Using twitter data networks to investigate italian constitutional referendum campaign

Lucio Palazzo *and* Francesco Santelli

DiSES, University of Salerno *and* Department of Social Sciences, University of Napoli

Abstract

In this poster we discuss about relationships between hashtags related to the 2016 Italian referendum campaign making use of hashtags co–occurences in tweets. We propose an application of a fast greedy algorithm in Tweet mining context in order to detect communities of hashtags. We observe that is possible to cluster group of hashtags that show similar features for both camps.

Introduction

The referendum took place on December 4^{th}, 2016 and it was the third constitutional referendum in Italy. Voters were asked whether they approve a constitutional law that modifies 47 articles of the Italian Constitution to reform the composition and powers of the Parliament of Italy as well as the division of powers between the State, the provinces and administrative entities. The debate between the two opposite factions arose strongly on microblogging message services, such as Twitter, where communication among users is coordinated by shared hashtags, that are keywords used like tags in order to *create or engage with discrete brand–like policy ideas* [1].

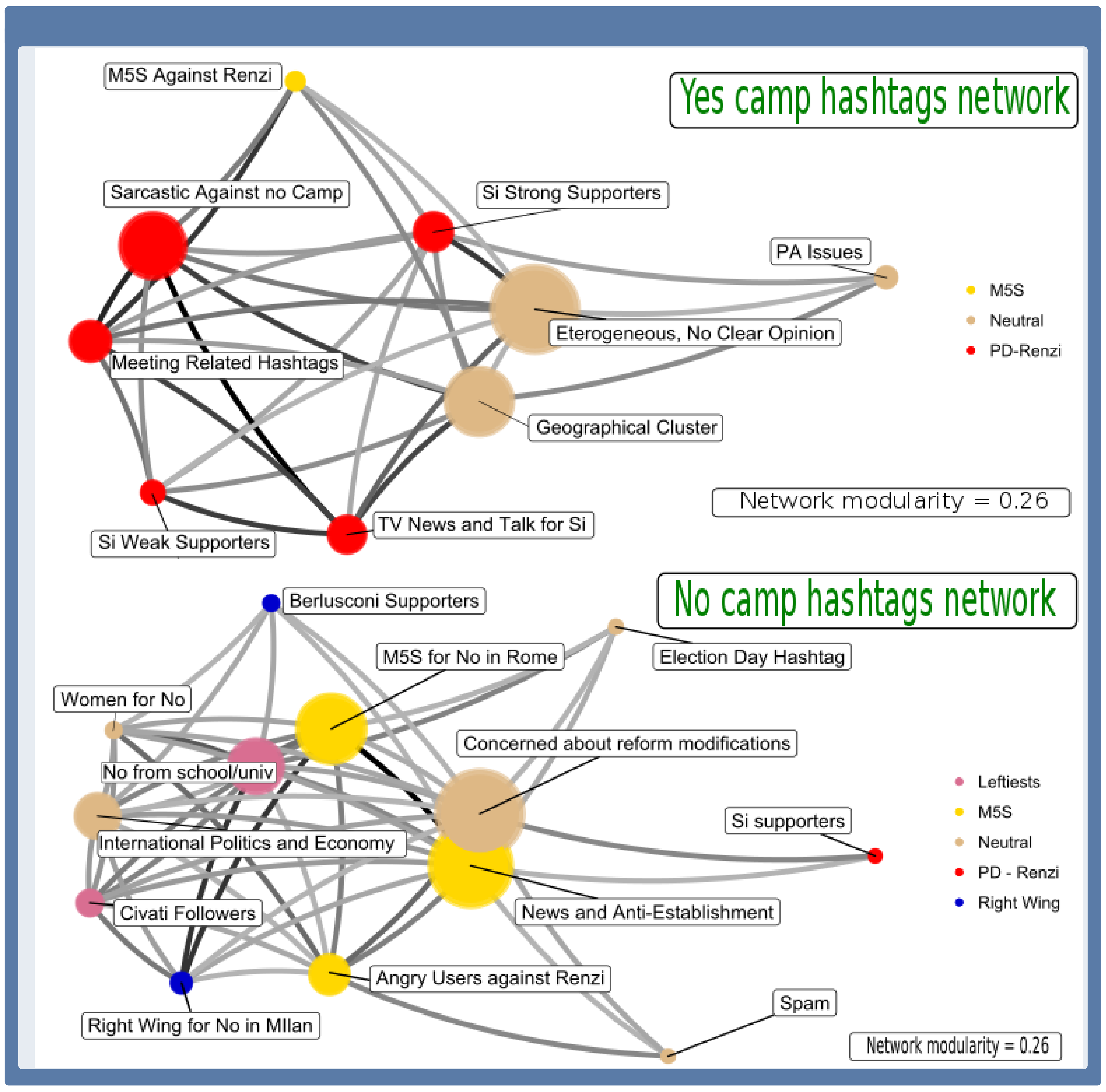
We discuss about relationships between hashtags related to the referendum campaign in order to detect communities, making use of hashtags co–occurences in tweets. In order to characterize the tweets related to both committees either the two official and the two most representative hashtags for both factions have been taken in account, that are *#bastaunsi*, *#iovotosi* e *#iovotono*, *#iodicono*.

Contact Information

- *Lucio Palazzo, PhD*
  lpalazzo@unisa.it
- *Francesco Santelli, PhD*
  francesco.santelli@unina.it

Data Webscraping for Community Mining

We gathered all tweets containing one of the key hashtags mentioned previously, posted during a time window starting from September, 1^{st} up to the December, 4^{th}. Then, two different weighted and undirected adjacency matrices have been produced, one for each camps, taking into account couples of hashtags that were used in the same twitter (*hashtag co–occurrencies*). The resulting large networks are centered around the originally selected key hashtags, that have been removed from the analysis to explore the relationships between the other hashtags involved. Finally, a fast greedy algorithm has been carried out in order to perform communities clustering.



Fast Greedy algorithm

Fast Greedy [2] algorithm (F–G) is a fast implementation of Newman–Girvan [3] algorithm (N–G), one of the first community detection algorithms applied in network analysis. Starting from a set of isolated nodes, the links of the original graph are iteratively added such to produce the largest possible increase of number of edges falling within groups, minus the expected number in an equivalent network with edges placed at random. Such value is known as Newman–Girvan modularity.

Let be A the (symmetric) adjacency matrix for an undirected network, the element $A_{v,w}$, for each couple of nodes v, w belonging to the network, represents the weight of the edge that connects those two nodes; if $A_{v,w} = 0$, that means there is no interaction between two nodes. Let also define k_v as the degree of each node. The Newman–Girvan modularity Q for weighted networks [4] is then defined as follows:

$$Q = \frac{1}{2m} \sum_{v,w} \left[A_{vw} - \frac{k_v k_w}{2m} \right] \delta(c_v, c_w) \qquad (1)$$

where n and m are the number of vertices and edges belonging to the network, c_v corresponds to the community to which the node v belongs and $\delta(\cdot, \cdot)$ is an indicator function with value 1 if $c_v = c_w$ and 0 otherwise.

N–G procedure updates the adjacency matrix A merging repeatedly pairs of columns or rows at each step. On the other hand, F–G algorithm improves computational cost replacing, after the first step, the original adjacency matrix with the matrix ΔQ, which elements $\Delta Q_{v,w}$ are the increased modularity values calculated after merging the two communities v and w. The estimated computational time for F–G is $O(n \log^2 n)$ while with N–G equals to $O(m^2 n)$, therefore F–G shows a considerable improvement in running time, especially in very large and sparse networks.

Conclusions

The network related to the *yes* camp hashtags is clustered in 9 communities, whereof 5 of them are strongly related to the Renzi political figure and to his party, the *Democratic Party*. One of them is marked by a strong use of several passionate hashtags against the opposite camp. Moreover, this community is linked to the only community clearly made by hashtags coming from the *no* side, marking a potential area of Twitter conflicts between users. In the center of the network are detected two clusters rather heterogeneous, one related to people expressing a certain degree of *indecisiveness about the topic* and the other made by hashtags with a wide expression of *geographical entities* (cities, regions etc...).

On the other hand, the *no* side network is made by 13 communities, showing an higher degree of political orientations heterogeneity: left wing (*Civati* cluster, *school/university* world) and also the clusters belonging to *Movimento 5 Stelle*, that is a political organization that plays by far the main role in the network. In the end, the right wing clusters (especially *Silvio Berlusconi*) show a marginal importance. Remarkably, the two wide communities in the center are expression of two distinct aspects of the referendum campaign: one related to the constitutional reform in itself (*technical cluster*) and the other one concerning the *anti–establishment* issue.

References

[1] Stephen Jeffares.
Interpreting Social Media Data.
In *Interpreting Hashtag Politics*, pages 112–134. Palgrave Macmillan UK, London, 2014.

[2] Aaron Clauset, M. E. J. Newman, and Cristopher Moore.
Finding community structure in very large networks.
Cond-Mat/0408187, 70:066111, aug 2004.

[3] M Girvan and M E J Newman.
Community structure in social and biological networks.
Proceedings of the National Academy of Sciences of the United States of America, 99(12):7821–6, jun 2002.

[4] M. E. J. Newman.
Analysis of weighted networks.
Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 70(5 2), jul 2004.