



**HAL**  
open science

## Analysis and Comparison of Interaction Patterns in Online Social Network and Social Media

Jiali Lin, Zhenyu Li, Dong Wang, Kavé Salamatian, Gaogang Xie

► **To cite this version:**

Jiali Lin, Zhenyu Li, Dong Wang, Kavé Salamatian, Gaogang Xie. Analysis and Comparison of Interaction Patterns in Online Social Network and Social Media. 21st International Conference on Computer Communications and Networks (ICCCN 2012), Jul 2012, Munich, Germany. pp.1-7. hal-00737767

**HAL Id: hal-00737767**

**<https://hal.science/hal-00737767>**

Submitted on 2 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Analysis and Comparison of Interaction Patterns in Online Social Network and Social Media

Jiali Lin<sup>\*†</sup>, Zhenyu Li<sup>\*</sup>, Dong Wang<sup>\*†</sup>, Kavé Salamatian<sup>‡</sup>, Gaogang Xie<sup>\*</sup>  
<sup>\*</sup>Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China  
{linjiali, zyli, wangdong01, xie}@ict.ac.cn  
<sup>†</sup>Graduate School of Chinese Academy of Sciences, Beijing, China  
<sup>‡</sup>Université de Savoie, France  
kave.salamatian@univ-savoie.fr

**Abstract**—In this work, we aim to analyze and compare interaction patterns in different types of social platforms. To this end, we measured Renren, the largest online social network in China, and Sina Weibo, the most popular microblog service in China. We model the interaction networks as unidirectional weighted graphs in light of the asymmetry of user interactions. Following this model, we first study the basic interaction patterns. Then, we examine whether weak ties hypothesis holds in these interaction graphs and analyze the impacts on information diffusion. Furthermore, we model the temporal patterns of user interactions and cluster users based on the temporal patterns. Our findings demonstrate that although users in the two platforms share some common interaction patterns, users in Sina Weibo are more popular and diverse. Moreover, analysis and simulation results show that Sina Weibo is a more efficient platform for information diffusion. These findings provide an in-depth understanding of interaction patterns in different social platforms and can be used for the design of efficient information diffusion.

**Index Terms**—Interaction patterns, Social networks, Information diffusion

## I. INTRODUCTION

In the past few years, the Internet has witnessed the unprecedented boom of online social services. Online social networks, such as Facebook <sup>1</sup>, and Renren <sup>2</sup>, have attracted millions of users. The key feature for their success is that they allow or even encourage users to interact with others by publishing their opinions, leaving messages, chatting, sharing content, and some other social interactions. Later, Microblog services, such as Twitter <sup>3</sup> and Sina Weibo <sup>4</sup>, emerge as social media systems [8] and quickly get extremely popular.

The level of interactions between users characterizes their popularity and their friendship, which can be further leveraged for information diffusion. The interaction graphs of different social platforms have been examined in [8] and [13]. Different characteristics of interaction graph and social graph have been identified. Viswanath *et al.* [12] demonstrated that links in Facebook interaction network fluctuates rapidly over time, suggesting that time should be taken into account when analyzing interaction networks. Jiang *et al.* [6] further

analyzed the latent interaction graph in Renren. These works mainly focus on the difference between interaction and social graphs and only examine the online social networks, such as Facebook, Cyworld and Renren. The interaction patterns in different types of social platforms are remained not well understood.

Our study in this paper focuses on a comprehensive understanding of interaction patterns over time and a comparison of the patterns between an online social network (*i.e.* Renren) and a social media platform (*i.e.* Sina Weibo). To this end, we collected two large datasets: one is for Renren and the other is for Sina Weibo. Given that the interaction between two users is not necessarily reciprocal and different users interact with different strengths, we model the interaction network as a unidirectional weighted graph.

We first analyze the basic interaction patterns in Renren and Sina Weibo, and then examine whether weak ties hypothesis [10] holds in interaction graphs. A hidden Markov model (HMM) is used to characterize temporal interaction behaviors. Then, we cluster users with similar HMM parameters using a self-organizing map (SOM) in order to investigate whether there exist groups of users with similar temporal behavior patterns. In particular, our findings can be summarized as follows:

- While the in-degree distributions in interaction graphs for both social platforms follow power-law modes, we find stretched exponential node strength distributions in both interaction graphs. The stretched factor characterizes the number of cascade stages for information spreading.
- By measuring the coupling between interaction strengths and the local structure in interaction graph, we find the weak ties hypothesis holds for Renren, but not for Sina Weibo. The results indicate that Sina Weibo as a social media is more efficient than Renren as an online social network when it comes to information diffusion.
- Users in Sina Weibo are generally more popular than users in Renren. In particular, the top popular users in Sina Weibo are with a much higher probability of receiving interactions than those in Renren.
- Users do show clustering behavior patterns in both social platforms, and users behave more diversely in Sina Weibo

<sup>1</sup><http://www.facebook.com>.

<sup>2</sup><http://www.renren.com>.

<sup>3</sup><http://www.twitter.com>.

<sup>4</sup><http://weibo.com>.

than in Renren. Moreover, users with similar behavior patterns do have some correlations in attributes.

Our findings demonstrate that although users in Renren and Sina Weibo share some common interaction patterns, users in Sina Weibo are more popular and diverse. Moreover, information diffusion in Sina Weibo is more efficient than that in Renren. These findings provide an in-depth understanding of interaction patterns in different social platforms and can be used for the design of efficient information diffusion. To the best of our knowledge, this work is the first to compare an online social network and a social media in terms of user interaction.

The rest of this paper is organized as follows: Section II describes the background and dataset on Renren and Sina Weibo. We analyze the interaction networks in both websites in Section III. In Section IV, we adopt hidden Markov model and self-organizing map to dig into the user behavior. Section V covers the related work. Finally, we conclude in Section VI.

## II. BACKGROUND AND DATASET

In this section, we first briefly introduce Renren and Sina Weibo. Then we describe the datasets.

### A. Renren and Sina Weibo

Renren, with a claimed 170 million registered users, is the largest online social network in China [6]. Renren can be described as the Facebook’s Chinese cloning since it is very similar with Facebook in both the user interfaces and the features. A mutual friendship between two users is built if and only if one sends a request and the other approves the request. It provides various interaction applications. Each user has a gossip wall where visitors can leave messages. Another useful interaction application is ‘Status’, which enables users inform their friends about the recent news or thoughts. The friends can see the status and eventually they may reply to it in thread.

Sina Weibo is launched in 2009, about three years after Twitter. It is now the most popular microblog service platform in China with more than 200 million registered users and 32 million daily active users. Like Twitter, Sina Weibo allows users to post 140-character messages (called tweet in Twitter), retweet and reply others’ tweets, follow whoever they are interested in. Besides texts and photos, Sina Weibo also allows users to post video clips and audio files. As a consequence, it broadens the boundary of microblog, making it a platform for information dissemination like a *social media* platform, social connection and entertainment.

### B. Datasets

We collect two datasets for analysis, one for Renren and one for Sina Weibo.

**1) Dataset of Renren.** Our Renren dataset was collected from May 10th, 2010 to July 10th, 2010. Initially, 200 random users were selected. Then the crawler followed the friend links to find more users. In total, we got over 3 million users, within which 1,267,731 (42.3%) users make their status

publicly accessible. We selected among them 191,121 users who updated their status information and got feedbacks in the eight weeks of our study. We chose to gather the interaction data from ‘Status’ feature as it had a detailed and temporal log of each user’s interactions. Moreover, a larger proportion of users are willing to make public their status; there are three times more interactions relative to status than relative to blogs or gossip wall.

**2) Dataset of Sina Weibo.** We crawled Sina Weibo from Oct 1st, 2010 to Oct 15th, 2010. We randomly chose 100 users as seed users then followed the followers and followings of users to get more users. In total, we got over 3.75 million users. We captured the profile of each user, *e.g.* the number of followers and followings, and the number of tweets he had posted. Then we collected the original tweets these users published between Aug 1st, 2010 and Sep 30th, 2010 and the comments to these tweets. We found that among 3.75 million users, 1,159,269 users published their tweets and got replies within the time window. We crawled Sina Weibo again to get latest profiles of the 3.75 million users on Dec 20th, 2011.

## III. INTERACTION NETWORK ANALYSIS

In light of the asymmetry and skewed strength of user interactions, we model the interaction network as a unidirectional weighted graph. In this graph nodes represent users, and a directed edge from  $A$  to  $B$  exists if and only if user  $A$  contacts directly with  $B$ . When we speak directly contacting, for Renren, we mean  $A$  replied  $B$ ’s statuses at least once, while we mean  $A$  replied  $B$ ’s tweets at least once for Sina Weibo. Throughout this paper, a reply means one interaction.

The edge is weighted with the number of interactions, denoted  $w_{AB}$ . The directed edge is weighted by the interaction times between users. For example,  $B$  replies three messages to  $A$ ’s statuses or tweets, then a directed edge from  $B$  to  $A$  is generated and the weight  $w_{BA} = 3$ . Here, the node is equivalent to the user.

### A. Power-law Distribution for Node In-degree

As the network is directed, there are two types of node degrees: *in-degree* and *out-degree*. Each reflects a distinct perspective: one’s in-degree characterizes his popularity whereas his out-degree represents his activity. In this paper, we are interested in user popularity. Thus, we analyze in-degree distribution for Renren and Sina Weibo.

Figure 1 plots the complementary cumulative distribution function (CCDF) for in-degree in the two interaction graphs. A look at the in-degree distribution shows that even if the large majority of nodes have very low in-degree, implying the possible existence of sockpuppets with low interactions, there are also very popular people with a lot of friends interacting with them. The in-degree distributions in both Renren and Sina Weibo roughly follow power-law. The power-law coefficient  $\alpha$  for Renren is 3.5, consistent with which was found by Jiang, J. *et al.* [6]. It can also be found that the in-degree distribution in Sina Weibo interaction graph is less skewed than that in Renren one.

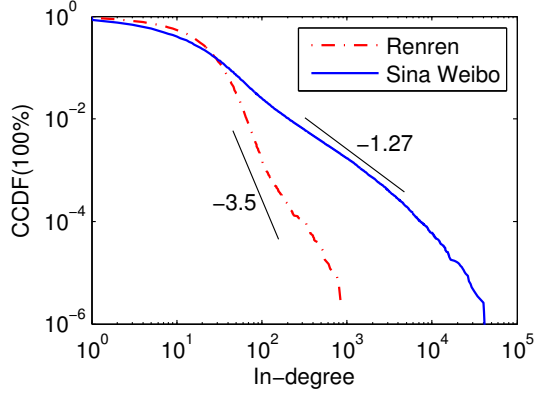


Fig. 1. Node in-degree distribution

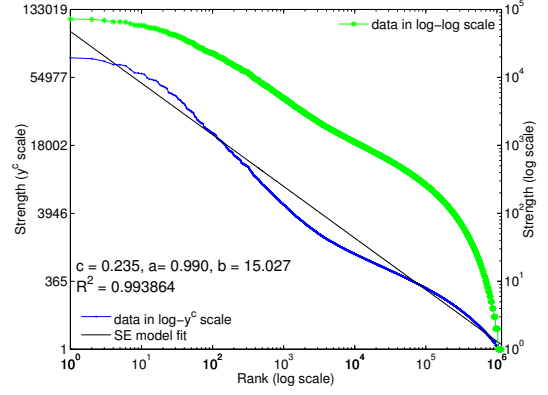


Fig. 3. Rank-ordering distribution of node strength in Sina Weibo

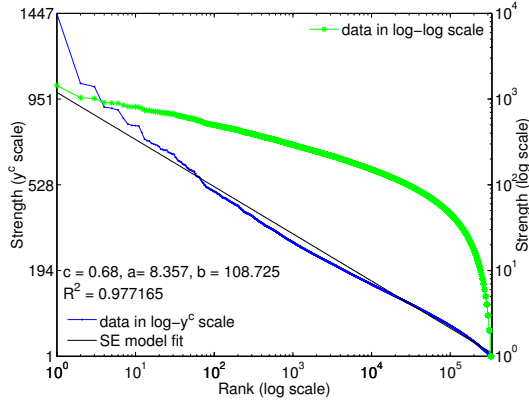


Fig. 2. Rank-ordering distribution of node strength in Renren

### B. Stretched Exponential Distribution of Node Strength

The node strength is defined as the sum of all weights of incoming edges in the weighted directed graph, *i.e.* the strength of node  $i$  with in-degree  $k$  is defined as:  $s_i = \sum_{j=1}^k w_{ji}$ . We are interested in whether the distribution of node strength follows power-law model or not.

We rank the users according to node strength, and plot the rank-ordering distribution in Figure 2 and Figure 3 for Renren and Sina Weibo, respectively. The first glance may lead us to believe that power law fits the node strength distribution well. However, in log-log scale the distribution curves in both figures are not straight lines, meaning that node-strength *does not* well follow the power-law model. However, instead of simply fitting the distributions with other models, we analyze the reason behind.

The statuses or tweets of a user may be forwarded by his friends to a large number of audiences. These audiences may or may not reply the statuses or tweets. The forwarding process can be modeled as a cascade process, which is formally defined as a random process  $X_n$  that can be described as a multiplication of  $n$  random variables  $m_1, \dots, m_n$ , *i.e.*  $X_n = m_1 \times m_2 \times \dots \times m_n$ . Among all users that see a status or tweet from a link made in  $(i-1)$ -th cascade stage

( $1 < i \leq n$ ), there is a percentage  $\alpha_i$  that will reply the content and make a link to it. This means that the number of replies can be related to a factor  $(1 + \alpha_1)(1 + \alpha_2) \dots (1 + \alpha_n)$ , representing the overall number of replies after  $n$  steps of the cascade process.

Quite interestingly, a limit theorem similar to the Central Limit Theorem can be derived for cascade processes [4]. These processes converge to a stretched exponential (SE) distribution defined as:

$$P(X \geq x) = e^{-\left(\frac{x}{x_0}\right)^c} \quad (1)$$

where the stretched factor  $c$  is the inverse of the number of multiplied random variables and represents the inverse of the number of cascade stages,  $x_0$  is a constant parameter. In a rank-ordering distribution,  $N$  objects are ranked in a descending order of their reference numbers. Then  $P(X \geq x_i) = i/N$ , where  $i$  ( $1 \leq i \leq N$ ) is the number objects with reference numbers larger or equal to  $x_i$ . That is  $\log(i/N) = -\left(\frac{x}{x_0}\right)^c$ . By substituting  $x_i$  for  $y_i$ , we have

$$y_i^c = -a \log i + b \quad (2)$$

where  $a = x_0^c$  and  $b = y_1^c$ . Hence, the rank-ordering distribution curve for data following a stretched exponential model should be a straight line in  $\log\log-y^c$ .

We thus fit the SE distribution to node strength using the fitting method proposed by Guo *et al.* in [5]. To gauge the fitting errors, we use the coefficient of determination of the data fit, also known as  $R^2$ . The closer  $R^2$  to 1, the better the model fits the empirical data. The results are plotted in Figure 2 and Figure 3. As expected, the SE models well fit the node strength distributions for both Renren and Sina Weibo.

An interesting finding is related to the stretched factor. The stretched factor  $c$  for Renren is 0.68, meaning that a status is on average forwarded by 1~2 hops. For Sina Weibo,  $c$  is 0.235, meaning an average of 4~5 hops forwarding. The above finding indicate that content in Sina Weibo can be forwarded to distant users. The reason behind is that Renren is friendship-based; users are almost only interested in interaction with their friends. However, Sina Weibo can be treated as a social media

[8] and formed by content interests. Users in Sina Weibo may reply to whatever they are interested.

### C. Weak Ties Hypothesis Analysis

Social platforms are now wildly used for information diffusion. The efficiency greatly depends on the structure of the graph on which information is diffused. It has been found that ties with different strengths play different roles in information diffusion [1]. We here examine whether weak ties hypothesis holds in the interaction graphs of Renren and Sina Weibo. Weak ties hypothesis states that the strength (*i.e.* weight) of a tie (*i.e.* edge) between  $A$  and  $B$  increases with the overlap of their friendship circles, resulting in the importance of weak ties in connecting communities. We define the friendship overlap of two users  $i$  and  $j$ , connected by an edge  $e_{ij}$  in weighted directed graph as:

$$O_{ij} = \frac{n_{ij}}{(d_{i\_out} - 1) + (d_{j\_in} - 1) - n_{ij}} \quad (3)$$

where  $n_{ij}$  is the number of the entire two-hop paths from node  $i$  to node  $j$  except the edge  $e_{ij}$ ,  $d_{i\_out}$  is the out-degree of node  $i$ , while  $d_{j\_in}$  is the in-degree of node  $j$ . If node  $i$  and node  $j$  have no common acquaintances, then we have  $O_{ij}=0$ , and if all the neighbors of node  $i$  have directed link to node  $j$ , then  $O_{ij}=1$ .

To quantify the correlation between tie strength and friendship overlap, we leverage Spearman rank correlation  $\rho$  which is defined as

$$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{n(n^2 - 1)} \quad (4)$$

where  $x_i$  and  $y_i$  are the ranks of edges according to the edge weight and friendship overlap for an  $n$ -edge system. It is a non-parametric measure of correlation, which shows how well an arbitrary monotonic function could describe the relationship between two variables. The coefficient lies in between  $[-1,1]$ , where “1” indicates perfect positive correlation and “-1” means perfect negative correlation.

For any pair of two users connected by an edge in the interaction graph, we compute their friendship overlap using Eq. 3, then use Spearman’s correlation measure Eq. 4 to compute the correlation between tie strength and friendship overlap. The results of both Renren and Sina Weibo are listed in Table I. To avoid the tied ranks among the edges with the least weight, we bin the ties based on the weight of edge.

TABLE I  
SPEARMAN’S RANK CORRELATION COEFFICIENTS

Tie Strength	All	[0,10)	[10,100)	[100,1000)	[1000,10000)
Renren	0.484	0.487	0.359	0.182	N/A
Sina Weibo	0.408	0.386	0.082	-0.013	0.055

It can be found that the correlations in Renren for all bins are relatively high, meaning that the hypothesis holds for Renren. But for Sina Weibo, the correlations for edges with high strengths are very close to 0, meaning the hypothesis does not

hold. Recall that, as Facebook, Renren is based on real social relationship, aiming at maintaining existing friendship and making new friends. This would yield users with close social relationship highly connecting with each other in clusters, where the links among different social clusters act as weak ties. On the other hand, Sina Weibo is a social media platform where users can follow whoever they like.

Onnela *et al.* [10] have found a social graph where weak ties hypothesis holds are not efficient when it comes to information diffusion. Next, we evaluate the performance of information diffusion in Renren and Sina Weibo interaction graphs.

### D. Information Diffusion Simulation

We follow the simulations in [10] to evaluate the performance of information diffusion in Renren and Sina Weibo interaction graphs. The spreading mechanism is similar to the susceptible-infected model of epidemiology, in which recovery is impossible, implying that an infected individual will continue transmitting information. We simulate two scenarios.

The first scenario is called *real simulation*, in which the infection probability  $P_{ij}$  from an infected node  $j$  to  $i$  is set according to the tie strength  $w_{ij}$  in interaction graph, *i.e.*  $P_{ij} = xw_{ij}$ . The parameter  $x$  controls the overall spread rate. Changing  $x$ ’s value does not change the qualitative nature of results [10]. The second scenario is called *control simulation*, where the network (*i.e.* tie) is the same and we set strength of edge as the average over all edges. That means the infection probability on any edge is the same.

We chose 10 nodes as initially infected nodes with novel message at time 0. The initial node sets in two scenarios are the same. Obviously, the infected nodes are bounded by the large connected components starting from the 10 initial nodes.

Figure 4 plots the real diffusion simulation versus the control simulation in Renren. Under no circumstance would the real Renren interaction network is more efficient for information diffusion than the control one. Before  $t = 125$  the coverage of real and control simulation are almost the same. When  $t$  is greater than 125, the control simulation covers more fractions of nodes. This is because the message is more likely to escape from the original communities in the control simulation as time increases due to equal weight of all the edges.

Figure 5 plots the results in Sina Weibo. Compared with Renren in Figure 4, the gap between the results of real and control simulations are small. Besides, before time  $t$  reaches 165, the number of infected nodes in the real simulation is slight bigger than that in control simulation. We can draw the conclusion that Sina Weibo is good for information diffusion, especially in the case of reaching the maximum persons within limited time.

Our simulation results on one hand indicate that a social graph where weak ties hypothesis holds are not efficient for information diffusion. On the other hand, they demonstrate that Sina Weibo as a social media platform is more efficient than

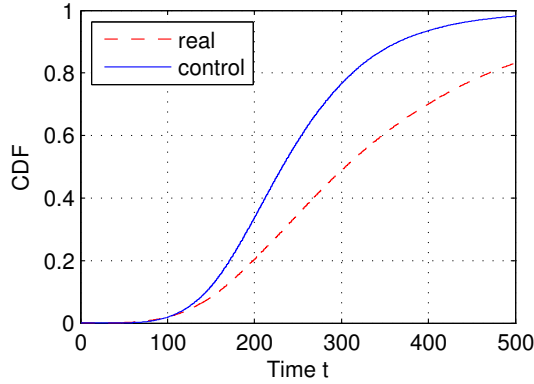


Fig. 4. Diffusion simulation in Renren

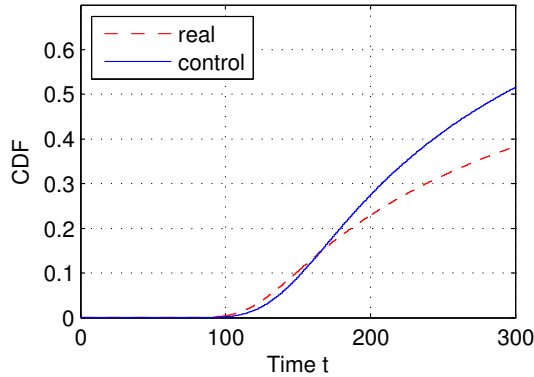


Fig. 5. Diffusion simulation in Sina Weibo

Renren as an online social network in terms of information diffusion.

#### IV. TEMPORAL USER INTERACTION PATTERNS ANALYSIS

In this section, we show an insight into the temporal user interaction patterns. By observing how users interact with each other, we analyze and model the user behavior patterns. Firstly, hidden Markov model is adopted to precisely depict how user behavior evolves. Then, we implement the self-organizing map to characterize the common patterns of user behavior. Finally, we discuss the applications based on temporal interaction patterns.

##### A. Modeling Temporal User Behaviors with Hidden Markov Model

We observe users of Renren and Sina Weibo for a period of time to study the temporal evolution of user interactions. In [12] an analysis based on *a priori* separation of users' popularity and unpopularity is defined and comparing consecutive state using a resemblance factor is proposed. In this paper we are using a more precise characterization method which is hidden Markov model (HMM).

A hidden Markov model (HMM) is a Markov chain in which the chain states are not observable directly [11]. However, in hidden Markov model, the outputs which statistically depend

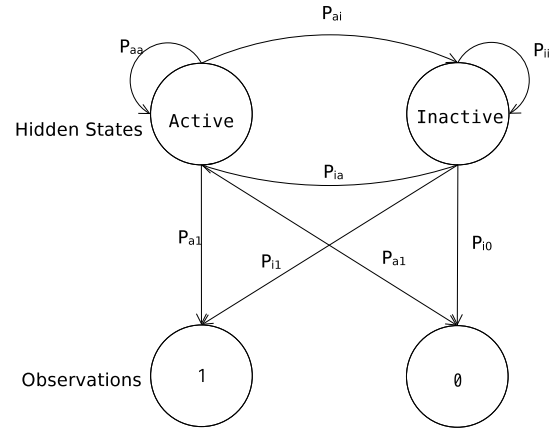


Fig. 6. Hidden Markov model in interaction networks

on the hidden states are visible, *i.e.* there is probability distribution of visible output for each hidden state in the HMM. Therefore the sequence of HMM output observations gives some information about the sequence of hidden states.

Intuitively, we assume that in interaction network, each user follows a HMM with two hidden states: *active* and *inactive* showed in Figure 6. The intuition of setting these two hidden states derives from the bi-model of human life. Then we define the visible output of each user with a two-state observation as well. The observation equals to 1 if he receives at least a message on the day and 0 otherwise. We extract eight consecutive weeks' interaction logs of the users in Renren, and 61 consecutive days' interaction logs of the users in Sina Weibo. Each user has therefore a sequence of 1/0 (popular/unpopular) values relative to a series of consecutive days.

For each sequence we calibrate a HMM, then we use Baum-Welch algorithm [2] to acquire the maximum likelihood estimate of four parameters: the state transition probabilities  $P_{ai}$ ,  $P_{ia}$  and the output probabilities  $P_{a1}$ ,  $P_{i1}$ .  $P_{ai}$  ( $P_{ia}$ ) indicates the transition probability from active (inactive) state to inactive (active) state. Meanwhile  $P_{a1}$  ( $P_{i1}$ ) represents the probability that a user in the active (inactive) state to receive at least an incoming interaction. These four values are the prominent profile of user behaviors. As a result, everyone has one HMM characterizing the incoming interactions behavior, resulting in 191,121 different HMMs for those who have at least one interaction in our capturing period of Renren and 1,160,798 that of Sina Weibo.

With these four values, one can obtain the probability of receiving at least an interaction as:

$$P_p = \frac{P_{ai}}{P_{ai} + P_{ia}} P_{i1} + \frac{P_{ia}}{P_{ai} + P_{ia}} P_{a1} \quad (5)$$

We suppose that a user's popularity is in proportion to the probability of receiving at least an interaction, namely  $P_p$  in formula (5). Then we plot in Figure 7 the CDF of the user popularity. About 95% of users in Renren are with a probability smaller than 0.4 and the distribution of these users

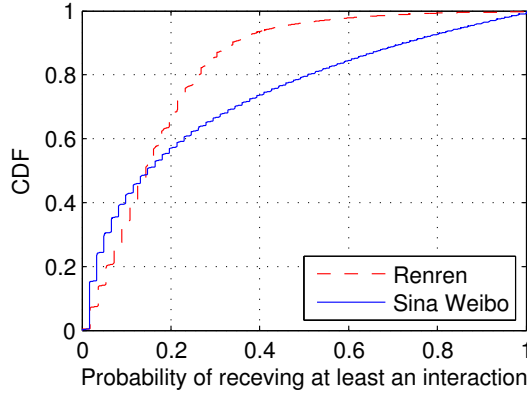


Fig. 7. The probability of receiving at least an interaction for all users

is almost uniform. Moreover, there are as many as 20% of users in Sina Weibo have a probability larger than 0.5 to receive an interaction, while this percentage is only 2% in Renren. The above results indicate that users in Sina Weibo are generally more popular than users in Renren, and the top popular users in Sina Weibo are with a much higher probability of receiving interactions than those in Renren.

### B. Clustering User Behavior Patterns by Self-organizing Map

In order to figure out whether there is the common characteristic interaction behavior pattern shared by users, we apply a clustering algorithm to the user behavior patterns, characterized by the 4 HMM parameters described in the previous subsection.

The clustering algorithm we applied is self-organizing map (SOM) [7]. It is a type of artificial neural network which is trained using unsupervised learning to produce maps, *i.e.* a low-dimensional (typically two dimensions) discretized representation of the input space of the training samples. SOM works by assigning observed samples to neurons in such a way that similar samples are placed in close-by neurons. SOM is well known to perform robust clustering over complex data in an agnostic way (without making strong assumption of the data nature) [7].

We have built HMMs for 191,121 users in Renren and 1,160,798 users in Sina Weibo. For each user, the input to SOM is the vector containing the four probabilities  $(P_{ai}, P_{ia}, P_{a1}, P_{i1})$ . We train a  $20 \times 20$  network over 200 epochs with these data.

We show in Figure 8 the result of clustering for users in Renren. Each node in the figure represents a group of users, thus we get 400 groups of users. In addition, the color of the link between two nodes indicates the Euclidean distance between them. The lighter the color is, the closer or more similar the nodes are. As can be seen, there are clearly three big clusters.

To explain the resulting clusters we examine weight planes showing the importance of each one of the four input probabilities on the users clustered in each node. The weight planes are shown in Figure 9. Four subgraphs correspond to the four

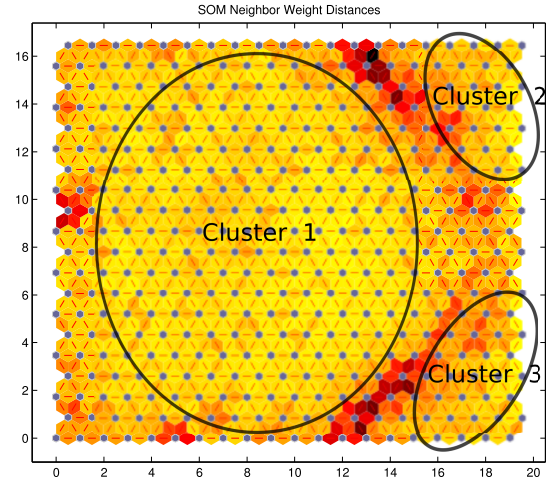


Fig. 8. SOM neighbor weight distance of Renren

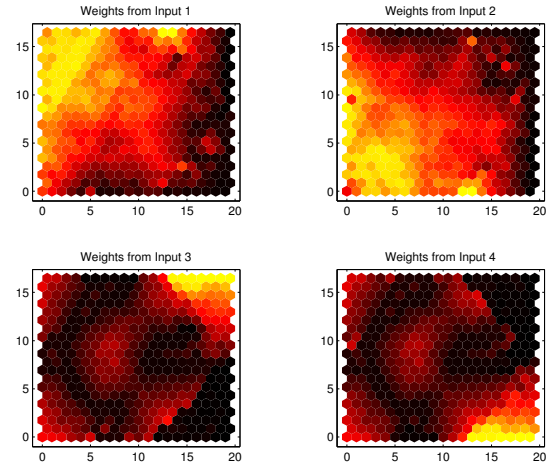


Fig. 9. SOM weight planes of Renren

input HMM parameters  $(P_{ai}, P_{ia}, P_{a1}, P_{i1})$  sequentially. The darker the node color is, the smaller the parameter represented by the node is.

It can be seen from Figure 9. Users in cluster 2 are stable in both states to some extent. They have a high probability to receive interaction in active state and a low probability in inactive state. Users in cluster 3 are also relatively stable, but they are more likely to receive interaction in inactive state. Thus, cluster 3 can be treated as a mirror cluster of cluster 2. Cluster 1 contains users with a low probability of getting interaction in either state.

We further estimate the popularity (measured by Eq. 5) for the users in the three clusters. The results demonstrate that users in cluster 2 and cluster 3 are more popular than users in cluster 1. Taking all the features of the three clusters into consideration, we conclude the properties of these clusters below. Users in cluster 2 seem to be content producers, and their activities in the network greatly contribute to their popularity. Cluster 3 contains celebrities, since they are less active but still can receive interactions, implying their popularity based

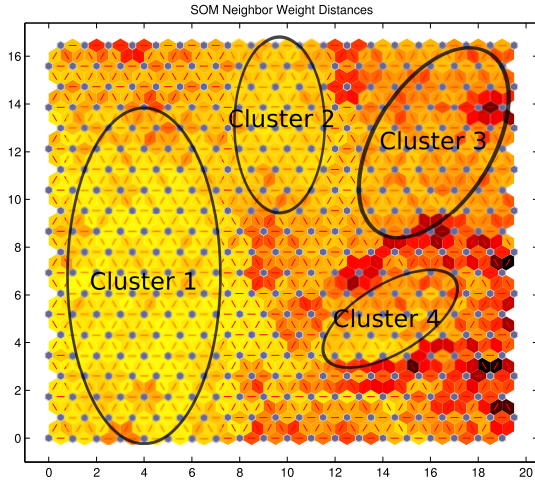


Fig. 10. SOM neighbor weight distance of Sina Weibo

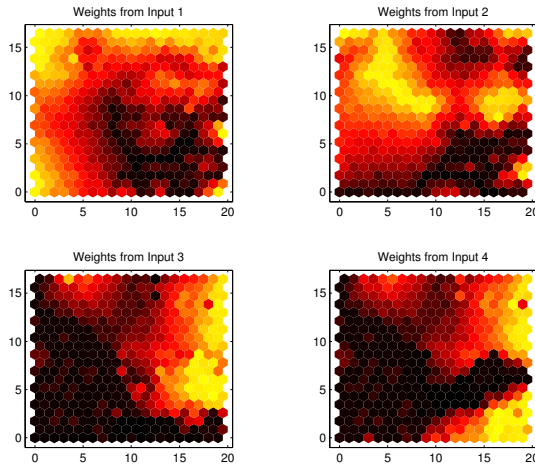


Fig. 11. SOM weight planes of Sina Weibo

on the social attentions to them. Cluster 1 simply represents ordinary users that have not a much differentiated behavior.

Figure 10 illustrates the clustering result of Sina Weibo. As can be seen, there are roughly four big clusters. However, compared with Renren, users behave more diversely in Sina Weibo. We plot the weight planes in Figure 11. Following the same way, we find that Cluster 1 represents the ordinary users with the least popularity, while cluster 2 consists of users with moderate popularity. Cluster 3 contains celebrities, since no matter which states they stay at, they always have a high probability to get contacted. Although users in cluster 4 are also popular ones, their popularity is largely based on their efforts: the more popular the users in cluster 4 want to be, the longer they should stay in active state.

Next, we are interested in the correlation between clustered user behavior and user attributes (*e.g.* number of friends/followers). Here, we focus on Sina Weibo. The attribute we look at is the number of new followers for users during Oct 15th, 2010 to Dec. 20th, 2011. We cluster users into 400

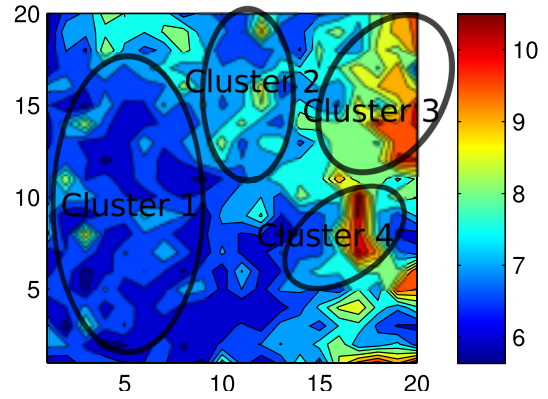


Fig. 12. Contour map of the increase of users' followers

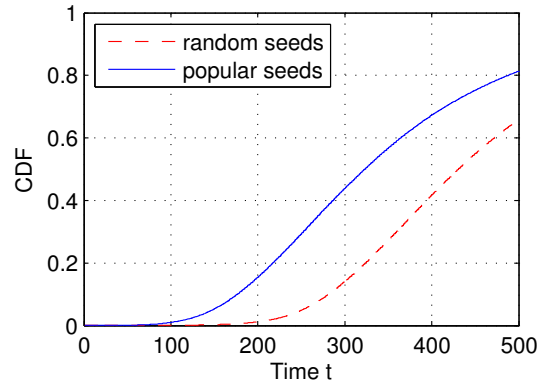


Fig. 13. Diffusion simulation in Renren

groups as in Figure 10, and then compute the average number of new followers per group. The results are plotted in Figure 12 as a contour map, where the height is the *logarithm* of the average value per group. It depicts that users in cluster 3 and cluster 4 attract more new followers than cluster 1 and cluster 2. This is due to the fact that users in cluster 3 and cluster 4 are popular ones. The above finding reveals that the users in same cluster surely share similar attributes.

### C. Applications of User Interaction Patterns

A direct application of the above findings is to choose popular users as seeds for information diffusion. We performed simulations on Renren interaction graph to show how seeds impact the performance of information diffusion. We simulated two scenarios. In the first one, we selected 10 popular users from cluster 2 as seeds, while in the second one we randomly select 10 users from all users as seeds. The settings of the simulation are similar with those introduced in Section III-D. Besides, the upper bounds of two scenarios are almost the same. The results are plotted in Figure 13. It can be seen that message starting from the popular seeds reaches a broader range within a shorter time lag. Clearly, the one with popular users as seeds is more efficient than that with random seeds.

Another application is from hidden Markov model of user temporal interaction patterns. We figure out a HMM for each



user. Using the parameters we have obtained, if one also knows the state a user stays at time  $t$ , one can predict the probability that he will receive an interaction at time  $t + 1$ . This can be used for user popularity prediction.

## V. RELATED WORK

The basic properties of online social networks and social media have been heavily studied. Mislove *et al.* [9] discover "small-world" properties in four online social networks after investigating the structure properties of these websites. Kwak, H. *et al.* [8] compare fundamental features between Twitter and traditional social networks, leading the conclusion Twitter is different with online social networks.

Besides, lots of research works focus on human interactions in daily life. Onnela, J.P. *et al.* [10] validate the weak ties hypothesis in mobile communication networks. They also verify both weak ties and strong ties are inefficient in information diffusion.

Given that the level of interactions characterizes user popularity and friendship, interaction graphs of different online social networks are examined in [3] [13] [6] [12]. Chun *et al.* [3] analyze Cyworld and find that the structure of the interaction network is similar to the social network. Wilson *et al.* [13] adopt an unweighted graph to model interactions in Facebook. They find that, in contrary to social networks, the interaction network over Facebook does not strongly exhibit "small-world" properties. Jiang *et al.* [6] investigate the latent interaction network of Renren. Viswanath *et al.* [12] study the evolution pattern of Facebook interaction graph. Their results suggest that time should be taken into account when analyzing interaction networks.

## VI. CONCLUSION

In this paper, we present the analysis and comparison of the interaction patterns in online social network and social media using unidirectional weighted graphs.

Our findings show that node strength follows stretched exponential distribution. Moreover, weak ties hypothesis holds in Renren, leading an inefficient information diffusion network. In contrast, Sina Weibo exhibits a good ability for spreading messages. We figure out that there are generally a larger proportion of popular users in Sina Weibo than that in Renren. After clustering the users' HMM parameters with self-organizing map, we find the user interaction patterns in Sina Weibo are more diverse. Besides, users in the same clusters show some common features.

The future work will be focused on understanding the incentives and the motivations of interaction patterns. We will analyze whether their interactions are affected by other attributes such as, social event, age, gender, hometown *etc.*

## REFERENCES

- [1] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World wide web*. ACM, 2012.
- [2] J.A. Bilmes. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4:126, 1998.
- [3] H. Chun, H. Kwak, Y. Eom, Y. Ahn, S. Moon, and H. Jeong. Comparison of Online Social Relations in Volume vs Interaction: a Case Study of Cyworld. *IMC*, October 2008.
- [4] U. Frisch and D. Sornette. Extreme deviations and applications. *Journal of Physics I France*, 7:1155–1171, September 1997.
- [5] L. Guo, E. Tan, S. Chen, X. Zhang, and Y.E. Zhao. Analyzing patterns of user content generation in online social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 369–378. ACM, 2009.
- [6] J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, and B.Y. Zhao. Understanding latent interactions in online social networks. In *Proceedings of the 10th annual conference on Internet measurement*, pages 369–382. ACM, 2010.
- [7] T. Kohonen. Self-organization and associative memory. *Self-Organization and Associative Memory*, 100 figs. XV, 312 pages.. Springer-Verlag Berlin Heidelberg New York. Also Springer Series in Information Sciences, volume 8, 1, 1988.
- [8] 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*, pages 591–600. ACM, 2010.
- [9] 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 conference on Internet measurement*, pages 29–42. ACM, 2007.
- [10] J.P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.
- [11] L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [12] B. Viswanath, A. Mislove, M. Cha, and K. Gummadi. On the Evolution of User Interaction in Facebook. *WOSN*, August 2009.
- [13] C. Wilson, B. Boe, A. Sal, K. Puttaswamy, and B. Zhao. User Interactions in Social Networks and their Implications. *Eurosys*, April 2009.