Classification of Social Network Sites based on Network Indexes and Communication Patterns

Fujio Toriumi, Isamu Okada, Hitoshi Yamamoto, Hirohiko Suwa, Kiyoshi Izumi and Yasuhiro Hashimoto
Background

- The trend of social networks changes

Research Question

- The essence of the web communications
- How to encourage the utilization of social media

What comes next?
Summary

• Analyze social networks
  – To find the types of communications on the web

• Focus on small-size SNSs
  – Use 615 small-size SNSs data

• We analyzed SNSs from view point of
  – Network structure
  – Communication
Why small SNS?

- Small SNS represents small community
  - Campus
  - Company
  and so on
- To know how people growth the relationships
- To know how people communicate on the web
From view point of Network
Social Network Data

• Data from a SNS provide service
  – So-net SNS beta
  – Number of SNSs: 615
    • Each with 50-10000 users

• Analyzed Data
  – User(user ID, date on registered)
  – Link(link ID, user id, user ID, date to created)
  – Blog entry(blog ID, user ID, date on entered)
  – Blog comment(comment ID, blog ID, comment user ID, comment date)
Features of Analyzed SNS

• Connection
  – Be friends with other users
  – Invite friends

• Communication
  – Entry the blog articles
  – Comment to the blog entry
Structures of SNSs

• Difficult to understand the characteristics from figure
• Use Network Analysis
Characteristics Parameters of Network

- Average path length
- Clustering coefficient
  - Density of network
- Power Law Determination coefficients
  - Scale-Free
- Power Index
- Assortativity
  - Degree correlation
# Network Indexes

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>C</th>
<th>$R^2$</th>
<th>$\gamma$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average</strong></td>
<td>2.13</td>
<td>0.377</td>
<td>0.631</td>
<td>-0.91</td>
<td>-0.471</td>
</tr>
<tr>
<td><strong>max</strong></td>
<td>4.26</td>
<td>0.894</td>
<td>1.00</td>
<td>0.161</td>
<td>0.130</td>
</tr>
<tr>
<td><strong>min</strong></td>
<td>1.41</td>
<td>0</td>
<td>0.01</td>
<td>-1.58</td>
<td>0</td>
</tr>
</tbody>
</table>

L: Average path length  
C: Cluster Coefficient  
$R^2$: Power Low Determination coefficients  
$\gamma$: Power Index  
r: Assortativity
Classification of SNSs by Clustering approach

• Classified the SNSs from viewpoint of the network structure
  – By using K-means

• Character vector

\[ v_i = \begin{bmatrix} \frac{L_i}{\sigma_L} & \frac{C_i}{\sigma_C} & \frac{R^2_i}{\sigma_{R^2}} & \frac{r_i}{\sigma_r} \end{bmatrix} \]

Note: We did not use \( \gamma \) because there are some SNSs which do not comply with power distribution
K-Means

- Non-hierarchical Clustering method
- Set similar users to same groups
## Four types of SNS

<table>
<thead>
<tr>
<th></th>
<th>No. User</th>
<th>No. Link</th>
<th>L</th>
<th>C</th>
<th>R²</th>
<th>r</th>
<th>No. SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>283.6</td>
<td>1001.1</td>
<td>2.095</td>
<td>0.436</td>
<td>0.713</td>
<td>-0.388</td>
<td>263</td>
</tr>
<tr>
<td>C2</td>
<td>236.9</td>
<td>287.5</td>
<td>2.025</td>
<td>0.163</td>
<td>0.573</td>
<td>-0.721</td>
<td>184</td>
</tr>
<tr>
<td>C3</td>
<td>87.9</td>
<td>743.0</td>
<td>1.833</td>
<td>0.686</td>
<td>0.380</td>
<td>-0.369</td>
<td>92</td>
</tr>
<tr>
<td>C4</td>
<td>423.7</td>
<td>1454.1</td>
<td>2.851</td>
<td>0.313</td>
<td>0.783</td>
<td>-0.280</td>
<td>76</td>
</tr>
</tbody>
</table>

- **General (C1)**
- **Star (C2)**
- **Aggregation (C3)**
- **Spread (C4)**
Feature of C1

<table>
<thead>
<tr>
<th></th>
<th>No. Users</th>
<th>No. Links</th>
<th>L</th>
<th>C</th>
<th>R²</th>
<th>r</th>
<th>No. SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>283.6</td>
<td>1001.1</td>
<td>2.095</td>
<td>0.436</td>
<td>0.713</td>
<td>-0.388</td>
<td>263</td>
</tr>
</tbody>
</table>

- General SNS
  - 40% of SNS classified to C1
  - Small World
  - Scale Free
  - Negative Assortativity
Feature of C2

<table>
<thead>
<tr>
<th></th>
<th>No. Users</th>
<th>No. Links</th>
<th>L</th>
<th>C</th>
<th>$R^2$</th>
<th>r</th>
<th>No. SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>236.9</td>
<td>287.5</td>
<td>2.025</td>
<td>0.163</td>
<td>0.573</td>
<td>-0.721</td>
<td>184</td>
</tr>
</tbody>
</table>

- **Star type SNS**
  - Smaller Clustering Coefficient
  - Smaller Assortativity
  - Average degree is 1.21
    - Most of nodes connected to only one node
  - Largest degree node has 78% of links in the network
    - Only one node has most of links in the network
# Feature of C3

<table>
<thead>
<tr>
<th></th>
<th>No. User</th>
<th>No. Link</th>
<th>L</th>
<th>C</th>
<th>R²</th>
<th>r</th>
<th>No. SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>87.9</td>
<td>743.0</td>
<td>1.833</td>
<td>0.686</td>
<td>0.380</td>
<td>-0.369</td>
<td>92</td>
</tr>
</tbody>
</table>

- **Aggregation SNS**
  - Very high clustering coefficient
  - Short average path length
  - Not scale-free
  - Roughly Complete network
    - Most of nodes connected to most of other nodes

[Network Diagram]
### Feature of C4

<table>
<thead>
<tr>
<th>No. User</th>
<th>No. Link</th>
<th>L</th>
<th>C</th>
<th>R²</th>
<th>r</th>
<th>No. SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C4</td>
<td>423.7</td>
<td>1454.1</td>
<td>2.851</td>
<td>0.313</td>
<td>0.783</td>
<td>-0.280</td>
</tr>
</tbody>
</table>

- **Spread SNS**
  - Longer path length
  - Many users
  - Growing out of the administrator’s hands
# Four types of SNS

<table>
<thead>
<tr>
<th></th>
<th>No. User</th>
<th>No. Link</th>
<th>L</th>
<th>C</th>
<th>R²</th>
<th>r</th>
<th>No. SNSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>283.6</td>
<td>1001.1</td>
<td>2.095</td>
<td>0.436</td>
<td>0.713</td>
<td>-0.388</td>
<td>263</td>
</tr>
<tr>
<td>C2</td>
<td>236.9</td>
<td>287.5</td>
<td>2.025</td>
<td>0.163</td>
<td>0.573</td>
<td>-0.721</td>
<td>184</td>
</tr>
<tr>
<td>C3</td>
<td>87.9</td>
<td>743.0</td>
<td>1.833</td>
<td>0.686</td>
<td>0.380</td>
<td>-0.369</td>
<td>92</td>
</tr>
<tr>
<td>C4</td>
<td>423.7</td>
<td>1454.1</td>
<td>2.851</td>
<td>0.313</td>
<td>0.783</td>
<td>-0.280</td>
<td>76</td>
</tr>
</tbody>
</table>

- **General (C1)**
- **Star (C2)**
- **Aggregation (C3)**
- **Spread (C4)**
From viewpoint of Communication
Summary

- Analyzed the relationships and communications
  - The activation of user behavior
  - The contribution of the core users
- RQ: Is there any relationships between comments on the blog entry and friend network structure?
  - Aggregation ratio for friends
  - Coverage ratio for friends
- RQ: Are high-degree users important for activation?
  - Index of degree contribution.
Aggregation Ratio (A)

- The ratio of comments to friends in all the comments.

Comments to Friends: 3
Comments          : 4
Aggregation Ratio: 3/4 = 0.75
Coverage Ratio (C)

- The ratio of friends in all the users who get comments.

Friendship with comments: 3
All Friends who entered the blog: 5
Aggregation Ratio: $3/5 = 0.60$
Communication Patterns

Inclusive friend

Parity friend

Independent friend

Partial friend

Coverage ratio

Aggregation ratio
Social Network Data

• Data from SNS provide service
  – Number of SNSs: 309
    • Each with more than 100 users

• Analyzed Data
  – User(user ID, date on registered)
  – Link(link ID, user id, user ID, date to created)
  – Blog entry(blog ID, user ID, date on entered)
  – Blog comment(comment ID, blog ID, comment user ID, comment date)
Cross table of Network and Communication classification

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial</td>
<td>43</td>
<td>7</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>Parity</td>
<td>41</td>
<td>14</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Inclusive</td>
<td>39</td>
<td>27</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Independent</td>
<td>32</td>
<td>31</td>
<td>2</td>
<td>9</td>
</tr>
</tbody>
</table>

- Star SNS: Inclusive, Independent
- Aggregation SNS: Partial
# Effect of Communication Pattern on Activation of User Behavior

- Active communication in high coverage ratio SNSs
  - Communications are based on friend networks.
  - No difference in aggregation ratios

<table>
<thead>
<tr>
<th>Type</th>
<th>N</th>
<th>Comments No. Posting</th>
<th>Blog entries No. Postings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial</td>
<td>81</td>
<td>135.556</td>
<td>145.111</td>
</tr>
<tr>
<td>Parity</td>
<td>73</td>
<td>186.178</td>
<td>170.918</td>
</tr>
<tr>
<td>Inclusive</td>
<td>81</td>
<td>182.296</td>
<td>178.926</td>
</tr>
<tr>
<td>Independent</td>
<td>74</td>
<td>115.649</td>
<td>123.932</td>
</tr>
<tr>
<td>Chi-Square or F-Value</td>
<td></td>
<td>34.642</td>
<td>18.066</td>
</tr>
</tbody>
</table>

**Significance Prob.**
- .000***
- .000***
Contributions of Degree

• Who must be the active user?
  – Hub user (High-degree user)
  – Fringe user (Low-degree user)

• Focus on the degree contribution
  – Whether degree contributes for communication
An index of degree contribution ($D_c$)

$$
D_c = \frac{1}{N} \sum_{i} c_i \cdot d_i
$$

$C_i$ is number of user $i$'s comments

$d_i$ is degree of user $i$

- High $D_c$
  - High-degree users communicates well

- Low $D_c$
  - High-degree users do not communicate well
## Correlation between Dc and Communication indexes

<table>
<thead>
<tr>
<th>Type</th>
<th>Comments No. Postings</th>
<th>Blog Entries No. Postings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial</td>
<td>.046</td>
<td>-.057</td>
</tr>
<tr>
<td>N=81</td>
<td>.541</td>
<td>.448</td>
</tr>
<tr>
<td>Parity</td>
<td>-.138</td>
<td>-.154</td>
</tr>
<tr>
<td>N=73</td>
<td>.085+</td>
<td>.054+</td>
</tr>
<tr>
<td>Inclusive</td>
<td>.180</td>
<td>.137</td>
</tr>
<tr>
<td>N=81</td>
<td>.018*</td>
<td>.070+</td>
</tr>
<tr>
<td>Independent</td>
<td>.051</td>
<td>.057</td>
</tr>
<tr>
<td>N=74</td>
<td>.517</td>
<td>.469</td>
</tr>
</tbody>
</table>

- Negative correlation in parity friend network
- Positive correlation in Inclusive friend network
In the parity friend network

- Negative correlation between Dc and Communication indexes
  - Active communication by high degree user
  - Decrease in total amount of communication

- Hub member have not necessity to communicate actively
  - Each members communicate each other
  - Effort of core user will not be rewarded
In the inclusive friend network

- Positive correlation between Dc and Communication indexes

  Active communication by high degree user
  Increase in total amount of communication

- Hub member must communicate actively
  - Efforts of core users to activate communications are effective
Conclusion

• Analyzed large number of small SNSs
• Classified SNSs by network structure
  – General SNS
  – Star SNS
  – Aggregation SNS
  – Spread SNS
• Classified SNSs by communication patterns
  – partial
  – parity
  – inclusive
  – independent
Conclusion

• Each types of SNSs have different characteristics
  – Different type of communication is found in different types of SNS

• Analysis of SNSs are important for administrators
  – To find current conditions
  – To find what to do
Future Works

• Clarify the trajectory of growing processes by using time analysis
  – How friend structures growth?
  – How communication relationships growth?

• Apply to actual SNS
  – How to manage SNS
  – How to manage web community
Thank you

• Contact to:
  tori@is.nagoya-u.ac.jp