On the Evolution of User Interaction in Facebook

Bimal Viswanath[†] Alan Mislove[†‡] Meeyoung Cha[†] Krishna P. Gummadi[†]

[†]MPI-SWS [‡]Rice University

ABSTRACT

Online social networks have recently exploded in popularity; numerous sites allow users to interact and share content. Users of these networks have been observed to establish many hundreds or even thousands of friendship links. Recently, researchers have suggested examining the *activity network*, or the network formed by users who actually interact, as way of discriminating between the strong and weak links. While the initial studies of the activity networks have shown that activity networks are structurally different from the social network itself, the studies have disregarded an important and unique aspect of the activity network: the fact that the activity links can strengthen and weaken over time.

In this paper, we take a first look at the evolution of activity between users in the Facebook social network. We find that links in the activity network tend to come and go rapidly over time, and the strength of ties exhibit a general decreasing trend of activity as the social network link ages. For example, only 30% of the Facebook user pairs interact consistently from one month to the next. Interestingly, even as the links of the activity network change rapidly over time, many graph-theoretic properties of the activity network remain unchanged.

1. INTRODUCTION

Online social networks have become a popular way for users to connect, express themselves, and share content. Popular sites have hundreds of millions of registered users and are growing at a rapid pace. As these networks grow and mature, users have been observed to possess many hundreds or even thousands of friendship links. For example, previous studies have shown that the average user degree in Orkut is over 100 [8] and the average user degree in Facebook is over 120 [3]. In fact, we observed one user in Flickr who possessed over 25,000 friends [8]!

While most social networking sites provide only a binary state of friendship, it has been unsurprisingly observed that not all links are created equal. A recent study [5] has shown that the "strength of ties" varies widely, ranging from "best friends" to pairs of users who even wished they weren't friends. In order to distinguish between these strong and weak links, researchers have suggested examining the *activity network*, or the network formed by users who actually interact. While the initial studies of the activity networks [2, 10] have examined the static structure of the activity network, they have disregarded an important aspect of user interaction: the fact that the level of interaction between two users can vary over time.

In this paper, we collect detailed data on a large subset of the Facebook social network, encompassing both friendship relationships and user interactions. In total, we examined information on 60,000 users with over 800,000 interactions over a period of two years. We make our data set available to the research community in an anonymized form. We examine how user activity evolves over time, looking at both microscope and macroscopic properties of the activity network. We investigate how pairs of users interact, and then examine how the varying patterns of interactions affect the overall structure of the activity network.

Our analysis shows a number of interesting findings. We find that the activity of the pairs of users who interact infrequently tends to be triggered by site mechanisms. For example, we find that the over 40% of the activity of the infrequently-interacting user pairs is directly attributable to Facebook's birthday-reminder feature. For the pairs of users who interact frequently, we find that user activity shows a marked decrease over the life of the link, implying that links tend to die out over time. However, we find that more-frequently interacting pairs of users show this trend to a lesser degree, further emphasizing the strength of these links.

These findings together suggest that the underlying activity network is rapidly evolving. Thus, we also examine how the macroscopic properties of the activity network vary over time. Surprisingly, we find that while the individual user pairs that compose the activity network changes rapidly over time (i.e., the activity network contains only 30% of the same links from one month to the next), many of the graph-theoretic properties show remarkable stability over the course of two years.

The remainder of this paper is organized as follows. Section 2 describes our data collection methodology and introduces the dataset. Section 3 presents an analysis of the evolution of pairwise user interactions, discriminating between high and low rate interactions. Section 4 provides an analysis of the evolution of network-level properties of the activity network over time. Section 5 discusses implications of our findings. We discuss related work in Section 6 and conclude in Section 7.

2. DATA COLLECTED

In this section, we introduce our dataset and the relevant characteristics of Facebook. We collected data on (a) the social network of Facebook users, and (b) interaction activity between these users.

2.1 Measurement methodology

The subset of the Facebook network that we focus on is the New Orleans regional network. We chose to crawl a regional network (rather than, say, a university network) because we could easily create additional account to crawl with; university networks require email addresses from the university's domain in order to create an account. Thus, we created a number of Facebook accounts and joined the New Orleans regional network, which allowed us to view most of the profiles of the users in the same regional network.

We collected our data set over two separate periods. First, in order to collect the social network, we conducted a crawl from December 29th, 2008 to January 3rd, 2009, where we only collected friendship link information (e.g., who is friends with who). We started the crawl with a single user in the New Orleans network and then conducted a breadth-first-search (BFS) crawl of all visible users, in the same manner as previous work [8]. In order to conduct the crawl, we used HTML screen-scraping of Facebook's public network.

Second, in order to collect a user interaction trace, we focused on the *wall* feature in Facebook. A user's friends can post comments to the user's wall, and these comments appear in the user's wall and can be seen by other when they visit the user's profile. Thus, the Facebook wall represents a broadcast-style messaging service for Facebook users. We conducted our second crawl conducted from January 20th, 2009 to January 22nd, 2009, and downloaded the entire wall history for all users who we had previously discovered.

2.2 Collected data

In total, we crawled 90,269 users in the New Orleans network connected by 3,646,662 links. However, we were not able to view the walls of all these users; we were only able to view the walls of 60,290 (66.7%) users. For the remainder of the paper, we only consider this subset of the social network. In total, these 60,290 users are connected together by 1,545,686 links in the social network, for an average user degree of 25.3.

Facebook introduced the wall feature in September of 2006, thus, the wall post data we obtained ranges between September 26th, 2006 to January 22nd, 2009. Each wall post contains the identity of the wall owner, the user who made the post, the time at which the post was made, and the post content. In total, we observed 838,092 wall posts, for an average of 13.9 wall posts per user. We observed communication between 188,892 distinct pairs of users, representing 12.2% of the links in the social network. Thus, 87.8% of the links in the social network are not backed by any wall activity.

All the data in this paper is available to the research community in anonymized form. Description of data format and downloading instructions are available at
`http://socialnetworks.mpi-sws.mpg.de`

2.3 Limitations

Our measurement methodology also presents a few limitations and caveats. First, in Facebook, users can interact in many ways (e.g., by sending private messages, through applications, through photo uploads, and though a chat feature). While wall posting is one of the most popular methods of interaction between users, we do not know if it is representative of other forms of interaction. Second, our Facebook crawl was limited to the number of people who made their profiles visible to the people in the same regional network, which is the default setting. However, since this covers a majority of the network (66.7%), we expect the results to be representative. Finally, while our data set covers only a single Facebook regional network, a recent study that showed a similarity across multiple Facebook networks [10] gives us confidence that our findings will generalize to these other networks.

2.4 Wall activity

We now give a high-level overview of the characteristics of the wall posting activity. We first use the time stamps of wall posts to calculate how the number of wall posts grows over time. Figure 1 shows the number of wall posts over the entire trace period. We observe a steady increase until approximately the middle of 2008. While this sudden growth in mid-2008 seems abrupt, we found that Facebook rolled out a new site design on July 20, 2008 that allows users to more easily view wall posts through friend feeds [4]. We speculate that this change lead to the spike in wall posts towards the end of the trace.

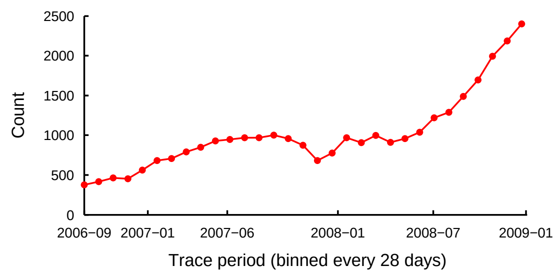

Figure 1: Growth in number of wall posts over time.

From the wall post data, we construct an *activity network*, which is an undirected graph where a link exists between users if they interacted at least once. Previous work [10] has shown that Facebook social links do not always represent real user interaction. We observe the same for our Facebook data set, as only 12.2% of the social network links are backed by any wall post activity. In more detail, Figure 2 shows the correlation between social network degree and activity network degree, with an error bar representing standard deviation. It can be clearly seen that users only interact over a small number of their social links.

3. PAIRWISE USER INTERACTIONS

In this section, we examine the patterns of wall post activity between pairs of Facebook users. Our goal is to understand the patterns of relationships backed by wall posts, based on the frequency and the pattern of wall posts exchanged between the user pairs. Towards this goal, we divide our analysis into two steps. First, we investigate the overall distribution of wall posts, looking to what extent certain

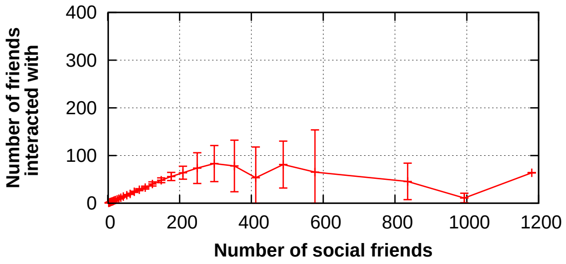

Figure 2: Correlation between the number of friends in the social network and the number of friends interacted with.

user pairs generate more posts than others. Second, we look in more detail at the patterns of communication of the low interacation user pairs and the high interaction user pairs.

3.1 Data used

In our trace, different user pairs formed their links and initiated their first interaction at different times. To make our analysis fair across user pairs with differing periods of activity, in this section, we focus only on the set of user pairs whom we were able to monitor for at least one year. Thus, we only consider user pairs for who (a) we could determine when the link was established, based on the wall post, and (b) created their link before January 22nd, 2008, and thus had at least one year of activity. In total, this set contains 59,916 (31.7%) of the user pairs. Finally, in order to compare this large set of links, we only consider the first year of activity for each of these links.

3.2 Distribution of wall activity

We first investigate the overall distribution of wall posts across links, looking at the extent to which certain links generate more activity than others. For this, we calculate the cumulative distribution function (CDF) of the number of wall posts per each user pair in Figure 3. We observe a significant skew in the distribution. The median number of posts per user pair is 2 and nearly 80% of the pairs exchange no more than 5 wall posts. However, some user pairs (<1%) exchanged more then 100 posts. In fact, the most active pair exchanged 767 total messages during the first year the link was established.

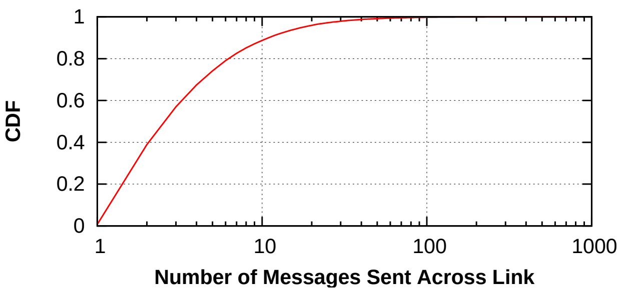

Figure 3: Distribution of the number of wall posts across user pairs.

For the remainder of this section, we take a closer look at the low interaction and high interaction links separately. User pairs who generated low rates of interaction may be casual acquaintances in Facebook, while user pairs who generated high rates of interaction may represent close friends. Thus, these two classes of users may show different patterns of activity. Following the observation above, we partition the links at the threshold of 5 wall posts in the first year of activity.

3.3 Patterns of low-rate interactions

We first examine the patterns of interaction for user pairs who sent no more than 5 wall posts across their link. This group represents a majority of the activity network, accounting for 48,689 (81%) of the user pairs. We are interested in understanding the reasons behind the low rate of interaction, and in particular, which events triggered messages to be sent across the link. Because this group shows an extremely low rate of traffic, we initially expected that these users will interact a few times after their social network link is established and then discontinue their interaction.

To investigate our hypothesis, we calculated the time between the establishment of the link and the time when the first message was sent. Figure 4 shows the CDF of this first-message-delay for all pairs in the low-rate interaction group. We find that 20% of the user pairs interacted the very same day they become friends in Facebook. However, the remaining pairs had an almost even distribution over the following year – in fact, more than half of the pairs exchanged their very first wall post message more than a month after they become friends! This demonstrates that most of the time, the interaction between these users was not triggered by the link creation event.

We then turned to investigate the cause of this unexpected distribution. For this, we examined the actual content of the wall messages in order to shed light on what triggered the interaction. We found that over 39% of the *first* wall posts contained either "birthday" or "bday", suggesting that they represented one user sending a birthday message to the other. Because the birthdays of Facebook users are likely to be spread out across the year, we expect that this underlies the smooth curve in the time to first interaction in Figure 4. In fact, messages related to birthday wishes are even more prevalent if we consider wall posts other than the first: 54% of the user pairs in low-rate interaction group exchanged at least one message related to birthday wishes. This phenomenon could be due to the fact that Facebook provides a friend birthday reminder feature when each user logs in, a mechanism which could trigger the interaction.

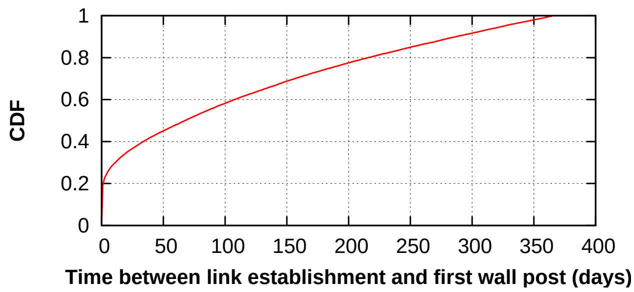

Figure 4: Distribution of time between link creation and the first wall post for pairs with low rates of interaction.

3.4 Patterns of high-rate interactions

Now we turn our focus to user pairs who sent more than 5 wall posts. In this group, the average number of wall posts exchanged is 16.2 and the median number of wall posts is 10. In contrast to the low-rate interaction group, user pairs in this group exhibit very different interaction patterns. The median duration for time to first interaction is 6 days, compared to 74 days in low-rate interaction group. This means that most users in the high-rate interaction group exchange their first wall message soon after friendship links are established. Put in another way, these users did not need a specific event (e.g., birthday) to initiate their conversation.

We first examine how wall posts are spread across the first year of activity for the high-rate interaction group. Are interactions evenly spread out in time? Does the overall interaction pattern shows an increase or decrease over time? We analyze the average rate of the user activity over time in Figure 5, which plots the fraction of total wall messages that are sent in each month, out of the total wall messages exchanged between each user pair. Figure 5 plots this value for a number of different subsets of the active users: pairs who exchanged more than 5, 10, 20, 30, and 100 wall messages.

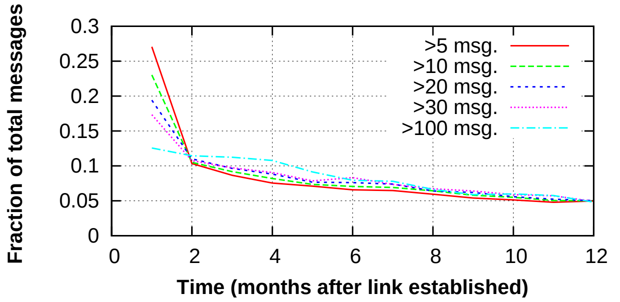

Figure 5: Aggregate user activity over a period of 1 year. Activity is binned by months. Y-axis shows the fraction of total messages sent in a particular month(averaged over all links). The plot is shown for user pairs with more than 5, 10, 20, 30 and 100 interactions.

We make two observations. First, the level of the pairwise user interaction is highest right after link establishment and decays over time. A detailed look into the trace shows within the first month, the interaction also drops largely by the week. This trend contrary to what we observed in the low-rate interaction group, where we find that the wall posts are more evenly spread across time because these wall posts are often driven by external events (e.g., birthdays). Second, we observe that user pairs who denote higher activity (e.g., user pairs with over 100 wall messages exchanged) show a more even distribution than the rest.

The dropping of wall post frequency in Figure 5 implies a general trend for users to reduce their amount of activity over time. Thus, we examine how long the interaction across each of these links lasts, focusing on how often activity ceases across a link (meaning the link is removed from the activity network). Figure 6 shows the fraction of user pairs that no longer interact after varying numbers of months. We again examine users with differing thresholds of activity. Across all activity bins, the largest fraction of user pairs discontinue their conversation after the first month of interaction, as expected. Interestingly, this is also true for pairs who denote higher activity (>100 wall messages). However, the drop rate decreases over time for all activity bins. This indicates a reinforcing relationship between the duration of interaction and the probability of continued interaction: the more one has engaged in an interaction, the more likely it is to continue. In total, 23% of the user pairs in the high-rate interaction group stayed active during the entire one year period.

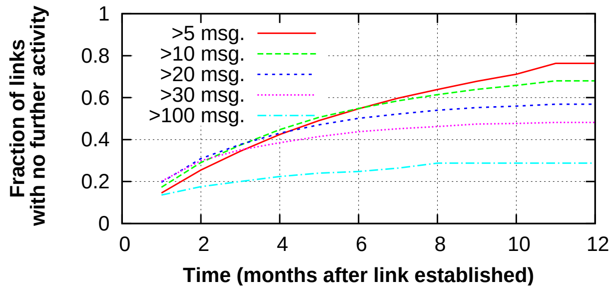

Figure 6: Fraction of links with no further activity after a given number of months.

3.5 Summary

In this section we characterized the patterns of interaction between pairs of users in Facebook. We observed a strong skew in the total number of wall posts exchanged between user pairs; a majority of the user pairs denoted very little activity. To better understand the behaviors of both the active and relatively inactive user pairs, we grouped user pairs into classes and characterized their interaction patterns in detail. The low-rate interaction group typically takes more than a month to initiate their wall interaction since the time they become friends, and often the ice-breaker was a popular common event (e.g., birthday). In contrast, the high-rate interaction group shows very different pattern. They tend to initiate conversation immediately after becoming friends and shows a decaying pattern in their rate of conversation. As a result, only very few user pairs continue to be active after a long period of time.

Thus, we observed that both the low-rate and the high-rate interaction groups showed significant variability, implying that the links in the activity network change rapidly. In the next section, we examine how this impacts the structure of the activity network over time.

4. NETWORK EVOLUTION OVER TIME

So far we examined the high-level characteristics of wall post activity in Facebook compared to size of the social network (Section 2) and the patterns of pairwise user interactions (Section 3). We made two key observations. First, while individuals have a large number of friends, they interact with very few of their friends. Second, on average, interactions between users tend to change rapidly over time. These findings indicate that the network of wall activity is highly dynamic in that links quickly come and go. In this section, we focus on understanding how the global structure of the activity network changes over time. In particular, we are interested in capturing (a) what fraction of the links change state from being active to inactive (or vice versa) and (b) to what extent overall activity network properties change over time.

We study evolution of the activity network using multiple snapshots of the network taken at different times. We generated snapshots of the activity network based on the wall posts at 90 day intervals during the trace period. In total, we generated 10 snapshots of the activity network between September 2006 to January 2009. Thus, each snapshot contains links for all pairs of users who exchanged at least one wall post during the 90 day period. Below, we examine the evolving activity network based on these 10 snapshots.

4.1 Structural change over time

First, we examine how many activity network links persist from one snapshot to the next. To measure the overlap in network links in two consecutive snapshots, we use the notion of *resemblance*. Resemblance is used to measure the quantitative overlap between two sets. In our context, resemblance is defined as the proportion of the network links that remain unchanged over two network snapshots. Resemblance $\mathcal{R}_t$ at time t is defined as

$$\mathcal{R}_t = |\frac{P_t \cap P_{t+1}}{P_t}|, \quad (1)$$

where P_t is the set of user pairs who interact at time t. The value of $\mathcal{R}_t$ varies between 0 and 1. If $\mathcal{R}_t$=1, then the entire set of pairs continued to interact at the next time step. Conversely, if $\mathcal{R}_t$=0, then it indicates that none of the user pairs who interacted in time t interacted in time $t+1$.

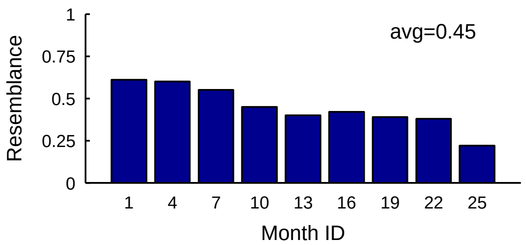

Figure 7: Resemblance of the activity network

Figure 7 shows the resemblance values between consecutive snapshots. The average resemblance across all snapshots is 0.45, indicating that, on average, 45% of the links remain active over time. Conversely, the remaining 55% of the links did not exist in the consecutive snapshot. Moreover, there is a general downward trend, whereby the final pair of snapshots only have a resemblance of 0.30, indicating that the activity network is becoming even more volatile over time.

The resemblance value is sensitive to the size of the window. When we use a smaller window of 30 days to generate activity network snapshots, the average resemblance is 0.29; 70% of the links refresh over time. However, even when we use a much larger window of 180 days, resemblance value only increased marginally to 0.53. This demonstrates that the links in the activity network changes dynamically even over a large time scale.

4.2 Network properties over time

We now examine to what extent this rapid fluctuation in the activity network links affect the overall network structure. To do so, we calculate network metrics across our 10 network snapshots. Consistent with prior work on social network analysis [8], we focus on three quantities: average node degree, clustering coefficient,[1] and average path length.[2] In each of the snapshots, the largest connected component typically accounted for 86% of the entire activity network.

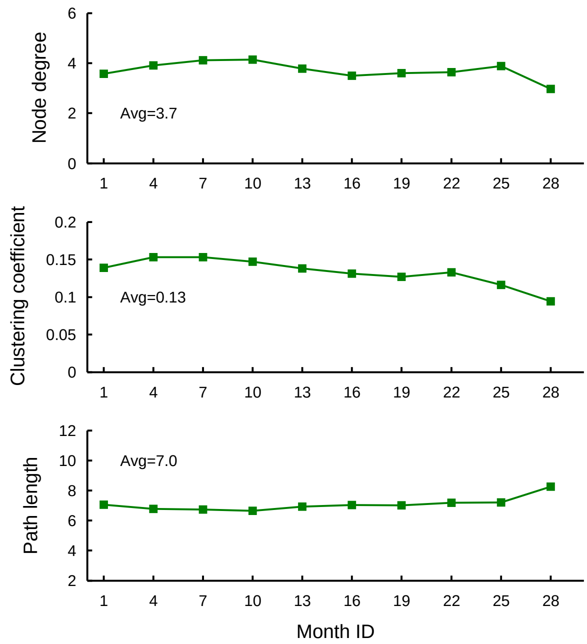

Figure 8: Network-level properties over time

Figure 8 presents how each of these three metrics vary over the consecutive snapshots. Surprisingly, the different network measures all exhibit relative stability over time. There is some fluctuation towards the end of the trace, which may be influenced by Facebook's launch of new web design in July 2008 (i.e., month 22).

Except for the last three months, the three network measures remain relatively stable over time. Combined with our finding about the fast refresh rate of the activity network, our results on the relatively stable network properties indicate a rather surprising pattern of network evolution. The set of active links change rapidly over both a short and a large time scale, while at any given period of time, the activity network exhibits strikingly constant structural properties!

4.3 Summary

In this section, we analyzed the patterns of the evolving activity network from two perspectives: the refresh rate of links (or the persistence of the user pairs over time) and the network-level properties. We found that while the individual user pairs that consist the activity network changes rapidly over time (i.e., instability in organizational structure), average network properties remained strikingly stable.

[1]The clustering coefficient quantifies to what extent a graph is a clique (i.e., complete graph). This quantity varies between 0 and 1, where 0 represents singleton(s) and 1 represents a complete graph.

[2]Because it is computationally expensive to find shortest paths between all user pairs, we randomly picked 1,000 user pairs that belong to the largest component of the activity network and computed their average path length.

5. DISCUSSION

One of the primary concerns about online social networks is the extent to which links in the social network represent actual trust relationships between users. The fact the many social networks contain users with hundreds or thousands of friends only makes this concern more worrisome. However, since online social networks are used to exchange information between friends, the patterns of activity offer a potential way to distinguish the strong links (representing users who interact often) from the weak links (representing users who become friends out of courtesy). Our work is the first to examine exactly how users are interacting on these sites.

One of the interesting findings of our study is that the mechanisms present on the online social networking sites can dramatically affect the activity network in unexpected ways. For example, we found that Facebook's birthday-reminder feature caused much of the activity that we observed, implying that this activity may not have occurred had that feature not been in place. This suggests that, going forward, one may have to examine the content and the cause of interaction in order to determine the nature of the link strength between users.

Many systems have been recently proposed which leverage the properties of social networks in various ways. For example, SybilGuard [12] and SybilLimit [11] use the difficulty of forming social network links to prevent Sybil attacks, and Ostra [9] uses the same property to block users who send large amounts of unwanted communication. However, one of the primary concerns with these systems is whether a social network exists which has the requisite difficulty in establishing links.

The activity network would seem like a natural fit to define a social network for such systems. Since a link in the activity network requires users to actually interact, malicious users cannot maintain an arbitrary number of links or obtain new links arbitrarily fast. Our findings strengthen this argument, showing that users tend to interact with a small number of other users over time and that the activity network shows remarkable stability in high-level properties. Moreover, our study provides a better understanding of the real interaction that takes place between users in the network, which could help designers of such systems to make the choice of which network to base their socially-enhanced applications on.

6. RELATED WORK

Recently, much work has focused on understanding the structure and evolution of large-scale online social networks [1, 6, 8]. Though this has thrown light on the high level growth and topological characteristics of social network, they do not examine how the social network links were being used by users to interact. This has lead to a new direction of research focusing on the activity network.

Hyunwoo et al. studied the activity network from the guest book logs of Cyworld online social network [2]. They calculated the activity network from the comments posted by users in each others' guestbooks. They observed that the structure of the activity network was similar to the social network graph and that user interactions tended to be bidirectional. However, Christo et al. [10] used user interactions to demonstrate the opposite trend in Facebook, where the structure of activity network differed significantly from the social network. For example, authors show that the activity network did not have "small world" properties. Leskovec et al. [7] analyze online interaction patterns among users in a large instant messaging trace found that the average path length was significant higher. The authors also discovered that the interaction network displayed strong influence of homophily in interaction, where similars users interact more often.

While all of these studies have examined the activity network in static form, our study is the first to examine the patterns of interaction, or how users are interacting over time and how the activity network is evolving.

7. CONCLUSION

In this paper, we studied the interaction between users in the Facebook online social network. We collected data on static friendship links, and activity data based on wall posts for a large subset of the Facebook network. Our analysis showed that there is a significant skew in the distribution of activity across links, meaning a minority of the user pairs generate a majority of the activity. We observed a general decay in the amount of interaction between pairs of users, suggesting that the activity network is rapidly changing over time. Analysis of the activity network as a whole revealed that even though there is a high churn in the user pairs that interact over time, many of the global network properties remained relatively constant over time.

8. REFERENCES

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of Topological Characteristics of Huge Online Social Networking Services. In *WWW*, 2007.

[2] H. Chun, H. Kwak, Y.-H. Eom, Y.-Y. Ahn, S. Moon, and H. Jeong. Online Social Networks: Sheer Volume vs Social Interaction. In *ACM IMC*, 2008.

[3] Facebook Statistics. `http://www.facebook.com/press/info.php?statistics`.

[4] Facebook Unveils Next Evolution of Site Design. `http://www.facebook.com/press/releases.php?p=47448`.

[5] E. Gilbert and K. Karahalios. Predicting Tie Strength With Social Media. In *CHI*, 2009.

[6] R. Kumar, J. Novak, and A. Tomkins. Structure and Evolution of Online Social Networks. In *ACM KDD*, 2006.

[7] J. Leskovec and E. Horvitz. Planetary-Scale Views on a Large Instant-Messaging Network. In *WWW*, 2008.

[8] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *ACM IMC*, 2007.

[9] A. Mislove, A. Post, K. P. Gummadi, and P. Druschel. Ostra: Leveraging Trust to Thwart Unwanted Communication. In *Usenix NSDI*, 2008.

[10] C. Wilson, B. Boe, A. Sala, K. P. N. Puttaswamy, and B. Y. Zhao. User Interactions in Social Networks and their Implications. In *ACM Eurosys*, 2009.

[11] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao. SybilLimit: A Near-Optimal Social Network Defense against Sybil Attacks. In *IEEE Symposium on Security and Privacy*, 2008.

[12] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilGuard: Defending Against Sybil Attacks via Social Networks. In *ACM SIGCOMM*, 2006.