Process-driven Analysis of Dynamics in Online Social Interactions

Zhi Yang
Computer Science Department
Peking University
Beijing, China
yangzhi@pku.edu.cn

Jilong Xue
Computer Science Department
Peking University
Beijing, China
xjl@net.pku.edu.cn

Christo Wilson
College of Computer and Information Science
Northeastern University
MA, USA
cbw@ccs.neu.edu

Ben Y. Zhao
Computer Science Department
U. C. Santa Barbara
CA, USA
ravenben@cs.ucsb.edu

Yafei Dai
Computer Science Department
Peking University
Beijing, China
dyf@pku.edu.cn

ABSTRACT

Measurement studies of online social networks show that all social links are not equal, and the strength of each link is best characterized by the frequency of interactions between the linked users. To date, few studies have been able to examine detailed interaction data over time, and studied the problem of modeling user interactions. A generative model can shed light on the fundamental processes that underlie user interactions.

In this paper, we analyze the first *complete* record of full interaction and network dynamics in a large online social network. Our dataset covers all wall posts, new user events, and new social link events during the first full year of Renren, the largest social network in China, including 623K new users, 8.2 million new links, and 29 million wall posts. Our analysis provides surprising insights into the evolution of user interactions over time. We find that users invite new friends to interact at a nearly constant rate, prefer to interact with friends with whom they share significant overlaps in social circles, and most social links drop in interaction frequency over time. We also validate our findings on Facebook, and show that they do generalize across OSNs.

We use our insights to derive a generative model of social interactions that accurately captures both our new results and previously observed network properties. Our model captures the inherently heterogeneous strengths of social links, and has broad implications on the design of social network algorithms such as friend recommendation, information diffusion and viral marketing.

1. INTRODUCTION

Without a doubt, online social networks (OSNs) have had an enormous impact on the lives of millions of people, and changed the way people communicate and interact online. For scientists in both engineering and social disciplines, OSNs offer a digital representation of human social behavior that is both overly-simplified and yet tantalizingly concrete, in the form of abstract social graphs that capture social network activity.

The last few years has seen the arrival of several measurement studies of user relationships and activities on popular OSNs, including Facebook [9, 25, 26], Twitter [7, 13], LinkedIn [18], Renren [15, 28] and others [5, 8, 22]. A common observation made across many platforms is that the presence of a social link connecting two users is a poor estimate of the "relationship strength" between them. Instead, many have proposed using the number or frequency of interactions on these networks to capture the strength of a social link [11, 27].

Despite this realization, we are still far from a real understanding of the processes that underlie user interactions. To capture strength of social links, recent studies proposed the use of weighted "interaction graphs" where each link is labeled with some measure of interaction frequency [8, 15, 26]. But these studies focus on a static view of interactions, and therefore only capture a small piece of the picture. Prior study [25] examined changing dynamics of user interactions on Facebook users, but was limited to a sample set of 60,000 users crawled from a single geographic network.

A deeper understanding of user interactions requires the formulation of a generative model, which can intuitively capture the processes that drive user interaction events. No generative graph model exists to explain properties observed in measured traces of user interactions, or to construct realistic arbitrary-sized user interaction traces. Not only would such a model advance our understanding of social networks, it would be immensely useful to a number of social network applications. For example, it can be used to make more accurate predictions in the link prediction problem [2], to reorder or filter user news feeds by accurately predicting the likelihood of specific user interactions, or improve resource planning by predicting about how data access patterns between users change over time.

In this paper, we seek to fill this void by building a model based on two large detailed traces of user interactions on *Renren* and *Facebook*. Our Renren trace covers over a year in length, and contains data on the creation of 600+K users, 8+Million new links, and 29+ Million interaction events. The Facebook dataset is a 1.6M node sample that includes 49M edges and 16M interactions. The

COSN'15, November 2–3, 2015, Palo Alto, California, USA.

DOI: http://dx.doi.org/10.1145/2817946.2817952.

core contribution of this paper is a new generative model that combines the growth of social links with the generation of user interaction events on those links. We use analysis of empirical data to understand and model the growth processes that lead to the observed network structure and link strength distributions.

We present detailed analysis of our growth data, and extract three processes that drive dynamics of social interaction during the network formation:

Forgetting process: A particular pair of users slowly decrease their interaction frequency over time. The potential reason is that users tend to forget each other as they cannot meet face to face on a regular basis, leading to the closeness between friends declined rapidly over time.

Reinforcement process: For each pair of users, the probability of continued interactions displays a memory reinforcement (inertia). In particular, the more two nodes interact with each other, the more it demonstrates a close relationship between them . Thus, the user are more likely to reinforce this relationship to counteract the forgetting process.

Exploration process: In order to replace existing ones which are no longer attractive, users continuously explore new interaction relationships at a nearly constant rate, irrespective of their age or degree. We find that users prefer to interact with friends with whom they share significant overlap in social circles (homophily). This reveals a positive correlation between social structure and interaction as complementary indicators of social closeness between individuals.

The above processes captures the fundamental fact that the interaction relationships require that we invest time to keep them alive, especially once it becomes physically difficult for friends to meet face to face on a regular basis.

Combining our new observations with previously studied processes of these networks, we propose a generative model for social and interaction networks. Our model is important for understanding how the pairwise user interaction and social network evolve together. It explains why the number of people a user communicate with does not scale linearly with the number of friends users declare, and also explains along which friendship links that interactions are more likely to occur. The model is directly useful in the future interaction prediction (*e.g.*, by taking a current existing network and further evolving it) and in the design of algorithms incorporating social influence and homophily effects (*e.g.*, by locating and highlighting stronger relationship).

In addition, our model can be used to construct interaction traces that can represent the full spectrum of relationship strengths (from weak to strong), which has not been captured by models before. This is an important application because real-world network datasets are often proprietary and hard to obtain. Controlling network parameters allows the generation of datasets with different properties which can be used for thorough exploration and evaluation of network analysis algorithms.

Our contributions include the following: First, we discover a number of new interrelated processes drive the evolution of social interactions. Second, we propose a co-evolution model that precisely captures both social link formation and user interactions afterwards. Finally, we provide a thorough evaluation of our model, showing that it produces realistic network evolution following the true evolution of network properties.

The remainder of this paper is organized as follows. Section 2 provides background and related work on the growth of social network and user interaction. Section 3 provides insights into the interaction evolution by observing Renren social network. Section 4 provides our evolution model that captures both social network and user interaction. Section 5 evaluates the accuracy of our model, and we finally conclude in Section 6.

	Renren	Facebook
# of nodes	623,511	1,600,214
# of edges	8,266,149	48,949,304
Mean node degree	13.2	27.3
Mean path length	4.2	5.0
Mean clustering coefficient	0.18	0.19

Table 1: Properties of Renren and Facebook network

2. PRELIMINARIES

In this section, we provide background on work related to the growth of OSNs, and then introduce our datasets.

2.1 Related Work

Previous studies on social network evolution mainly focus on friendship relations, and attempt to discover the underlying processes that produce properties observed in real networks. For example, the preferential attachment model [3] captures power-law degree distributions. The forest fire model [19] captures the densification and shrinking diameters over time. A recently proposed, microscopic evolution model [18] provides insights into the node and edge arrival processes, and confirms preferential attachment and triangle closure features. Similar conclusions were reached by studies on Flickr [20] and a social network aggregator [10]. Zhao et al. [31] study the early evolution of the Renren social network, and analyze its network dynamics at different granularities to determine their influence on individual users.

Another set of works begin to investigate the effect of node attributes on social network evolution. For example, Allamanis et al. [1] examine influence of spatial factors on the temporal evolution of online social ties. Gong et al. [12] study the influence of four attributes including school, major, employer and city. They found users share attributes are more likely to be connected, augmenting structure-based triangle closing.

However, these works on evolutionary process or growth models treat all friendship links as equal. In fact, a recent study [11] has demonstrated the strength of links varies widely, ranging from users' best friends to acquaintances. To differentiate links, interaction data has been utilized in predicting relationship strength [11,16,27].

While the recent studies [8, 21, 26] brought great insights into the structural difference between the interaction network the social network, little attention has been paid to the temporal evolution of pairwise user interactions. The study [25] examined user interactions dynamics on Facebook users, but no generative model has been developed to reveal the underlying processes driving user interaction dynamics. With respect to these results, our work provides a more systematic understanding of the evolution of user interaction behavior.

Prior works [4, 29] provide some models of traffic networks, whereas others [24,30] present a model for face-to-face interactions of users. Although these models generate interaction network, they are not suitable in the context of today's OSNs due to different underlying dynamics and network properties. Our work attempts to fill this void.

2.2 Social Dynamics and Interaction Data

To construct the interaction evolution model, the dataset should contain the information on both topology and interaction dynamics.

	Renren	Facebook
Period	2005.11∼2006.12	2008.1∼2009.6
# of wall posts along edges	23,000,141	16,313,273
# of interactions	7,697,270	3,233,780
# of users having interactions	420,978	324,430
# of edge having interactions	2,623,040	1,695,448

Table 2: Summary of Renren and Facebook interaction data

However, to our knowledge, there are no publicly available datasets satisfying this requirement. To fill this lack, this paper presents two datasets: [1]

Renren Dataset. With 120 million users, Renren is the largest and oldest online social network in China, and provides functionality and features similar to Facebook. Like Facebook, Renren first started in 2005 as a social network for college students in China, then saw its user population grow exponentially once it opened its doors to the non-student population. Like Facebook, Renren users maintain personal profiles and establish bidirectional friendship links with others. Below we use the term *edge* to mean a friendship link.

To study user interactions, Renren provides us two ground-truth datasets. The first dataset encompasses the timestamped creation events of all users and edges in the first year of Renren's growth. The dataset starts on Nov. 21, 2005 (when the first edge was created) and ends on Dec. 30, 2006. In all, it includes the creation times of 623,511 nodes and 8,266,149 edges. Table 1 shows the statistics of the social graph formed at the end of 2006.

The second dataset includes all 29,506,068 wall posts that occurred in our measurement period. To guarantee user privacy, we only get the anonymized IDs of sender and receiver for each wall post, without knowing the content. Since our goal is to characterize edge strength based on user interactions, we ignore the wall posts not along edges (e.g., greeting messages between strangers). As a result, we focus on the remaining 23,000,141 wall posts created along edge, representing the friendship maintenance effort of users (accounting for nearly 80% of the total wall posts).

Facebook Dataset. Our Facebook dataset comes from a complete crawl of a large regional network conducted in 2009 [26]. This crawl visited the 1.6M users in the region with default privacy settings (roughly half of the total population of the region). Each user's friend list and all interactions in the user's news feed between Jan. 1, 2008 and Jun. 30, 2009 were downloaded. These interactions cover a broad range of activities, with the most popular by far being wall posts and photo tags. Each interaction includes ther sender, receiver, and a timestamp. Table 1 lists the statistics of the dataset.

Unfortunately, the Facebook data is not as comprehensive as the Renren data. The Facebook data does not include creation timestamps of social links, thus we only focus on analyzing user interaction patterns on Facebook, not graph structural dynamics. Furthermore, the Facebook data includes social links and interactions with users outside the target regional network. Because these users were not crawled, our data on them is incomplete. Thus, in our analysis we focus exclusively on social links and interactions between users in the region.

Other Types of Interactions. Although our datasets focus on Wall posts and photo comments, modern OSNs may have many additional types of interactions, *e.g.* retweets, shares, *etc.* When our datasets were collected, Wall posts and photo comments were the most popular types of interactions on Renren and Facebook by a large margin, which is why we focus on them [26]. However, our model is general enough to incorporate other types of interactions.

[1]Available at `http://net.pku.edu.cn\%7ezzh/data`.

2.3 Definitions and Dataset Cleaning

To better measure mutual relationship (tie strength) between users, we refer to a pair of reciprocal wall posts (or photo comments) as an *interaction*. For example, if node u sends m messages to v but receives n messages from v, the number of interactions between them is $\min(m, n)$. The wall posts that have not been replied are pruned. This definition means that u and v cannot be supposed to have strong mutual relationship if one sends many messages to the other but rarely receives replies (e.g., u trusts user v, but not necessarily vice versa). So we use the the number of interactions as a conservative estimate on the edge strength, instead of the total number of wall posts over the edge.

The interaction definition allows us to represent the interaction network evolution as a series of undirected, edge-weighted graphs $G_1, \ldots, G_T$, so that a snapshot G_t consists of the nodes, edges, and interactions that have arrived by time t. The term **interaction edge** represents the friendship edge along which at least one interaction is generated. We say a node u creates an interaction edge with a node v when u interacts with v for the first time, and we say v becomes one of u's interaction partners. We use the timestamp of an interaction as the creation time of the corresponding interaction edge.

Table 2 summarizes our Renren and Facebook interaction datasets. We see that Facebook users produce less than half as many interactions as Renren users, and fewer users and social edges are interactive. The reduced interactivity of Facebook users versus Renren users is very clear in our analysis in Section 3. However, as we will show, interactions on both OSNs still exhibit the same overall trends. Table 2 also shows that interaction edges only cover 32% and 3% of social edges on Renren and Facebook, respectively. This means that users only interact with a small subset of their friends.

3. ANALYSIS OF USER INTERACTIONS

In this section, we analyze the Renren and Facebook interaction data to uncover temporal patterns of user interactions. In particular, we want to answer three questions: First, at what rate do users choose new friends to interact with? Second, how do users choose which particular friend to begin interacting with? Finally, once a pair of users begin to interact, what are the temporal dynamics of the relationship? We shall leverage the answers to these questions to motivate the design of our generative model of user interactions.

3.1 Interaction Partners Addition

The first question we address is: what is the rate at which users add new interaction partners? Prior works [3, 18, 20] demonstrate that users accelerate the creation of social relations as their node degree (or age) increases, i.e., "rich-get-richer" type. Since users get friends more quickly, we want to examine whether they also ad-

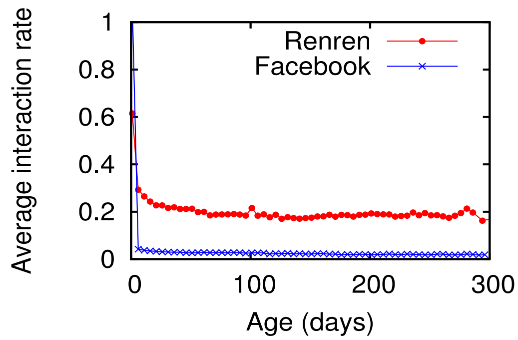


Figure 1: The average rate of adding new friends into interaction for nodes of different age.

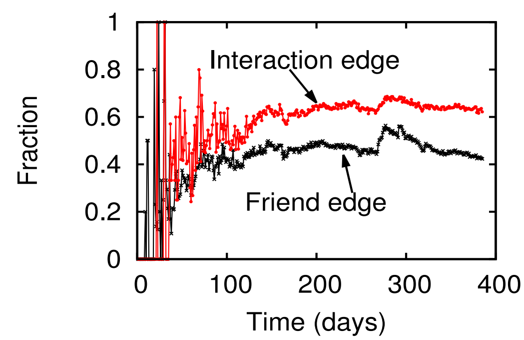


Figure 2: The fraction of friend and interaction edge creations targeting users with mutual friends in Renren.

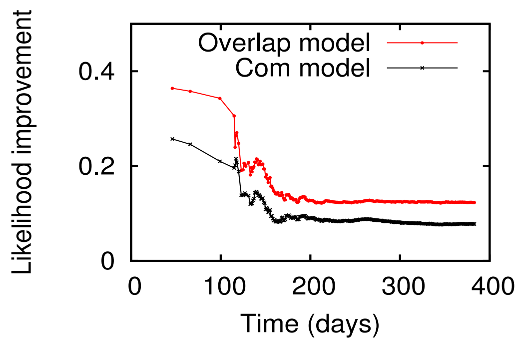


Figure 3: The likelihood improvement over the random selection in Renren.

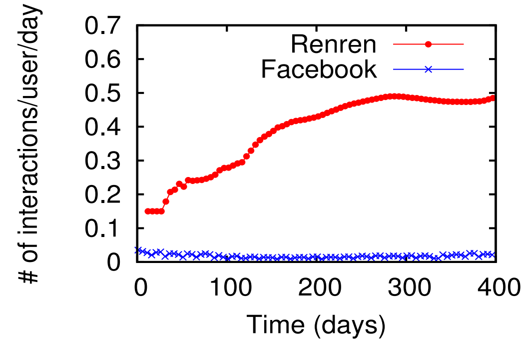


Figure 4: The average node-interaction frequency over time.

d interaction partners in an accelerate manner. Intuitively, making friends in an OSN is very easy, since the click of "add as friend" button does not need any energy cost. In contrast, interaction relationship requires more effort to create and maintain, e.g., a certain amount of time and energy used for reading and writing wall posts. Such energy cost will limit the rate at which users add new interaction partners, since they only have a finite amount of resources (e.g., time and energy).

To test this hypothesis, we examine the growth pattern of interaction edges. We define *node interaction rate* $\lambda_u(a)$ as the ratio of the number of interaction edges $n_u(a)$ that a node u has created to its current age a_u, i.e., $\lambda_u(a) = n_u(a)/a_u$. Interaction rate measures the speed at which users begin interacting with new friends. To examine the temporal pattern of interaction edge initiations, $\lambda(a)$, we calculate the average interaction rate of nodes achieving age a during our measurement period:

$$\lambda(a) = \frac{\sum_{t=1}^{T} \sum_{u \in S_t(a)} \lambda_u(a)/|S_t(a)|}{T} \quad (1)$$

where $S_t(a) = \{u|t - t_0 = a\}$ is the set of nodes achieving age a at time t. Here, t_0 is the arrival time of node.

As shown in Fig. 1, users in both Renren and Facebook are more interactive immediately after they join. However, the effect quickly wears off. For example, in Renren, the $\lambda(a)$ converges to a constant after only a week. The Pearson correlation coefficient between $\lambda(a)$ and node age is only -0.102, showing interaction rate is nearly independent of node age. This observation means that a node invites new interaction partners at a constant speed due to the interaction cost and limited time resource.

3.2 Interaction Partners Selection

The next question we address is: which friends do users select as interaction partners, given that only a small fraction of friends interact. Intuitively, strong relations are more likely to develop between socially similar people (*i.e.* homophily effects). We hypothesize that sharing common neighbors may have a strong impact on interaction partner selection.

To test this hypothesis, we unroll the evolution of the Renren network. Fig. 2 plots the fractions of friendship edge and interaction edge creation events occurring between users with mutual friends on each day. About 63% of interaction edge creation events occur between friends with mutual neighbors, as compared with 47% of friendship edge creation events. Thus, the common friend factor has a stronger influence on the creation of interaction relationship than on friendship.

After confirming the influence of common neighbors, we aim to understand the way that it affects how interaction targets are selected. Consider the case when a source node u initiates interaction with a friend v by sending the first wall post. We define *com* as the number of common neighbors and *overlap* as the Jaccard coefficient between u and v (i.e., $\frac{\Gamma_u \cap \Gamma_v}{\Gamma_u \cup \Gamma_v}$ where Γ_u and Γ_v are the sets of users connect to u and v, respectively). We examine the following alternatives for choosing node v: i) *com*: proportional to the number of common neighbors; ii) *overlap:* proportional to the Jaccard coefficient.

We apply the maximum likelihood principle to examine which model better explains the observed interaction data. Estimating the likelihood of a model M involves considering each interaction edge $s_t = (u, v)$ and computing the likelihood $P_M(s_t)$ that the

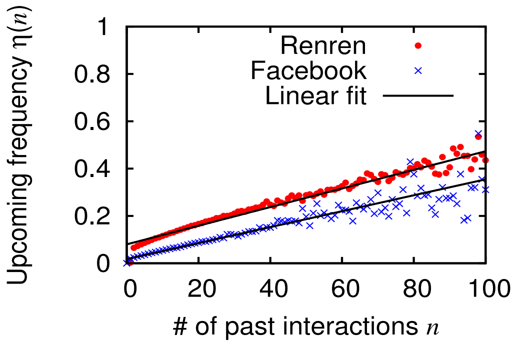


Figure 5: Upcoming interaction frequency $\eta(n)$ for an interaction edge that already has n interactions.

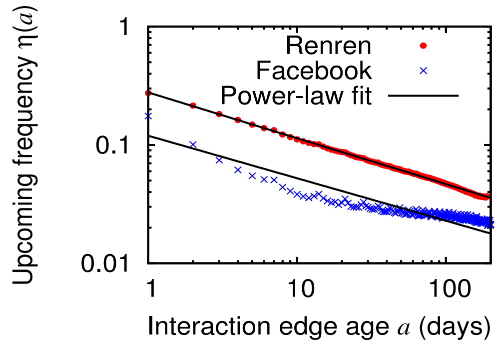


Figure 6: Upcoming interaction frequency $\eta(a)$ for an interaction edge of age a.

source u chooses v according to the model. In our case, $P_M(s_t) = c_v / \sum_{i=1}^{d} c_i$ where c_v (or c_i) is the numbers of common neighbors or neighborhood overlap between u and v (or ith friend of node u). The likelihood P_M that model M reproduces actual interaction edges across the graph is given by the product of the individual likelihoods: $P_M = \prod_t P_M(s_t)$. We use the log-likelihood $\log P_M$ for better numerical accuracy.

We compute the log-likelihood of `com` and `overlap` on each day. Fig. 3 shows the likelihood improvement of `com` and `overlap` over the random target selection, respectively. The `overlap` model shows improvement of 12.5%, compared to 8.5% for the `com` model. This result demonstrates that interactions are more likely to occur between friends with high neighborhood overlap, and that this effect is sustained over time, as the OSN matures.

Because the Facebook dataset does not include social link creation times, we cannot analyzing the correlation between social graph structure and user interaction over time. Instead, we analyze the neighborhood overlap between interactive (or non-interactive) friends on Facebook. We find that interactive friends have 0.085 neighborhood overlap on average, versus only 0.065 on average for non-interactive friends. The relatively low overlap numbers on Facebook are not surprising given that mean degree on Facebook is more than double mean degree on Renren (see Table 2). This result also demonstrates that users tend to initiate interactions with friends with high neighborhood overlap.

3.3 Interaction Generation and Distribution

The final question we ask is: how do users generate new interactions across existing interaction edges? We begin by analyzing the interaction creation process in absolute time, focusing on the speed that nodes generate interactions over their interaction edges.

We define the *node-interaction frequency*, as its total number of interactions averaged over time. Fig. 4 plots the average node-interaction frequency on each day of our dataset. In Renren, we see that the frequency increases early in its existence, but converges to a constant after day 230. Note that most users (and thus interactions) arrive after day 230 since Renren grows exponentially (as we shown later in Fig. 7). Our Facebook data was collected after the OSN had matured, and Fig. 4 shows that node interaction frequency is essentially constant. This result implies that an user spends a constant amount of time per day on interaction.

Next, we examine how users distribute new interactions over their existing interaction partners. We analyze the interaction distribution from two perspectives: first, what is the effect of *intensity*, *i.e.* is their correlation between the number of times friends have interacted in the past, and the number of times they will interact in the future? Fig. 5 plots $\eta(n)$, the average number of new interactions between friends that already have n interactions:

$$\eta(n) = \frac{\sum_{t=0}^{T} \sum_{e \in S_n(t)} I_e(t) / |S_n(t)|}{T} \qquad (2)$$

where $S_n(t) = \{e | \sum_{k=0}^{t-1} I_e(k) = n\}$ is the set of interaction edges that already have n interaction before time t. We observe that $\eta(n)$ is proportional to the number of past interactions across the edge in both networks. Intuitively, this means that the interactions between friends *reinforce* their relationship, leading to more future interactions.

Second, what is the effect of *time*, *i.e.* do friends tend to interact more or less over time? Fig. 6 plots $\eta(a)$, the average number of new interactions created along edges of age a:

$$\eta(a) = \frac{\sum_{t=0}^{T} \sum_{e \in S_a(t)} I_e(t) / |S_a(t)|}{T} \qquad (3)$$

where $S_a(t) = \{e | t - t_0(e) = a\}$ is the set of interaction edges with age a at time t, and $I_e(t)$ is the number of new interactions generated along edge e at time t. We see that $\eta(a)$ is inversely proportional to edge age a in both networks. Intuitively, this means that a given pair of users tends to interact less over time. One possible explanation for this is users tend to forget each other as they cannot meet face to face on a regular basis, leading to the closeness between friends declined rapidly over time.

4. A SOCIAL CO-EVOLUTION MODEL

In this section, we introduce our *generative* model for creating interaction graphs that takes into account the coupled evolution in time of topology and user interaction. The model is based on the insights about user interactions derived in the previous section. Intuitively, a co-evolution model has two complementary processes: one concerned with forming social links (the *social graph model*), while the other generates interactions along the links (the *interaction model*). Although many social graph models exist [1, 3, 12, 18, 19], these models do not include an interaction model.

4.1 Social Link Generation

Before we introduce our interaction model, we need to first choose an underlying social graph model to build upon. Rather than attempting to invent and justify a new social graph model, we choose to use the microscopic evolution model [18] for social link generation, which is based on observing the temporal properties of large social networks.

Algorithm 1 Social Co-evolution model.

```
Node set V = ∅
for each time step t ∈ T do
    Node arrival. V = V ∪ V_{t,new}
    for each new node u ∈ V_{t,new} do
        Lifetime sampling
        First social linking
    end for
    for each living node u ∈ V_t do
        if u wakes up then
            Social linking
            Sleep time sampling
        end if
        if rand() ≤ γ_u then
            if rand() ≤ p then
                Requests a friend v it has not interacted with
            else
                for each its interaction partner v do
                    Update the weight w_uv ← n_uv / a_uv^τ
                end for
                Pick an existing partner v with prob. ∝ w_uv
                Generate an interaction with v
            end if
        end if
    end for
end for
```

The main ideas behind the microscopic evolution model are that nodes join the social network following a node arrival function, and each node has a lifetime l, during which it wakes up multiple times and creates new edges by closing triads two random steps away (*i.e.* befriending friends of friends). The key functions and parameters needed for the microscopic evolution model are:

Node arrival. New nodes $V_{t,new}$ arrive at time t according to a pre-defined arrival function $N(.)$.

Lifetime sampling. At arrival time t, node u samples lifetime a from $\lambda e^{-\lambda a}$, and becomes inactive after time $t + a$. Let V_t be the set of active nodes at time t.

First social linking: Node u declares its first friend v based on the preferential attachment model (*e.g.* connecting to v with probability proportional to v's degree).

Sleep time sampling: After creating an edge, node u goes to sleep for δ time steps, where δ is sampled from a power law with exponential cut-off distribution given by $p(\delta) = \frac{1}{Z}\delta^{-\alpha}e^{-\beta \cdot degree(u) \cdot \delta}$.

Social linking: When node u wakes up, it creates a new edge by befriending a two-hop neighbor (a friend of a friend).

These are the set of parameters needed for the microscopic evolution model: $N(.)$ is the node arrival function, λ is the parameter of the exponential distribution of the lifetime, and α, β are the parameters of the power law with exponential cut-off distribution for the node sleep time gap. Further details of the model can be found in the paper [18].

4.2 User Interaction Generation

Besides befriending with others, nodes also request a certain number of friends to interact with, and distribute interactions over their interaction friends. Based on the insights on user interaction behavior, we now introduce our interaction model. Algorithm 1 presents our co-evolution model. Its interaction evolution part mainly consists of the following processes:

Intuitively, not all the users are simultaneously present in system. Thus we assume that users can be in an active or an inactive state. If an user is active, she interacts with her friends; otherwise she simply rests without interacting. According to empirical observations (e.g., constant interaction frequency), we assume that, at each time step, one inactive user can become active with a probability r, while one active user can become inactive with probability $1-r$. In practice this means that the user activity pattern, while stochastic, will display some regularity in time, interaction events following each other on average at $1/r$ steps, very long inter-event times are exponentially rare.

Once an user u is active, she would selects a target node v from her friends to make an interaction event, which increases the number of their interactions by one. The empirical observations show that an user invites new friends to interact at a constant rate, irrespective of node age or social degree. Therefore, we assume that, once a node is active, she selects the interaction target either from friends without any interaction with her yet or from existing interaction partners with probabilities p or $1 - p$, respectively. In other words, users are free to establish new interaction relationships with their social friends, and they are also responsible for maintaining existing interaction relationships with their partners, the degree of which is controlled by p.

In the case of selecting the target from its existing partners, interactions are biased by the interpersonal attraction built up over time. The more interest she raises in a partner, the more likely she will interact with this partner (inertia). Based on the empirical observations (e.g., effects of intensity and time), we measure the appeal η_{uv} of a partner v to an user u by $\eta_{uv} = n_{uv}/a_{uv}^{\tau}$, where n_{uv} is the current number of interactions between users u and v, a_{uv} is the current age of this interaction relationship and τ is the decay factor. Thus, if an user u chooses to interact with an existing partner, she will choose the partner v with a probability proportional to η_{uv}. In the other case, the probability to choose a friend as the target is proportional to their neighborhood overlap (homophily).[2] The new target node would be added into the set of existing interaction partners.

The interaction model captures the fundamental fact that the interaction relationships require that we invest time to keep them alive, especially once it becomes physically difficult for friends to meet face to face on a regular basis. In particular, each user has a *forgetting* behavior: the attraction between a pair of users declined rapidly when they lose contact (captured by the decay factor τ). Interestingly, our model on online relationships is consistent with the ecology model on real-life relationships. Prior work [6] investigated four annual surveys of colleague relationships for 345 bankers in a large financial organization, and found that the liveness of relationships decay over time and decay is also a power function of time. To counteract the effects of forgetting, each user exhibits a *reinforcing* behavior: she wants to keep the important relationships alive. Thus, with limited time to use, she biases towards relationships of more interactions. Also, each user has a *exploring* behavior: she continuously explores new interaction relationships (captured by the probability p), in order to replace existing ones which are no longer attractive.

4.3 Extension of the Model

The activation probability r represents user activeness in the social interaction. To this end, we have assumed that all users have the same tendency to be active, that is, the activation probability r does not depend on the user who is interacting. Real social systems display however additional complexity since the social behavior of individuals may vary significantly across the population. For example, individuals vary widely in the total time spent accessing OSNs [5], and may devote different amount of energy to interaction.

[2]One could further explore other social similarities, such as profile and geographic similarities, in choosing the new partner.

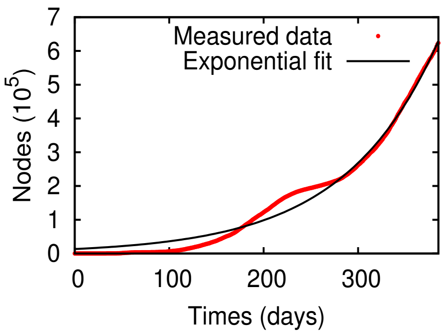


Figure 7: Nodes arrival over time, beginning Nov. 21, 2005.

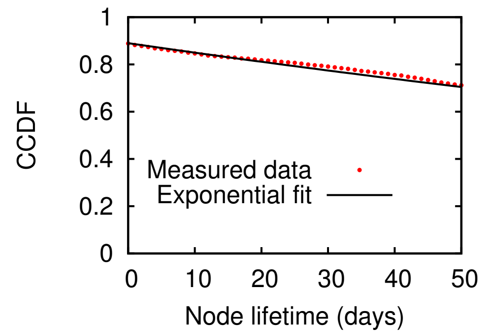


Figure 8: Complementary cumulative distribution of node lifetimes.

A natural extension of the model presented above consists therefore of making the probability r dependent on the user who is interacting. To this aim, we assign to each user u a parameter r_i that characterizes his/her propensity to form social interactions. In real networks this propensity will depend on the features of the users. In the model we assume that this propensity, that we call "sociability", is a quenched random variable randomly chosen from a prefixed distribution $\zeta(r)$ characterizing the system's heterogeneity, which is assigned to each agent at the start of the dynamical evolution and remains constant.

4.4 Desired Property

Before we evaluate our co-evolution model on real data, we first seek to demonstrate its theoretical soundness. To evaluate this, we focus on the relationship between social degree and interaction degree. Several studies of different social networks have all observed that the number of people a user communicates with (their interaction degree) *does not scale linearly* with social degree [14, 25, 26]. This means the social graph grows at a faster rate than the interaction graph. Given how universal this property is, the question becomes *does our model successfully capture this phenomenon*?

In on our co-evolution model, the growth of social degree and interaction degree are governed by different processes. Specifically, interaction edges are created at a constant rate, whereas the creation rate of friendship edges accelerates with social degree (*i.e.* the edge gap gets shorter). Thus, as the graph grows, the former quickly falls behind the latter, leading to a non-linear relationship.

We now formalize this relationship. Recall the friendship edge creation process: given the edge gap distribution $p(\delta|d;\alpha,\beta) = \frac{1}{Z}\delta^{-\alpha}e^{-\beta d\delta}$, a node creates the first edge and sleeps $\delta(1)$ time units sampled from $p(\delta|d=1;\alpha,\beta)$, creates the second edge and sleeps for $\delta(2)$ time units sampled from $p(\delta|d=2;\alpha,\beta)$, and so on. Thus, the time duration T needed by a user to achieve degree D is:

$$T = \sum_{d=1}^{D}\delta(d) \tag{4}$$

To get the edge gap $\delta(d)$, we first compute the normalizing constant Z for the edge gap:

$$Z = \int_0^\infty \delta^{-\alpha}e^{-\beta d\delta}d\delta = \frac{\Gamma(1-\alpha)}{(\beta d)^{1-\alpha}} \tag{5}$$

With the expression of constant Z, we obtain the expected time gap for a node to create its d_{th} edge:

$$E[\delta|d;\alpha,\beta] = \int_0^\infty \frac{1}{Z}\delta^{-\alpha}e^{-\beta d\delta}d\delta = \frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}(\beta d)^{-1} \tag{6}$$

Equation (6) shows that the gap gets shorter as the node's degree increases (*i.e.* $E(\delta)\propto 1/d$), manifesting the "rich-get-richer" phenomenon. On the other hand, the node creates interaction edges at a constant rate r. Hence, the average interaction degree K the node accumulates during T is:

$$K = r\sum_{d=1}^{D}E[\delta|d;\alpha,\beta] = r\sum_{d=1}^{D}\frac{\Gamma(2-\alpha)}{\Gamma(1-\alpha)}(\beta d)^{-1} = \Theta(lnD) \tag{7}$$

Therefore, the co-evolution model produces a *logarithmic* relationship between average interaction degree K and social degree D.

5. MODEL EVALUATION

We now perform simulations based on our Renren dataset to validate the accuracy of our model. We focus on the Renren dataset because it is *complete*, unlike the Facebook data which is missing edge creation times. We fit the parameters of our interaction model to the interactions that occurred during the first 320 days of our Renren dataset. This time period corresponds to half of the overall nodes joining the social graph. Later, we evaluate the ability of the interaction models to generate synthetic data that captures the interaction characteristics of the full 385-day Renren graph. We observe that the co-evolution model accurately captures the characteristics of the 385-day Renren data, even when it is only trained on half the dataset.

5.1 Social Graph Parameter Fitting

We analyze the Renren network to get the values for model parameters. The *social* graph model [18] needs the following parameters to generate the underlying social graph:

Node arrival function $N(.)$: We start by modeling the node arrival process. Fig. 7 measures the number of nodes in the network $N(t)$ on each day t. We see that Renren gets a burst of growth around day 200 (June 4, 2006) due to launching a network campaign at some biggest universities in Beijing. And after that, the network maintains the stable growth from day 287 (the end of August, 2006). We focus on the stable growth stage since it captures the arrival of most nodes, and fit the node arrival process by exponent function $N(t) = a\exp(bt)$, where $a = 13,200$ and $b = 0.01$. Thus, Renren grows exponentially over much of our network. *The parameter λ for node lifetime distribution:* We define node arrival as the time when a node creates its first social edge, and departure if the node does not create an edge for 100 days. Node lifetime is the time between node arrival and departure. Fig. 8

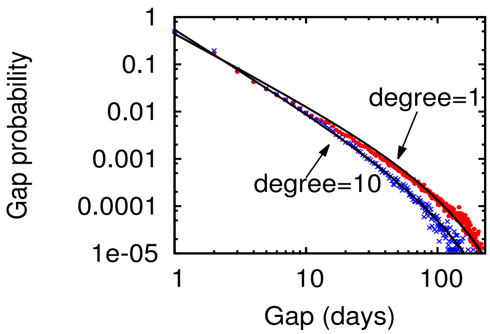


Figure 9: Edge gap distribution for a node to obtain the second edge.

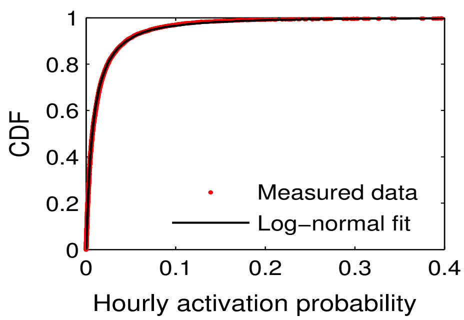


Figure 10: The distribution of user activation probability at an hourly time step.

shows the distribution of node lifetimes before day 230. Although about 10% nodes create an edge and never return, the lifetime of most nodes is fit by an exponential distribution with shape parameter $\lambda = 0.004$.

Parameters α, β of the sleep time gap distribution: The microscopic evolution model [18] defines edge gap δ as the time elapsing between the edge initiations from a node. In our case, we modify the model slightly since our dataset does not record which node initiates edge creation. Specifically, in our case, when edge (u, v) is created, both u and v perform sleep sampling according to their respective degrees. Although we modify the edge gap definition, we find that the edge gap distribution p_g can still be modeled by a power law with exponential cutoff: $p_g(\delta) \propto \delta^{-\alpha} \exp(-\beta d \delta)$. Fig. 9 confirms this model by showing the gap distributions and corresponding fittings ($\alpha = 1.735, \beta = 0.0008$) for nodes with different degree d.

5.2 Interaction Parameter Fitting

On the other hand, the *interaction* model needs the following parameters to generate interactions over the underlying social graph.

Decay factor τ: We compute $\eta(a)$, the interaction frequency for edges of age a, with equation (3) (note that T is limited to 320), and get the exponent $\tau = 0.4$.

Exploring probability p: We compute p by ratio of the existing interaction edges to the number of existing interactions by the end of training data (day 320), and get $p = 0.32$.

Activation probability r: In our evaluation, we use the heterogeneous model, which assumes that the activation probability r of an individual is randomly chosen from a prefixed distribution $\zeta(r)$. We compute the hourly activation probability r of an individual user by the average number of her interactions per hour by the end of training data. Fig. 10 shows that $\zeta(r)$ is a log-normal distribution, *i.e.* $\zeta(r) = \frac{1}{r\sigma\sqrt{2\pi}} e^{-(\ln r - \mu)^2/2\sigma^2}$, with $\mu = -4.9$ and $\sigma = 1.5$.

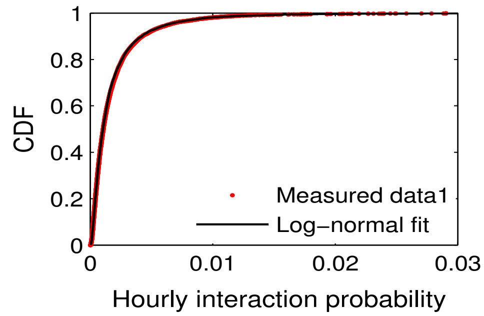


Figure 11: The distribution of edge interaction probability at an hourly time step.

Algorithm 2 Social Naïve model.

```
Node set V = ∅
for each time step t ∈ T do
    Node arrival. V = V ∪ V_{t,new}
    for each new node u ∈ V_{t,new} do
        Lifetime sampling
        First social linking
    end for
    for each living node u ∈ V_t do
        if u wakes up then
            Social linking (u, v)
            Sleep time sampling
            if rand() ≤ θ then
                Adds v into interaction partner group
                Samples probability η_{uv} from P(η)
            end if
        end if
        for each partner v do
            Generate a new interaction with η_{uv}
        end for
    end for
end for
```

5.3 Baseline Model

We propose a *naïve model*, which assumes the evolution of social network structure and user interaction are independent processes, and a stable interaction probability between a pair of users. This naïve model serves as a baseline for comparison.

At each time step, an node u creates social links using the microscopic evolution model (with the parameters outlined above). Once creating a new link, she selects this friend as interaction partner with a probability θ. At each time step, she interacts with each of her interaction partners with their own probability η_{uv} sampled from a given distribution $P(\eta)$ characterizing heterogeneous link strength. Algorithm 2 presents the naïve model in detail.

Using our training data, we get $\theta = 0.29$ as only about 29% of social edges have interactions. We quantify the hourly interaction probability of a pair of partners by their average number of interactions per hour by the end of training data. Fig. 11 shows that $P(\eta)$ is well-fit by a log-normal distribution, *e.g.* $P(\eta) = \frac{1}{\eta\sigma\sqrt{2\pi}} e^{-(\ln \eta - \mu)^2/2\sigma^2}$, with $\mu = -6.9$ and $\sigma = 1.1$.

5.4 Evaluation of Interaction Models

We now evaluate our interaction models. Using the parameter-

	Real Network	Co-evolution Model	Naïve Model
# of interactions	7,697K	7,979K	11,822K
# of users have interactions	421K	452K	470K
Mean # of interactions/user	18.3	17.7	25.1
# of edge have interactions	2,623K	2,654K	2,285K
Mean # of interactions/edge	2.9	3.0	5.2

Table 3: Statistics of a real network vs. synthetic ones

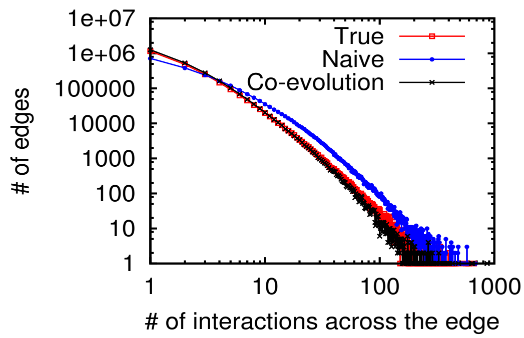


Figure 12: The number of interactions across interaction edges.

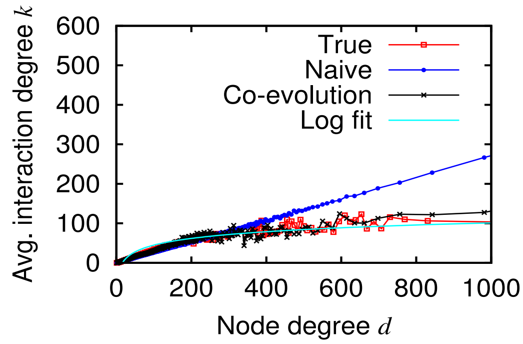


Figure 13: The correlation between node degree and interaction degree.

ization outlined above, we generate two social graphs with associated interactions (one from the naïve model, one from the co-evolution model). Each graph is evolved for 385 days, and then compared to the complete 385 days of ground-truth data from Renren. We focus on analyzing the interactions generated by the two models, rather than the social graph generated by the microscopic evolution model, because the microscopic model has already been thoroughly evaluated by prior work [18].

We analyze the generated interactions at two levels of granularity. First, we evaluate high-level interaction characteristics (*e.g.* total interactions, interactions per user/edge). Second, we examine the structural features of interaction graphs (*e.g.* degree distribution and clustering coefficient). In both cases, the output of the naïve model serves as a baseline for gauging the improvement of the co-evolution model.

We now present our evaluation results by comparing both the high-level characteristics (such as the total number of interaction edges and interactions) and microscopic structure features (e.g., clustering coefficient, degree distribution, and pairwise distance) of real and synthetic interaction graphs. The performance of the naïve model serves as a baseline that will help us to confirm the necessity of each process in the co-evolution model.

Interaction Analysis. Table 3 shows the overall interaction statistics of the real and synthetic graphs. The output of the co-evolution model is very close to the true data in every category. In contrast, the naïve model generates 54% too many interactions overall and 12% too many users become interactive. By capturing the fraction of interaction edges, the naïve model could roughly predict the total number of interaction edges, but severely overestimates the total number of interactions due to ignoring the recency effect in user interactions (i.e., the interaction frequency of user pairs tends to decrease markedly over time). As a comparison, the co-evolution model rightly captures the rates that users create both interaction edges and interactions, thus generating the very similar statistics to those of the real graph.

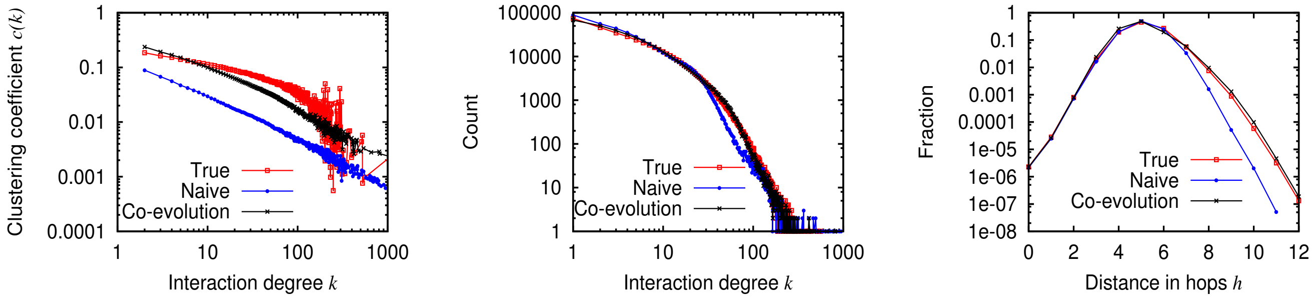
Next, we examine the number of interactions across interaction edges (*i.e.* link strength) in Fig. 12. Since the naïve model does not take interaction recency into account (*i.e.* interactions along an edge tend to decrease over time), it overestimates link-strengths. In contrast, the co-evolution model takes interaction intensity and recency into account, and thus generates nearly the same link-strength distribution as the real graph.

Finally, we examine the correlation between social degree and interaction degree, since this non-linear correlation is an important observation in many social networks [14, 25, 26]. Fig. 13 shows the correlation for the Renren network, which exhibits the expected sub-linear relationship. The naïve model cannot generate this non-linear correlation. In contrast, the co-evolution model does generate the expected sub-linear relationship, and fits the ground-truth Renren data very closely. We further find that the logarithmic function fits true correlation curve well, validating empirically the derivation of a logarithmic correlation in Section 4.4.

Interaction Graph Analysis. To further evaluate our interaction models, we construct undirected *interaction networks*, where a link exists between a pair of users that interacts at least once (*i.e.* interaction edge). We then compare the structural properties of the 385-day Renren interaction network to the synthetic networks from the naïve and co-evolution models.

Fig. 14 shows the clustering coefficient distribution, degree distribution, and pairwise distance histogram for the Renren and synthetic interaction networks. We observe that both models produce accurate interaction degree distributions. However, the co-evolution model is much more accurate at reproducing accurate clustering and distance characteristics. The weakness of the naïve model in these two cases is due to the fact that it ignores the correlation between social structure and interactions. In contrast, the co-evolution model is designed to capture this correlation, and consequently generates interaction networks that closely match the true network.

We note that the true network has a slightly higher clustering coefficient than that generated by the co-evolution model. This indicates that the ground-truth interaction partners are more densely



(a) Clustering Coefficient. (b) Degree distribution. (c) Geodesic distance.

Figure 14: Comparison between the real and synthetic Renren interaction graphs.

clustered than that predicted by the model (where the partner selection is proportional to the neighborhood overlap). To improve the accuracy of our model with respect to clustering coefficients, we could add an additional parameter to the model that increases its bias towards selecting interaction partners with high neighborhood overlap. However, this additional parameter adds complexity to the model, and we leave the study of how to tune this new parameter as future work.

6. CONCLUSION AND DISCUSSION

In this paper, we develop and evaluate a co-evolution model that generates interactions across social links. The insights behind our model are derived from large scale datasets from Renren and Facebook. This data reveals that users invite new friends to interact at a nearly constant rate, prefer to interact with friends with whom they share significant overlaps in social circles, and gradually lose interest in interacting with old friends. We believe that these observations not only affect the design of network interaction models but also have broader implications in other areas, such as friend recommendation, information diffusion, and news feed ranking.

Our co-evolution interaction model captures the important statistical properties of interaction networks, and provides new insights into the evolution of user interaction during network formation. To our knowledge, the co-evolution model is the first generative model for interactions on OSNs, and our evaluation shows that is it very accurate at capturing the observed properties of real OSN data. Although, we only evaluate our co-evolution interaction model when paired with the microscopic social graph evolution model, one of the strengths of our interaction model is that it can be paired with any underlying model for generating the social graph structure.

Another strength of the co-evolution model is that it is scalable, because individual nodes act locally (i.e., focusing on their neighbors) and independently (i.e., no coordination of one's own actions with those of others). Thus, the model can easily be parallelized, where each machine is responsible for performing the social linking and interaction processes of a subset of nodes.

Bursty Dynamics. In this paper, our interaction model mainly focuses on capturing user behavior in distributing interactions among friends as OSN structure evolves. However, our model simplifies reality by assuming that each interaction is immediately generated between a node and her target, so the fine-grained temporal features of interactions (such as bursty dynamics) are not captured by the model. One way to mitigate this issue would be to introduce response delay into the model, e.g., an user could respond only when she is active. This delay would control the speed at which nodes respond to interactions; two nodes that both have low delay would thus generate fast bursts of interactions. Also, response delay could be various from the per edge perspective. Intuitively, this would capture cases where users quickly respond to interactions from strong friends, while delaying responses to acquaintances (or even ignoring these interactions entirely). We leave this extension of the model as future work.

Future Work. There are several directions to extend the current study. First, we can study the interaction graph evolution at the community level. Seshadhri et al. [17, 23] have proposed scalable models for reproducing social graphs with community structure. However, for interaction graphs, one needs to further study the correlation between link weight and community evolution, since links of various strength might play different roles on the community formation. Another direction is to accommodate more attributes of nodes to improve the accuracy of the model. Recent works [1, 12] begin to examine influence of spatial and profile attributes on the temporal evolution of friendship links, but how these factors affect interaction evolution remains unknown.

7. ACKNOWLEDGMENTS

We thank the anonymous reviewers for their comments, as well as our shepherd Cecilia Mascolo. This work was supported by the National Basic Research Program of China (Grant No. 2014CB340 400), the National Science Foundation for Young Scholars of China (Grant No.61202423), and the NSF under grant CNS-1319019, IIS-1321083 and CNS-1527939. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

8. REFERENCES

[1] ALLAMANIS, M., SCELLATO, S., AND MASCOLO, C. Evolution of a location-based online social network: Analysis and models. In *Proc. of IMC* (2012).

[2] BACKSTROM, L., AND LESKOVEC, J. Supervised random walks: Predicting and recommending links in social networks. In *Proc. of WSDM* (2011).

[3] BARABÁSI, A., AND ALBERT, R. Emergence of scaling in random networks. *Science 286*, 5439 (1999), 509–512.

[4] BARRAT, A., BARTHÉLEMY, M., AND VESPIGNANI, A. Modeling the evolution of weighted networks. *Physical review E* (2004).

[5] BENEVENUTO, F., RODRIGUES, T., CHA, M., AND ALMEIDA, V. Characterizing user behavior in online social networks. In *Proc. of IMC* (2009).

[6] BURT, R. S. Emergence of scaling in random networks. *Social Networks 1*, 22 (2000), 1–28.

[7] CHA, M., HADDADI, H., BENEVENUTO, F., AND GUMMADI, K. P. Measuring user influence in twitter: The million follower fallacy. In *Proc. of ICWSM* (May 2010).

[8] CHUN, H., KWAK, H., EOM, Y., AHN, Y., MOON, S., AND JEONG, H. Comparison of online social relations in volume vs interaction: a case study of Cyworld. In *Proc. of IMC* (2008).

[9] GAO, H., HU, J., WILSON, C., LI, Z., CHEN, Y., AND ZHAO, B. Y. Detecting and characterizing social spam campaigns. In *Proc. of IMC* (2010).

[10] GARG, S., GUPTA, T., CARLSSON, N., AND MAHANTI, A. Evolution of an online social aggregation network:an empirical study. In *Proc. of IMC* (2009).

[11] GILBERT, E., AND KARAHALIOS, K. Predicting tie strength with social media. In *Proc. of CHI* (2009).

[12] GONG, N. Z., XU, W., HUANG, L., MITTAL, P., STEFANOV, E., SEKAR, V., AND SONG, D. Evolution of social-attribute networks: Measurements, modeling, and implications using google+. In *Proc. of IMC* (2012).

[13] GRIER, C., THOMAS, K., PAXSON, V., AND ZHANG, M. @spam: the underground on 140 characters or less. In *Proc. of CCS* (2010).

[14] HUBERMAN, B. A., ROMERO, D. M., AND WU, F. Social networks that matter: Twitter under the microscope. *CoRR* (2008).

[15] JIANG, J., WILSON, C., WANG, X., HUANG, P., SHA, W., DAI, Y., AND ZHAO, B. Y. Understanding latent interactions in online social networks. In *Proc. of IMC* (2010).

[16] KAHANDA, I., AND NEVILLE, J. Using transactional information to predict link strength in online social networks. In *Proc. of ICWSM* (2009).

[17] KOLDA, T. G., PINAR, A., PLANTENGA, T., AND SESHADHRI:, C. A scalable generative graph model with community structure. arXiv:1302.6636, 2013.

[18] LESKOVEC, J., BACKSTROM, L., KUMAR, R., AND TOMKINS, A. Microscopic evolution of social networks. In *Proc. of KDD* (2008).

[19] LESKOVEC, J., KLEINBERG, J. M., AND FALOUTSOS, C. Graph evolution: Densification and shrinking diameters. *TKDD* (2007).

[20] MISLOVE, A., KOPPULA, H. S., GUMMADI, K. P., DRUSCHEL, P., AND BHATTACHARJEE, B. Growth of the flickr social network. In *Proc. of WOSN* (Seattle, WA, August 2008).

[21] NAZIR, A., WAAGEN, A., VIJAYARAGHAVAN, V. S., CHUAH, C.-N., D'SOUZA, R., AND KRISHNAMURTHY, B. Beyond friendship: Modeling user activity graphs on social network-based gifting applications. In *Proc. of IMC* (2012).

[22] SCHNEIDER, F., FELDMANN, A., KRISHNAMURTHY, B., AND WILLINGER, W. Understanding online social network usage from a network perspective. In *Proc. of IMC* (2009).

[23] SESHADHRI, C., KOLDA, T. G., AND PINAR, A. Community structure and scale-free collections of erdos-renyi graphs. *Physical Review E* (2012).

[24] STARNINI, M., BARONCHELLI, A., AND PASTOR-SATORRAS, R. Modeling human dynamics of face-to-face interaction networks. *Physics Review Letter* (2013).

[25] VISWANATH, B., MISLOVE, A., CHA, M., AND GUMMADI, K. P. On the evolution of user interaction in facebook. In *Proc. of WOSN* (2009).

[26] WILSON, C., BOE, B., SALA, A., PUTTASWAMY, K. P. N., AND ZHAO, B. Y. User interactions in social networks and their implications. In *Proc. of EuroSys* (2009).

[27] XIANG, R., NEVILLE, J., AND ROGATI, M. Modeling relationship strength in online social networks. In *Proc. of World Wide Web Conference* (2010).

[28] YANG, Z., WILSON, C., WANG, X., GAO, T., ZHAO, B. Y., AND DAI, Y. Uncovering social network sybils in the wild. In *Proc. of IMC* (2011).

[29] YOOK, S. H., JEONG, H., AND BARABÁSI, A.-L. Weighted evolving networks. *Physical review letters* (2001).

[30] ZHAO, K., STEHLÉ, J., BIANCONI, G., AND BARRAT, A. Social network dynamics of face-to-face interactions. *Physical review E* (2011).

[31] ZHAO, X., SALA, A., WILSON, C., WANG, X., GAITO, S., ZHENG, H., AND ZHAO, B. Y. Multi-scale dynamics in a massive online social network. In *Proc. of IMC* (2012).