Computing with Social Networks on the Web

Jennifer Golbeck
College of Information Studies
University of Maryland, College Park
Web-Based Social Networks

• What are they?
• How do they grow and change?
• Challenges
• Applications
Web-Based Social Networks

• What are they?
• How do they grow and change?
• Challenges
• Applications
What are Web-based Social Networks

- Websites where users set up accounts and list friends
- Users can browse through friend links to explore the network
- Some are just for entertainment, others have business/religious/political purposes
- E.g. MySpace, Friendster, Orkut, LinkedIn
Growth of Social Nets

- *The* big web phenomenon
- About 230 different social networking websites
- Over 675,000,000 user accounts among the networks
- Number of users has more than doubled in the last 18 months
- Full list at [http://trust.mindswap.org](http://trust.mindswap.org)
<table>
<thead>
<tr>
<th>Rank</th>
<th>Network</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MySpace</td>
<td>250,000,000</td>
</tr>
<tr>
<td>2</td>
<td>ChinaRen Xiaonei</td>
<td>60,000,000</td>
</tr>
<tr>
<td>3</td>
<td>Orkut</td>
<td>60,000,000</td>
</tr>
<tr>
<td>4</td>
<td>Friendster</td>
<td>53,000,000</td>
</tr>
<tr>
<td>5</td>
<td>zoominfo</td>
<td>35,000,000</td>
</tr>
<tr>
<td>6</td>
<td>Adult Friend Finder</td>
<td>26,000,000</td>
</tr>
<tr>
<td>7</td>
<td>Bebo</td>
<td>25,000,000</td>
</tr>
<tr>
<td>8</td>
<td>Facebook</td>
<td>24,000,000</td>
</tr>
<tr>
<td>9</td>
<td>Cyworld</td>
<td>21,000,000</td>
</tr>
<tr>
<td>10</td>
<td>Tickle</td>
<td>20,000,000</td>
</tr>
</tbody>
</table>
Structure of Social Nets

• Small World Networks
  – AKA Six degrees of separation (or six degrees of Kevin Bacon)
  – Term coined by Stanley Milgram, 1967

• Math of Small Worlds
  – Average shortest path length grows logarithmically with the size of the network
  – Short average path length
  – High clustering coefficient (friends of mine who are friends with other friends of mine)
Web-Based Social Networks

• What are they?
• **How do they grow and change?**
• Challenges
• Applications
Behavior and Dynamics

• Social networks are not static.
  – Relationships constantly change, are formed, and are dropped.
  – New people enter the network and others leave

• Do people behave the same way in social networks on the Web?
Questions

• How do these networks grow (and shrink)?
• How are relationships added (and removed)?
• What affects social disconnect?
• What affects centrality?
Methodology

• 24 month study
• Automatically collected adjacency lists (everyone and who they know), join dates, and last active dates for all members.
  – December 2004
  – December 2006
• For 7 networks, I collected adjacency lists every day for 7 weeks.
  – Who joined or left
  – What relationships were added or removed
Network Growth

• People do not leave social networks
  – On sites with a clear simple process, less than a dozen members leave per day
  – In some networks, essentially no one has ever left

• Lots of people join social networks
  – For ten networks we knew the date that every member joined the network
  – Networks tend to show linear growth
  – The slope can shift
    • Usually occurs suddenly
    • Explained by some event
Relationships

- Forming relationships is the basis for social networking
- Almost all networks are growing denser
  - Relationships grow at approximately 1.7 - 2.7 times the rate of membership
- There is a strong social disincentive to remove relationships

<table>
<thead>
<tr>
<th>Network</th>
<th>Removed</th>
<th>Added</th>
<th>Net Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buzznet</td>
<td>6,418</td>
<td>135,158</td>
<td>128,740</td>
</tr>
<tr>
<td>Ecademy</td>
<td>1,774</td>
<td>26,751</td>
<td>24,977</td>
</tr>
<tr>
<td>Fototthing</td>
<td>2,059</td>
<td>11,314</td>
<td>9,255</td>
</tr>
<tr>
<td>Tribe</td>
<td>11,211</td>
<td>180,936</td>
<td>169,725</td>
</tr>
<tr>
<td>FilmTrust</td>
<td>1</td>
<td>399</td>
<td>398</td>
</tr>
</tbody>
</table>
Friendless and the Outsiders

- Friendless have no social connections
- Outsiders have social connections but are independent from the major connected component of the network
- Important because if we are using the social network for information access, these people will get little benefit.
Centrality

• Other than having lots of friends, what makes people more central?
  – Average shortest path length as centrality measure

• Activity
  – Consider join date, last active date, and length of activity (last active date - join date)
  – Compute rank correlation with centrality
  – Medium strength correlation (~0.5) between duration and centrality
Results

• Networks follow a linear growth pattern, where the slope shifts in response to events
  – People rarely leave networks
• Networks grow denser, with relationships added more frequently than members
  – People will delete relationships, but orders of magnitude less frequently than they add them
• Websites with more non-social features tend to have more friendless and disconnected users
• Users with longer periods of activity tend to be more central to the network
Web-Based Social Networks

• What are they?
• How do they grow and change?
• **Challenges**
• Applications
Challenges

• Aggregation
  – People have accounts in multiple places
  – What if we want to see all that data together

• Size
  – Scalability of algorithms is important when working with hundreds of millions of nodes in a graph
Social Networks on the Semantic Web

- FOAF (Friend Of A Friend)
  - A simple ontology for representing information about people and who they know
- About 20,000,000 social network profiles are available in FOAF format
- Approximately 60% of all semantic web data is FOAF data
FOAF for Aggregation

- Semantic Web Vocabulary for describing people and social networks
- Automatically generated by many social networking websites
  - Advocato
  - Buzznet
  - DeadJournal
  - eCademy
  - FilmTrust
  - GreatestJournal
  - InsaneJournal
  - LiveJournal
  - LJ.Rossia.org
  - Minilog.com
  - Tribe
<foaf:Person>
   <foaf:nick>golbeck</foaf:nick>
   <foaf:name>golbeck</foaf:name>
   <foaf:mbox>4d14fc9da1d0929dae3cde648ae4a7195d120bae</foaf:mbox>
   <foaf:openid rdf:resource="http://golbeck.livejournal.com/">
   </foaf:openid>
   <foaf:page>
      <foaf:Document rdf:about="http://golbeck.livejournal.com/profile">
         <dc:title>LiveJournal.com Profile</dc:title>
         <dc:description>Full LiveJournal.com profile, including information such as interests and bio.</dc:description>
      </foaf:Document>
   </foaf:page>
   <foaf:weblog rdf:resource="http://golbeck.livejournal.com/">
</foaf:weblog>
</foaf:Person>
Semantics of FOAF

- Inverse Functional Properties
  - foaf:aimChatID
  - foaf:homepage
  - foaf:icqChatID
  - foaf:jabberID
  - foaf:mbox
  - foaf:mbox_sha1sum
  - foaf:msnChatID
  - foaf:weblog
  - foaf:yahooChatID

- Two people who share a common value for one of these properties are inferred to be the SAME person
Do People Have Multiple Accounts?

- FOAF is fine and good, but can we take advantage of the reasoning to merge networks?
- Yes!
Table 1: The social networks used in this study, including the average shortest path length pre and post-reasoning.

<table>
<thead>
<tr>
<th>Network</th>
<th>Purpose</th>
<th>Members Studied</th>
<th>Avg. Degree</th>
<th>APL (Pre)</th>
<th>APL (Post)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advocato</td>
<td>Business</td>
<td>2,778</td>
<td>13.51</td>
<td>2.17</td>
<td>2.15</td>
</tr>
<tr>
<td>Buzznet</td>
<td>Photos</td>
<td>208,324</td>
<td>1.00</td>
<td>4.43</td>
<td>2.76</td>
</tr>
<tr>
<td>DeadJournal</td>
<td>Blogging</td>
<td>9,801</td>
<td>3.74</td>
<td>3.19</td>
<td>3.23</td>
</tr>
<tr>
<td>eCademy</td>
<td>Business</td>
<td>61,242</td>
<td>3.08</td>
<td>2.20</td>
<td>2.19</td>
</tr>
<tr>
<td>FilmTrust</td>
<td>Social/Entertainment</td>
<td>1,250</td>
<td>1.06</td>
<td>3.75</td>
<td>3.84</td>
</tr>
<tr>
<td>GreatestJournal</td>
<td>Blogging</td>
<td>36,862</td>
<td>33.36</td>
<td>2.25</td>
<td>2.31</td>
</tr>
<tr>
<td>InsaneJournal</td>
<td>Blogging</td>
<td>1,410</td>
<td>13.36</td>
<td>3.19</td>
<td>3.26</td>
</tr>
<tr>
<td>LiveJournal</td>
<td>Blogging</td>
<td>3,563,267</td>
<td>8.38</td>
<td>2.85</td>
<td>2.83</td>
</tr>
<tr>
<td>Minilog.com</td>
<td>Blogging</td>
<td>119</td>
<td>1.63</td>
<td>3.66</td>
<td>3.66</td>
</tr>
<tr>
<td>Rossia.org</td>
<td>Blogging</td>
<td>4,180</td>
<td>9.65</td>
<td>2.33</td>
<td>2.36</td>
</tr>
<tr>
<td>Tribe</td>
<td>Social/Entertainment</td>
<td>218,694</td>
<td>9.93</td>
<td>2.74</td>
<td>2.69</td>
</tr>
<tr>
<td>Networks</td>
<td>Advocato</td>
<td>Buzznet</td>
<td>DeadJournal</td>
<td>eCademy</td>
<td>FilmTrust</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Advocato</td>
<td>x</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Buzznet</td>
<td></td>
<td></td>
<td>x</td>
<td>53</td>
<td>89</td>
</tr>
<tr>
<td>DeadJournal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>eCademy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FilmTrust</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GreatestJournal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>InsaneJournal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiveJournal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rossia.org</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minilog.com</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tribe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Web-Based Social Networks

• What are they?
• How do they grow and change?
• Challenges
• Applications
Information Access

• Aggregate, Sort, Filter information
• A user’s social relationships inform what they want to see and how important it is
• Use data from web-based social networks to build intelligent applications
• My focus is specifically on trust
Example: Filtering

![Image of TrustMail software interface]

The image shows the TrustMail software interface with a list of emails in the inbox. The email subjects and senders are visible, and the interface includes options for managing emails, such as moving them to the outbox or sent items folder.
Example: Aggregating

• If we have numeric data, a simple average only shows what the overall population thinks
  – Not so useful, e.g. politics

• What if we weight the values in the average by how much we trust the people who created them?

• FilmTrust
  – http://trust.mindswap.org/FilmTrust
A Clockwork Orange (1971)

User Options

Your Rating: 7

You have reviewed this film:
I read the book before seeing the movie, and the book had a point ...
Edit Review - Delete Review

Ratings of A Clockwork Orange (1971)

Number of Ratings 266
Average User Rating 3
Your Recommended Rating 7
Your Rating 7

Reviews of A Clockwork Orange (1971)

I read the book before seeing the movie, and the book had a point. There was a deep social commentary there, amidst the violence and torture. The movie, however, lost much of the meaning of the book. The violence was not justified by the plot, and the message was left very shallow.

A lot of people who love this movie love it because they know it is supposed
Example: Sorting

- When many users create information, we want to see the data from people like us first
Reviews of *A Clockwork Orange* (1971)

I read the book before seeing the movie, and the book had a point. There was a deep social commentary there, amidst the violence and torture. The movie, however, lost much of the meaning of the book. The violence was not justified by the plot, and the message was left very shallow.

A lot of people who love this movie love it because they know it is supposed to be deep and important, not because there is actually a lot to love here. Read the book instead.
- by Jen Golbeck

This movie sucked! It was probably the worst movie I have ever seen. As I left the theatre, I remember thinking, "what was that all about?"
- by john golbeck

Absolutely the worst movie ever. Very weird, hard to follow and disturbing. Perhaps if the book had been followed better it would have been more tolerable to sit through.
- by irene golbeck

One of Stanley Kubrick's science fiction classics. Staring Malcolm McDowell as "Alex", a young man whose principle interests are rape, ultra-violence, and Beethoven.

The first half of this film is a disturbing no-holds-barred look at the life of Alex. Kubrick seems to want to make the audience dislike the principle character, with little sympathy for what will ultimately happen to him.

When Alex goes to jail, he appears unrepentent and determined to subvert the system. Instead, he is subjected to a treatment to "cure" him of his ways. After his cure and subsequent release, his past revisits him, causing the audience to
tolerable to sit through
- by irene

One of Stanley Kubrick's science fiction classics. Staring Malcolm McDowell as "Alex", a young man whose principle interests are rape, ultra-violence, and Beethoven.

The first half of this film is a disturbing no-holds-barred look at the life of Alex. Kubrick seems to want to make the audience dislike the principle character, with little sympathy for what will ultimately happen to him.

When Alex goes to jail, he appears unrepentent and determined to subvert the system. Instead, he is subjected to a treatment to "cure" him of his ways. After his cure and subsequent release, his past revisits him, causing the audience to question the attitudes they have formed about this character.

This is a dark and cynical film about the way society treats its fringe elements. It contains numerous dichotomies, including an incredibly funny sex scene, as well as the most disturbing rape scene I've ever witnessed. The violence is stylized rather than authentic, but Kubrick makes great use of psychological elements with a much greater effect than the highly explicit elements used in more modern films.
- by Paul Gearon

Notwithstanding the movie's poor showing in the Golbeck demographic, this is actually an interesting --- if at times difficult to watch --- film. Kubrick does a pretty amazing job of making the audience feel compassion for a sociopath. That said, the film is not 100% faithful to the book. Most notably, the final (redemptive) chapter is omitted. But Kubrick never seemed all that much into redemption anyway, and so the ending is not necessarily what we'd hope for, but since when have Kubrick's films been described as 'life-affirming'? Worth watching, even through hands over your eyes in parts...
- by Brian Shields
Conclusions

• Social networks are growing exponentially in number and size
• Web standards and technologies are available that let us access them easily and aggregate them
• There is a wide range of applications that could benefit from taking users’ social relationships into consideration.
Ongoing Work

• Social Networks and trust for disaster response and recovery
  – Alerting people of emergencies taking place
  – Helping people find information in the aftermath of disasters
Info

• Jennifer Golbeck
• jgolbeck@umd.edu
• http://www.cs.umd.edu/~golbeck
• http://trust.mindswap.org