User Interaction in OSNs

User behavior and interactions in online social networks

Yoonji Shin
Motivation
Why do we study user interactions in OSNS?

Socially Enhanced Applications
"Use social links make better applications"

Socially Enhanced Applications are emerging class of applications that leverage relationships from social networks to improve security and performance in applications such as email, web browsing and overlay routing.
Motivation
Why do we study user interactions in OSNS?

"But, do they really work?"

**Questioning effectiveness of Social Links**
Researchers in psychology and sociology have repeatedly cast doubt on the practice of inferring meaningful relationships from social network connections alone.
Motivation
Why do we study user interactions in OSNS?

Q. Are social links valid indicators of real user interaction?

Q. Can we quantify these factors to form a more accurate model for evaluating socially enhanced applications?

A. Use Interaction Graphs!
Motivation
Why do we study user interactions in OSNS?

Misleading Assumption
Unfortunately, these socially enhanced applications assume that all social links denote a uniform level of real-world interpersonal association, an assumption disproven by social science.

Familiar Stranger
The prevalence of low-interaction social relationships such as Milgram's "Familiar Stranger" [Milgram 1977] shows that uniform-level interpretation of all social links is misleading.
Method
How we studied user interactions in OSNs

1. Analyze interaction graphs derived from Facebook user traces.
2. Observe the levels of "small-world" properties shown
3. Compare with other other social network models
4. Analyze the number of "supernodes" and the overall network diameter.
5. Propose a new model
Study
Analyze interaction graphs from Facebook

1. Analyze interaction graphs derived from Facebook user traces.
2. Observe the levels of "small-world" properties shown.
3. Compare with other other social network models.
4. Analyze the number of "supernodes" and the overall network diameter.
5. Propose a new model.
Study

Analyze interaction graphs from Facebook

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>London, UK</td>
<td>1,241K (50.8)</td>
<td>30,725K (26.5)</td>
<td>11</td>
<td>15</td>
<td>5.09</td>
<td>0.170</td>
<td>0.23</td>
</tr>
<tr>
<td>Australia</td>
<td>1,215K (61.3)</td>
<td>121,271K (71.4)</td>
<td>10</td>
<td>14</td>
<td>5.13</td>
<td>0.175</td>
<td>0.17</td>
</tr>
<tr>
<td>Turkey</td>
<td>1,030K (55.5)</td>
<td>42,799K (56.7)</td>
<td>13</td>
<td>17</td>
<td>5.10</td>
<td>0.133</td>
<td>0.06</td>
</tr>
<tr>
<td>France</td>
<td>728K (59.3)</td>
<td>11,219K (34.6)</td>
<td>10</td>
<td>13</td>
<td>5.21</td>
<td>0.172</td>
<td>0.11</td>
</tr>
<tr>
<td>Toronto, ON</td>
<td>483K (41.9)</td>
<td>11,812K (21.9)</td>
<td>10</td>
<td>13</td>
<td>4.53</td>
<td>0.158</td>
<td>0.21</td>
</tr>
<tr>
<td>Sweden</td>
<td>575K (68.3)</td>
<td>17,287K (44.8)</td>
<td>8</td>
<td>11</td>
<td>4.55</td>
<td>0.157</td>
<td>0.18</td>
</tr>
<tr>
<td>New York, NY</td>
<td>378K (45.0)</td>
<td>7,225K (15.7)</td>
<td>11</td>
<td>14</td>
<td>4.80</td>
<td>0.146</td>
<td>0.18</td>
</tr>
<tr>
<td>Colombia</td>
<td>565K (71.7)</td>
<td>10,242K (31.7)</td>
<td>9</td>
<td>12</td>
<td>4.94</td>
<td>0.136</td>
<td>0.08</td>
</tr>
<tr>
<td>Manchester, UK</td>
<td>395K (55.5)</td>
<td>11,120K (35.2)</td>
<td>11</td>
<td>15</td>
<td>4.79</td>
<td>0.195</td>
<td>0.21</td>
</tr>
<tr>
<td>Vancouver, BC</td>
<td>314K (45.1)</td>
<td>35,518K (59.3)</td>
<td>9</td>
<td>14</td>
<td>4.71</td>
<td>0.170</td>
<td>0.23</td>
</tr>
<tr>
<td>Total/Average [Std. Dev.]:</td>
<td>10,697K (56.3)</td>
<td>408,265K (43.3)</td>
<td>9.8 [1.34]</td>
<td>13.4 [1.84]</td>
<td>4.8 [0.41]</td>
<td>0.164</td>
<td>0.17 [0.07]</td>
</tr>
<tr>
<td>Orkut [Mislove 2007]</td>
<td>1,846K (26.9)</td>
<td>22,613K</td>
<td>6</td>
<td>9</td>
<td>4.25</td>
<td>0.171</td>
<td>0.072</td>
</tr>
</tbody>
</table>

Table 1. High level statistics and social graph measurements for the ten largest regional networks in our Facebook data set.

Facebook: 500 million users, biggest social network, biggest photo sharing
Study was done by crawling wall & photo comments
Crawled roughly 10 million users from 22 largest regional network
**Study**

*Analyze interaction graphs from Facebook*

---

**Figure 1.** Comparing social degree in Facebook to those of Orkut, YouTube and LiveJournal.

---

**Social Degree Analysis**

We compare the social degree of Facebook to other social networks.

As expected, social degrees on Facebook follows power law.

Connectivity among Facebook users most closely resembles those of users in Orkut, likely because both are sites primarily focused on social networking. In contrast, YouTube and LiveJournal are content distribution sites with social components, and exhibit much lower social connectivity.
Study

Analyze user interaction from Facebook

**Figure 5.** Normalized Wall post distribution of the users with top total Wall interaction.

**Figure 6.** Normalized photo comments distribution of the users with top total photo interaction.

User Interaction Analysis

The goal of our analysis of Facebook user interactions is to understand how many social links are actually indicative of active interactions between the connected users.

We first examine the difference in size between interaction graphs and social graphs for users in our data set. We compute for each user a distribution of the user’s interaction events across the user’s social links. We then select several points from each distribution (70%, 90%, 100%) and aggregate across all users the percentage of friends these events involved.
Study
Analyze user interaction from Facebook

**Figure 7.** The contribution of different users to total interactions in Facebook.

**Figure 8.** Plot of top % of users ordered by social degree and the interaction contributed by them.

**Distribution of User Activities**

For both Wall posts and photo comments, we plot the contribution of different users sorted by each user’s interaction in that application.

The bulk of all Facebook interactive events are generated by a small, highly active subset of users, while a majority of users are significantly less active, which lends credence to our assertion that not all social links are equally useful when analyzing social networks, since only a small fraction of users are actively engaged with the network.
Study

Interaction Graph

Figure 11. Deviations in pairwise interaction patterns on Facebook.

Interaction Graph

To reasonably model directed Facebook interaction events as an undirected interaction graph, we must first demonstrate that pairwise sets of social friends perform reciprocal interactions with each other.

Figure 11 shows the length of the set resulting from the symmetric set difference of each user’s incoming and outgoing interaction partners plotted.
Study
Interaction Graph

**Figure 12.** Percentage of nodes remaining in interaction graphs WCC as n and t vary.

**Interaction Graph**

Figure 12 shows the size of the weakly-connected components for interaction graphs as t and n change.

Larger t and lower n are less restrictive on links, therefore allowing for more nodes to remain connected. Based on Figure 12, we chose several key interaction graphs for further study, including those with n > 1 at the 1 year, 6 months, and 2 months time periods.
Study
Interaction Graph Analysis

Figure 14. Graph measurements for four interaction graphs compared to the entire Facebook social network.

Interaction Graph Analysis

Figure 14 (b) shows the average radius, diameter, and path lengths for all of the interaction graphs, as well as for the social network. These measures all display the same upward trend as the interaction graphs become more restricted.

As the average number of links per node and the number of high-degree “super-nodes” decreases (see Figure 15) the overall level of connectivity in the graph drops.
Interaction Graph Analysis

As the average number of links per node and the number of high-degree “super-nodes” decreases (see Figure 15) the overall level of connectivity in the graph drops. This causes average path lengths to rise.
RE: Reliable Email

"RE" [Garriss 2006] is a white-listing system for email based on social links that allows emails between friends and Friends-of-Friends (FoFs) to bypass standard spam filters.

The presence of small-world clustering and scale-free behavior in social graphs translate directly into short average path lengths between nodes. For RE, this means that the set of friends and FoFs that will be white-listed for any given user is very large.

In this situation, a single user who sends out spam email is likely to be able to successfully target a very large group of recipients via the social network.
Apply

Applying Interaction Graphs

![Graph showing spam penetration vs. percentage of users spamming for different interaction graphs.]

**Figure 17.** Spam penetration as the number of spammers is varied for the Reliable Email [Garriss 2006] system.

**RE: Reliable Email**

For social graph and interaction graphs, we randomly choose a percentage of nodes to act as spammers. In the RE system, all friends and FoFs of the spammer will automatically receive the spam due to white-listing.

Figure 17 plots the percentage of users in each graph receiving spam versus the percentage of users who are spamming. In contrast, spam penetration is reduced by 40% over the social graph when the number of spammers is low, and 20% when the number of spammers is high when RE is run on the interaction graphs.
Contributions
Contributions of the study

1. The first large-scale study of Facebook social network

Users tend to interact mostly with up to 50% of their Facebook friends, which casts doubt on the practice of extracting meaningful relationship from social graphs, and suggests an alternative model for validating user relationships in OSNs.
Contributions
Contributions of the study

1. The first large-scale study of Facebook social network

2. Introducing "interaction graph"

Interaction graph is a model for representing user relationships based on user interactions.

Interaction graph contains all nodes from its social graph counterpart, but only a subset of links. This link exists if and only if its connected users have interacted directly through communication or an application.

Shows larger network diameters, lower clustering coefficients, and higher assortativity.
Contributions
Contributions of the study

1. The first large-scale study of Facebook social network

2. Introducing "interaction graph"

3. Showing the impact of using different graph models in evaluating socially-enhanced applications.

Conducted simulated experiments of the Reliable Email [Garriss 2006] and SybilGuard [Yu 2006] on both social and interaction graphs, and demonstrated that difference in the two graph models translate into significantly different performance results.