Social Networks and Social Media
Social Media: Many-to-Many
Characteristics of Social Media

- “Consumers” become “Producers”
- Rich User Interaction
- User-Generated Contents
- Collaborative environment
- Collective Wisdom
- Long Tail

Broadcast Media
Filter, then Publish

Social Media
Publish, then Filter
Top 20 Websites at USA

<table>
<thead>
<tr>
<th></th>
<th>Website</th>
<th></th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Google.com</td>
<td>11</td>
<td>Blogger.com</td>
</tr>
<tr>
<td>2</td>
<td>Facebook.com</td>
<td>12</td>
<td>msn.com</td>
</tr>
<tr>
<td>3</td>
<td>Yahoo.com</td>
<td>13</td>
<td>Myspace.com</td>
</tr>
<tr>
<td>4</td>
<td>YouTube.com</td>
<td>14</td>
<td>Go.com</td>
</tr>
<tr>
<td>5</td>
<td>Amazon.com</td>
<td>15</td>
<td>Bing.com</td>
</tr>
<tr>
<td>6</td>
<td>Wikipedia.org</td>
<td>16</td>
<td>AOL.com</td>
</tr>
<tr>
<td>7</td>
<td>Craigslist.org</td>
<td>17</td>
<td>LinkedIn.com</td>
</tr>
<tr>
<td>8</td>
<td>Twitter.com</td>
<td>18</td>
<td>CNN.com</td>
</tr>
<tr>
<td>9</td>
<td>Ebay.com</td>
<td>19</td>
<td>Espn.go.com</td>
</tr>
<tr>
<td>10</td>
<td>Live.com</td>
<td>20</td>
<td>Wordpress.com</td>
</tr>
</tbody>
</table>

40% of websites are social media sites
What is Social Network and Social Media

• Social Network
  – The networks formed by individuals

• Social Media
  – social network + media
    • media = content of twitter, tag, videos, photos ....
Statistical Properties of Social Networks
Why do statistics

• To understand the networks
  – Understand their topology and measure their properties
  – Study their evolution and dynamics
  – Create realistic models
  – Create algorithms that make use of the network structure
Interesting Questions: demonstration

• Some interesting questions
  – What do social networks look like, on a large scale?
  – How do networks behave over time?
  – How do the non-giant weakly connected components behave over time?
  – What distributions and patterns do weighted graphs maintain?
Networks and Representation

Social Network: A social structure made of nodes (individuals or organizations) and edges that connect nodes in various relationships like friendship, kinship etc.

- Graph Representation

- Matrix Representation

<table>
<thead>
<tr>
<th>Node</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>
Basic Concepts

• A: the adjacency matrix
• V: the set of nodes
• E: the set of edges
• \( v_i \): a node \( v_i \)
• \( e(v_i, v_j) \): an edge between node \( v_i \) and \( v_j \)
• \( N_i \): the neighborhood of node \( v_i \)
• \( d_i \): the degree of node \( v_i \)
• geodesic: a shortest path between two nodes
  – geodesic distance
Statistical Properties

• **Static analysis**
  – Static snapshots of graphs

• **Dynamic analysis**
  – A series of snapshots of graphs
Some famous properties

1. ‘Small-world’ phenomenon
   – An Experiment by Milgram (1967)
     • Asked randomly chosen “starters” to forward a letter to the target
     • Name, address, and some personal information were provided for the target person
     • The participants could only forward a letter to a single person that he/she knew on a first name basis
     • Goal: To advance the letter to the target as quickly as possible
The Milgram Experiment (Wikipedia)

Detailed procedure

1. Milgram typically chose individuals in the U.S. cities of Omaha, Nebraska and Wichita, Kansas to be the starting points and Boston, Massachusetts to be the end point of a chain of correspondence because they were thought to represent a great distance in the United States, both socially and geographically.

2. Information packets were initially sent to "randomly" selected individuals in Omaha or Wichita. They included letters, which detailed the study's purpose, and basic information about a target contact person in Boston. It additionally contained a roster on which they could write their own name, as well as business reply cards that were pre-addressed to Harvard.
The Milgram Experiment (cont.)

3. Upon receiving the invitation to participate, the recipient was asked whether he or she personally knew the contact person described in the letter.
   – If so, the person was to forward the letter directly to that person. For the purposes of this study, knowing someone "personally" was defined as knowing them on a first-name basis.

4. In the more likely case that the person did not personally know the target, then the person was to think of a friend or relative they know personally that is more likely to know the target.
   – A postcard was also mailed to the researchers at Harvard so that they could track the chain's progression toward the target.
The Milgram Experiment

5. When and if the package eventually reached the contact person in Boston, the researchers could examine the roster to count the number of times it had been forwarded from person to person.
   – Additionally, for packages that never reached the destination, the incoming postcards helped identify the break point in the chain.
Result of the Experiment

• However, a significant problem was that often people refused to pass the letter forward, and thus the chain never reached its destination.

• In one case, 232 of the 296 letters never reached the destination.[3]

• However, 64 of the letters eventually did reach the target contact.

• Among these chains, the average path length fell around 5.5 or six.
Some famous properties con’t

1. ‘Small-world’ phenomenon
   – Property
     • Any two people can be connected within 6 hops

\textit{six degrees of separation}

– Verified on a planetary-scale IM network of 180 million users (Leskovec and Horvitz 2008)
  • The average path length is 6.6
Some famous properties con’t

2. ‘Power-law’ degree distributions

- Degree distributions

\[ f_k = \text{fraction of nodes with degree } k \]
\[ = \text{probability of a randomly selected node to have degree } k \]

\( f_k \text{ vs. } k \)
log-log plot

- Power law distribution becomes a straight line if plot in a log-log scale
Some famous properties con’t

2. ‘Power-law’ degree distributions

– The degree distributions of most real-life networks follow a *power law*

\[ p(x) = C x^{-\alpha}, \quad x \geq x_{\text{min}}, \quad \alpha > 1 \]
Some other properties

3. Triangle Power Law

\[ f(\Delta) \propto \Delta^\alpha \quad \alpha < 0 \]
Some other properties

4. Eigenvalue Power Law

- The 20 or so largest eigenvalues of the adjacency matrix are power law distributed
- This is consequence of the “Degree Power Law”
Some other properties

5. Community Structure

– Social networks are modular
  • i.e nodes form communities
Statistical Properties

- **Static analysis**
  - Static snapshots of graphs

- **Dynamic analysis**
  - A series of snapshots of graphs
Properties

1. Shrinking Diameter
Properties con’t

2. Densification Power Law (DPL)
   - $E(t)$: the number of edges
   - $N(t)$: the number of nodes

$$E(t) \propto N(t)^\beta$$
Properties con’t

3. Diameter-plot and Gelling point
   – Graph forming
     • Establishment period
     • Stable period
Properties con’t

4. Constant/Oscillating Connected Components (CC)
   – Largest: constant
   – Second/Third: oscillation
5. Principal eigenvalue over time

- $E(t)$: the number of edges
- $\lambda_1(t)$: the largest eigenvalue

\[ \lambda_1(t) \propto E(t)^\alpha \quad \alpha \leq 0.5 \]
Conclusion

• Usefulness of the statistical properties
  – Understanding human behaviors
  – Anomalous graphs/subgraphs detection
  – Identifying authorities and search algorithms
  – Prepare resources based on the prediction
  – ...

2/11/2014