Invisible Interactions: What Latent Social Interaction Can Tell Us about Social Relationships in Social Networking Sites

Miriam J. Metzger[§]

Christo Wilson[†]

Rebekah A. Pure[§]

Ben Y. Zhao[†]

[§]Department of Communication

[†]Department of Computer Science

University of California, Santa Barbara, USA

Invisible Interactions: What Latent Social Interaction Can Tell Us about Social Relationships in SNSs

Abstract

A deep understanding of social interaction in social networking sites (SNSs) can provide important insights into questions of human social and relational behavior, as well as shape the design of new social platforms and applications. Recent studies have shown that a majority of user interactions on SNSs are latent interactions—passive actions such as profile browsing that cannot be observed directly by traditional research methods. This paper presents a new technique to capture natural latent social interaction in SNSs, which offers a better understanding of both visible (e.g., comments and wall posts) and latent (e.g., passive profile browsing) user interactions in SNSs than has been possible before. Our data, collected from over 42 million users of Renren, the most popular SNS in China, shed new light on human social and relational behavior in SNS contexts, in some cases supporting prior work on phenomena such as lurking, interpersonal electronic surveillance, and social capital, and in other cases challenging past research findings.

Introduction

Popular social networking sites (SNSs) like Facebook and Twitter are changing the way people communicate and interact with each other, affecting how users manage both their own identity and their relationships with others. Today's social networks already count close to one billion members worldwide. Facebook, the most popular SNS, has more than 500 million active users (Sorkin, 2010), and has surpassed Google as the most visited site on the Internet (Yarow, 2010). Increasingly, Facebook and Twitter are replacing email and search engines as users' primary interfaces to the Internet (Gannes, 2010; Kirkpatrick, 2009). 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 (http://opengraphprotocol.org).

A deep understanding of user interactions in social networks can provide important insights into questions of human social and relational behavior, as well as shape the design of social platforms and applications. For example, gauging the level of reciprocity in social interactions on SNSs can shed light on the factors that motivate social interactions. In addition, understanding how interactions are distributed between linked friends can assist in understanding information dissemination in social networks, thus identifying "popular" or "influential" users within the network (Chen, Wang, & Yang, 2009; Gruhl, Guha, Liben-Nowell, & Tomkins, 2004; Kempe, Kleinberg, & Tardos, 2003), as well as how information flows (Cha, Mislove, Adams, & Gummadi, 2008; Galuba, Aberer, Chakraborty, Despotovic, & Kellerer, 2010; Yoneki & Crowcroft, 2009). Moreover, lessons from studying how users interact through these communication tools can guide the design of new, more engaging mechanisms for social interaction and relationship management.

Initial studies of SNSs (Ahn, Han, Kwak, Moon, & Jeong, 2007; Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007; Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009) focused on topological characteristics of the *social graph*, which reveal patterns of explicit digital relationships (i.e., 'friend' links) between users, and have been used to approximate real-world social relationships that exist between users. However, researchers in psychology and sociology have cast doubt on the practice of inferring meaningful relationships from friend linkages alone, given how easy it is to link to others in the network whom one does not know at all or whom one barely knows (e.g., Bilge, Strufe, Balzarotti, & Kirda, 2009; boyd, 2006; Sophos, 2007). To better capture the true nature of social relationships between SNS users, recent work has shifted to measuring observable social interactions (e.g., wall posts and photo comments) between users in these networks (Chun, Kwak, Eom, Ahn, Moon, & Jeong, 2008; Leskovec & Horvitz, 2008; Viswanath, Mislove, Cha, & Gummadi, 2009; Wilson et al., 2009). By constructing and then analyzing the *visible interaction graph* to examine who actually interacts with whom across different friend links in SNSs, these studies can better distinguish close-knit, active social relationships from weak or dormant relationships, and thus derive a more accurate representation of true social relationships between social network users.

All of these studies take advantage of the fact that SNSs represent powerful environments to observe social relationships at a scale never before possible by making such interactions visible, traceable, and machine-parsable, and thus easy for researchers to capture and study. While this offers exciting promise for developing new and better models of human socio-relational behavior, some recent studies (Benevenuto, Rodrigues, Cha, & Almeida, 2009; Schneider, Feldmann, Krisnamurthy, & Willinger, 2009) are revealing that many user interactions on SNSs consist of *latent* social interactions, which involve passive actions such as profile browsing (i.e., looking at someone's profile without leaving any messages or posts).

Latent interactions within SNSs have been difficult to study because they cannot be observed by traditional network measurement techniques, including those described above, because they leave no traceable data. Moreover, most SNS privacy policies make these latent interactions invisible to researchers. The few studies that do exist (Benevenuto et al., 2009; Schneider et al., 2009) rely on clickstream data that are constrained by properties of complexity and verboseness that make it extremely difficult to gather enough data to study user latent interactions comprehensively at scale. However, a large-scale study is necessary to answer deeper questions about user behavior and social interactions in SNSs that go beyond what existing studies of social, visible, and latent interaction graphs have allowed thus far. For example, some of the questions that can be answered with comprehensive data on latent social interactions in SNSs include: To what degree are user interactions in SNS reciprocal? Do latent interactions such as profile browsing occur with the same frequency as visible actions such as user comments? What can users do to become "popular" and draw more visitors to their pages? In this paper we begin to answer these, as well as related questions by discussing new research that looks at both visible and latent social interactions in SNSs.

Answering these sorts of questions has theoretical import as well. In fact, examining latent social interaction in SNSs can contribute new understandings to current theoretical debates surrounding social and relational behavior online. These debates include such topics as the prevalence and impacts of interpersonal electronic surveillance (including 'virtual people watching' and 'cyberstalking', for example, see boyd, 2008; Joinson, 2008, and Tokunaga, 2011 for discussions of these issues), motives for lurking behavior in online communities (e.g., Preece, Nonnecke, & Andrews, 2004; Rau, Gao, & Ding, 2008; Soroka & Rafaeli, 2006), as well as the role that SNS play in enhancing users' self-identity (e.g., Valkenburg, Peter, & Shouten, 2006) and social capital (e.g., Ellison, Steinfeild, & Lampe, 2007). For example, a great deal of concern has focused on the issue of whether SNS usage facilitates stalking (e.g., Gross & Acquisti, 2005). Recent work has suggested that this concern may be overblown because most social interaction on SNS users takes place between users who know and trust one another (Joinson, 2008; Lampe, Ellison, & Steinfield, 2006; Lenhart & Madden, 2007; Pempek, Yermolayeva, & Calvert, 2009). Data on latent social interactions by both friends and strangers within SNSs can shed new light on the prevalence and types of actual social interactions that do take place between all network users. Also, to the extent that physical stalking may be initiated by profile browsing, these data can help us better understand the extent to which SNSs play a role in stalking.

Similarly, knowing the prevalence and nature of latent social interaction in SNSs can help to better understand the ways in which people use these venues for building social capital, which is another major area of theoretical concern within SNS research (boyd & Ellison, 2007). Social capital refers to the potential benefits derived from having connections to others within a network that one can draw resources from, such as information diffusion and linking to external assets. While several studies have theorized and found evidence that visible social interactions in SNSs can increase users' social capital (e.g., Ellison, Lampe, & Steinfield, 2008; Ellison et al., 2007; Steinfield, Ellison, & Lampe, 2008), little is known whether more passive forms of social interaction in SNSs have similar effects, although latent tie theory (Haythornwaite, 2005) and theories of information propagation within networks (see Jiang, Wilson, Wang, Huang, Sha, Dai, & Zhao, 2010) suggest that they should.

Examinations of latent social interactions in SNS can also help to flesh out theories connecting SNS use and its impacts on identity and impression management. Most studies in this domain have only looked at the effects of visible social interactions or size of SNS users' social networks (i.e., number of linked friends) on impression management and identity construction (e.g., Tong, van der Heide, Langwell, & Walther, 2008; Valkenburg et al., 2006). However, according to social information processing theory (Walther, 1992), people will use any available cues in computer-mediated environments to form impressions, suggesting that latent social interactions showing one's profile popularity (i.e., knowing how many people and who has viewed your profile), may have equally important effects on users' social judgments and self-esteem.

As we shall show in this paper, our findings on latent social interactions in SNSs in some cases support the theoretical claims made in these literatures, but also extend and challenge existing theory and research in some of these domains as well. Before reaching those conclusions, however, we describe our approach to measuring latent social interactions in SNSs in detail.

Capturing Latent Social Interactions in SNSs

Our quest for a deeper understanding of user interactions in SNSs begins by addressing the challenge of gathering data on latent interactions. To do this, we performed a large-scale crawl of the Renren social network (www.renren.com). Renren is the largest SNS in China with more than 150 million users to date (Jiang et al., 2010). Functionally, Renren 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, however, Renren has two unique features that make it a particularly attractive platform to study user social interactions.

First, while Renren users have full privacy control over their personal profiles, their friend lists are public and unprotected by privacy mechanisms, which means that they can be crawled to produce an exhaustive snapshot of Renren's largest connected component, producing an extremely large social graph. For example, the data described in this paper are based on a social graph with 42 million users and 1.66 billion explicit social ('friend') links. Second, and perhaps more important, Renren user profiles make a variety of information visible to both the profile owner and her visitors. 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 nine visitors ordered from most to least recent. This list is updated in real time. In addition, each photo and diary entry also has its own page that includes a count of visits by users other than the owner. These records are extremely valuable, in that they expose latent browsing events (i.e., profile visits) to crawlers. With this information, we are able to construct *latent interaction graphs* of user profile browsing behavior, granting us a unique opportunity to gather and analyze large-scale naturalistic data on latent social interaction within the Renren social network.[1]

As mentioned earlier, these data allow us to answer a number of questions about the nature of social interaction within SNSs that have been as-yet unavailable to researchers. For example, this paper will compare the structural properties of the latent social interaction graph (i.e., profile views) against those of both visible interaction graphs (i.e., written posts and comments) and social graphs (i.e., confirmed 'friend' linkages) to answer the following questions:

1. What proportion of social interactions in SNSs is latent versus visible?
2. Do latent social interactions follow similar distribution patterns in terms of participation across users as visible interactions within the SNS user population?
3. How do social, visible, and latent interaction graphs differ in terms of (a) their social connection (average degree of separation between users), (b) how tightly or loosely

connected friend relationships are within each type of network, and (c) the likelihood of users interacting with others similar to themselves in their degree of social connectedness?

A second set of questions to be answered in this paper revolves around a detailed examination of latent social interactions (i.e., profile visits) specifically. These include:

1. What factors (e.g., number of linked friends, account longevity, frequency of profile content updates, amount of comments) influence latent social interactions?
2. While visible social interactions are typically between linked friends, are profile visitors mostly friends or are they strangers?
3. What portion of profile visitors are repeat visitors and, of the repeat visitors, are most profile viewers friends or strangers?
4. Visible interactions tend to be highly reciprocal because social norms compel people to reply to one another when contacted via visible interactions, but to what extent are latent interactions reciprocated?

Answers to these questions will reveal new information about human behavior within SNSs environments, and will thus offer potential to advance current understandings of social interactions within online social networks that are useful to social and computer scientists alike.

Crawling the Renren Social Network

As mentioned previously, Renren is 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 (Wilson, Boe, Sala, Puttaswamy, & Zhao, 2009). Fortunately, profiles of users in authenticated networks are public by default to other members of the same network, which is a feature that allowed us to access user profiles within the Peking University (PKU) network, by creating nearly unlimited authenticated accounts using our own block of IP addresses.

We crawled the entire Renren network from April 2009 to June 2009, and again from September to November of 2009.[2] 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. In addition to the complete crawl, we also performed smaller, more detailed crawls of the PKU network between September and November of 2009 (90 days) to collect information about users' profiles and interaction patterns. 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 have crawled SNS regional networks (Wilson et al., 2009).

As part of our PKU crawls, we gathered all comments generated by users in message board posts, diary entries, photos, and status updates. These data form the basis of our analysis of visible social 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.[3]

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 via successive crawls. These data form the basis of our study of latent social interactions.[4] Overall, from the 61,405 user profiles we crawled, we obtained a total of 8,034,664 total records of visits to user profiles in the PKU network, but after integrating the raw results, we were left with 1,863,168 unique profile visit events. The 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.[5]

Our Renren dataset is larger than most previously studied SNS datasets, the exception being recent studies of the Twitter network (Cha, Haddadi, Benevenuto & Gummadi, 2010; Kwak, Lee, Park, & Moon, 2010). However, our dataset shares similar properties with social network datasets collected in prior studies (Chun, Kwak, Eom, Ahn, Moon, & Jeong, 2008; Mislove, Marcon, Gummadi, Druschel, & Bhattacharjee, 2007; Wilson et al., 2009). This suggests that the Renren network is quite similar to other popular social networks in terms of the network's structural properties, which are derived from user behavior. For greater detail on the methods used to extract the data for this study, please see Jiang et al. (2010).

Participation in Latent versus Visible Social Interactions Across Users

We compared latent (i.e., viewing others' profiles) and visible (i.e., leaving posts and comments on others' profiles) social interactions to see how they differ in terms of prevalence. To understand the level of participation of users in both latent and visible interactions, we examined 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: 28% of users account for nearly 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 (see Figure 1). This confirms that latent interactions are more prevalent in Renren than visible interactions—users are more active in viewing profiles than they are in leaving comments—potentially because of a sense of anonymity in profile browsing.

Figure 1. Distribution of visible and latent interactions

Next, for each user, we analyzed the distribution of their latent and visible interactions across their social links. Aggregating across all users, the percentage of friends involved in these events show that roughly 80% of users only interact visibly with 5% of their friends, and no users interact with more than 40% of their friends. By 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 do visible interactions.

To get a sense of how many visible comments are generated by latent interactions, we examined the average number of comments per page view for a variety of pages on Renren, including profiles, diary entries, and photos. Recall that along with visible comments, Renren keeps a visitor counter for each photo and diary entry. For diary entry and photo views, 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. By 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 analyzed the repeat activity frequency for latent and visible interactions on Renren. In particular, we examined the likelihood that users will repeatedly interact with the same profile page after they have viewed or commented on it once. The data show that 80% of users view a given profile fewer than two times. However, 80% of users leave an average of 3.4 comments, almost twice the number of latent interactions. This result makes sense intuitively: for most types of information, users only need to view it once to consume the data. However, comments can stimulate flurries of dialog on a given page, resulting in many consecutive interactions.

Understanding Latent Social Interactions: Profile Views and Viewing Behavior

Because our work focuses on the analysis of latent interactions and the role they play in SNSs, we examined profile viewing behavior by analyzing the histories of visits to user profiles in Renren. This allowed an understanding of the popularity of latent social interaction both in terms of profile views of individual profiles and their viewing of others' profiles. It also enabled us to examine the composition of profile visitors and reciprocity in latent social interactions.

Popularity. Our analysis of the distribution of latent interactions across the Renren user base showed that the number of views a user's profile receives (i.e., profile "*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. Moreover, 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 indicates that latent interactions are highly skewed towards a very popular subset of the population.

Consumption. Our analysis also showed that the popularity of users does not necessarily correspond to their profile viewing behavior. We define *consumption* as the number of others' profiles a user views. In Renren, the top 1% most popular users have 9% overlap with the top 1% biggest profile consumers. Those 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 means that many (presumably average) users actively visit others, but are not visited in return. The degree of reciprocation in latent interactions is explored in greater detail below.

Composition of profile visitors: Friends or strangers? While visible social interactions are typically between linked friends, we sought to understand whether this was also true for latent social interactions. To answer this question, we analyzed the composition of visitors to user profiles to determine whether profile visitors were mostly friends or strangers. We define a stranger as any user who is not a direct linked 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. The results are fairly evenly divided: roughly 45% of users receive less than half of their profile visits from strangers. Or conversely, a slight majority of the population receives most of their profile views from strangers (see Figure 2, which shows the percentage of visitors that are strangers).

Figure 2. Percentage of strangers in profile visitors

To understand what component of a profile's visitors are strangers, and how far away they are from the profile owner in the social network, we grouped the owners of profiles together by their number of friends ("social degree") and computed 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 visiting. For well-connected users with 100–1,000 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 1,000 friends, their notoriety is such that they start to attract more strangers to visit their profiles. Figure 3 illustrates these results. These results confirm findings from previous work that discovered many Orkut users browse profiles two or more hops away on the social network (Benevenuto, Rodrigues, Cha, & Almeida, 2009).

Figure 3. Breakdown of profile visitors by owner's social degree

---

Repeat visitors. The portion of profile visitors that are repeat visitors was calculated from the percentage of repeated visitors for each profile. Roughly 70% of users have less than 50% repeat visitors, meaning that the majority of visitors do not browse the same profile twice (see Figure 4). This indicates that the long tail of latent interactions is generated by users randomly browsing others in the social network.

Figure 4. Ratio of repeat profile visitors

Most repeat visits occur on the same day as the initial visit, as shown by the probability density function (PDF) of the interval time between repeat visits (see Figure 5). 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 a later section of this paper. The probability for repeated views decreases with time, 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.

Figure 5. Interval time between repeat profile visits

Unlike friends, strangers do not tend to 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 computed the average number of visits for strangers and friends for each profile. Surprisingly, our results indicate that the repeat profile viewing behavior for friends and strangers is very similar, with friends only edging out strangers by only a small margin. This demonstrates that when considering information dissemination via latent interactions, the significance of non-friend strangers should not be overlooked.

Reciprocity in latent social interaction. Social norms compel users to reply to one another when contacted via visible interactions (e.g., comments). Prior work has shown that these visible interactions are largely reciprocal between linked friends on SNSs (Wilson et al., 2009). 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 users' profiles actually trigger reciprocal visits?

As the first step towards looking more deeply at reciprocity of latent interactions in Renren, we constructed the set of visitors who view each user profile, and the set of people who are visited by each user. Then, we computed 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—in other words, the latent interactions are reciprocated. Unions contain all latent relationships for a given user, including all users who viewed them, or they viewed. The ratio of intersection size to union size for each user 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. Friends had higher probability of reciprocal visits than did strangers and, in comparison to visible interactions, latent interactions showed much less reciprocity.[6]

We next examined how reciprocal profile visits vary over time, for both strangers and friends. We compute the number of reciprocal visits that take place within 90 days after the initial visit. Not surprisingly, more profile visits are reciprocated as more time elapses. However, reciprocity remains low overall. Even across the entire 90-day 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.

Factors Impacting Latent Social Interactions

As discussed earlier, not all users in Renren are the target of equal numbers of latent interactions. Understanding the factors that may impact latent interactions can help us to understand the popularity of individual users (i.e., the number of views a profile receives). This is important because popularity is likely to be correlated with sources of significant information dissemination and because, as an underlying social goal, popularity can drive individuals' behavior within SNSs. There are several potential factors that could influence latent social interaction by attracting profile visitors, for example, the degree to which a user maintains an active presence in the SNS by adding or updating their profile content, including status updates, diary entries, photos, and shared links to content on the web; how often a user leaves comments on others' profiles; the number of linked friends they have in the network; and how long a user has been active in the network. To examine these, we analyzed whether (a) the number of friends a user has, (b) the longevity of a user's SNS account, (c) the frequency of uploading new user-generated content (i.e., status updates, diary entries, photos, and shared links); and (d) the frequency of leaving comments influence the number of latent social interactions (i.e., profile views) a user has.[7]

The data show that all factors increase with popularity such that users with the most profile views (latent interactions) also have the most friends, the oldest accounts, and generate the largest amounts of content/visible interactions. To reach this conclusion, we divided users into four groups based on their profile viewership "popularity," and calculated both the average value of these factors in each group, as well as the correlation between each factor and user popularity within each group (see Table 1, correlations are shown in parenthesis beside the average value for each factor).[8] Although all factors exhibit high correlation with the low viewership popularity (those whose profiles were rarely viewed by others) and "all users" categories, this is an artifact of the tied ranks among the many low activity users (0-100 profile views category). All of these users exhibit very low interactivity and number of friends, thus leading to high levels of correlation. Previous work has observed similar artifacts when analyzing all users in a large SNS dataset (Cha et al., 2010).

Table 1. Average value of factors affecting user popularity and correlations of factors with each user popularity category

---

For the two medium popularity groups (100-1,000 and 1,000-10,000 profile views), number of friends had the highest correlation with profile views. Users in these categories can be broadly defined as normal social network users. They are not celebrities; they simply use the SNS 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 1,000-10,000 popularity range, given the ease with which users can quickly amass hundreds of friends on SNSs.

No factor has strong correlation with profile views for users in the high popularity group. Correlations 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 study that contribute to determine users' profile 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 viewership popularity.

Comparing Social, Visible, and Latent Interaction Graph Properties

Thus far, our results have 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 likely to have profound implications on applications that leverage social graphs (e.g., knowing what percentage of people are passive consumers versus active producers of information in social review systems such as Yelp, Delicious, or other recommendation systems could be useful in enhancing user trust of these systems, as well as in designing new systems to encourage active participation), and thus warrant the construction of a new model to capture the properties of latent social interactions. We call this new model a latent interaction graph.

While previous studies have constructed visible interaction graphs by connecting users from the social graph who have visibly interacted one or more times (Wilson et al., 2009), a *latent interaction graph* is defined as a set of users that are connected via links representing latent interaction events between them. We constructed 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.[9] Finally, we constructed the *social graph* for Renren, which reflects explicit friendship linkages between users in the network. We use these data to understand how the three types of graphs differ in terms of (a) how tightly or loosely connected friend relationships are within each type of network, (b) the likelihood of users interacting with others similar to themselves in their degree of social connectedness, and (c) their social connection, defined in terms of the average degree of separation between users.

Relationship strength. To determine relationship strength in the three graphs, a clustering coefficient was calculated which shows the level of local connectivity between nodes (i.e., users or events) in a network. The data show that the average clustering coefficients for the latent interaction graph (.03) and for the visible interaction graph (.05), are both much less than that of the social graph (.18). This is because not all friend links are accurate indicators of active social relationships between those linked users, and friend links with no interactions are removed in interaction graphs. This produces loose connections between neighbors, and thus relatively low clustering coefficients in the interaction graphs. Because a portion of the latent interactions to a profile is from strangers who randomly browse the network, it is unlikely that a user will share mutual friends with strangers, which further lowers the clustering coefficients in latent interaction graphs. These findings make sense since many people in SNSs do not actually talk to all of their linked friends, and linked friends who do talk to each other are more likely to have a closer relationship than those who merely browse others users' profiles.

Heterogeneity of association. Assortativity is the likelihood that users establish links to other users of similar social degree (i.e., users with few friends connect to other users with few friends, and users with many friends connect to others who have many friends). Renren latent interaction graph is disassortative (-.06). This makes sense intuitively because latent social interactions are highly skewed towards a small subset of extremely popular users (e.g., celebrities, physically attractive users), as described earlier. In contrast, the other two graphs are both assortative, with the social graph (.23) being more assortative than the visible interaction graph (.05). This also makes sense because, unlike latent interactions, explicit social friend links and visible interactions tend to be based on relationships that are bidirectional (e.g., users typically befriend and write wall messages to those whom they have an established real-world relationship with). Generally speaking, relationships tend to form between similar, rather than dissimilar, people (Fehr, 2008; McPherson, Smith-Lovin, & Cook, 2001).

Social Connection. The degree of social connection within these graphs is reflected in the average path length, which is the average of all-pairs-shortest-paths in the network. In the case of Renren, the average path length of the latent interaction graph (4.02) falls between that of the visible interaction graph (5.43) and the social graph (3.64). Average path length of a graph decreases when the number of connections inside the graph increases, because more connections means it is easier to find shorter paths between users. The social graph is the most highly connected of the three types of graphs studied here, and thus has the shortest average path length. Because in the visible interaction graph, links between friends who do not talk to one another are removed, the sheer number of connections in this graph is lower than in the social graph and thus the average path length is higher. The average path length for latent interactions falls between these two other graphs because people browse profiles a lot more than they write visible comments, but still not quite enough to cover the original social graph (i.e., most people do not view the profiles of all of their friends within a given timeframe such as a week or month).

Conclusion and Theoretical Implications of the Research

Our study of Renren reveals a new perspective on social interaction within online social networks. This paper uses novel methodology in computer science to capture latent social interactions in SNSs. Our data include detailed visit histories to the profiles of 61,405 Renren users over a 90-day period. We compute a single visitor history for each profile by using a new technique to merge visitor logs from multiple consecutive crawls. We then analyze profile visit histories to study important social scientific questions concerning user popularity and reciprocity for profile browsing behavior, and the link between passive profile browsing and (inter)active comments.

Indeed, the concept of latent social interaction overlaps with two widely studied phenomena in the social sciences: lurking and interpersonal electronic surveillance (IES). Lurking occurs when people use an online community to read messages posted by others without posting messages themselves. Lurkers have thus been defined as “a persistent, yet silent audience” (Soroka & Rafaeli, 2006). Interpersonal electronic surveillance is defined as “surreptitious strategies individuals use over communication technologies to gain awareness of another user’s offline and/or online behaviors” (Tokunaga, 2011, p. 706). IES is a general term closely related to concepts such as horizontal or participatory surveillance (Albrechtslund, 2008), peer-to-peer monitoring (Andrejevic, 2005), social surveillance (Steinfield et al., 2008), and social searching (Lampe et al., 2006).

Although lurking is common and expected in traditional communication systems (e.g., one-way mass media), it would not seem to be a hallmark of newer interactive media. Indeed, lurking seems to fly in the face of the participatory culture of Web 2.0. Soroka and Rafaeli (2006) argue that newer forms of communication, including SNSs, are fundamentally about interactivity (see also Pempek et al., 2009). It is interesting, then, that studies of many online communities find a great deal of lurking (Katz, 1998; Pempek et al., 2009; Preece et al., 2004). Our data on the prevalence of latent versus visible interactions confirm that many SNS users engage in lurking. The amount of lurking in online environments is unknown and estimates vary widely. Some studies report lurking by about 50-60% of community or network members, whereas others put the number at about 90% (Soroka & Rafaeli, 2006). These studies tend to be based on rather small sample sizes. Moreover, estimates based on social scientific studies may systematically underestimate the amount of lurking, and particularly the amount of viewing strangers' profiles in SNSs due to stigma associated with 'cyberstalking' (or virtual people-watching) and resulting social desirability response biases in survey-based studies.

Our study provides another view of the prevalence of lurking, based on actual SNS user behavior, rather than self-reports. Similar to prior research in other contexts, we find that many more users engage in latent social interaction (lurking via profile viewing) than in visible social interaction in Renren. Furthermore, we see that most users (80%) interact visibly with only a very small subset (5%) of their friends, and yet a majority of users view a sizable portion of their friends' profiles on at least a weekly basis. Of course, there are two types of lurking possible in SNSs: viewing friends' profiles and viewing strangers' profiles. Pempek et al.'s (2009) survey of undergraduate Facebook users found that 70% report viewing friends' profiles, which is very similar to our results. However, their estimate of stranger profile browsing was only 9%, whereas we found that most Renren users received a majority of their profile views from strangers. This was especially true for Renren users with few friends and those with many friends, while the middle groups got most of their profile views from direct friends. Interestingly, we also found that the frequency of repeat profile views from friends and strangers was very similar. Our data offer nuance to, and in some ways refute, Pempek et al. as well as others who claim that interactions in SNS are predominantly between people who know and trust each other, as this does not appear to be the case for latent social interactions. We will return to this point later.

Explanations for lurking in online communities include things such as needing to learn about the community before actively participating, allowing participants to gain a sense of belonging at low cost, fear of negative reaction from community, and information overload (Preece et al., 2004; Soroka & Rafaeli, 2006). The logic of information overload theory is that, as the number of interactive posters within a community increases, it is difficult for individuals to keep up with and respond to the flow of messages. Limitations in both human cognitive ability to process massive amounts of information and time constraints lead to a preference for lurking. As mentioned earlier, our data on latent social interaction in Renren certainly support the notion that a great deal of lurking takes place in SNSs, which is consistent with information overload theory. The theory further implies that users with more friends will lurk more than users with fewer friends, given that each friend connection opens exponentially more opportunities for social interaction with that person that can be monitored or responded to on the SNS. Although our data do not speak to this directly, we find that having more friends results in a higher number of incoming views to one's own profile, but that this does not necessarily trigger reciprocal latent (or visible) social interaction. It may be the case that lurking itself contributes to overload, and thus is not reciprocated. These findings suggest that further investigation of information overload theory, as well as other motives for lurking in SNSs, are warranted.

Rau et al. (2008) indeed point out that motivations for lurking on SNSs may be different than in other types of online communities. Online communities (e.g., online forums, USENET groups, etc.) differ from SNSs in that SNSs satisfy social-emotional needs, rather than informational needs, and people on SNSs are connected to individual users, rather than a large community. Rau et al. found that SNS users lurk on SNSs because they believe that their social-emotional needs may not be satisfied even if they were to post. Further, they found that SNS users are not motivated to post on the profiles of users whom they seldom talk to, or with whom they only discuss shallow topics. Rau et al. also find that affective intimacy with friends in the network negatively influences lurking behavior. These findings are consistent with our data that show differences in lurking behavior among users with greater numbers of profile views. Users who get a lot of profile views may not feel intimately connected to their viewers and are thus not compelled to interact with them visibly or otherwise. Together, these results call for future social science research to build on our findings on latent social interactions by focusing on explanations for lurking in SNS environments.

As with lurking, naturalistic behavioral data on latent social interactions can go beyond what has been revealed by social science research methods about SNS interaction in terms of interpersonal electronic surveillance. Early concerns of SNS use focused on the dangers of social surveillance and 'stalking' by strangers on these sites (see boyd, 2008 for a discussion of the "moral panic" surrounding sexual predators and SNS usage). However, more recent studies based on user surveys found that most self-reported SNS social interactions are between real-world friends or contacts rather than strangers (Joinson, 2008; Lampe et al., 2006; Lenhart & Madden, 2007; Pempek et al., 2009). Our data contradict these findings to show that a sizable portion of social interactions on SNSs do take place between strangers and that passive social browsing is an important component of SNS use. Our data also add nuance to the largely survey-based literature by showing that profile views of Renren users who possess a middle range of friends on the network (between 100-1,000) are mostly from their known friends, whereas profile views of users with very few (< 100) and very many (over 1,000) linked friends are mostly from strangers. Overall, then, our data indicate that the moral panic surrounding cyberstalking may be more justified for some types of SNS users than for others.

The negative connotation surrounding IES is not surprising because, according to latent tie theory (Haythornwaite, 2005), SNSs are a communication channel that open possibilities for greater numbers of connections to others. The connotation may not be warranted, however, because latent social interaction is mostly between friends for a sizable portion of the user base, and because even weak connections can lead to positive outcomes, such as greater social capital (Albrechtslund, 2008; Donath & boyd, 2004; Haythornwaite, 2005; Steinfield et al., 2008). The connection between social capital and SNSs is based on Granovetter's (1982) "strength of weak ties" argument (Ellison et al., 2007), which says that personal and social benefits can accrue from even loose connections between people, as they can provide access to new information and resources that are unavailable through one's more intimate network. Steinfield et al. (2008) found this to be the case in that more avid users of Facebook (in terms of time on the SNS and number of friends) reported greater social capital.

Latent social interactions between friends or strangers in SNSs may be considered a form of 'weak ties' in that mere profile browsing might also activate previously dormant or non-existent ties between users, which the user could then benefit from. For example, via profile browsing, a user could learn about common concerns, efforts, or interests of her fellow SNS members, which could motivate her to activate a relationship. A form of social capital may even exist when latent social interaction events produce no action or reciprocation, simply by the transmission of information from one user to another in the network. That is, when a user browses a profile, information on the browsed page is naturally propagated to the viewer. In an experiment comparing information dissemination across social, visible, and latent interaction graphs, Jiang et al. (2010) found that the latent interaction graph was most efficient for transmitting information between users. This is due to the fact that latent interactions require less time and effort to perform than visible interactions, and so they happen more frequently. Thus, the potential for activation between users is high. Also, social graphs can only spread information between linked friends, and so they are less efficient for information dissemination because there are far fewer formal connections (i.e., friend links) than there are informal latent interactions between users.

Understanding latent social interactions may also shed light on processes of identity management in SNSs. Donath and boyd (2004) and boyd and Heer (2006) found that impression management motivated Friendster users to establish formal friend links to certain others on the network (e.g., to interesting or physically attractive people) because such public displays of connection serve as important identity signals to others, as well as provide self-validation. Although it is obvious why having a lot of friends on the SNS can make users feel good about themselves, it could also be that seeing that one's profile attracts a lot of visitors may similarly enhance users' self-esteem. Of course, this effect would be limited to SNSs that make or allow profile visitors visible to users, such as Renren. As mentioned at the outset of this paper, Facebook itself does not afford this capability, and nor does its terms of service allow third party applications that make latent social interactions visible to Facebook users. This highlights the need for researchers to collect data beyond those available in Facebook to understand how processes of identity management may vary across different SNSs.[10]

To conclude, our study reveals interesting insights into the nature of user behavior in SNSs. We observe that user behavior is different for latent compared to other types of social interactions via SNSs: more users participate in latent social interaction, users do not feel the need to reciprocate profile 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 peer social interactions. Analysis of these graphs derived from our Renren dataset 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 than the formal ‘friend’ linkages that are provided by the social graph alone.

Finally, our study 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 a 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.

References

Ahn, Y. -Y., Han, S., Kwak, H., Moon, S. B., & Jeong, H. (2007). Analysis of topological characteristics of huge online social networking services. *Proceedings of the 16th International Conference on World Wide Web (WWW'*07), Banff, Alberta, Canada.

Albrechtslund, A. (2008). Online social networking as participatory surveillance *First Monday*, *13*(3).

Andrejevic, M. (2005). The work of watching one another: Lateral surveillance, risk, and governance. *Surveillance & Society, 2*(4), 479-497.

Benevenuto, F., Rodrigues, T., Cha, M., & Almeida, V. (2009). Characterizing user behavior in online social networks. *Proceedings of the ACM Internet Measurement Conference (IMC'09)*, Chicago, IL.

Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009). All your contacts belong to us: Automated identity theft attacks on social networks. *Proceedings of the 18th International Conference on World Wide Web (WWW'*09), Madrid, Spain.

boyd, d. (2006). Friends, Friendsters, and MySpace Top 8: Writing community into being on social network sites. *First Monday, 11*(12).

boyd, d. (2008). *Taken out of context: American teen sociality in networked publics* (Doctoral dissertation). University of California, Berkeley, CA.

boyd, d., & Heer, J. (2006). Profiles as conversation: Networked identity performance on Friendster. *Proceedings of the 39th Hawaiian International Conference on System Sciences*, Kauai, HI.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010). Measuring user influence in twitter: The million follower fallacy. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM),* Washington, DC.

Cha, M., Mislove, A., Adams, B., & Gummadi, K. (2008). Characterizing social cascades in Flickr. *Proceedings of the ACM Workshop on Online Social Networks* (*WOSN'08*), Seattle, WA.

Chen, W., Wand, Y., & Yang, S. (2009). Efficient influence maximization in social networks. *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining,* Paris, France.

Chun, H., Kwak, H., Eom, Y. H., Ahn, Y.-Y., Moon, S. B., & Jeong, H. (2008). Comparison of online social relations in volume vs. interaction: a case study of Cyworld. *Proceedings of the ACM Internet Measurement Conference (IMC'08)*, Vouliagmeni, Greece.

Donath, J., & boyd, d. (2004). Public displays of connection. *BT Technology Journal, 22*(4), 71-82.

Ellison, N., Lampe, C., & Steinfield, C. (2008, May). Net worth: Facebook use and changes in social capital over time. Paper presented at the annual meeting of the International Communication Association, TBA, Montreal, Quebec, Canada.

Ellison, N., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends": Exploring the relationship between college students' use of online social networks and social capital. *Journal of Computer-Mediated Communication, 12*(3).

Fehr, B. (2008). Friendship formation. In S. Sprecher, A. Wenzel, & J. Harvey (Eds.), *Handbook of relationship initiation* (pp. 29-54). New York: Taylor & Francis.

Galuba, W., Aberer, K., Chakraborty, D., Despotovic, Z., & Kellerer, W. (2010). Outtweeting the Twitterers? Predicting information cascades in microblogs. *Proceedings of the ACM Workshop on Online Social Networks* (*WOSN'10*), Boston, MA.

Gannes, L. (2010). *When social replaces search, what can you do to monetize?* Retrieved October 29, 2010, from http://gigaom.com/2010/03/24/when-social-replaces-search-what-can-you-do-to-monetize/.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory, 1*(1), 201-233.

Gruhl, D., Guah, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. *Proceedings of the 13th International Conference on World Wide Web (WWW'04)*, New York, NY.

Haythornthwaite, C. (2005). Social networks and Internet connectivity effects. *Information, Communication & Society, 8*(2), 125-147.

Jiang, J., Wilson, C., Wang, X., Huang, P., Sha, W., Dai, Y., & Zhao, B. Y. (2010). Understanding latent social interactions in online social networks. *Proceedings of the ACM Internet Measurement Conference (IMC'10)*, Melbourne, Australia.

Joinson, A. N. (2008). 'Looking at', 'Looking up' or 'Keeping up with' people? Motives and uses of Facebook. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05-10 2008), CHI '08, ACM, New York, (pp. 1027-1036).

Katz, J. (1998). *Luring the lurkers*. Retrieved October 30, 2010, from http://slashdot.org/features/98/12/28/1745252.shtml.

Kempe, D., Kleinberg, J. M., & Tardos, Ï. (2003). Maximizing the spread of influence through a social network. *Proceedings of the ACM Conference on Knowledge Discovery and Data*

Mining, Washington, DC.

Kirkpatrick, M. (2009). *Social networking now more popular than email, report finds*. Retrieved August 12, 2010, from http://www.readwriteweb.com/archives/social_networking_now_more_popular_than_email.php.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is twitter, a social network or a news media? Proceedings of the *19th international conference on World Wide Web (WWW'10),* Raleigh, North Carolina.

Lampe, C., Ellison, N. & Steinfield, C. (2006). A Face(book) in the crowd: Social searching vs. social browsing. *Proceedings of the Conference on Computer-Supported Cooperative Work* (CSCW'06), 167-170, Banff, Alberta.

Lenhart, A., Madden, M., Macgill, A., & Smith, A. (2007, December 19). *Teens and social media*. Pew Internet & American Life Project. Retrieved October 28, 2010, from http://www.pewinternet.org/Reports/2007/Teens-Privacy-and-Online-Social-Networks.aspx.

Leskovec, J., & Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. *Proceedings of the 17th International Conference on World Wide Web (WWW'08),* Beijing, China.

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophility in social networks. *Annual Review of Sociology, 27*, 415-444.

Mislove, A., Marcon, M., Gummadi, P. K., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. *Proceedings of the ACM Internet Measurement Conference (IMC'07)*, San Diego, CA.

Pempek, T. A., Yermolayeya, Y. A., & Calvert, S. L. (2009). College students' social networking experiences on Facebook. *Journal of Applied Developmental Psychology, 30*(3), 227-238.

Preece, J., Nonnecke, B., & Andrews, D. (2004). The top five reasons for lurking: Improving community experiences for everyone. *Computers in Human Behavior, 20*(2), 201-223.

Rau, P.-L. P., Gao, Q., & Ding, Y. (2008). Relationship between the level of intimacy and lurking in online social network services. *Computers in Human Behavior, 24*(6), 2757-2770.

Schneider, F., Feldmann, A., Krishnamurthy, B., & Willinger, W. (2009). Understanding online social network usage from a network perspective. *Proceedings of the ACM Internet Measurement Conference (IMC'09)*, Chicago, IL.

Sophos (2007). *Facebook ID probe shows 41% of users happy to reveal all to potential identity thieves*. Retrieved October 15, 2010, from http://www.sophos.com/pressoffice/news/articles/2007/08/facebook.html/

Sorkin, A. R. (2010, June). *Facebook tops 500 million users*. New York Times Blog. Retrieved August 11, 2010 from http://dealbook.blogs.nytimes.com/2010/07/22/facebook-tops-500-million-users/

Soroka, V. & Rafaeli, S. (2006). Invisible participants: How cultural capital relates to lurking behavior. *Proceedings of the 15th International Conference on World Wide Web (WWW'06)*, Edinburgh, Scotland.

Steinfield, C., Ellison, N., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology, 29*(6), 434-445.

Tokunaga, R. S. (2011). Social networking site or social surveillance site? Understanding the use of interpersonal electronic surveillance in romantic relationships. *Computers in Human Behavior, 27*(2), 705-713.

Tong, S. T., van der Heide, B., Langwell, L., & Walther, J. B. (2008). Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. *Journal of Computer-Mediated Communication, 13*(3), 531-549.

Valkenburg, P. M., Peter, J., & Schouten, A. P. (2006). Friend networking sites and their relationship to adolescents' well-being and social self-esteem. *CyberPsychology & Behavior, 9*(5), 584-590.

Viswanath, B., Mislove, A., Cha, M., & Gummadi, K. P. (2009). On the evolution of user interaction in facebook. P*roceedings of the ACM Workshop on Online Social Networks (WOSN'*09), Barcelona, Spain.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective*. Communication Research, 19,* 52–90.

Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P. N., & Zhao, B. Y. (2009). User interactions in social networks and their implications. *Proceedings of the ACM SIGOPS Eurosys'*09, Nuremberg, Bavaria.

Yarow, J. (2010). *Facebook was more popular in the U.S. than Google last week*. Retrieved October 30, 2010, from http://www.businessinsider.com/facebook-was-more-popular-in-the-us-than-google-last-week-2010-3.

Yoneki E., & Crowcroft, J. (2009). GIS: Geographical Information Cascade in Online Social Networks. *Proceedings of the International Conference on Network Science (NetSci),* Venice, Italy.

Notes

[1] Like Facebook, a Renren user's homepage includes a number of friend recommendations that encourage formation of new friend relationships. Renren lists three users with the most number of mutual friends on each page. In addition, Renren shows a list of eight "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.

Friend lists in Renren are always public, which contrasts with most other SNSs, where full social graph crawls are prevented by user privacy policies that hide friendship links from the public. Moreover, comments in Renren are threaded, which enabled us to precisely distinguish the intended target of each comment. One other 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 data, we saw that very few users (0.3%) took advantage of this feature.

[2] We seeded crawlers with the 30 most popular users' profiles, and performed a breadth-first traversal of the social graph. During the crawl, we collected unique userIDs, network affiliations, and friendship links to other users.

[3] This study focused on the structure of social graphs and interactions between users. Since we did 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 were replaced with relative sequence numbers. Our group has visited and held research meetings with technical teams at Renren, and they are aware of our ongoing research.

[4] Crawling Renren for recent visitor records is complicated by two things. First, each user's profile only lists the last nine 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.

[5] We were 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. Most users (99.3%) receive 8 or fewer 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 computed the number of visits missed by the crawler when we reduce the frequency, beginning with visits every 15 minutes. This shows that 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. Moreover, less than 0.7% of all users receive more than nine 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 the rest), we are relatively confident that very few visits are missed by our crawls.

[6] To calculate this, we quantified the lack of reciprocity for latent interactions by calculating a reciprocity coefficient (see Jiang et al., 2010). 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 is 0.78.

[7] For account longevity we measure the number of days 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 (Wilson et al., 2009). 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.

[8] Given the drastic differences in size of each popularity group, we used Spearman’s rank correlations (Spearman’s ρ), which is a nonparametric measure of the correlation between two variables and is thus more appropriate for this analysis.

[9] 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 we have limited information. Also note that because non-friend strangers can view user's profiles, the latent interaction graph will contain links 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.

[10] Of course, user behavior in other SNSs may vary due to sociopolitical circumstances (e.g., the extent of government control on expression), cultural norms (e.g., collectivism versus individualism, conceptions of personal privacy, etc.), and as a result of different SNS systems' interfaces, user etiquettes, and data control mechanisms. That said, it is likely that our findings hold true for all SNSs that allow users to passively browse profiles of non-friends and see some of their personal information. Most systems do this, including Facebook, although the amount of viewable information there depends on each profile owner's privacy settings. In systems that do not reveal who has browsed a user's profile, such as Facebook, latent social interaction could be *more* prevalent than in Renren since they pose no risk of exposure to users, and yet less reciprocated since they are invisible to users. Recall, however, that even in Renren reciprocity in profile views was low. At the same time, however, in systems like Facebook where users know they can see only limited profile information for network members they are not linked to, latent social interactions could be *less* prevalent than in Renren.

The foregoing points highlight the need for research across different SNS systems and across diverse users situated in culturally diverse settings. This is not easy, however, as Facebook and some other SNSs do not release visitor log data even to academic researchers. One way that future research could try to circumvent this problem is to get data from a major social game company like Zynga, the creator of Farmville. Close to a fifth of the Facebook population plays Zynga games, and the games actively encourage browsing behavior (e.g., "visit my farm!"). Of course, the context for interactions is different here from profile browsing, but Zynga's user population is both very large and highly motivated. It would be interesting to examine whether their data refute or support our Renren findings.

Figure 1: Distribution of visible and latent interactions.

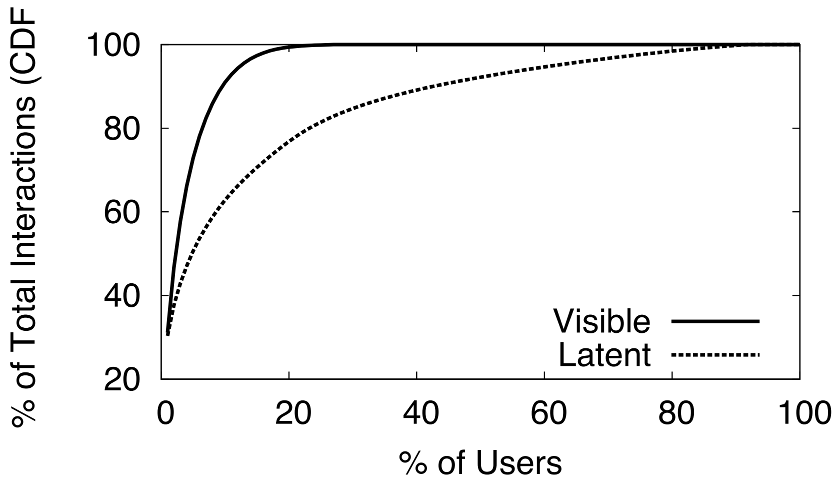

Figure 2: Percentage of strangers in profile visitors.

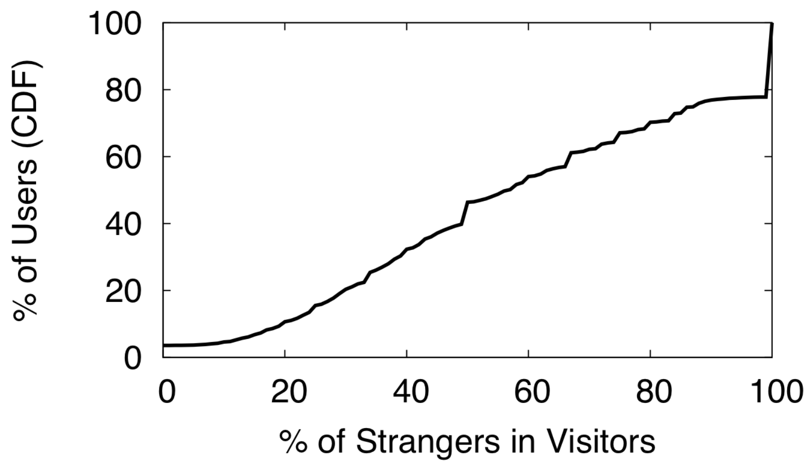

Figure 3: Breakdown of profile visitors by owner's social degree.

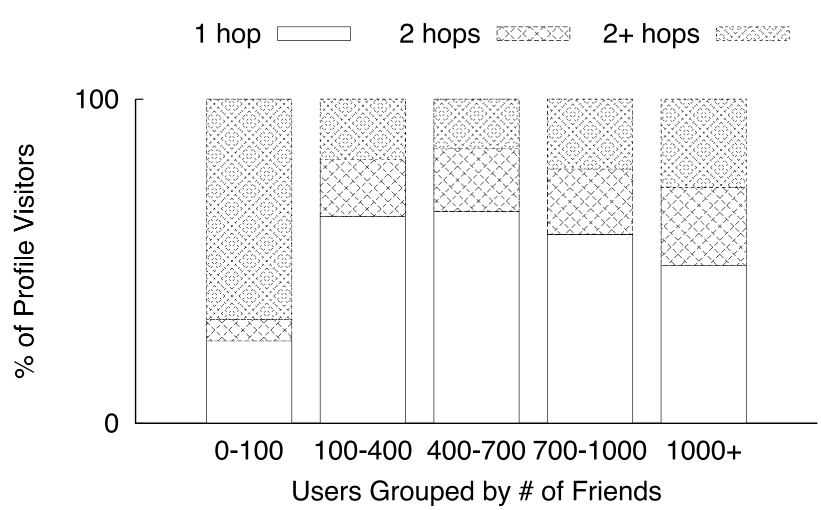

Figure 4: Ratio of repeat profile visitors.

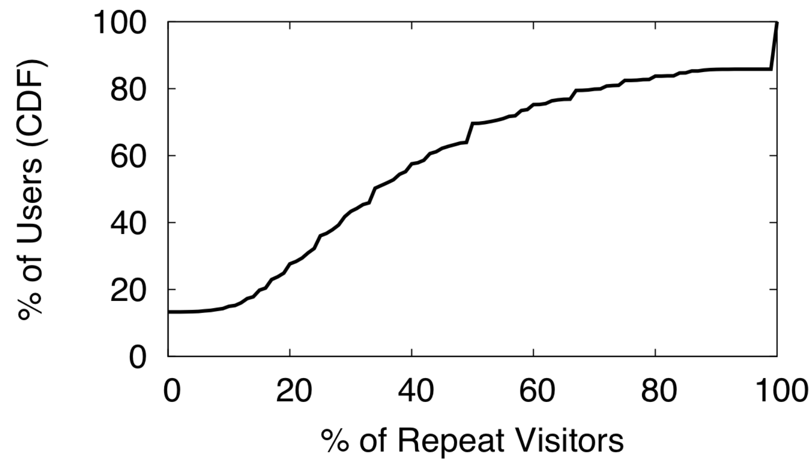

Figure 5: Interval time between repeat profile visits.

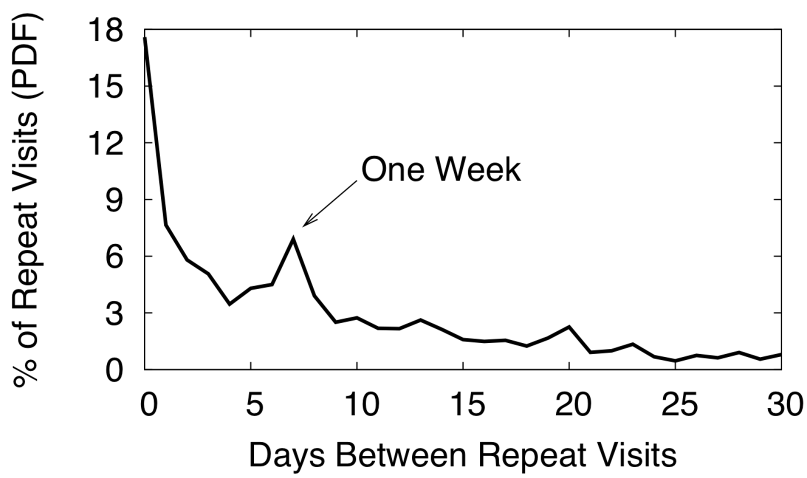

Table 1. Average value of factors affecting user popularity and correlations of factors with each user popularity category

User Popularity (i.e., profile views)	Friends	Lifetime	Diary	Photo	Status	Share	Comment
0-100	16 (.15)	35 (.55)	1 (.51)	3 (.47)	1 (.50)	1 (.50)	2 (.54)
100-1,000	131 (.56)	423 (.41)	11 (.33)	41 (.24)	27 (.36)	43 (.41)	96 (.45)
1,000-10,000	401 (.43)	792 (.24)	50 (.18)	125 (.10)	115 (.18)	155 (.23)	596 (.28)
>10,000	708 (.02)	869 (.02)	117 (.02)	251 (-.03)	236 (.01)	273 (-.05)	1581 (.01)
all users	112 (.73)	263 (.75)	12 (.70)	34 (.61)	28 (.69)	39 (.72)	134 (.76)

Note: Average values are shown in the table cells; Spearman rank correlation coefficients for each average value are shown in parentheses.