Mapping online political talks through network analysis
A case study of the website of Italy’s Five Star Movement

FRANCESCO BAILO
The University of Sydney

IPP2014: Crowdsourcing for Politics and Policy
Oxford Internet Institute, University of Oxford

26 September 2014
Framework

Hypothesis

Five Star MoVement: A brief history

Forum statistics

Network analysis

Debate analysis

Conclusion
WHAT is going on and WHO is talking

- Everyday political talk
  - Mansbridge (1999)
  - J. Kim and E. J. Kim (2008)
WHAT is going on and WHO is talking

- *Everyday political talk*
  - Mansbridge (1999)
  - J. Kim and E. J. Kim (2008)

- *Networked self*
  - Rainie and Wellman (2012)
Hypothesis

- **H1**: There is a correlation between public events and online participation;
- **H2**: Online political participation decreases gender imbalances;
- **H3**: Users engage in discussions on limited topics;
- **H4**: Online political participation fragments in conversations among like-minded;
- **H5**: Forum activity shows a property of preferential attachment with a ‘the-rich-get-richer’ effect.
Five Star MoVement: A brief history

- January 2005: www.beppegrillo.it launched
- September 2007: First *Fuck Off Day* (Vaffanculo Day or V-Day)
- Mid-2009: Movement participation to first elections
- October 2009: Movement named ‘Five Star Movement’ (M5S)
- February 2013: M5S won 25.5 percent popular vote (1st party with 8,689,168 votes)
# Five Star MoVement: A brief history

## Table: Provenance of the electorate of the M5S in 2013 Italian general election by party voted in 2008 (%) - (Source: Tronconi, 2013)

<table>
<thead>
<tr>
<th>City</th>
<th>Far left</th>
<th>Idv</th>
<th>Pd</th>
<th>Udc</th>
<th>Pdl</th>
<th>Mpa</th>
<th>Lega</th>
<th>Far right</th>
<th>Did not vote</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turin</td>
<td>6</td>
<td>20</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>10</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Milan</td>
<td>10</td>
<td>13</td>
<td>13</td>
<td>4</td>
<td></td>
<td></td>
<td>29</td>
<td>9</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Bologna</td>
<td>5</td>
<td>12</td>
<td>48</td>
<td>4</td>
<td></td>
<td></td>
<td>8</td>
<td>12</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Florence</td>
<td>5</td>
<td>2</td>
<td>58</td>
<td>5</td>
<td></td>
<td></td>
<td>2</td>
<td>5</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>Rome</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>26</td>
<td></td>
<td></td>
<td>10</td>
<td>35</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Naples</td>
<td>8</td>
<td>9</td>
<td>44</td>
<td>3</td>
<td>26</td>
<td></td>
<td>2</td>
<td>7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Catania</td>
<td>6</td>
<td>18</td>
<td>11</td>
<td></td>
<td></td>
<td>34</td>
<td>2</td>
<td>27</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
An integrated media system
# Forum statistics

Table: Users, proposals and comments published per week between 11 February 2009 and 31 January 2014

<table>
<thead>
<tr>
<th></th>
<th>Users</th>
<th>Proposals</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84,203</td>
<td>86,943</td>
<td>461,297</td>
</tr>
</tbody>
</table>

francesco.bailo@sydney.edu.au
Forum statistics

Figure: Forum weekly activity

Comments

National
classsembly

Regional
elections

Local
elections

Local
elections

Sicilian
elections

General
elections

Proposals

2009 2010 2011 2012 2013 2014
Forum statistics

**Figure: Gender distribution**

(a) Users ($N = 84,202$)

(b) Proposals ($N = 86,943$)

(c) Comments ($N = 461,297$)
Forum statistics

**Figure:** Elected local administrators in Italy ($N = 145,256$) (Source: Ministero Dell’Interno, 2013)

---

francesco.bailo@sydney.edu.au
Forum statistics

Figure: User locations \((N = 364)\)

(a) M5S users

(b) Population density \(^\text{Source: Istat, 2012}\)
Forum statistics

Figure: Proposals by category

development
development / city planning
development / jobs
development / public contracting
development / sustainable construction
development / territory
development / trade
energy
energy / alternative energy
energy / energy saving
energy / production
environment
environment / LNG terminal
environment / incinerators
environment / recycling
services / children
services / elders
services / youth
services to citizens
transport / public transport
transport / road conditions
water / public-water
water / saving
Forum
None

francesco.bailo@sydney.edu.au

@FrBailo
Network analysis

Figure: Bipartite directed network
Network analysis

Figure: Indegree distribution

(a) Forum
(b) Random bipartite Erdős-Rényi model
(c) Random Barabási-Albert model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>12,606</td>
</tr>
</tbody>
</table>

francesco.bailo@sydney.edu.au
Network analysis

Figure: Outdegree distribution

(a) Outdegree distribution

(b) Common outdegree

<table>
<thead>
<tr>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>18,519</td>
</tr>
</tbody>
</table>
Network analysis

Figure: Network components

(a) Before

(b) After

Figure: Distribution cluster size
One-mode projection: threads by category

(a) Bipartite network

(b) One-mode projection

(c) Contraction on thread theme
One-mode projection: threads by category

Figure: Relations between discussion themes

Development
Energy
Environment
Forum
No label
Services
Transport
Water
One-mode projection: users

(a) Bipartite network

(b) One-mode projection

(c) Community detection

(d) Contraction on communities
One-mode projection: users

Figure: Relations between user communities

@FrBailo
One-mode projection on threads

(a) Bipartite network
(b) One-mode projection

@FrBailo
Gravity model

For each pair of discussion threads, given the time difference between the first two postings and the product of size of the two threads, how many users may be expected to take part in both discussions?

\[
\log_e(\text{shared_users}) = \beta_0 + \beta_1 \times \text{mult_size} + \beta_2 \times \text{time_diff}
\]
Immigration

Figure: Relations between discussion threads on immigration
Coalition government

Figure: Relations between discussion threads on supporting a coalition government

@FrBailo

francesco.bailo@sydney.edu.au
<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Supported by the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: There is a correlation between public events and online participation</td>
<td>Yes</td>
</tr>
<tr>
<td>H2: Online political participation decreases gender imbalances</td>
<td>No</td>
</tr>
<tr>
<td>H3: Users engage in discussions on limited topics</td>
<td>No</td>
</tr>
<tr>
<td>H4: Online political participation fragments in conversations among like-minded</td>
<td>No</td>
</tr>
<tr>
<td>H5: Forum activity shows a property of preferential attachment with a ‘the-rich-get-richer’ effect</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Bibliography I


Bibliography II


francesco.bailo@sydney.edu.au
Forum architecture

Figure: Forum user structure.

- Unique ID users
- Non-unique ID users

Unrelated

Not allowed

Proposals

Comments
Forum user interface

Figure: Forum’s homepage (‘Forum: Movimento 5 Stelle’, 2009)
Forum user interface

Figure: The page of a Forum’s proposal (‘Forum: Movimento 5 Stelle’, 2013)
Data collection

- Google Custom Search Engine
- Disqus API
- Data retrieval
Data cleansing

Table: Forum raw data gathered until 19 February 2014 and published between 11 February 2009 and 31 January 2014

<table>
<thead>
<tr>
<th>Proposals</th>
<th>Comments</th>
<th>Unique users</th>
<th>Non-unique users</th>
</tr>
</thead>
<tbody>
<tr>
<td>108,269</td>
<td>510,909</td>
<td>30,921</td>
<td>121,603</td>
</tr>
</tbody>
</table>

Table: Suspect spam

<table>
<thead>
<tr>
<th>Spam users</th>
<th>Spam comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>6618</td>
<td>41,353</td>
</tr>
</tbody>
</table>
Data integration

Figure: User unification

Unique ID users

Alice

Bob

Non-unique ID users

Alice

Bob

Carol

Comments

Figure: User unification

No unique ID users

Diana

Eddy

Diana

Carol

Comments

francesco.bailo@sydney.edu.au
**Table: Regression summary**

<table>
<thead>
<tr>
<th></th>
<th>Naive SE (1)</th>
<th>Robust SE (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mult_size</strong></td>
<td>0.000003***</td>
<td>0.000003***</td>
</tr>
<tr>
<td>(0.0000005)</td>
<td>(0.000001)</td>
<td></td>
</tr>
<tr>
<td>t = 70.960000</td>
<td>t = 5.862000</td>
<td></td>
</tr>
<tr>
<td>p = 0.000000</td>
<td>p = 0.000000</td>
<td></td>
</tr>
<tr>
<td><strong>time_diff</strong></td>
<td>-0.000000***</td>
<td>-0.000000***</td>
</tr>
<tr>
<td>(0.000000)</td>
<td>(0.000000)</td>
<td></td>
</tr>
<tr>
<td>t = -13.590000</td>
<td>t = -56.88000</td>
<td></td>
</tr>
<tr>
<td>p = 0.000000</td>
<td>p = 0.000000</td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.073010***</td>
<td>0.073010***</td>
</tr>
<tr>
<td>(0.001196)</td>
<td>(0.000382)</td>
<td></td>
</tr>
<tr>
<td>t = 61.040000</td>
<td>t = 191.30000</td>
<td></td>
</tr>
<tr>
<td>p = 0.000000</td>
<td>p = 0.000000</td>
<td></td>
</tr>
</tbody>
</table>

**Observations** 980,929 980,929

**Log Likelihood** -1,028,512.000000 -1,028,512.000000

**Akaike Inf. Crit.** 2,057,030.000000 2,057,030.000000

**Residual Deviance (df = 980926)** 57,501.000000 57,501.000000

**Null Deviance (df = 980928)** 59,396.000000 59,396.000000

**Note:** *p<0.1; **p<0.05; ***p<0.01