Social features of online networks: The strength of intermediary ties in online social media

J.J. Ramasco, P. Grabowicz, E. Moro, J.M. Pujol, V.M. Eguiluz
Outline:

- Online Social Networks: Twitter
- Sociological theories
- Network clustering
- Results
- Conclusions
Online Social Networks

by R. Munroe  http://xkcd.com/256
Online Social Networks

by R. Munroe     http://xkcd.com/256
Online Social Networks
Online Social Networks

---

The Joy of Tech™

by Nitrozac & Snaggy

©2007 Geek Culture

Signs of the social networking times.

Unemployable due to stupid personal stuff I put on my Facebook page.

Me too!

FOR ME, IT WAS AN EMBARRASSING YOUTUBE VIDEO.

http://ifisc.uib-csic.es
Online Social Networks
Twitter

Starting date **2006-2007**

**Twitter - Total Registered Users**

- Sept 2011: 375,000,000
- Nov 2011: 375,000,000
- Jan 2012: 500,000,000
- Mar 2012: 500,000,000

**Twitter - Total Active Users**

- Sept 2011: 105,000,000
- Nov 2011: 105,000,000
- Jan 2012: 140,000,000
- Mar 2012: 140,000,000
Starting date 2006-2007

K. Ryan, ed. (2009-), "Twitter Study - August 2009", Pair Analytics
Twitter: structure & use

Tweets as basic interactions
Twitter: structure & use

Tweets as basic interactions

**Distribution of Follower Counts**

<table>
<thead>
<tr>
<th>Number of Followers</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>1,000</td>
</tr>
<tr>
<td>4</td>
<td>10,000</td>
</tr>
<tr>
<td>5</td>
<td>100,000</td>
</tr>
<tr>
<td>6</td>
<td>1,000,000</td>
</tr>
<tr>
<td>7</td>
<td>10,000,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Username</th>
<th>Followers</th>
<th>Following</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lady Gaga (ladygaga)</td>
<td>7,394,362</td>
<td>145,898</td>
</tr>
<tr>
<td>Britney Spears (britneyspears)</td>
<td>6,416,696</td>
<td>416,140</td>
</tr>
<tr>
<td>Justin Bieber (justinbieber)</td>
<td>6,355,747</td>
<td>95,167</td>
</tr>
</tbody>
</table>

Twitter: structure & use

Mentions & Retweets
Twitter: structure & use

Mentions & Retweets

Mentions: @john
Twitter: structure & use

Mentions & Retweets

RT: @sam
Our Twitter database

Year: 2008

<table>
<thead>
<tr>
<th>Property</th>
<th>Follower links</th>
<th>Links with mentions</th>
<th>Links with retweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>2 408 534</td>
<td>377 760</td>
<td>26 480</td>
</tr>
<tr>
<td>Links</td>
<td>48 776 888</td>
<td>1 224 484</td>
<td>32 169</td>
</tr>
<tr>
<td>Reciprocity</td>
<td>27%</td>
<td>14%</td>
<td>0.7%</td>
</tr>
</tbody>
</table>

Table 1: Overall characteristics of the follower network and of the interactions taking place on it.
Groups and sociological theories:

Strength of the weak ties

M. Granovetter, Am. J. Sociology 78, 1360 (1973).
Groups and sociological theories:

Strength of the weak ties

M. Granovetter, Am. J. Sociology 78, 1360 (1973).
Groups and sociological theories:

Tradeoff diversity-bandwidth

Groups and sociological theories:

Tradeoff diversity-bandwidth


http://ifisc.uib.csic.es
Groups and sociological theories:

Tradeoff diversity-bandwidth


http://ifisc.uib-csic.es
Communities?

Cluster detection

We are applying a number of clustering methods to the follower network of Twitter. The aim is to validate the main results with the groups obtained with different techniques.

Just to name a few:

- Oslom
- Infomap

Results

internal links

between groups

intermediary links

no-group links

http://ifisc.uib-csic.es
Results

General statistics

(A) \( P(S) \) vs. Group size

\[ N_{\text{groups}} = 92,062 \]

(B) \( P(N) \) vs. Number of groups per node

\[ N_{\text{users}} = 2,408,534 \]

(C) Bar chart:
- Followers
- Retweets
- Mentions

http://ifisc.uib-csic.es
Results

![Graph showing link fraction vs. group size](http://ifisc.uib.csic.es)
Results

Internal links

\[ N_{\text{mentions}} = 866,264 \]

![Graph showing the link fraction and group size distribution.]

http://ifisc.uib.csic.es
Results

Internal links

\[ N_{\text{mentions}} = 866,264 \]
Results

Between-groups links

![Graph showing link fraction against size group of origin and size of targeted group](http://ifisc.uib-csic.es)
Between-groups links

Results

![Graph showing link fraction vs. group similarity and size](http://ifisc.uib.csic.es)
Intermediary links

Results

- between groups
- bridging
- internal

% of links

- mentions
- retweets
Results

Intermediary links

Graph showing the probability of non-shared groups ($P(N-\text{sg})$) against the number of non-shared groups. The graph includes two curves, one in red and another in green, with a secondary axis showing a ratio that peaks around 8 non-shared groups.
Table 1: Summary of the results regarding internal connections when the groups are obtained with several clustering algorithms for different samples of the network. We measure the trend of the mentions to concentrate in internal connections. Legend: \( w \) - weak signal, \( sg \) - signal only for small groups, typically smaller than 10 members, a hyphen is inserted if we have no results.

<table>
<thead>
<tr>
<th>Network sample</th>
<th>Nodes</th>
<th>Edges</th>
<th>Oslom</th>
<th>Infomap</th>
<th>Moses</th>
<th>Real-time</th>
<th>Louvain</th>
<th>Radatools</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole network</td>
<td>2 408 534</td>
<td>48 776 888</td>
<td>✔</td>
<td>✔</td>
<td>-</td>
<td>✔</td>
<td>w</td>
<td>-</td>
<td>S6</td>
</tr>
<tr>
<td>Snowball 2 hops</td>
<td>61 492</td>
<td>5 558 036</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>x</td>
<td>x</td>
<td>-</td>
<td>S7</td>
</tr>
<tr>
<td>Snowball 3 hops</td>
<td>175 078</td>
<td>10 356 020</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>w</td>
<td>✔</td>
<td>-</td>
<td>S8</td>
</tr>
<tr>
<td>Random 200k</td>
<td>200 000</td>
<td>346 578</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>sg</td>
<td>sg</td>
<td>sg</td>
<td>S9</td>
</tr>
<tr>
<td>No hubs</td>
<td>2 395 415</td>
<td>23 404 103</td>
<td>-</td>
<td>✔</td>
<td>✔</td>
<td>w</td>
<td>x</td>
<td>-</td>
<td>S10</td>
</tr>
<tr>
<td>Oslom groups</td>
<td>99 832</td>
<td>1 216 942</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>S11</td>
</tr>
</tbody>
</table>
Other clustering methods

Table 1: Summary of the results regarding interna are obtained with several clustering algorithms fo work. We measure the trend of the mentions to nections. Legend: w - weak signal, sg - signal or smaller than 10 members, a hyphen is inserted if

<table>
<thead>
<tr>
<th>Network sample</th>
<th>Nodes</th>
<th>Edges</th>
<th>Oslo</th>
<th>Infomap</th>
<th>Moses</th>
<th>Real-time</th>
<th>Louvain</th>
<th>Radatools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole network</td>
<td>2 408 534</td>
<td>48 776 888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snowball 2 hops</td>
<td>61 492</td>
<td>5 558 036</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snowball 3 hops</td>
<td>175 078</td>
<td>10 356 020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random 200k</td>
<td>200 000</td>
<td>346 578</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No hubs</td>
<td>2 395 415</td>
<td>23 404 103</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oslom groups</td>
<td>99 832</td>
<td>1 216 942</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- We have analyzed a large-scale online social network correlating groups and user activity.
- The way the activity localizes with respect to the groups follows rules similar to offline social networks.
- Users belonging to several groups play an important role for information diffusion.