Understanding Latent Interactions in Online Social Networks

Jing Jiang[§†], Christo Wilson[†], Xiao Wang[§], Peng Huang[§], Wenpeng Sha[§], Yafei Dai[§] and Ben Y. Zhao[†]
[§]Peking University, Beijing, China
[†]U. C. Santa Barbara, Santa Barbara, CA, USA
{jiangjing,wangxiao,huangpeng,swp,dyf}@net.pku.edu.cn, {bowlin,ravenben}@cs.ucsb.edu

ABSTRACT

Popular online social networks (OSNs) like Facebook and Twitter are changing the way users communicate and interact with the Internet. A deep understanding of user interactions in OSNs can provide important insights into questions of human social behavior, and into the design of social platforms and applications. However, recent studies have shown that a majority of user interactions on OSNs are *latent interactions*, passive actions such as profile browsing that cannot be observed by traditional measurement techniques.

In this paper, we seek a deeper understanding of both visible and latent user interactions in OSNs. For quantifiable data on latent user interactions, we perform a detailed measurement study on Renren, the largest OSN in China with more than 150 million users to date. All friendship links in Renren are public, allowing us to exhaustively crawl a connected graph component of 42 million users and 1.66 billion social links in 2009. Renren also keeps detailed visitor logs for each user profile, and counters for each photo and diary/blog entry. We capture detailed histories of profile visits over a period of 90 days for more than 61,000 users in the Peking University Renren network, and use statistics of profile visits to study issues of user profile popularity, reciprocity of profile visits, and the impact of content updates on user popularity. We find that latent interactions are much more prevalent and frequent than visible events, non-reciprocal in nature, and that profile popularity are uncorrelated with the frequency of content updates. Finally, we construct *latent interaction graphs* as models of user browsing behavior, and compare their structural properties against those of both visible interaction graphs and social graphs.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and behavioral sciences; H.3.5 [**Information Storage and Retrieval**]: Online Information Services

General Terms

Human factors, Measurement, Performance

IMC'10, November 1–3, 2010, Melbourne, Australia.

Keywords

Latent interactions, Online social networks

1. INTRODUCTION

Not only are online social networks (OSNs) popular tools for interaction and communication, but they have the potential to alter the way users interact with the Internet. Today's social networks already count close to one billion members worldwide. Facebook, the most popular OSN, has more than 500 million active users [26], and has surpassed Google as the most visited site on the Internet [30]. Increasingly, Facebook and Twitter are replacing email and search engines as users' primary interfaces to the Internet [9, 15]. This trend is likely to continue, as networks like Facebook seek to personalize the web experience by giving sites access to information about their visitors and their friends, through new platforms such as OpenGraph [23].

A deep understanding of user interactions in social networks can provide important insights into questions of human social behavior, as well as the design of social platforms and applications. For example, gauging the level of reciprocity in social interactions can shed light on the factors that motivate interactions. In addition, understanding how interactions are distributed between friends can assist in tracking information dissemination in social networks, thus identifying "popular" or "influential" users to target in branding and ad campaigns [6, 11, 14]. Finally, lessons from studying how users interact through different communication tools can guide the design of new, more engaging mechanisms for social interaction.

Initial measurement studies [1, 22, 29] of OSNs focused on topological characteristics of the social graph, the underlying structures of these services that capture explicit relationships between users. To better understand the true nature of relationships between OSN users, more recent work has shifted focus to measuring observable social interactions [7, 19, 28, 29]. By examining records of interaction events across different links, the studies distinguish close-knit, active relationships from weak or dormant relationships, and derive a more accurate predictive model for social behavior.

Recently, two significant studies [3, 25] used clickstream data at the network level to capture the behavior of OSN users, and revealed that passive or *latent interactions* such as profile browsing often dominate user events in a social network [3].

Unfortunately, these studies have been constrained by several limitations of clickstream data. First, the type of data captured in a clickstream is highly dependent on the time range of the clickstream. Captured events are also from the perspective of the active user, making it challenging to correlate events across time and users. Second, clickstream data is also highly dependent on the structure of the OSN site, and can be extremely challenging to reduce large volumes of data to distinct user events. Finally, each application-level user event generates a large volume of clickstream data, and extremely large clickstreams are needed to capture a significant number of user events. These properties of verboseness and complexity mean that it is extremely difficult to gather enough clickstream data to study user interactions comprehensively at scale. However, a comprehensive and large study is necessary to answer many of the deeper questions about user behavior and interactions, such as: are user interactions reciprocal, do latent interactions such as profile browsing reflect the same popularity distributions as visible actions like user comments, what can users do to become “popular” and draw more visitors to their pages?

In this paper, we seek to answer these and other questions in our search for a deeper understanding of user interactions in OSNs. To address the challenge of gathering data on latent interactions, we perform a large-scale, crawl-based measurement of the Renren social network [24], the largest and most popular OSN in China. Functionally, it is essentially a clone of Facebook, with similar structure, layout and features. Like Facebook, Renren also evolved from a university-based social network (a predecessor called Xiaonei). Unlike Facebook, Renren has *two unique features* that make it an attractive platform to study user interactions.

First, while Renren users have full privacy control over their private profiles, their friend lists are public and unprotected by privacy mechanisms. This allowed us to crawl an exhaustive snapshot of Renren’s largest connected component, producing an extremely large social graph with 42.1 million nodes and 1.66 billion edges. Second, and perhaps more importantly, Renren user profiles make a variety of statistics visible to both the profile owner and her visitors. Each user profile keeps a visible list of “recent visitors” who browse the profile, sorted in order, and updated in real-time. Each photo and diary entry also has its own page with a count of visits by users other than the owner. These records are extremely valuable, in that they expose latent browsing events to our crawlers, granting us an unique opportunity to gather and analyze large scale statistics on latent browsing events.

Our study. Our study of latent user interactions includes three significant components. *First*, we begin by characterizing properties of the large Renren social graph, and compare them to known statistics of other OSNs, including Facebook, Cyworld, Orkut and Twitter. Our *second* component focuses on questions concerning latent interactions, and constitutes the bulk of our study. We describe a log reconstruction algorithm that uses relative clocks to merge visitor logs from repeated crawls into a single sequential visitor stream. We repeatedly crawl users in the Peking University Renren network over a period of 90 days, extract profile visit history for 61K users, and examine issues of popularity, visitor composition, reciprocity, and latency of reciprocation. We compare user popularity distributions for latent and visible interactions, and use per-object visit counters to quantify the level of user engagement generated from user profiles, photos, and diary entries. We also study correlation of different types of user-generated content with a user’s profile popularity. Finally, in our *third* component, we build *latent interaction graphs* from our visitor logs, and compare their structure to those of social graphs and interaction graphs. We also revisit the issue of experimental validation for social applications, and perform a case study of the impact of different social graphs on evaluating information dissemination algorithms.

Our study provides a number of insights into user behavior on online social networks, including:

- Users’ profile popularity varies significantly across the population, and closely follows a Zipf distribution.
- Profile visits have extremely low reciprocity, despite the fact that Renren users have full access to the list of recent visitors to their profile.
- Compared to visible interactions, latent profile browsing is far more prevalent and more evenly distributed across a user’s friends. Profile visits are less likely to be repeated than visible interactions, but are more likely to generate visible comments than other content such as photos and diary entries.
- For all users, regardless of their number of friends, profile popularity is not strongly correlated with frequency of new profile content.

Finally, we use our data to construct *latent interaction graphs* that capture browsing activity between OSN users. Our analysis finds that latent interaction graphs exhibit general Power-law features, fall between social and visible interaction graphs in terms of connectivity, but show less local clustering properties than both.

2. METHODOLOGY AND INITIAL ANALYSIS

Before diving into detailed analysis of user interaction events, we begin by providing background information about the Renren social network and our measurement methodology. We then give more specifics on our techniques for reconstructing profile browsing histories from periodic crawls. Using a random subset of user profiles, we perform sampling experiments to quantify the expected errors introduced by our approach. Finally, we analyze characteristics of the Renren social graph, and compare it to known graph properties of existing social graph measurements.

2.1 The Renren Social Network

Launched in 2005, Renren is the largest and oldest OSN in China. Renren can be best characterized as Facebook’s Chinese twin, with most or all of Facebook’s features, layout, and a similar user interface. Users maintain personal profiles, upload photos, write diary entries (blogs), and establish bidirectional social links with their friends. Renren users inform their friends about recent events with 140 character status updates, much like tweets on Twitter. Similar to the Facebook news feed, all user-generated updates and comments are tagged with the sender’s name and a time stamp.

Renren organizes users into membership-based networks, much like Facebook used to. Networks represent schools, companies, or geographic regions. Membership in school and company networks require authentication. Students must offer an IP address, email address, or student credential from the associated university. Corporate email addresses are needed for users to join corporate networks. Renren’s default privacy policy makes profiles of users in geographic networks private. This makes them difficult to crawl [29]. Fortunately, profiles of users in authenticated networks are public by default to other members of the same network. This allowed us to access user profiles within the Peking University network, since we could create nearly unlimited authenticated accounts using our own block of IP addresses.

Like Facebook, a Renren user’s homepage includes a number of friend recommendations that encourage formation of new friend relationships. Renren lists 3 users with the most number of mutual friends in the top right corner of the page. In addition, Renren shows a list of 8 “popular users” at the very bottom of the page. These popular users are randomly selected from the 100 users with the most friends in the university network.

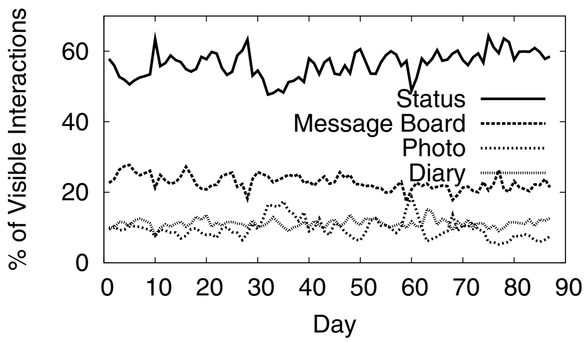

Figure 1: Daily distribution of comments across applications.

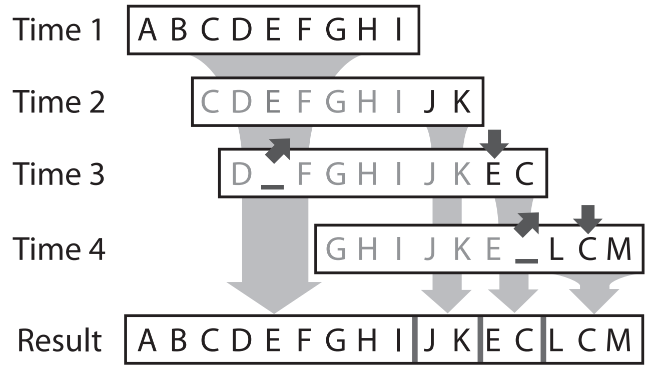

Figure 2: Integrating multiple visitor lists captured by multiple crawls of the same profile into a single history.

Unique features. Renren differs from Facebook in several significant ways. First, each Renren user profile includes a box that shows the total number of visitors to the profile, along with names and links to the last 9 visitors ordered from most to least recent. In addition, Renren also keeps on each individual photo and diary page a visible counter of visitors (not including the user himself). These lists and counters have the same privacy settings as the main profile. They have the unique property of making previously invisible events visible, and are the basis for our detailed measurements on latent user interactions.

A second crucial feature is that friend lists in Renren are always public. Users have no way to hide them. This allowed us to perform an exhaustive crawl of the largest connected component in Renren (42.1 million users). This contrasts with other OSNs, where full social graph crawls are prevented by user privacy policies that hide friendship links from the public. The exception is Twitter, which behaves more like a public news medium than a traditional social network [16].

In addition, comments in Renren are threaded, *i.e.* each new comment is always in response to one single other event or comment. For example, user A can respond to user B's comment on user C's profile, and only B is notified of the new message. Thus we can precisely distinguish the intended target of each comment. One final difference between Renren and Facebook is that each standard user is limited to a maximum of 1,000 friends. Users may pay a subscription fee to increase this limit to 2,000. From our measurements, we saw that very few users (0.3%) took advantage of this feature.

2.2 Data Collection and General Statistics

Like Facebook, Renren evolved from a social network in a university setting. Its predecessor was called Xiaonei, literally meaning "inside school." In September 2009, Renren merged with Kaixin, the second largest OSN in China, and absorbed all of Kaixin's user accounts.

Crawling the Renren social graph. We crawled the entire Renren network from April 2009 to June 2009, and again from September to November of 2009. We seed crawlers with the 30 most popular users' profiles, and proceeded to perform a breadth-first traversal of the social graph. During the crawl, we collect unique userIDs, network affiliations, and friendship links to other users. For our study, we use data from our last crawl, which was an exhaustive snapshot that included 42,115,509 users and 1,657,273,875 friendship links. While this is significantly smaller than the 70 million users advertised by Renren in September 2009, we believe the discrepancy is due to Kaixin users who were still organized as a separate, disconnected subgraph. We describe properties of the social graph later in this section.

Crawling the PKU network. We performed smaller, more detail oriented crawls of the PKU network between September and November of 2009 (90 days) to collect information about user's profiles and interaction patterns. This methodology works because the default privacy policy for authenticated networks is to make full profiles accessible to other members of the same network. Since we collected the network memberships of all users during our complete crawl, we were able to isolate the 100,973 members of the PKU network to seed our detailed crawl. Of these users, 61,405 users had the default, permissive privacy policy, enabling us to collect their detailed information. This covers the majority of users (60.8%) in the PKU network, and provides overall network coverage similar to other studies that crawled OSN regional networks [29].

As part of our PKU crawls, we gathered all comments generated by users in message board posts, diary entries, photos, and status updates. This data forms the basis of our experiments involving visible interactions. Our dataset represents the complete record of public visible interactions between users in the PKU network. In total, 19,782,140 comments were collected, with 1,218,911 of them originating in the September to November 2009 timeframe.

Figure 1 plots the percentage of comments in various applications each day. The most popular events to comment on are status updates, which accounts for roughly 55% of all daily comments. Message boards cover 25%, while diary and photo each account for roughly 10%.

Privacy and data anonymization. Our study focuses on the structure of social graphs and interaction events between users. Since we do not need any actual content of comments, photos, or user profiles, we waited for crawls to complete, then went through our data to anonymize userIDs and strip any private data from our dataset to protect user privacy. In addition, all user IDs were hashed to random IDs, and all timestamps are replaced with relative sequence numbers. We note that our group has visited and held research meetings with technical teams at Renren, and they are aware of our ongoing research.

2.3 Measuring Latent User Interactions

In addition to visible interactions generated by users in the PKU network, we also recorded the recent visitor records displayed on each user's profile. This data forms the basis of our study of latent interactions.

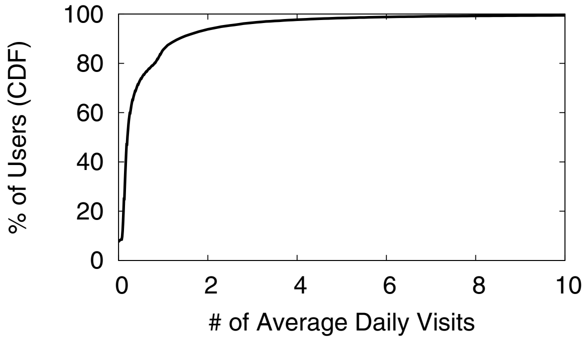

Figure 3: Average daily visit counts of user profiles.

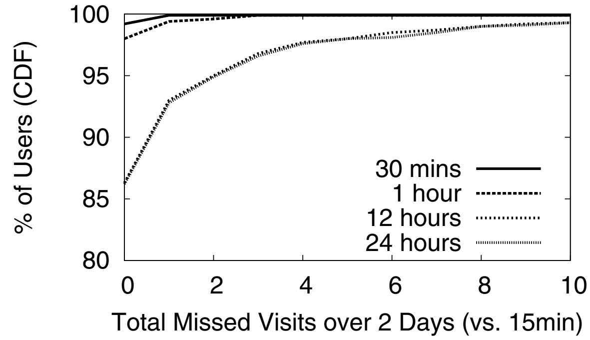

Figure 4: Number of visits missed when we lower crawler frequency from a high of once every 15 minutes.

Reconstructing Visitor Histories. Crawling Renren for recent visitor records is complicated by two things. First, each user's profile only lists the last 9 visitors. This means that our crawler must be constantly revisiting users in order to glean representative data, as new visitors will cause older visitors to fall off the list. Clearly we could not crawl every user continuously. Frequent crawls leave the ID of our crawler on the visitor log of profiles, which has generated unhappy feedback from profile owners. In addition, Renren imposes multiple per-account rate limits that slow our crawler significantly despite our large number of crawler accounts. Thus, we designed our crawler to be self-adapting. This means that we track the popularity and level of dynamics in different user profiles, and allocate most of our requests to heavily trafficked user profiles, while guaranteeing a minimum crawl rate (1/day) for low traffic users. The individual lists from each crawl contain overlapping results, which we integrate into a single history.

The second challenge to crawling recent visitor records is that each visitor is only shown in the list once, even if they visit multiple times. Repeat visits simply cause that user to return to the top of the list, erasing their old position. This makes identifying overlapping sets of visitors from the iterative crawls difficult.

To solve these two challenges, we use a log-integration algorithm to concatenate the individual recent visitor lists observed during each successive crawl. More specifically, some overlapping sets of visitors exist in successive crawl data, and our main task is to find new visitors and remove overlaps. There are two kinds of incoming visitors: new users, who do not appear in the previous list, and repeat users, who appear in the prior list at a different relative position. The first kind of incoming visitor is easily identified, since his record is completely new to the recent visitor list. New visitors provide a useful checkpoint for purposes of log-integration, since other users behind them in the list are also necessarily new incoming visitors. The second type of incoming visitor, repeat users, can be detected by looking for changes in sequence of the recent visitor list. If a user repeatedly visits the same profile in-between two visits of other users, nothing changes in the recent visitor list. Therefore, consecutive repeat visits are ignored by our crawler.

Figure 2 demonstrates our integration algorithm. We observe that visitors `ABCDEFGHI` viewed a user's profile at some time before our first crawl. New users view the profile and are added to the recent visitor list by the second crawl at time 2. We re-observe the old sequence `CDEFGHI`, and identify `JK` as new visitors, since `JK` do not exist in the previous visitor list. Next, we compare recent visitor lists at time 2 and 3. We find that `E` is before `K` in the recent visitor list crawled at time 2, but this order is changed at time 3. This means that at some time before the third crawl user `E` revisited the target and changed positions in the list. Thus we identify `E` as a new visitor. Since `C` is behind `E` at time 3, `C` is also identified as a new visitor. Our integration algorithm also works correctly at time 4. User `L` has not been observed before, and thus `L`, plus subsequent visitors `C` and `M`, are all classified as new visitors.

Overall, from the 61,405 user profiles we continuously crawled, we obtained a total of 8,034,664 total records of visits to user profiles in the PKU network. After integrating these raw results, we are left with 1,863,168 unique profile visit events. This high reduction (77%) is because most profiles receive few page views, and thus overlaps between successively crawled results are very high. Although Renren does not show individual recent visitors of user diaries and photos, it does display the total number of visits, which we crawled as well.

Impact of Crawl Frequency. We are concerned that our crawls might not be frequent enough to capture all visit events to a given profile. To address this concern, we took a closer look at the impact of crawler frequency on missing visits. First, we take all of the profiles we crawled for visit histories, and computed their average daily visit count between September and November 2009. We plot this as a CDF in Figure 3. Most users (99.3%) receive $<= 8$ visits per day on average. Since Renren shows 9 latest visitors, crawling a profile once every day should be sufficient to capture all visits. While our crawler adapts to allocate more crawl requests to popular, frequently visited profiles, we guarantee that every profile is crawled at least once every 24 hours.

Next, we select 1,000 random PKU users and crawl their recent visitors every 15 minutes for 2 days. We use the data collected to simulate five frequencies for crawling process, namely 15 minutes, 30 minutes, 1 hour, 12 hours and 1 day. Then we use the log-integration algorithm to concatenate the individual recent visitor lists at different crawling frequencies. For every person, we compute the number of visits missed by the crawler when we reduce the frequency, beginning with visits every 15 minutes. We plot CDF of these deviations in Figure 4. As we see, for 88% of users, there are no additional visits missed when we reduce the crawler rate from once every 15 minutes to once per day. Only for a very small group of users (0.7%) is the number of missing visits greater than 10 when crawling at once per day. Figure 3 also shows that less than 0.7% of all users receive more than 9 visits per day. Only these users would require more than one crawl per day to collect a full history of their visits. Since we allocate the bulk of our crawler requests to these high popularity users (and crawl once per day for

Network	Users Crawled	Links Crawled	Avg. Degree	Cluster Coef.	Assortativity	Avg. Path Len.
Renren	42,115K	1,657,273K	78.70	0.063	0.15	5.38
Facebook [29]	10,697K	408,265K	76.33	0.164	0.17	4.8
Cyworld [1]	12,048K	190,589K	31.64	0.16	-0.13	3.2
Orkut [22]	3,072K	223,534K	145.53	0.171	0.072	4.25
Twitter [13]	88K	829K	18.84	0.106	0.59	N/A

Table 1: Topology properties of social networks

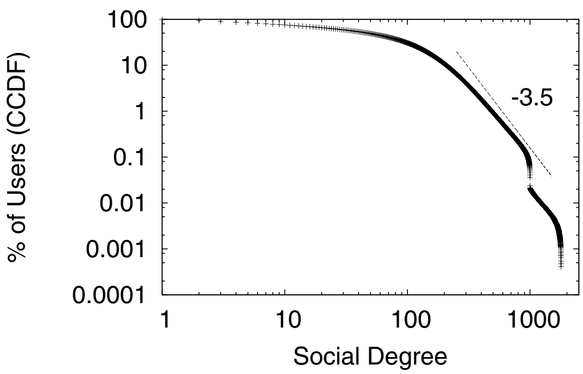

Figure 5: Node Degree Distribution in the Renren network

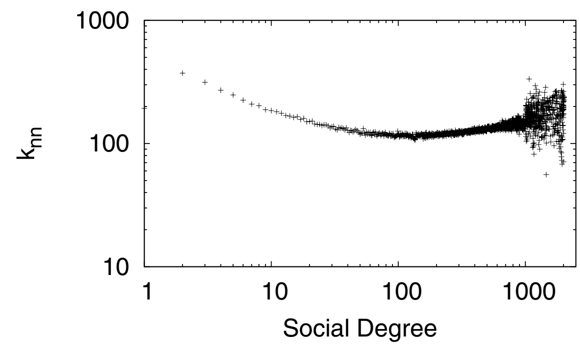

Figure 7: Renren k_{nn} Distribution

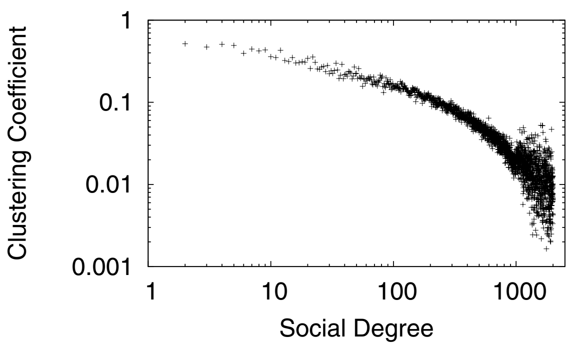

Figure 6: Renren Clustering Coefficient Distribution

the rest), we are relatively confident that very few visits are missed by our crawls.

2.4 Social Graph Analysis

In this section, we analyze the topological properties of the entire Renren social graph by focusing on salient graph measures. Table 1 shows some general properties of Renren, such as average degree, clustering coefficient, assortativity, and average path length as compared to other social networks. Our Renren dataset is larger than most previously studied OSN datasets, the exceptions being recent measurements of the Twitter network [4, 16]. However, our dataset shares similar properties with prior studies [7, 22, 29]. This confirms that Renren is a representative social network, and the behavior of its users is likely to be indicative of users in other OSNs like Facebook.

Figure 5 plots the complementary cumulative distribution function (CCDF) of user social degrees in the Renren network, and shows that Renren's network structure roughly follows a power-law [2]. We compute that the power-law curve fitting the social degree CCDF has an alpha value of 3.5, with fitting error of 0.02. This alpha value is significantly higher than that found for Facebook (1.5), Orkut (1.5), and Twitter (2.4).

We also look at Renren's clustering coefficient, which assesses the level of local connectivity between nodes. In undirected graphs, the clustering coefficient of a person is defined as the ratio of the number of links over all possible connections between one's friends. The clustering coefficient of the entire network is defined by the average of all individual clustering coefficients. Renren's average clustering coefficient is only 0.063, implying that Renren friend relationships are more loosely connected than the other social networks studied in Table 1. Figure 6 displays the distribution of clustering coefficient versus node degree. As expected for a social network, users with lower social degrees have higher clustering coefficients, thus demonstrating high levels of clustering at the edge of the social graph.

The assortativity coefficient measures the probability for users to establish links to other users of similar degree [29]. It is calculated as the Pearson correlation coefficient of the degrees of node pairs for all links in a graph. A positive assortativity coefficient indicates that users tend to connect to other users of similar degree, and a negative value indicates the opposite trend. Renren's assortativity is 0.15, implying that connections between like-degree users are numerous. Figure 7 displays node degree correlation (k_{nn}) versus node degree. k_{nn} is a closely related metric to assortativity. The positive correlation starting around degree 100 demonstrates that higher-degree users tend to establish links with other high-degree users. These chains of well-connected super-users form the backbone of the social network.

Average path length is the average of all-pairs-shortest-paths in the social network. It is simply not tractable to compute shortest path for all node pairs, given the immense size of our social graph. Instead, we choose 1000 random users in the network, perform Dijkstra to build a spanning tree for each user in the social graph, and compute the length of their shortest paths to all other users in the network. As shown in Table 1, the average path length in Renren is 5.38, which agrees with the six-degrees of separation hypothesis [20]. It is not surprising to see that average path length is

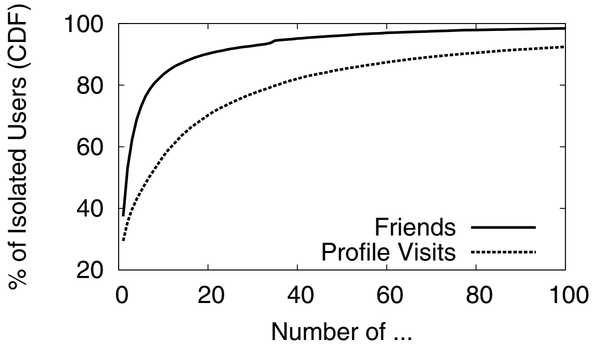

Figure 8: Social degree and total number of visits for isolated users

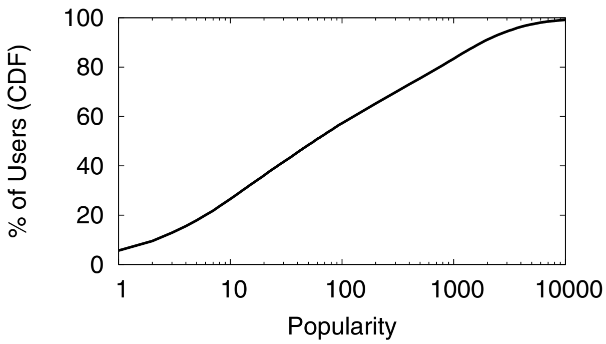

Figure 9: CDF of user profile popularity defined by visits

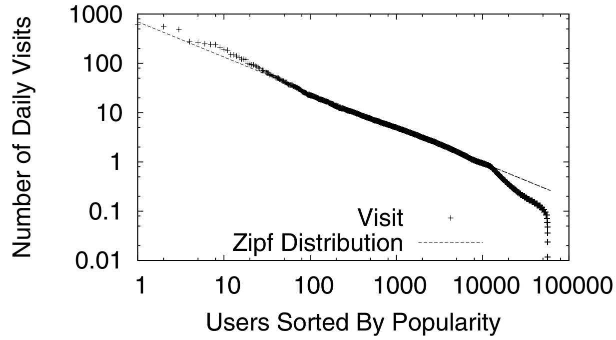

Figure 10: Average number of visits per day per user

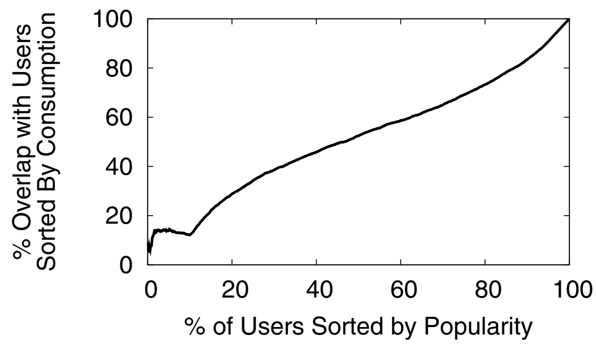

Figure 11: Overlap between users sorted by popularity vs. sorted by consumption

longer in Renren than prior results from Facebook [29], since those path lengths were computed over smaller subgraphs representing regional networks. It is surprising, however, when we compare it to results from Cyworld [7], since the two networks are similar in size and completeness. One reasonable explanation is that connectivity between Internet users in the Cyworld network is much stronger than connectivity between users in the Renren network.

Strongly Connected Component Analysis. User's online friendship links often correspond closely with their offline relationships [17]. Thus, it is natural to assume that college students would have many online friends in the same campus network. This behavior should manifest itself as a single, large, strongly connected component (SCC) that includes most users in the PKU network social graph. Surprising, we find that 23,430 (23.2%) of users in the PKU network have no friends in the PKU campus network, and are therefore disconnected from the SCC. We refer to these as *isolated users*. To confirm these results, we measured the SCC of 9 other large university networks and discovered similar numbers of isolated users.

Figure 8 shows social degrees and total number of profile visits for these isolated users. 83% of isolated users have social degrees less than 10. In addition, 70% of isolated users have less than 20 total profile visits, meaning their profiles are rarely browsed by others. These isolated users likely abandoned their accounts soon after registration, and did not have enough time to establish links with friends within the campus network. Further analysis into the behavior and activities of these isolated users is the subject of ongoing work.

3. PROPERTIES OF INTERACTION EVENTS

Our work focuses on the analysis of latent interaction events and the role they play in OSNs. In our measurement of the Renren OSN, we use histories of visits to user profiles to capture latent interactions. In this section, we take a closer look at latent interactions and compare them with visible interactions from a variety of perspectives.

3.1 Popularity and Consumption

We begin by analyzing the distribution of latent interactions across the Renren user-base. We define *popularity* as the number of views a user's profile receives; this is equivalent to the user's in-degree of latent interactions. Figure 9 shows the distribution of user popularity. As expected, popularity is not evenly spread across the population: only 518 people (1%) are popular enough to receive more than 10,000 views. Conversely, the majority of users (57%) exhibit very low popularity with less than 100 total profile views.

Figure 10 shows the average number of visits users receive on a daily basis. The distribution is fitted to a Zipf distribution of the form $\beta x^{-\alpha}$ where $\alpha = 0.71569687$ and $\beta = 697.4468225$. Popular users receive many more views per day: 141 users (0.2%) are viewed more than 20 times a day on average, with the most popular profile being viewed more than 600 times a day. Most users (85.5%) receive less than one visit per day on average. This reinforces our finding that latent interactions are highly skewed towards a very popular subset of the population.

Finally, we examine whether the popularity of users corresponds to their profile viewing behavior. We define *consumption* as the number of other profiles a user views; this is equivalent to the

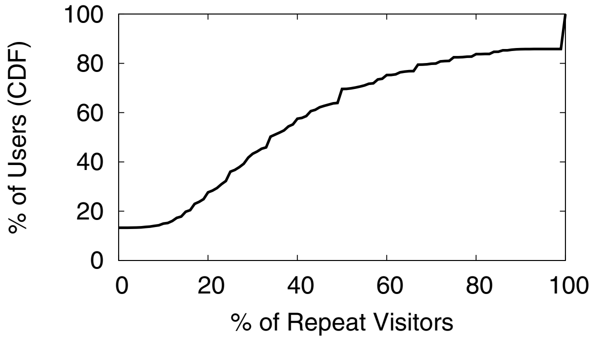

Figure 12: Ratio of repeat visitors

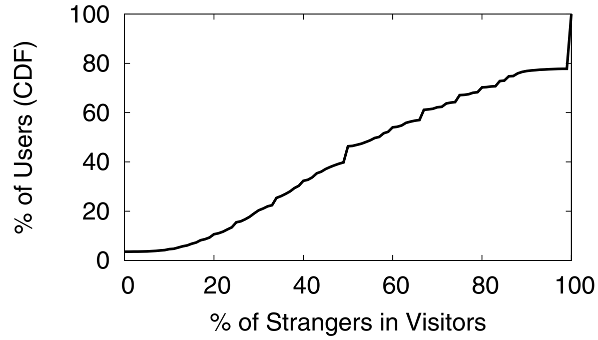

Figure 14: Percentage of strangers in visitors

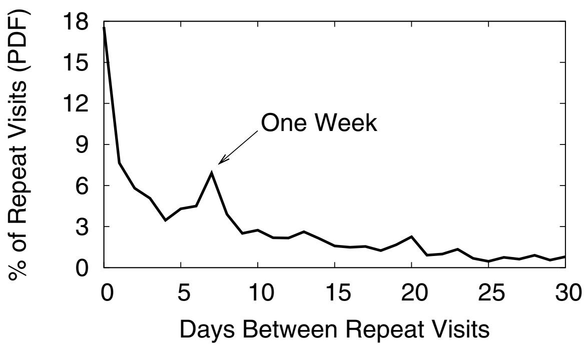

Figure 13: PDF of interval time between repeat visits

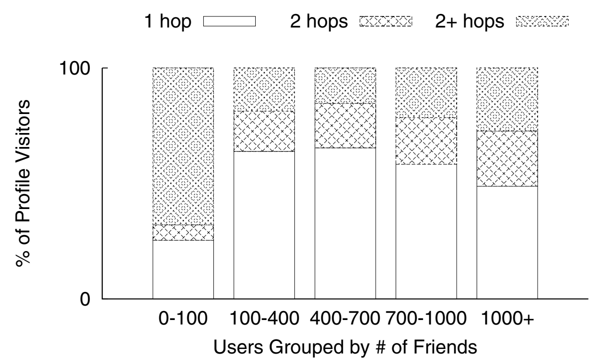

Figure 15: Breakdown of visitors by owner's social degree

user's out-degree of latent interactions. Figure 11 plots the overlap between the top users sorted by popularity, and top users sorted by consumption. The graph shows that the top 1% most popular users have 9% overlap with the top 1% biggest consumers. These users represent a hard-core contingent of social network users who are extremely active. For the most part however, users with high numbers of incoming latent interactions do not overlap with the people generating those interactions, *e.g.* profiles of celebrities are viewed by many users, but they are inactive in viewing others' pages. This necessarily means that many (presumably average, low-degree) users actively visit others, but are not visited in return. We examine the reciprocity of latent interactions in more detail in Section 3.3.

3.2 Composition of Visitors

Next, we want to figure out the composition of visitors to user profiles. We pose two questions: first, what portion of profile visitors are repeat visitors? Second, are visitors mostly friends of the profile owner, or are they unrelated strangers?

We begin by addressing the first question. We calculate the percentage of repeated visitors for each profile, and report the distribution in Figure 12. Roughly 70% of users have less than 50% repeat visitors, meaning that the majority of visitors do not browse the same profile twice. This seems to indicate the long tail of latent interactions is generated by users randomly browsing the social graph.

Next, we take a closer look at repeat profile visits. Figure 13 shows the probability density function (PDF) of the interval time between repeat visits. The graph peaks on day 0, meaning that users are most likely to return to a viewed profile on the same day. We will examine the causes for this behavior more closely in Section 4. The probability for repeated views decreases as the time delta expands, except for a noticeable peak at day 7. Interestingly, this shows that many users periodically check on their friends on a weekly basis. We confirmed that this feature is not an artifact introduced by our crawler or the use of RSS feeds by Renren users. Instead, we believe it may be due to the tendency for many users to browse their friends' profiles over the weekend.

We now move on to our second question: what users are generating latent interactions, friends of the profile owner, or *strangers*. We define a stranger as any user who is not a direct friend of the target user. Like Facebook, Renren's default privacy settings allow users in the same campus network to browse each other's profiles. Renren automatically recommends popular profiles to other users in the same network, which motivates people to view non-friends' profiles. However, they only recommend the 100 users in the network with the most users. Thus this feature should have minimal impact on our analysis of visitors to the average Renren user.

In order to answer our question, we calculate the percentage of visitors that are strangers and display the results in Figure 14. The results are fairly evenly divided: roughly 45% of users receive less than 50% of their profile visits from strangers. Or conversely, a slight majority of the population does receive a majority of their profile views from strangers.

We want to take a closer look at what component of a profile's visitors are strangers, and how far are they from the profile owner in the social graph. In Figure 15, we group the owners of profiles together by their social degree, and compute the average breakdown of their visitors into users who are friends (1-hop), friends-of-friends (2-hop) and other visitors (2+ hops). We see that for users with relatively few (< 100) friends, the large majority of their visitors are complete strangers, with very few friends-of-friends visit-

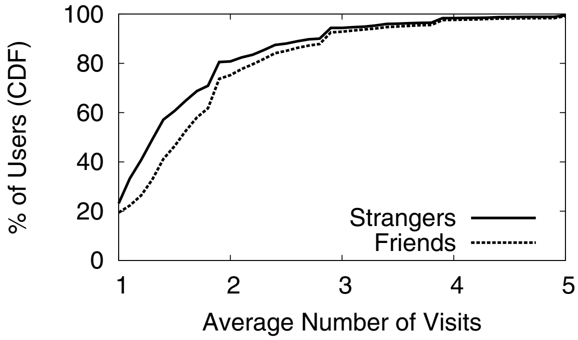

Figure 16: Average number of visits for friends and strangers

Figure 17: Ratio of reciprocated latent interactions over total latent relationships

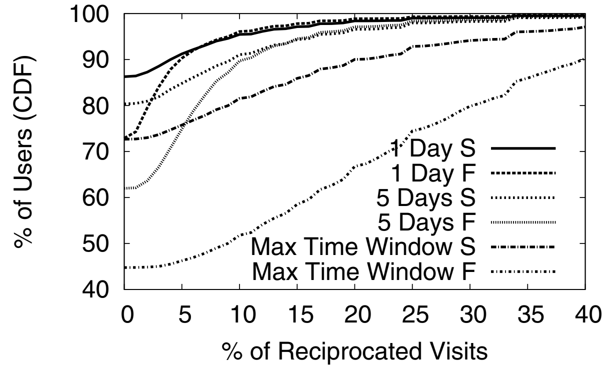

Figure 18: Probability of reciprocated profile views over various time windows for both strangers (S) and friends (F)

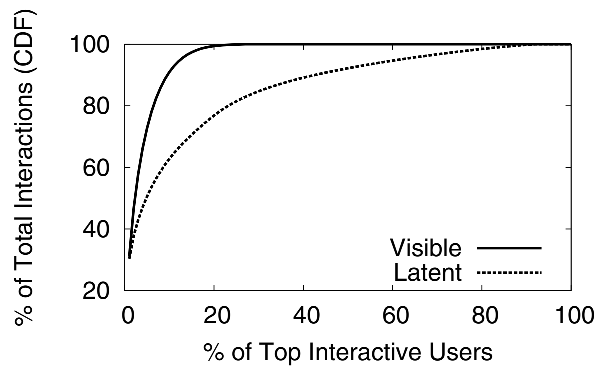

Figure 19: Distribution of total interactions

ing. For well-connected users with 100–1000 friends, the majority of their visitors are direct friends, and also a significant number of friends-of-friends. Finally, for extremely popular users with more than 1000 friends, their notoriety is such that they start to attract more strangers to visit their profiles. These results confirm those from previous work that discovered many Orkut users browse profiles 2 or more hops away on the social graph [3].

Unlike friends, strangers do not build long-term relationships with profile owners. Intuitively, this would seem to indicate that repeat profile viewing behavior should favor friends over strangers. To investigate this we compute the average number of visits for strangers and friends for each profile and plot the distribution in Figure 16. Surprisingly, our results indicate that the repeat profile viewing behavior for friends and strangers is very similar, with friends only edging out strangers by a small margin. This result demonstrates that when considering information dissemination via latent interactions, the significance of non-friend strangers should not be overlooked.

3.3 Reciprocity

Social norms compel users to reply to one another when contacted via visible interactions. Prior work has shown that these interactions are largely reciprocal on OSNs [29]. However, is this true of latent interactions? Since Renren users have full access to the list of recent visitors to their profile, it is possible for people to pay return visits to browse the profiles of their visitors. The question is, does visiting other user profiles actually trigger reciprocal visits?

As the first step towards looking at reciprocity of latent interactions, we construct the set of visitors who view each user profile, and the set of people who are visited by each user. Then, we compute the intersection and union of these two sets for every user. Intuitively, intersections include people who view a given user profile and are also visited by that user, *i.e.* the latent interactions are reciprocated. Unions contain all latent relationships for a given user, *i.e.* all users who viewed them, or they viewed. We calculate the ratio of intersection size to union size for each user and then plot the results in Figure 17. The ratio represents the number of reciprocated latent interactions divided by the total number of latent relationships. For more than 93% of users, less than 10% of latent relationships are reciprocated. This demonstrates that incoming profile views have little influence on user's profile browsing behavior. This is surprising, especially considering the fact that users know that their visits to a profile are visible to its owner through the visitor history feature.

Next we examine the time-varying characteristics of reciprocal profile visits for both strangers and friends. We compute the number of reciprocal visits that take place within t days after the initial visit. Figure 18 shows the results for threshold t values of 1 and 5 days plus the entire 90 days. As we look at increasingly larger window sizes, we see more profile visits being reciprocated. However, reciprocity remains low overall. Even across the entire measurement period, 73% of users receive no reciprocal page views from strangers, and 45% of users obtain no reciprocal page views from friends. This demonstrates that even with Renren's visitor history feature, visiting other user profiles is not sufficient to generate reciprocal visits. Compared to strangers, friends have relatively higher probability of reciprocal visits.

We take a further step and quantify the lack of reciprocity for latent interactions. For a data set of n users, if user i visits user j, then

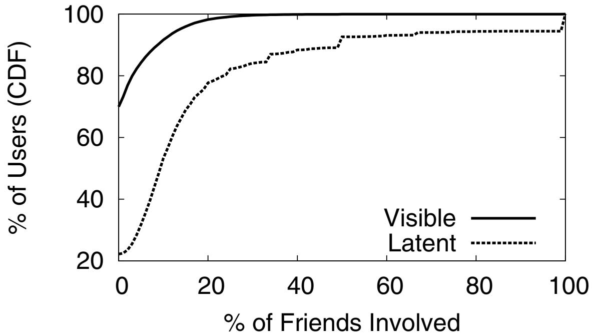

Figure 20: Distribution of interactions among each user's friends

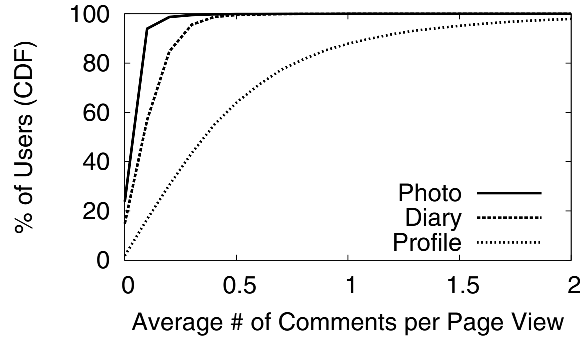

Figure 21: Average number of comments per page view for different types of pages

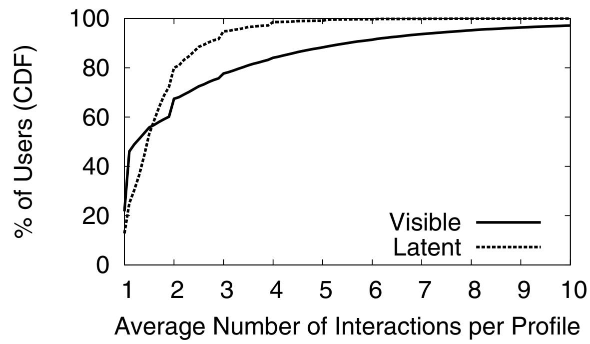

Figure 22: Average number of interactions per profile

$v_{ij} = 1$, otherwise $v_{ij} = 0$. The reciprocity coefficient [7] is defined as $\frac{\sum_{i \neq j}(v_{ij}-\bar{v})(v_{ji}-\bar{v})}{\sum_{i \neq j}(v_{ij}-\bar{v})^2}$, where $\bar{v} = \frac{\sum_{i \neq j} v_{ij}}{n(n-1)}$. The reciprocity coefficient is measured between -1 and 1, where positive values indicate reciprocity, and negative values anti-reciprocity. The reciprocity coefficient of profile visits on Renren is only 0.23. In contrast, reciprocity of visible comments on Renren is 0.49, and the reciprocity of visible interactions on Cyworld [7] is 0.78. Compared to these visible interactions, latent interactions show much less reciprocity.

3.4 Latent vs. Visible Interactions

In this section, we compare the characteristics of latent and visible interactions. To understand the level of participation of different users in both latent and visible interactions, Figure 19 plots the contribution of different users to both kinds of interactions. The bulk of all visible interactions can be attributed to a very small, highly interactive portion of the user-base: the top 28% of users account for all such interactions. In contrast, latent interactions are quite prevalent across the entire population, with more than 93% of all users contributing to latent interaction events. This confirms our original hypothesis that users are more active in viewing profiles than leaving comments, potentially because of a sense of anonymity in profile browsing. Given its widespread nature, this result also underscores the importance of understanding latent interactions as a way of propagating information across OSNs.

Next, we compare latent and visible interactions in coverage of friends. We compute for each user a distribution of their latent and visible interactions across their social links. We then aggregate across all users the percentage of friends involved in these events and plot the results in Figure 20. We see that roughly 80% of users only interact visibly with 5% of their friends, and no users interact with more than 40% of their friends. In contrast, about 80% of users view 20% or more of their friends' profiles, and a small portion of the population views all of their friends' profiles regularly. Thus, although not all social links are equally active, latent interactions cover a wider range of friends than visible interactions.

To get a sense of how many visible comments are generated by latent interactions, we examine the average number of comments per page view for a variety of pages on Renren, including profiles, diary entries, and photos. Figure 21 plots the results. Recall that along with visible comments, Renren keeps a visitor counter for each photo and diary entry. For diary entries and photos, the conversion rate is very low: 99% of users have less than 0.2 comments for every photo view, and 85% people have less than 0.2 comments for every diary view. This indicates that most users are passive information consumers: they view/read content and then move on without commenting themselves. In contrast, profile views have a higher conversion rate. Interestingly, 13% of users have a view/comment ratio greater than 1. This is because these users use profile comments as a form of instant messaging chat, leaving multiple responses and replies upon each visit.

Finally, we analyze the repeat activity frequency for latent and visible interactions on Renren. In particular, we want to examine the likelihood that users will repeatedly interact with the same page once they have viewed or commented on it once. Figure 22 plots the average number of interactions each user has with profile pages. 80% of users view a given profile <2 times. However, 80% of users leave 3.4 comments, almost twice the number of latent interactions. This result makes sense intuitively: for most types of data users only need to view them once to consume the data. However, comments can stimulate flurries of dialog on a given page, resulting in many consecutive interactions.

4. FACTORS INFLUENCING LATENT INTERACTIONS

As shown in Section 3.1, not all users in Renren are the target of equal numbers of latent interactions. In this section, we analyze factors that may impact latent interactions in order to quantify their effect on the popularity of individual users. We examine the following factors:

Friends. Does social degree directly correlate with popularity?

Lifetime. Are long lived accounts more likely to be popular than newer, less active accounts? We measure the number of days

Popularity	Friends	Lifetime	Diary	Photo	Status	Share	Comment
0-100	16 (0.15)	35 (0.55)	1 (0.51)	3 (0.47)	1 (0.5)	1 (0.5)	2 (0.54)
100-1000	131 (0.56)	423 (0.41)	11 (0.33)	41 (0.24)	27 (0.36)	43 (0.41)	96 (0.45)
1000-10000	401 (0.43)	792 (0.24)	50 (0.18)	125 (0.1)	115 (0.18)	155 (0.23)	596 (0.28)
>10000	708 (0.02)	869 (0.02)	117 (0.02)	251 (-0.03)	236 (0.01)	273 (-0.05)	1581 (0.01)
all users	112 (0.73)	263 (0.75)	12 (0.70)	34 (0.61)	28 (0.69)	39 (0.72)	134 (0.76)

Table 2: Average value of factors affecting user popularity. Spearman's rank correlation coefficient for each value is shown in parentheses.

in between a user joining and leaving Renren. Neither of these pieces of information is provided by Renren, and thus must be estimated. Join date can be approximated by the timestamp of the first comment received by a user, since the comment is likely to be a "welcome message" from a friend greeting the new user [29]. Because abandoned and inactive accounts can still receive comments, the best estimate of departure time is the timestamp of the last comment left by a user.

User-generated content. Do users who frequently update their profiles and upload new content attract the most visitors? This includes user's status updates, diary entries, photos, and shared links to content on the web.

Comments. Do users who comment frequently attract more users? Comments are snippets of text that user's can attach to other pieces of user-generated content.

We divide users into 4 groups based on their popularity, calculate the average value of these factors in each group, and display the results in Table 2. All factors increase along with popularity, *i.e.* the most popular users also have the most friends, the oldest accounts, and generate the largest amounts of content/visible interactions.

Given the drastic differences in size of each popularity group, and the average nature of the values in Table 2, it is difficult to infer definite correlations between any one factor and popularity. To analyze these correlations more specifically, we leverage a technique from prior work [4] called Spearman's rank correlation coefficient (Spearman's ρ). Spearman's ρ is a non-parametric measure of the correlation between two variables that is closely related to Pearson's correlation coefficient [18]. It is defined as $\rho = 1 - \frac{6\sum(x_i - y_i)^2}{n(n^2-1)}$, where x_i and y_i are the ranks of two different features in a data set of n users. $\rho > 0$ indicates positive correlation, while <0 indicates negative correlation.

Table 2 shows Spearman's ρ in parenthesis beside the average value for each factor. Although all factors exhibit high correlation with the low popularity and "all users" categories, this is an artifact of the tied ranks among the (numerous) low activity users. All of these users exhibit very low interactivity and social degree, thus leading to high levels of correlation. Previous work has observed similar artifacts when analyzing all users in a large OSN dataset [4].

For the two median popularity groups (100-1000 and 1000-10000), number of friends has the highest correlation with popularity. Users in these categories can be broadly defined as normal social network users. They are not celebrities; they simply use the OSN for its intended purpose of sharing information with friends. This is reflected in the fact that users in these categories show relatively high levels of correlation across all user-generated content categories. Account lifetime is a less important factor for users in the 1000-10000 popularity range, given the ease with which users can quickly amass hundreds of friends on OSNs.

No factor has strong correlation with popularity for users in the high popularity group. Spearman's ρ for photos and shared links are even negative. This is an important finding, as it shows popularity is not trivially gained simply by having lots of friends, or producing copious amounts of user-generated content. Therefore, there must be other factors outside the scope of our measurements that contribute to determine user popularity. One possibility is that quality, rather than quantity, of content may be a significant draw to attract visitors. Another possibility is that real-world celebrity status is the most important determining factor of online popularity. Unfortunately, we cannot quantify these factors at present.

Recall that 100 of the most popular users in the university network are recommended to users by Renren. These 100 users account for less than 19.3% of the total users in the high popularity group, so the recommendation mechanism has limited impact on the high popularity group results.

5. LATENT INTERACTION GRAPHS

Previous studies have demonstrated that taking visible interactions into account has important implications for applications that leverage social graphs [29]. These changes can be modeled by *interaction graphs*, which are constructed by connecting users from the social graph who have visibly interacted one or more times.

We have already demonstrated significant differences between latent and visible interaction patterns on Renren. To summarize these key differences briefly, latent interactions are more numerous, non-reciprocal, and often connect non-friend strangers. These results are also likely to have profound implications on applications that leverage social graphs, and thus warrant the construction of a new model to capture the properties of latent interactions. We call this new model a *latent interaction graph*. In this section we formally define latent interaction graphs, analyze their salient properties, and compare them to the Renren social and visible interaction graphs.

5.1 Building Latent Interaction Graphs

A latent interaction graph is defined as a set of users (nodes) that are connected via edges representing latent interaction events between them. Unlike the social graph and visible interaction graph, we have shown that latent interaction is non-reciprocal (Section 3.3). Thus, we use directed edges to represent user's page views, unlike the social and visible interaction graphs, which are both undirected. The set of users (61,405 total) remains unchanged between the social and interaction graphs. We define *latent interaction in-degree* of a node as the number of visitors who have visited that user's profile; while *out-degree* is the number of profiles that user has visited.

We construct latent interaction graphs from our Renren data using profile views as the latent interactions. We use user comments as the visible interaction data to construct visible interaction graphs for Renren. In this paper we restrict our social, latent, and visible interaction graphs to only contain users from the PKU network, since these are the only users for which we have complete interaction records. Note that we only consider interactions that occur between users in the PKU network, as it is possible for interactions to originate from or target users outside the network for whom

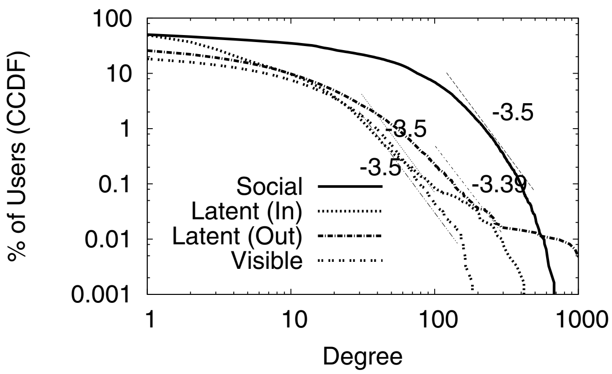

Figure 23: CCDF of node degree for latent interaction graph, visible interaction graph and social graph.

we have limited information. Also note that because non-friend strangers can view user's profiles, the latent interaction graph will contain edges between users who are not friends in the social graph.

Our formulation of interaction graphs uses an unweighted graph. We do not attempt to derive a weight scheme for interaction graphs analyzed in this paper, but leave exploration of this facet of latent interaction graphs to future work.

5.2 Comparing Social, Visible Interaction and Latent Interaction Graphs

In this section we compare the salient characteristics of the Renren social, visible, and latent interaction graphs using common graph metrics.

Degree distribution. Figure 23 plots the CCDFs of node degree for the three types of graphs. Since the latent interaction graph is directed, we plot both in-degree and out-degree. In Section 3.4 we show that latent interactions are more prevalent than visible interactions. This is reflected in the relative number of edges in the two interaction graphs, as shown in Table 3. This also leads to nodes in the latent graph having a noticeably higher degree distribution in Figure 23. However, neither of the interaction graphs have as many edges as the raw social graph, which leads to the social graph having the highest degree distribution. Interestingly, because a small number of Renren users are frequent profile browsers, *i.e.* they like to visit a large number of profiles (far greater than their circle of friends), the distribution of latent out-degrees flattens out at the tail-end and never approaches 0%.

Despite these differences, latent interaction graphs still exhibit power-law scaling. Table 3 shows similar in- and out-degree power-law fit alpha values for latent interaction graphs compared to the visible interaction graph and the social graph. Renren limits users to of 1,000 friends (2,000 for paying users), so all power-law curves decline quickly to zero, and alpha values are high.

Clustering coefficient. Table 3 shows that the average clustering coefficient is 0.03 for the latent interaction graph, and 0.05 for the visible interaction graph, which are both much less than that of the social graph. This is because not all social links are accurate indicators of active social relationships, and these links with no interactions are removed in interaction graphs. This produces loose connections between neighbors, and low clustering coefficients in these graphs. A portion of the latent interactions to a profile is from non-friend strangers who randomly browse the network. Thus, links between visitors in the latent interaction graph are less intensive than friends exchanging messages, which further lowers the clustering coefficients in latent interaction graphs.

Assortativity. Table 3 shows that the Renren latent interaction graph is slightly disassortative. This makes sense intuitively, as latent interactions are highly skewed towards a small subset of extremely popular users. In contrast, the other two graphs are both assortative, with the social graph being more so. This result contrasts with previous studies in which the interaction graph was more assortative than the social graph [29].

Average path length. The average path length of the latent interaction graph is between that of the visible interaction graph and the social graph. As the average number of links per node and the number of high-degree "super-nodes" decreases, the overall level of connectivity in the graph drops. This causes average path lengths to rise, especially in the visible interaction graph.

6. IMPACT ON SOCIAL APPLICATIONS

In the previous section, we analyzed salient properties of the Renren social, visible interaction, and latent interaction graphs. However, it remains unknown how different graph metrics impact the performance of real-world social applications. Social application tests present new perspectives since their performance on a particular graph cannot be easily correlated with a single graph metric.

Changes in user connectivity patterns can produce significantly different results for social network applications. In order to validate how much impact the choice of graphs can make on socially enhanced applications, we implement techniques from the paper "Efficient influence maximization" [6]. This work leverages graph properties in social networks to address information dissemination problems. We compare the effectiveness of these algorithms on the raw social graph, visible interaction graph, and latent interaction graph that are derived from our Renren measurements of the PKU network. These results allow us to validate whether socially enhanced applications require a model that takes latent interactions into account.

6.1 Efficient Influence Maximization

As OSNs become increasingly popular worldwide, they also become more critical platforms for information dissemination and marketing. Understanding how to fully utilize OSNs as marketing and information dissemination platforms is a significant challenge. The influence maximization problem seeks to determine the most influential individuals who will maximum the spread of information in an OSN.

Given the lack of publicly available social influence datasets, previous work [6, 14] builds statistical models based on raw social graph topologies, and designs algorithms to address influence maximization problems within these models.

Weighted cascade model. The weighted cascade model relies on social links for information propagation [14]. After receiving information, a user has a single chance of activating each currently inactive neighbor. In this model, the activation probability is related to node's degree: if a person w has d_w neighbors, it is activated by neighbors with probability $1/d_w$. Chen et al. propose instantiating the MixedGreedyWC algorithm to implement the weighted cascade model [6].

Latent cascade model. The weighted cascade model is based on social graphs, and relies on social relationships plus the activation probability to simulate the information dissemination process.

Network	Edges	Power-Law Fit Alpha	Cluster Coef.	Assortativity	Avg. Path Len.
Social Graph	753,297	3.5	0.18	0.23	3.64
Visible Interaction Graph	27,347	3.5	0.05	0.05	5.43
Latent Interaction Graph	240,408	3.5(in) 3.39(out)	0.03	-0.06	4.02

Table 3: Topology measurements for latent interaction, visible interaction and social graphs.

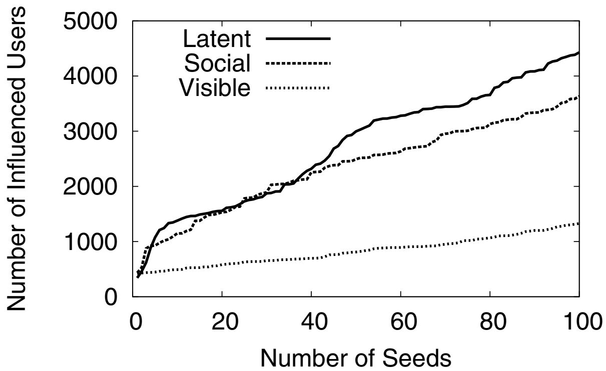

Figure 24: Influence spread using the MixedGreedyWC algorithm on different graphs.

However, we observe that information is not disseminated equally through social links. User profiles are usually only viewed by a small portion of friends, thus not all social links are active in information propagation. Moreover, a portion of the visits to profiles can be attributed to people not connected to the user, *i.e.* strangers.

When a user browses a profile, information in the web page is naturally propagated to that user. It is more accurate to represent information dissemination by links in latent interaction graphs than by passive friendship links in the social graph. Thus, to take latent interaction graphs into account we define the latent cascade model. When a person receives information, they have a single chance to activate inactive neighbors, who have directed links pointing to that person in the latent interaction graph. To transform the visitor counts from our month-long measurement into activation probabilities, we calculate the number of unique visitors for every user and determine the most popular user. We then compute the activation probability of every user by normalizing that user's number of unique visitors by the maximum value. This approximates an activation probability using the "visit rate" normalized to the most popular profile. We use visits from unique users and exclude repeated visits here, because the first visit contributes the most to the information propagation, and repeated follow-up visits are less likely to propagate information.

Visible cascade model. The visible cascade model is built atop visible interaction graphs, in a similar fashion to the latent cascade model. The information dissemination path is decided by visible interaction events, and the per node activation probability is determined using the same method as in the latent cascade model.

6.2 Experimental Results

For our experiments, we use the MixedGreedyWC algorithm [6] to find the most influential individuals in each of our three models, and then compute the number of people influenced. We vary the set of seed users to the MixedGreedyWC algorithm from 1 to 100 in our tests and observe the effects on influence spread.

Figure 24 shows influence spread versus seed set size for our three models. Influence in the weighted and latent cascade model both increase quickly in the beginning. Eventually, as more seeds are selected, the number of people influenced in latent cascade model surpasses the weighted model. The reason for this is that although social graphs have a large amount of links, the activation probability is generally low. This limits the spread of information to well connected components of the graph. Conversely, latent interaction graphs are constructed by page viewings, which correspond to higher activation probability. This is in accordance with real situations: people are likely to receive information when they browse profiles, while information cannot be disseminated through inactive social links. These high probabilities guarantee that information is progressively disseminated through the limited number of links in the latent interaction graph. Previous sections show that visible interaction graphs have the least number of links. Thus, although activation probabilities are relatively high, the visible cascade model reaches the smallest influence spread due to the dearth of links.

7. RELATED WORK

Much effort has been put into understanding the structure of large-scale online social networks [8]. Ahn et al. analyze topological characteristics of Cyworld, MySpace and Orkut [1]. Mislove et al. measure the structure of Flickr, YouTube, LiveJournal, and Orkut [22], and observe the growth of the Flickr social network [21]. Java et al. study the topological and geographical properties of Twitter [13]. Huang et al. measure user prestige and visible interaction preference in Renren [12]. To the best of our knowledge, our measurement of the Renren network provides the largest non-Twitter social graph to date, with 42,115,509 users and 1,657,273,875 friendship links. Most of Renren's topological properties are similar to those of other OSNs, including power-law degree distribution and small world properties.

Researchers have also studied the visible interaction network. Leskovec et al. analyze the instant messaging network, which contains the largest amount of user conversations ever published [19]. Valafar et al. characterize indirect fan-owner interactions via photos among users in Flickr [27]. Chun et al. observe that visible interactions are almost bidirectional in Cyworld [7]. Wilson et al. show the structure of the interaction graph differs significantly from the social network in Facebook [29]. Viswanath et al. observe that social links in the activity network tend to come and go quickly over time [28]. Finally, a recent study from Northwestern and UC Santa Barbara quantified the role of spam and phishing attacks in Facebook wall posts [10].

Benevenuto et al. collect detailed click-stream data from a Brazilian social network aggregator, and measure silent activities like browsing [3]. Schneider et al. extract click-streams from passively monitored network traffic and make similar measurements [25]. We analyze latent interactions from a different perspective than these existing works by leveraging data that is intrinsic to the OSN and not inferred from a third-party. Ideally, we would like to perform a detailed comparison between our dataset and prior studies using click-stream datasets. Unfortunately, the sensitive nature of these datasets make their distribution challenging. At publication time, we are currently unaware of any publicly available click-stream dataset.

Some researchers have performed initial studies on information propagation and user influence in OSNs. Cha et al. present a detailed analysis of popularity and dissemination of photographs on Flickr [5]. They find that popular users with high in-degree are not necessarily influential in terms of spawning subsequent, viral interactions in the form of retweets or mentions on Twitter [4]. Our Renren data confirms these results, as we show that factors like number of friends and amount of user-generated content produced are not strongly correlated with popularity.

8. CONCLUSIONS

Latent user interactions make up the large majority of user activity events on OSNs. In this paper, we present a comprehensive study of both visible and latent user interactions in the Renren OSN. Our data includes detailed visit histories to the profiles of 61,405 Renren users over a 90-day period (September to November 2009). We compute a single visitor history for each profile by using a novel technique to merge visitor logs from multiple consecutive crawls. We analyze profile visit histories to study questions of user popularity and reciprocity for profile browsing behavior, and the link between passive profile browsing and active comments.

Our analysis reveals interesting insights into the nature of user popularity in OSNs. We observe that user behavior changes for latent interactions: more users participate, users do not feel the need to reciprocate visits, and visits by non-friends make up a significant portion of views to most user profiles. We also see that visits to user profiles generate more active interactions (comments) than visits to photos or diary pages. Using profile browsing events, we construct *latent interaction graphs* as a more accurate representation of meaningful peer interactions. Analysis of latent interaction graphs derived from our Renren data reveal characteristics that fall between visible interaction graphs and social graphs. This confirms the intuition that latent interactions are less limited by constraints such as time and energy, but more meaningful (and thus sparser) than the social graph.

Finally, our measurement study also includes an exhaustive crawl of the largest connected component in the Renren social graph. The resulting graph is one of the biggest of its kind, with more than 42 million nodes and 1.6 billion edges. Other than the proprietary Cyworld dataset, this is the only social graph we know of that covers 100% of a large social graph component. Given its size and comprehensiveness, we are currently investigating different options for sharing this dataset with the research community.

Acknowledgments

We thank the anonymous reviewers for their valuable comments. This work is supported in part by National Science Foundation of China under project #60673183, #60873051, the National Basic Research Program of China under Grant No.2011CB302305, and Doctoral Funding of MOE #20060001044. The UCSB team is supported by the NSF under CNS-0546216, IIS-847925, and IIS-0916307. 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 National Science Foundation.

9. REFERENCES

[1] AHN, Y.-Y., HAN, S., KWAK, H., MOON, S. B., AND JEONG, H. Analysis of topological characteristics of huge online social networking services. In *Proc. of World Wide Web Conference* (2007).

[2] BARABASI, A.-L., AND ALBERT, R. Emergence of scaling in random networks. *Science 286* (1999).

[3] BENEVENUTO, F., RODRIGUES, T., CHA, M., AND ALMEIDA, V. Characterizing user behavior in online social networks. In *Proc. of ACM Internet Measurement Conference* (2009).

[4] CHA, M., HADDADI, H., BENEVENUTO, F., AND GUMMADI, K. Measuring user influence in twitter: The million follower fallacy. In *Proc. of International AAAI Conference on Weblogs and Social Media (ICWSM)* (May 2010).

[5] CHA, M., MISLOVE, A., AND GUMMADI, K. A measurement-driven analysis of information propagation in the flickr social network. In *Proc. of World Wide Web Conference* (2009).

[6] CHEN, W., WANG, Y., AND YANG, S. Efficient influence maximization in social networks. In *Proc. of ACM Conference on Knowledge Discovery and Data Mining* (June 2009).

[7] CHUN, H., KWAK, H., EOM, Y. H., AHN, Y.-Y., MOON, S. B., AND JEONG, H. Comparison of online social relations in volume vs interaction: a case study of cyworld. In *Proc. of ACM Internet Measurement Conference* (2008).

[8] FU, F., LIU, L., AND WANG, L. Empirical analysis of online social networks in the age of web 2.0. *Physica A 387*, 2–3 (2008), 675–684.

[9] GANNES, L. When social replaces search, what can you do to monetize? GigaOM, March 2010.

[10] GAO, H., HU, J., WILSON, C., LI, Z., CHEN, Y., AND ZHAO, B. Y. Detecting and characterizing social spam campaigns. In *Proc. of ACM Internet Measurement Conference* (Melbourne, Australia, November 2010).

[11] GRUHL, D., GUHA, R., LIBEN-NOWELL, D., AND TOMKINS, A. Information diffusion through blogspace. In *Proc. of World Wide Web Conference* (2004).

[12] HUANG, L., AND XIA, Z. Measuring user prestige and interaction preference on social network site. In *Proc. of ACIS-ICIS* (2009).

[13] JAVA, A., SONG, X., FININ, T., AND TSENG, B. L. Why we twitter: An analysis of a microblogging community. In *Proc. of WebKDD/SNA-KDD* (2007).

[14] KEMPE, D., KLEINBERG, J. M., AND TARDOS, Ï. Maximizing the spread of influence through a social network. In *Proc. of ACM Conference on Knowledge Discovery and Data Mining* (2003).

[15] KIRKPATRICK, M. Social networking now more popular than email, report finds. ReadWriteWeb, March 2009.

[16] KWAK, H., LEE, C., PARK, H., AND MOON, S. B. What is twitter, a social network or a news media? In *Proc. of World Wide Web Conference* (Raleigh, NC, April 2010).

[17] LAMPE, C., ELLISON, N., AND STEINFIELD, C. A familiar face(book): profile elements as signals in an online social network. In *Proc. of CHI* (2007).

[18] LEHMANN, E. L., AND D'ABRERA, H. J. M. *Nonparametrics: Statistical Methods Based on Ranks*. Prentice-Hall, 1998.

[19] LESKOVEC, J., AND HORVITZ, E. Planetary-scale views on a large instant-messaging network. In *Proc. of World Wide Web Conference* (2008).

[20] MILGRAM, S. The small world problem. *Psychology Today 1* (1967).

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

[22] MISLOVE, A., MARCON, M., GUMMADI, P. K., DRUSCHEL, P., AND BHATTACHARJEE, B. Measurement and analysis of online social networks. In *Proc. of ACM Internet Measurement Conference* (2007).

[23] Opengraph. `http://opengraphprotocol.org`.

[24] The renren social network. `http://www.renren.com`.

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

[26] SORKIN, A. R. Facebook tops 500 million users. New York Times Blog, June 2010.

[27] VALAFAR, M., REJAIE, R., AND WILLINGER, W. Beyond friendship graphs: A study of user interactions in flickr. In *Proc. of WOSN* (2009).

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

[29] 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 ACM Eurosys* (2009).

[30] YAROW, J. Facebook was more popular in the u.s. than google last week. BusinessInsider.com, March 2010.