Understanding Sina Weibo Online Social Network: A Community Approach

Kai Lei*‡, Kai Zhang*, Kuai Xu†
*Shenzhen Key Lab for Cloud Computing Technology & Applications (SPCCTA)
School of Electronics and Computer Engineering
Peking University, Shenzhen 518055, P.R. China
‡Corresponding author: leik@pkusz.edu.cn
†School of Mathematical and Natural Sciences, Arizona State University

***Abstract*—Sina Weibo, one of the most popular online social networks in China, has recently become a critical medium for Internet users to disseminate and discuss breaking news, social events and other information. Although online social networks and social media have received significant attention from the research community, few studies have focused on Sina Weibo due to the lack of data collection. Given the sheer size of Sina Weibo online social network and vast amount of tweets, retweets and comments, this paper introduces a novel *community* approach for understanding Sina Weibo online social network. Specifically, we collect all Weibo users registered with Shenzhen as primary geographic location, and build a *Shenzhen Weibo community graph* based on their following or follower relationships. Our experimental results describe interesting graphical characteristics such as clustering coefficients of this community graph, and reveal the impact of user popularity on tweet influence. Through modeling interactions of Shenzhen Weibo users and their tweeted messages with bipartite graphs and one-mode projections, we analyze the similarity of retweeting and commenting activities among these users, and discuss the implications of the findings on understanding different types of user accounts and the motivations of their following and retweeting behaviors. To the best of our knowledge, this study is the first effort to introduce a community approach for understanding the community characteristics of Sina Weibo and characterizing the similarity of retweeting behaviors and following relationships.**

I. Introduction

Sina Weibo, one of the most popular online social networks in China, has recently become a critical medium for over 500 million registered users to disseminate and discuss breaking news, social events and other information. Although online social networks and social media have received significant attentions from the research community [1]–[5], few studies have focused on Sina Weibo due to the lack of data collection.

Given the sheer size of Sina Weibo and vast amount of tweets, retweets and comments, this paper introduces a novel *community* approach for understanding Sina Weibo online social network. Specifically, we collect all Weibo users registered with Shenzhen as primary geographic location, and build a *Shenzhen Weibo community graph* based on their following relationships. Our data collection effort uses a simple yet effective algorithm to crawl over 2 million users from Shenzhen Weibo community, and obtains over 75 million following relationships among these users. To understand user influence and interactions, we also collect 47 million original tweets, 4 million retweets, and 10 million comments posted by all Shenzhen users during a one-month time span.

The availability of *Shenzhen Weibo community graph* enables us to explore graphical characteristics of this community such as clustering coefficients and the distribution of user popularity. To quantify the impact of user popularity on tweet influence, we present an empirical analysis on the influence of following relationships on the retweeting and commenting activities, and characterize temporal patterns of retweeting and commenting activities on the original tweets posted by a variety of users with diverse popularity.

Through modeling interactions of Shenzhen Weibo users and their tweeted messages with bipartite graphs and one-mode projections, we also analyze the similarity of retweeting and commenting activities among these users, and discuss the implications of the findings on understanding different types of user accounts and the motivations of their following and retweeting behaviors. To the best of our knowledge, this study is the first effort to introduce a community approach for understanding the community characteristics of Sina Weibo and characterizing the similarity of retweeting behaviors and following relationships.

The contributions of this paper are three-fold:

- This paper introduces a novel approach to characterize graphical structure and small-world nature of a community online social network in Sina Weibo;
- This paper discusses the impact of user popularity on the intensity of retweeting and commenting activities on their retweeted messages and on temporal patterns of information spreading over Sina Weibo;
- This paper sheds light on the similarity of retweeting and commenting behaviors among users in Shenzhen Weibo community, and reveals the weak correlations between following behaviors and retweeting and commenting activities.

The remainder of this paper is organized as follows. Section II describes our data collection effort. Section III presents our findings of community structure of Shenzhen Weibo community, while Section IV focuses on understanding the impact of user popularity on tweeted contents. Section V is devoted to characterizing the similarity of following behaviors and retweeting activities for Shenzhen users. Section VI discusses related work, and Section VII concludes this paper.

978-1-4799-1353-4/13/$31.00 ©2013 IEEE

II. Data Collection

Our data collection efforts focus on two types of information from Sina Weibo: relationship and content. The relationship information captures the following and follower activities among users and leads to a social network graph of Weibo users, while the content consists of tweeted messages and the corresponding retweets and comments, capturing the process of information cascading over this social network as well as the influence of its users.

In October 2012 we started to crawl Sina Weibo social network using Weibo API and selected Shenzhen as the community due to its technology-savvy population. A Weibo user is considered as a Shenzhen user if he or she claims Shenzhen as the primary residence location in the user profile. As described in [6], there are two approaches for crawling online social networks: using forward links and using a combination of forward links and reverse links. A forward link exists from user A to user B if A follows B in the social network, and we say A is a *follower* of B and B is a *friend* of A. The reverse link exists from user A to user B if user A is followed by user B. The experiments in [6] show that both approaches derive similar online network graphs, but the forward-link approach is much more cost-effective thanks to its simplicity. Thus in this work, we adopt the forward-link approach for crawling Shenzhen users from Sina Weibo social network. Algorithm 1 illustrates our algorithm of crawling all Shenzhen users. Note that this algorithm could be applied to crawl other social community graphs such as a professional society, a university, and an organization. Using this algorithm we successfully obtain over 2 million Shenzhen Weibo users and over 75 million direct following relationships among these users.

Algorithm 1 The algorithm of crawling all Shenzhen users of Sina Weibo online social network.

Require: a set of seed users S: the top N users of Sina Weibo social network; all Shenzhen users U, initially is empty;

```
push S into Q_tasks;
while Q_tasks is not NULL do
    u = pop Q_tasks;
    find all the friends of u: friend_u;
    for v ∈ friend_u do
        if v is from Shenzhen and v hasn't been crawled then
            U = U + u;
            push v into Q_tasks;
        end if
    end for
    mark u as being crawled.
end while
```

The second phase of our data collection effort is to crawl all the original tweets posted by Shenzhen Weibo users, and collect all the retweets and comments on these tweets. In November 2012 we used Weibo API to continuously collect over 47 millions original messages tweeted by Shenzhen users, and obtained over 4 million retweets and over 10 million comments on these tweets written by other users in Shenzhen Weibo community.

III. Graphical Characteristics of Sina Weibo Community Graphs

In this section we characterize the community graph that consists of all Shenzhen users from Sina Weibo online social network. Specifically, we present the reciprocal analysis of the community graph based on the symmetry of the following relationships, and study the distribution of user popularity measured by the number of follower count. Subsequently, we use clustering coefficients of the community graph to study the small-world nature of this community.

A. Building Sina Weibo Community Graphs

In this paper we consider all Shenzhen users in Sina Weibo to form a social "community" based on the same geographical location. These users share many social interests such as local news, weather, and events, thus analyzing the online behavior of this community could have a broad range of applications such as spreading emergency information and understanding the latest social trends. Combining the following and follower relationships, we could build a community social network graph that consists of these users and their relationships. For ease of presentation, we use "Shenzhen Weibo community" and "Shenzhen Weibo community graph" to represent all Shenzhen users and their social network graph, respectively. The availability of *Shenzhen Weibo community graph* enables us to explore the graphic characteristics of this representative Sina Weibo community.

Let $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$ represents the community graph, where $\mathcal{N}$ denotes the set of users in the community and $\mathcal{E}$ denotes the set of following and follower relationships among these users. Our crawling effort in the experiment leads to a community graph $\mathcal{G}$ with over 2 million users and 75 million following relationships (edges) among these users. Specifically, $|\mathcal{N}| = 2,234,591$, and $|\mathcal{E}| = 75,974,458$.

Unlike the friend relationship in Facebook online social network, the following or follower relationships in Sina Weibo and Twitter are uni-directional. In other words, a user A in Sina Weibo follows another user B, but the user B might not necessarily follow A. A popular celebrity might have millions of followers on Sina Weibo, but does not follow his or her followers. On the other hand, it is also common to observe two users following each other, which is referred to as *reciprocal* following relationship. An important metric to measure the degree of reciprocal followings in a social network is *reciprocal rate*: $rr = \frac{|\mathcal{E}_{\mathcal{R}}|}{|\mathcal{E}|}$, where $\mathcal{E}_{\mathcal{R}}$ represents the set of edges in $\mathcal{G}$ that have corresponding reciprocal edges. As $43,557,116$ following relationships are reciprocal, the reciprocal rate of Shenzhen Weibo community graph is 0.57, which is higher than the reciprocal rates of the entire Sina Weibo online social network [7] and Twitter [6].

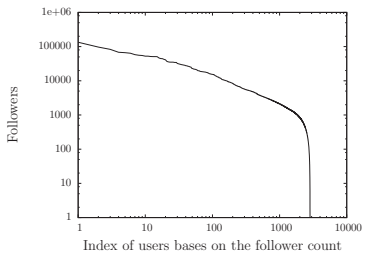


Fig. 1. The distribution of followers for Shenzhen users

Next we study the distribution of *user popularity* measured by the follower count. Note that we only consider followers

that are also in the Shenzhen Weibo community. As illustrated in Figure 1, there exist a small number of highly popular users such as sport stars and CEOs of large local companies, who have thousands of Shenzhen followers. We find that less than 1% of users in the community have over 1000 followers. On other other hand, the majority of users in the community have only a small number of followers, and nearly 60% of users have less than 10 followers.

B. Clustering Coefficient of Weibo Community Graphs

Clustering coefficient is a widely used measure to quantify the "closeness" of nodes in a graph, and has recently been used to reflect the "small world" nature of social networks and communication networks. Specifically, the clustering coefficient of user u in the community social graph $\mathcal{G}$ captures the number of following relationship among the u's followers. For example, given the user u with m followers, the possible number of following relationships among these m followers is $m \times (m-1)$. Let λ denote the number of observed following relationships among these m followers. Then the clustering coefficient of the user u, CC_u is calculated as $\frac{\lambda}{m \times (m-1)}$. Clearly, $0 \leq CC_u \leq 1$. A clustering coefficient CC_u of 0 indicates that the followers of the user u do not have any direct following relationships, while a clustering coefficient CC_u of 1 tells that any two followers of the user u follow each other. Thus a larger value of CC_u reveals a cluster of users with close following or follower relationships.

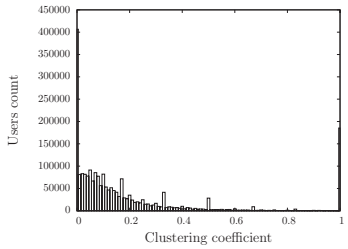


Fig. 2. The distribution of clustering coefficients for Shenzhen users.

Figure 2 shows the distribution of clustering coefficients for users in Shenzhen Weibo community. In average, the clustering coefficient of all users is 0.203. In general, as shown in Figure 2 most of the users have a small clustering coefficient, while a certain number of users have a coefficient clustering of 1 indicating the existence of close groups of users within Shenzhen Weibo community.

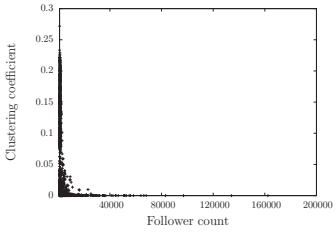


Fig. 3. The correlation between follower counts and clustering coefficients.

The varying degree of clustering coefficients motivates us to further understand the correlations between user popularity in the social community and their clustering coefficients. As illustrated in Figure 3, we find that i) in general, users with lower popularity tend to have higher clustering coefficients, which suggest that the followers of these users could possibly know each other or these users have similar interests in certain topics, ii) users with higher popularity have lower clustering coefficients, indicating that the followers of these popular users do not necessarily know each other or share common interests, except following the popular users in Shenzhen Weibo community.

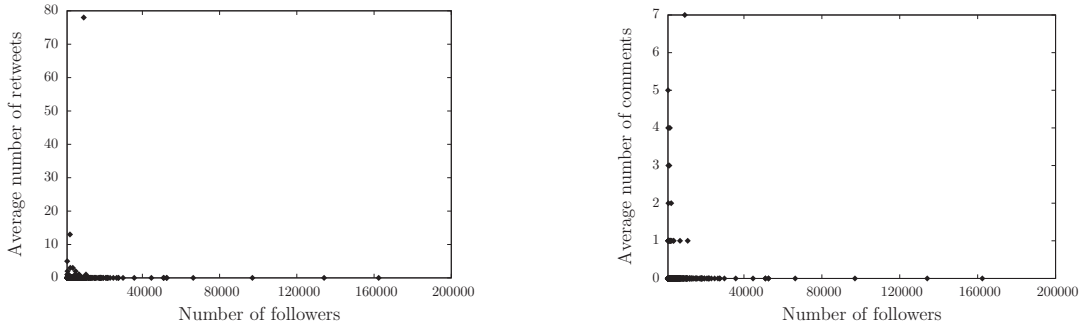
IV. Correlation Analysis of User Popularity and Tweet Influences

In this paper we explore the impact of user popularity in Shenzhen Weibo community on the tweeted messages. In particular, we attempt to understand: i) whether tweets from users with high popularity lead to significant retweeting and commenting activities; ii) whether most of the retweeting and commenting activities come from the followers; and iii) whether the tweets from different users have similar or different spreading speeds over online social networks.

Figure 4[a][b] illustrate the correlations between follower counts of *popular* Shenzhen users (users with at least 1000 followers) and the average numbers of retweets and comments received by their tweeted messages, respectively. As shown in these figures, the high user popularity does not always lead to the high retweeting and commenting activities. However, the tweets of a few users with relatively smaller numbers of followers actually receive a surprisingly high number of retweets or comments. For example, the official Weibo account for Shenzhen University, the largest college in Shenzhen, has only 9627 followers, but each tweet posted by this account receives an average of 78 retweets, the highest retweet counts across all users in Shenzhen Weibo community. The in-depth analysis of these retweets discover that most of the retweets come from the college students of this university, who are very active on Weibo and other online social networks.

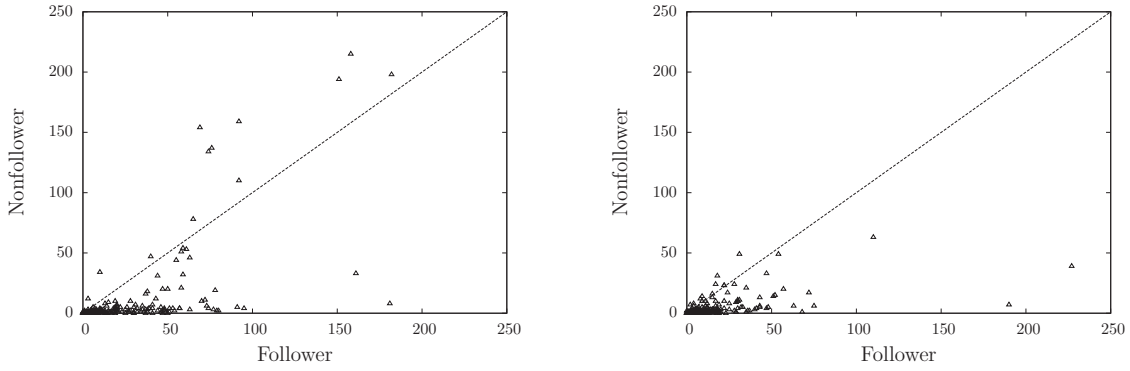
Next we explore whether the retweeting and commenting activities mostly come from the followers of the users who posted the original tweets. Figure 5[a] shows the difference between the average number of retweets from the followers of popular Shenzhen users and the average number of retweets from the non-followers. Our general observations are i) both followers and non-followers of the users who tweeted contribute to retweeting and commenting activities; and ii) in general the majority of retweets indeed come from the followers of the users. As illustrated in Figure 5[b], similar observations hold as well for the commenting behaviors of the tweets posted by users in Shenzhen Weibo community.

Characterizing the speed of retweeting and commenting behaviors of Sina Weibo is important for understanding user interactions and influences. In this study, we use case studies to analyze temporal patterns of retweeting and commenting activities of tweets from users with a variety of follower counts. Specifically, we select five users which rank 1st, 5th, 10th, 15th and 50th based on the user popularity, and measure the percentage growth of retweeting and commenting activities over time. Our experimental results reveal that the retweeting activities of tweets from users with higher user popularity tend to spread much faster than those posted by users with lower



(a) Retweeting activities (b) Commenting activities

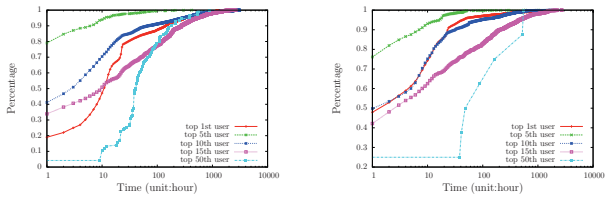
Fig. 4. The correlation between follower counts and the volumes of retweeting and commenting activities



(a) Retweets activities (b) Commenting activities

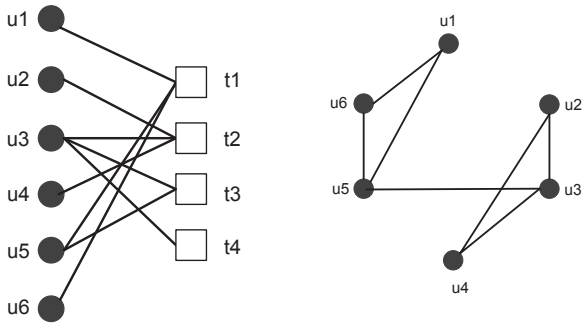
Fig. 5. The difference between followers and non-followers who retweet and comment on the tweeted messages from popular users in Shenzhen Weibo community.

popularity. For example, as shown in Figure 6[a], in average over 50% of retweets on the tweets from the top 10th user happen within 10 hours after the original message was posted, while 50% retweets of the tweets from the top 50th happen after nearly two days. Similarly, Figure 6[b] shows that tweets from users with higher user popularity receive comments at a much faster pace than those with lower popularity.



(a) Temporal pattern of retweets (b) Temporal pattern of comments

Fig. 6. Temporal patterns of retweeting and commenting activities.



(a) Bipartite graph (b) One-mode projections

Fig. 7. Modeling users interactions of Shenzhen Weibo community with bipartite graphs and one-mode projection: (a) the retweeting interactions between six Weibo users in Shenzhen Weibo community (u_1, ..., u_6) and four tweets (t_1, ..., t_4); (b) the similarity of retweeting activities among four users.

V. Similarity Analysis of Retweeting and Commenting Behaviors

In this section, we study the retweeting and commenting similarity of users to study i) whether Weibo users share similar interests in retweeting or commenting original tweets? and ii) whether such users have a similar set of friends in the social network graph? Towards answering these questions, we first build a bipartite graph, as illustrated in Figure 7[a], between Weibo users and the tweets they have retweeted or commented, since bipartite graphs are widely used to model bipartite interactions such as actors and movies, authors and publications [8], [9].

Subsequently, we build one-mode projection graphs to connect Weibo users based on the bipartite graphs between users and retweets. In Figure 7[b], an edge of one-mode projection graph between a pair of Weibo users represents that both users have retweeted or commented at least one common

tweet. The weight of the edge captures the number of the shared tweets both user have retweeted or commented.

We first study the correlation between the number of shared friends and similar retweeting and commenting behaviors for Shenzhen users, and find that a high number of shared friends do not often lead to the similar behavior of retweeting and commenting activities. For example, as shown in Figures 8[a][b], many users share hundreds of friends, but do not retweet or comment on a single common tweet. This finding suggests that many of these accounts are not actively posting tweets, retweets or comments, and they simply follow a number of popular Weibo users and passively view the tweets.

To validate the above conjecture, we further explore whether Weibo users with a large number of friends actively retweet or comment on his or her friends' tweets. As shown in Figure 9, we find that a large number of users with many friends do not retweet or comment as aggressively as their following activities. In other words, most of these Weibo users simply read or ignore these tweets, but do not participate the process of information spreading in Sina Weibo online social network. On the other hand, some users with less friends actively participate information spreading via retweeting and commenting the tweeted messages from their friends. Such findings have important applications of classifying Weibo user accounts and predicting information spreading over Sina Weibo online social network.

VI. Related Work

In recent years, online social networks such as Facebook and Twitter have received significant attention from the research community. A significant body of research has been devoted to studying the friendship and following relationships on these social networks and characterizing the patterns of information diffusion over such networks. However, few systematic studies have been done to study Sina Weibo, a very popular online social network in China with over 500 million registered users.

In [7] Guo et al. studied the topology of Sina Weibo social network graph and discovered the asymmetric nature of the following relationships among Sina Weibo users as well as a densely connected core network formed by popular and active users, while [10] characterized user behaviors and information spreading over this social network. By modeling user interactions in Sina Weibo and Renren, [11] revealed that compared with Renren, a Facebook-like social network, Sina Weibo is a more efficient platform for information diffusion. The same research team also presented an analysis on "Hall of Fame" of Sina Weibo, a group of popular users verified and recommended by Weibo, and discovered that relationships between Weibo users are much looser than friendship-based online social networks such as Facebook or Renren [12].

Several recent studies have focused on the content such as tweets, retweets and comments posted by Weibo users, and the spreading process of these contents on Sina Weibo. For example, [13] performed a measurement study on video tweeting on Sina Weibo, and showed that nearly 80% of the tweeted videos have less than ten minutes, and some interesting videos attracted hundreds of, or even millions of views in a short time duration causing the phenomenons of flash crowds. In light of rumor circulation on Sina Weibo, [14] proposed a framework to differentiate rumors from normal tweet contents. In addition, [15] analyzed a variety of aspects of information diffusion over Sina Weibo based on a case study of 2010 Yushu Earthquake event. Unlike these studies on Sina Weibo, this paper focuses on characterizing Sina Weibo social network via a community approach, analyzing the impact of user popularity on user influence, and studying the similarity of retweeting and commenting activities of users within Shenzhen Weibo community.

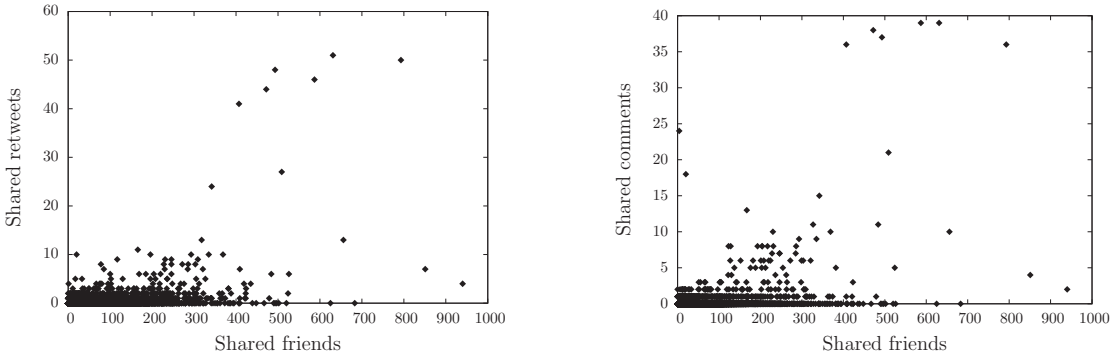
Understanding the graph structure and user interactions of online social networks has a broad range of applications. For example, [16] explored the underlying structure of social network graphs to design and implement a social partitioning and replication middle-ware to mediate the transparently between the applications and the back-end databases. [17] analyzed user workloads and interactions on online social networks, and showed that silent interactions such as browsing friends' pages account for over 90% of all user activities, but such interactions are often missed from publicly available data via Web crawling. Several recent papers have studied user influence over online social networks, mostly Twitter and Digg networks. For example, in [18] Meeyoung et al. showed that Twitter users with high in-degree are not necessarily influential measured by the retweeting or commenting activities. However, this study also discovered that for certain specific topics, some users could hold significant influence, e.g., the tweets of basketball stars on the latest games.

VII. Conclusions and Future Work

The explosive growth of users, tweets and social importance of Sina Weibo makes it imperative to gain an in-depth understanding of this large online social network in China. However, it remains a daunting task to collect and analyze all the data including users, relationships and content of Sina Weibo due to its sheer scale. This paper introduces a novel community approach for understanding Sina Weibo by focusing on a particular geographical location. Specifically, we crawl all Shenzhen Weibo users, collect their following and follower relationships, and analyze graphical structure of this community social network graph. To understand the impact of user relationships on retweeting and commenting activities, we perform correlation analysis of user popularity and tweet influence. In addition, we characterize the similarity of following patterns and user interactions in Shenzhen Weibo community. Our analysis sheds light on the community structure of social network graphs, reveals the influence of user popularity on information spreading over online social networks, and discovers the similarity of users' retweeting and commenting activities via exploiting the interactions between users and tweets. Our future work lies in understanding the difference and similarity of graphical structures for diverse online social network communities.

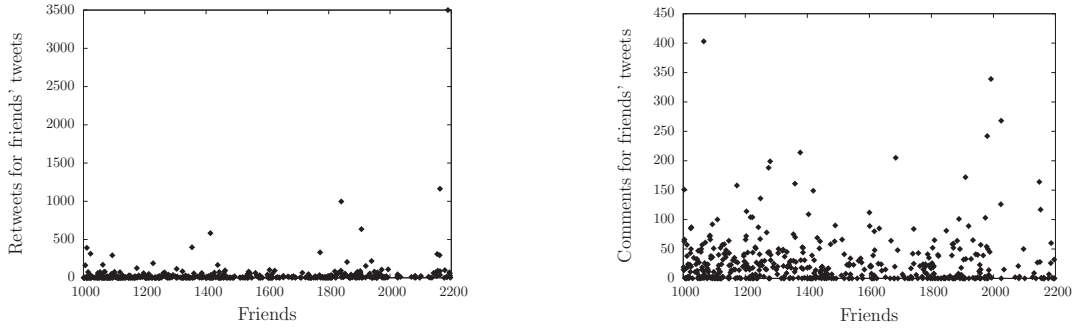
Acknowledgement

Kai Lei and Kai Zhang were financially supported by NSFC (No: 61103027), 973 Project (No: 2011CB302305), Shenzhen Gov Projects (No: JCYJ20120829170028558 and JCYJ20130331144541058). Kuai Xu was supported in part by the NSF grant CNS-1218212 and an ASU SRCA grant.



(a) Shared retweeting behavior (b) Shared commenting behavior

Fig. 8. Analysis of shared friends and common retweeting and commenting behaviors for users in Shenzhen Weibo community.



(a) Retweeting activities (b) Commenting activities

Fig. 9. Following behaviors and retweeting and commenting activities for users with a high number of friends in Shenzhen Weibo community.

REFERENCES

[1] J. Cheng, D. Romero, B. Meeder, and J. Kleinberg, "Predicting reciprocity in social networks," in *Proceedings of IEEE Conference on Social Computing*, October 2011.

[2] J. Yang and J. Leskovec, "Community-affiliation graph model for overlapping network community detection," in *Proceedings of International Conference on Data Mining (ICDM)*, December 2012.

[3] D. Ediger, K. Jiang, J. Riedy, D. Bader, C. Corley, R. Farber, and W. Reynolds, "Massive social network analysis: Mining twitter for social good," in *Proceedings of International Conference on Parallel Processing (ICPP)*, September 2010.

[4] Z. Guo, Z. Li, H. Tu, and D. Xie, "Detecting and modeling the structure of a large-scale microblog," *Lecture Notes in Electrical Engineering*, vol. 164, pp. 151–160, 2012.

[5] F. Wang, H. Wang, K. Xu, J. Wu, and X. Jia, "Characterizing information diffusion in online social networks with linear diffusive model," in *Proceedings of IEEE International Conference on Distributed Computing Systems (ICDCS)*, July 2013.

[6] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of Internet Measurement Conference*, 2007.

[7] Z. Guo, Z. Li, H. Tu, and L. Li, "Characterizing user behavior in weibo," in *Proceedings of International Conference on Mobile, Ubiquitous, and Intelligent Computing (MUSIC)*, June 2012.

[8] J.-L. Guillaume and M. Latapy, "Bipartite graphs as models of complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 371, pp. 795–813, 2006.

[9] J. J. Ramasco, S. N. Dorogovtsev, and R. Pastor-Satorras, "Self-organization of collaboration networks," *Physical review E*, vol. 70, pp. 036–106, 2004.

[10] Z. Guo, Z. Li, and H. Tu, "Sina microblog: An information-driven online social network," in *Proceedings of International Conference on Cyberworlds*, October 2011.

[11] J. Lin, Z. Li, D. Wang, K. Salamatian, and G. Xie, "Analysis and comparison of interaction patterns in online social network and social media," in *Proceedings of International Conference on Computer Communications and Networks (ICCCN)*, August 2012.

[12] G. Hao, L. Yu-Liang, W. Yu, and Z. Tong-tong, "Measurement of the weibo hall of fame network," in *Proceedings of International Conference on Instrumentation, Measurement, Computer, Communication and Control*, October 2011.

[13] Z. Guo, J. Huang, J. He, X. Hei, and D. Wu, "Unveiling the patterns of video tweeting: A sina weibo-based measurement study," in *Proceedings of Passive and Active Measurement Conference (PAM)*, March 2013.

[14] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, 2012.

[15] Y. Qu, C. Huang, P. Zhang, and J. Zhang, "Microblogging after a major disaster in china: a case study of the 2010 yushu earthquake," in *Proceedings of the ACM conference on Computer supported cooperative work*, 2011.

[16] J. M. Pujol, V. Erramilli, G. Siganos, X. Yang, N. Laoutaris, P. Chhabra, and P. Rodriguez, "The little engine (s) that could: scaling online social networks," *ACM SIGCOMM Computer Communication Review*, 2010.

[17] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of ACM SIGCOMM conference on Internet measurement conference*, 2009.

[18] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," in *Proceedings of International Conference on Weblogs and Social Media*, 2010.