} c(v,d)r(v,d)$ is the number of users who commented and reposted on the message d at the same time. Comments and reposts tend to be irrelevant when $J(d)$ is small, $0 \leq J(d) \leq 1$.

Figure 4 shows the CDF of the Jaccard Similarity Coefficient between repost and comment, and the correlation coefficient between comment and repost is low. For most of the message Jaccard Cofficient is less than 0.4, which means repost and comments are independent of each other, and cannot replace one another.

B. Single-Dimension Influence Measurement

WeiboRank algorithm consists of two parts: The single-dimension influential measurement and identifying influential people via Skyline. The first part is based on PageRank-like, which evaluates the influence score under a single-dimension and the second part extends the first part into multiple dimensions.

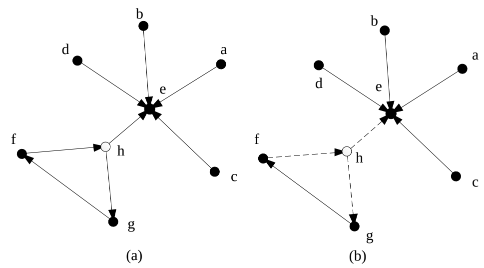


Fig. 5. An Example of Dependence

	a	b	c	d	e	f	g	h
A(*,G)	2.4	2.4	2.4	2.4	27.3	20.7	16.5	25.4
A(*,G_a)		2.4	2.4	2.4	25.2	21.4	17.1	26.3
A(*,G_b)	2.4		2.4	2.4	25.2	21.4	17.1	26.3
A(*,G_c)	2.4	2.4		2.4	25.2	21.4	17.1	26.3
A(*,G_d)	2.4	2.4	2.4		25.2	21.4	17.1	26.3
A(*,G_e)	2.0	2.0	2.0	2.0		29.9	29.9	29.9
A(*,G_f)	5.1	5.1	5.1	5.1	58.3		11.0	5.1
A(*,G_g)	4.2	4.2	4.2	4.2	61.7	4.2		12.5
A(*,G_h)	5.1	5.1	5.1	5.1	52.4	16.9	5.1	

TABLE I
AN EXAMPLE OF AUTHORITY RANK IN A SOCIAL NETWORK

The Weibo network contains four elements $G = \{V, E, F, A\}$, where V is the set of users. v a particular Weibo user, $F = \{follow, repost, comment\}$ the three types of operations. E is the directed edge set. $e = \langle v, u \rangle \in E$, if and only if there exists the above operations between v and u, that is $f(u, v) = 1$,$f \in F$. The edge is directed from user v to u, if user v reposts a post from user u. The edge is directed from user u to v, if user v has a comment on the post of user u. $A : V \rightarrow R$ is the influence measurement function which maps the node set to the set of real numbers. Since we consider that, the same Weibo user can reflect different aspects of social influence depending on the angles it is viewed from. We choose $A_f(v)$ which represents the influence value of user v under the f dimension which can be follow, repost and comment dimensions.

Definition 1 (Top-k Queries): Finding k persons that have the highest overall scores.

A_f is the influence measurement function and in our work we choose PageRank-like as A_f. The formulation of single-dimension influence measurement in Sina Weibo is shown as follows:

$$A_f(u) = \epsilon + \sum_{v \in M(u)} \frac{A_f(v)}{L(v)}, \tag{3}$$

where ϵ is the constant to ensure that the result is not negative. Where $M(u) = \{v | \langle u, v \rangle \in E\}$ represents the neighbour of user v, and $L(v)$ means the number of outlinks which user v has. In Sina Weibo, $L(v)$ presents the number of received posts. We randomly assign a score value to each node in the graph, and then apply the above equation (3) iteratively until the influence scores converge.

CPU	Intel i7 4 Cores 2.83GHz
Memory	48GB
OS System	KylinOS kernel version 2.6.18
Run Environment	Java Runtime Environment 1.6

TABLE II
TEST ENVIRONMENT

Follow	Repost	Comment	Integration
1182389073	ID:1182389073	ID:1880094403	ID:1182389073
1861636934	ID:1110687265	ID:1660452532	ID:1880094403
1618051664	ID:2286908003	ID:1914281205	ID:1702549133
1642088277	ID:2459942972	ID:1409871493	ID:1654592030
1644307787	ID:2054302531	ID:333598818	ID:1737650455
1702549133	ID:2308961447	ID:1989660417	ID:1737650455
1059018092	ID:61897999	ID:2104965662	ID:1810632930
1189591617	ID:1840604224	ID:1660252010	ID:1723981287
2096085361	ID:1959210547	ID:1799675814	ID:1763978027
2738419862	ID:2267147875	ID:1654592030	ID:2308961447

TABLE III
THE RESULTS OF INFLUENCE RANK ON THE OPERATION NETWORKS

C. Identifying Influential People via Skyline

The above subsection introduces a single-dimension influence measurement through the PageRank-like method. In this subsection, we describe how to combine a multi-dimension result together to return the final result. Some existed works only based on one dimension such as follower or retweet to evaluate influential users. Another previous works which consider two dimensions, but they cannot decide a reasonable weight in their work. In our work, we choose R-tree as the data structure and base it on the Skyline method to combine the three-dimension: follow, repost and comment and to avoid any dimension weight decision. The definition of Skyline is shown below:

Definition 2 (Domination): For two vertices $u, v \in V$, u **dominates** v, denoted by $u \succ v$, if $\forall f \in F$, $\mathcal{A}_f(u) \geq \mathcal{A}_f(u)$, and not all the three equalities hold.

Definition 3 (Skyband): In the Weibo network $G = \langle V, E \rangle$, a vertex $v \in V$ is a **m-skyline vertex** if there exists no other **m** vertices $u \in V$ such that $u \succ v$.

m-skyband [3] computation is a kind of query, which returns all m-skyband vertices.

IV. TRACING THE SOURCE OF INFLUENCE

The previous section describes how to evaluate a user's influence and in this section, and we analyze why a particular user is influential. We also propose a new concept "dependence" to trace the source of the influence.

In Figure 5, h represents the induced subgraph of G(Figure 5) by voiding the links connecting to a. In order to analyze the dependency of a vertex, we separate the vertex from the network by disconnecting its links instead of deleting it. In Table I, the first row shows the PageRank-like scores of vertices in graph G. $\mathcal{A}_G(v)$ is the influence measurement score of v in G. G_v represents the induced subgraph of G by separating vertex v. Apparently, if a, b or c are separated from G, e's pagerank-like score does not change a lot. However, if f, g or h are separated from G, e's pagerank-like score changes to greater than 50%, even f and g are not connected to e directly. Intuitively, for a vertex v, if a substructure is separated from the original network, and v's authority changes a lot, v can be regarded closely dependant on this substructure.

Definition 4 (Dependency): Given a four-tuple set $G = \{V, E, F, A\}$ and an individual $u, v \in G(V)$, we define the dependence $dep(u \to v)$ from u to v as:

$$dep(u \to v) = \log \frac{A(v, G)}{A(v, G_u)}. \quad (4)$$

$dep(u \to v)$ shows the ratio of the importance of v and the original score when graph G is disconnected from u. G_u denotes the induced subgraph of $G \backslash u$, and $A(v, G)$ denotes the importance value of an individual v in network G. $A : V \to R^+$, when $dep(u \to v) = 0$, $A(v, G_u) = A(v, G)$, hence $u \to v$ does not have dependence; when $dep(u \to v) > 0$, the dependence is positive, otherwise, it is negative.

For an individual u, if removing a part of the network significantly affects the importance of v, we say v is highly dependent on that part, and this dependence can be considered to be a standard of the analysis of why a particular user is important. We can also analyze the relationship between nodes; the higher the dependence, the closer the relationship, and vice versa.

V. Experimental Evaluation

In this section, we evaluate, WeiboRank using a real Sina Weibo dataset. We collected the Sina Weibo data through the API provided by Sina Weibo for preparing work. The Sina company provides about 20% of the most reliable data available in China. Our work chooses a subset of a full Sina Weibo dataset, which includes 22,514,394 users, whom 4,387,532 users published at least one post from April 15, 2011 to July 15, 2011. As for the repost and comment network, we obtained all the reposts and comments in the three months and found 22,620,281 records. We also analyzed the dependence result and evaluated the efficiency of the WeiboRank which is implemented using Java. The test environment is shown in Table II.

A. The Results of WeiboRank

Table III shows the top 10 users in Sina Weibo under different networks. Follow column presents the result of indegree influence, which evaluates user's influence only based on the number of followers. Repost column presents the result of retweet influence, which measure the influence through the number of retweets. Comment column based on the number of comments for an individual user to find influential user. Integration column is our Weibo Rank method, which considers follow, repost and comment three dimensions together through Skyline method to find the top 10 influenced user. Since comment-only operation is an unique feature in Sina Weibo comparing with Twitter, the result of comment column can help researchers to find some new influential users which are ignored by previous works. Moreover, Weibo Rank applies Skyline method to combine three dimensions together, the final result will not miss the best influenced user in each dimension. For example, the user ID 1182389073 is Zhiqiang Ren, which is not only a famous estate agent, but also a member of the CPPCC(Chinese People's Political Consultative Conference). Many people prefer to follow or repost his posts, since house price is one of the most important problem and normal people want to understand the trend of house strategy in the future.

B. Dependence Analysis

We choose two famous users, user A and user B, in Sina Weibo to analysis their dependence rank under a repost network. As show in Figure 7, there are a few particular users have a very high dependence scores compared to user B, which means a strong relationships exists among a few particular users to user B . If we break the connections between user B and these important and influential source users, the influence of user B will decrease a lot. However, user A as shown in Figure 6 is not affected a lot under the same circumstances.

Figure 8 and 9 present the dependence results under repost and comment dimensions and each node represents a Sina Weibo user. We find that some users who have high repost ranks may have a low comment dependence score, which means even for the same user, the dependent users have different dependence ranks under different dimensions.

Figure 10 and Figure 11 show that a set of users who have the same comment dependence score has a large rang of repost dependence score. It verifies again that, repost and comment have weak correlations and different operations cannot replace each other.

C. Efficiency of Algorithm

To validate the efficiency and scalability of WeiboRank, we use several subsets with different numbers of users in a high-dimension environment, where d=1 means WeiboRank runs in the follower network, d=2 means WeiboRank runs in the follower and repost networks, and d=3 means WeiboRank runs in follower, repost and comment network. The results are shown in Figure 12. We find that, with the increasing data size, the running time of the algorithm also increases linearly. The calculation time of each subset is proportional to dimension d. The larger the value of d, the larger the dominated space we need to search, and thus the longer the runtime. Based on the above analysis, the algorithm is effective and scaleable, which is suitable for large scale networks and parallel processing.

VI. Related Work

Several works have measured influence in OSN. Some works are simply based on the in-degree method [1][4][5][6] to identify influence user, however, [1][4] analysis found that using in-degree to identify influencers is not good enough, because it does not accurately capture the notion of influence. H.Kwak et al [5] compares three different measurements of influence: the number of followers, page-rank, and number of retweets. M.Cha et al [6] also compares three different measures of influence: the number of followers, the number of retweets, and the number of mentions. Both works find

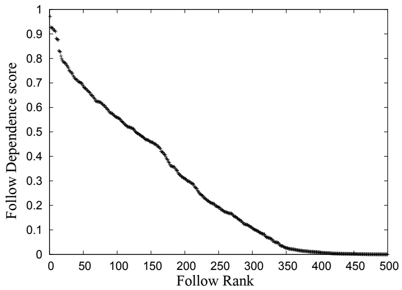


Fig. 6. Dependence Rank of User A Under Follow Dimension

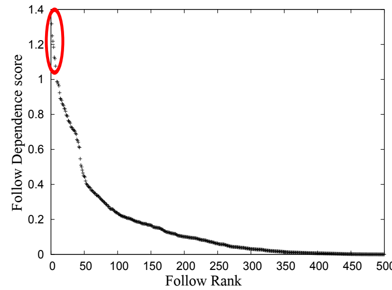


Fig. 7. Dependence Rank of User B Under Folloer Dimension

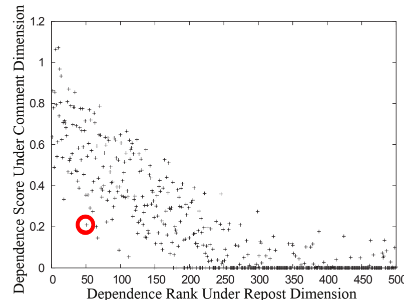


Fig. 8. Dependence Result of User A Under Repost and Comment Dimensions

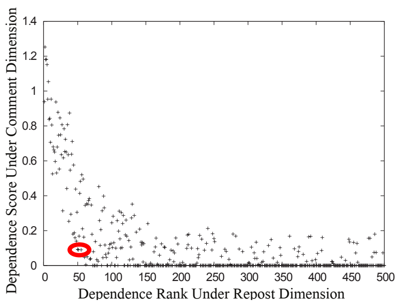


Fig. 9. Dependence Result of User B Under Repost and Comment Dimensions

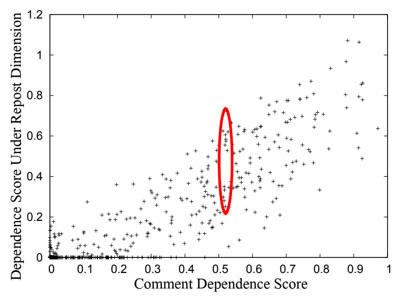


Fig. 10. Dependence Score of User A Under Repost and Comment Dimensions

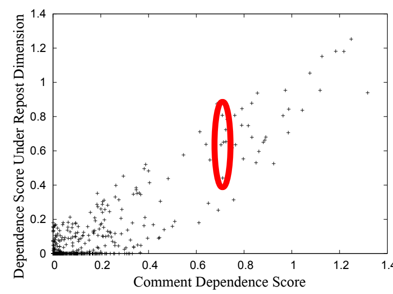


Fig. 11. Dependence Score of User B Under Repost and Comment Dimensions

that the most influential user may not have the highest influence score via the other measurements. Other works consider the link structure to evaluate influence [7][8][9][10][2]. T.H.Haveliwala [7] and J. Weng et al [8] extend the PageRank algorithm on the topic-level and proposed Topic-sensitive PageRank (TSPR) and TwitterRank algorithm to measure the influence of users on Twitter. The evaluation of [8] compared the result of TwitterRank and TSPR which showed that the different ranking results are dependent on the different measurement methods. Tang et al [9], propose the Topical Affinity Propagation(TAP) approach to model the topic-level social influence on a large network. E. Bakshy et al [2], they propose a narrow definition of influence and quantifies the influence of a given post by the number of users who subsequently repost the URL. N. Agarwal et al [10] adopts four features: inlink, comment, outlink and the length of post, to identify an influential blogger in an online community. However it has been proven that the structure of micro-bloggging and blogger platforms are different, therefore micro-blogging methods are not suitable for blogger platforms. Most of the works above are based on Twitter to evaluate influence which all do not consider the *comment-only* operation. Furthermore, existing works focus on how to identify influential users and the efficiency of algorithms, but they do not theoretically analyze why a particular user is important.

VII. Conclusion

This work studies the influence of online social network services from two aspects. First we proposed a novel method called WeiboRank based on PageRank-like and Skyline to evaluate influential users, on a novel dataset, Sina Weibo.

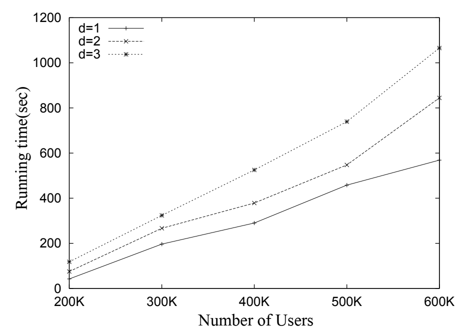


Fig. 12. Run Time of WeiboRank

Then we traced the source of influence under three-operation networks: follow, repost and comment. We also compared the differences between Sina Weibo and Twitter through quantitative analysis to proof that existing Twitter work does not relate to Sina Weibo. Experimental results show that our method can find the most influential users among previous influence evaluation methods by considering all available aspects. How to use dependence analysis to find the retweet-buying case will be researched more deeply and we will extend our work to the topic level in the future.

References

[1] J. Kleinberg, “Authoritative sources in a hyperlinked environment,” *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[2] E. Bakshy, J. Hofman, W. Mason, and D. Watts, “Everyone’s an influencer: quantifying influence on twitter,” in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 65–74.

[3] D. Papadias, Y. Tao, G. Fu, and B. Seeger, "Progressive skyline computation in database systems," *ACM Transactions on Database Systems (TODS)*, vol. 30, no. 1, pp. 41–82, 2005.

[4] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107–117, 1998.

[5] 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 conference on World wide web*. ACM, 2010, pp. 591–600.

[6] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010, pp. 10–17.

[7] T. H. Haveliwala, "Topic-sensitive pagerank," in *Conference on World Wide Web (WWW)*, 2002, pp. 517–526.

[8] J. Weng, E. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 261–270.

[9] 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*. ACM, 2009, pp. 807–816.

[10] N. Agarwal, H. Liu, L. Tang, and P. Yu, "Identifying the influential bloggers in a community," in *Proceedings of the international conference on Web search and web data mining*. ACM, 2008, pp. 207–218.