Real-World Popularity Estimation from Community Structure of Followers on SNS

Shuhei Kobayashi, Keishi Tajima
Kyoto University
Does the number of followers accurately reflect the real-world popularity?

- The number of followers is a metric of popularity on SNS.
- But followers on SNS are a sampling from real-world fans and are often biased.
- Therefore, the number of followers does not necessarily directly reflect real-world popularity.
- Can we better estimate real-world popularity by using the community structure of followers?

Research Background

- Do different community structures reflect the target user’s real-world popularity?
- Do SNS users with followers more distributed over the graph have more real-world popularity?
PageRank

- A method for determining the value of a node in a network
- Need a whole graph

In our research, we use only 2-hop graph
Calculate a score of each user

- Create the follower-neighbor graph
  - Retrieve a target user’s ID, followers’ IDs, and IDs of followers/friends of each follower

- Calculate the score of each follower using the method on the next page

- Sum up the scores of all the followers, and the result is the metric for estimating real-world popularity
  - The higher the score, the higher the popularity

User score = a + b + c

u is a target user, red nodes are followers, and blue nodes are friends/followers of each follower. a, b, c is each follower’s score.
Scoring the followers by the proposed method

- **Score** = \( \frac{\log_2(\text{# of blue nodes adjacent to red nodes} + 2)}{\log_2(\text{# of red nodes adjacent to red nodes} + 2)} \)

- The more the number of adjacent **red** nodes, the smaller the score
  - Followers from a same community (e.g. real friends) score lower
  - Smaller distribution when there are many adjacent red nodes

- The more the number of adjacent **blue** nodes, the larger the score
  - The more connections, the higher the score
  - More adjacent blue nodes are more valuable because followers are more likely to be influential or belong to various communities

- The above score is calculated for each follower, and then summed up to obtain the target user’s score
Example of score calculation

- Each user’s follower has two red adjacent nodes and two blue adjacent nodes.
- Each follower’s score is 1
  - Each follower’s score: \( \frac{\log_2(2+2)}{\log_2(2+2)} = 1 \)
- Therefore, the score of this target user is 3
  - User score = \( 1 + 1 + 1 = 3 \)
Predicting Ms/Mr University Competitions Standings

- Predicting the winner and the runners-up of nine competitions
- Compare the accuracy of proposed methods and the method using the number of followers
- Four evaluation metrics

<table>
<thead>
<tr>
<th>Combination</th>
<th>Ordered</th>
<th>Winner only</th>
<th>Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2</td>
<td>1 2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Calculate the score of each target user

- Use Twitter and Instagram
- Construct the follower-neighbor graph on Twitter and calculate the score
- Instagram scores are estimated from Twitter scores
Obtain twitter graph only

- Although it is desirable to obtain the follower structure for both Twitter and Instagram, the Instagram follower structure is difficult to obtain due to restriction.

- Instagram scores are estimated by converting scores obtained from Twitter.

  \[
  \text{Insta-score} = \text{Twitter-score} \times \left( \frac{\# \text{Instagram followers}}{\# \text{Twitter followers}} \right) \times \left( \frac{0.106}{0.42} \right)
  \]

  0.106 is the clustering coefficient of Twitter social graph (A. Java, 2007), 0.42 is that of Instagram social graph (L. Manikonda, 2014).

- Experiments with two patterns: Twitter+Instagram and Twitter only


Example of score converting

- Estimate Instagram score

<table>
<thead>
<tr>
<th>Twitter score</th>
<th>Instagram score</th>
<th># Twi followers</th>
<th># Ins followers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>?</td>
<td>1,000</td>
<td>2,000</td>
</tr>
</tbody>
</table>

- Instagram score is 504

\[
Insta\text{-score} = 1000 \times \left(\frac{2000}{1000}\right) \times \left(\frac{0.106}{0.42}\right) \approx 504
\]

\[
Twi + Insta\text{-score} = 1000 + 504 = 1504
\]
The number of total followers shows the best accuracy on Twitter+Instagram

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Combination</th>
<th>Ordered</th>
<th>Winner only</th>
<th>Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td># Twi followers</td>
<td>3/9</td>
<td>1/9</td>
<td>6/9</td>
<td>7/18</td>
</tr>
<tr>
<td># Ins followers</td>
<td>5/9</td>
<td>3/9</td>
<td>6/9</td>
<td>9/18</td>
</tr>
<tr>
<td>#Twi + Ins followers</td>
<td>5/9</td>
<td>3/9</td>
<td>7/9</td>
<td>10/18</td>
</tr>
<tr>
<td>Proposed method</td>
<td>4/9</td>
<td>2/9</td>
<td>6/9</td>
<td>8/18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Combination</th>
<th>Ordered</th>
<th>Winner only</th>
<th>Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td># Twi followers</td>
<td>3/9</td>
<td>1/9</td>
<td>6/9</td>
<td>7/18</td>
</tr>
<tr>
<td>Proposed Method</td>
<td>5/9</td>
<td>4/9</td>
<td>6/9</td>
<td>10/18</td>
</tr>
</tbody>
</table>

Proposed method shows the best accuracy on Twitter only
Follower structure may be used to estimate real-world popularity

- The proposed method shows a lower accuracy on Twitter+Instagram, but shows a higher accuracy on Twitter only.

- If the follower-neighbor graph of the Instagram could be obtained, the proposed method could show higher accuracy on Twitter+Instagram.

- The characteristics of the proposed method suggest that the more distributed the followers are, the higher the real-world popularity is, and the higher the value of the followers, the higher the real-world popularity.