



THE UNIVERSITY OF
SYDNEY

LNGS2628

Digital tools for the humanities

FRANCESCO BAILO

The University of Sydney

5 September 2018

Network analysis

A very short introduction

Tools

Resources

Online deliberative discourse

Network analysis

Why to think about networks?

Let's think about a group of interesting research subjects (e.g. people, documents, states, organisations, tweets). We have at least two approaches:

1. We can study each *component* (e.g. person, document, state, organisation, tweet) of our group as a separate individual;
2. We can study the group as a **system**, focusing on the **relations** between components and on the **role** played by each component *within* the system.

Network analysis is not a *theory* but an *holistic* and *relational* approach to research in multiple disciplines (in fact, it is discipline-neutral).

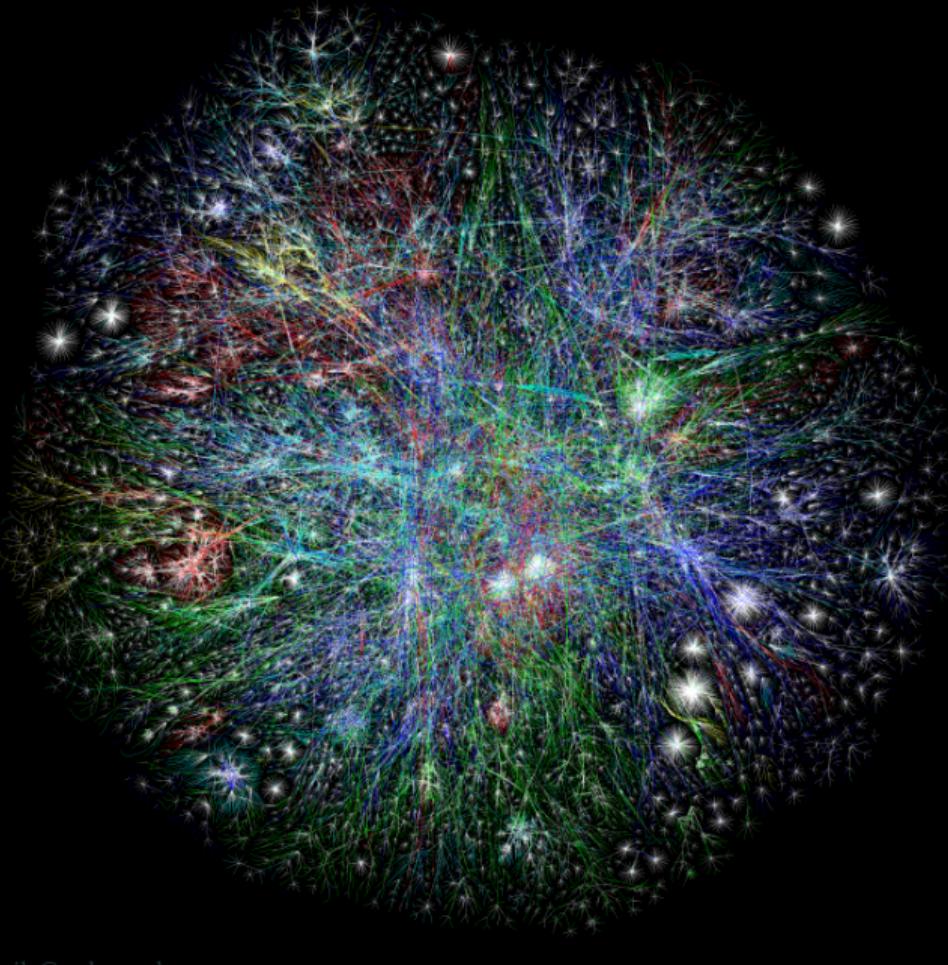


Figure 1: The Internet in 2003, each node represents groups of computers and each connection the route used by data packages (Credits: The Opte Project)

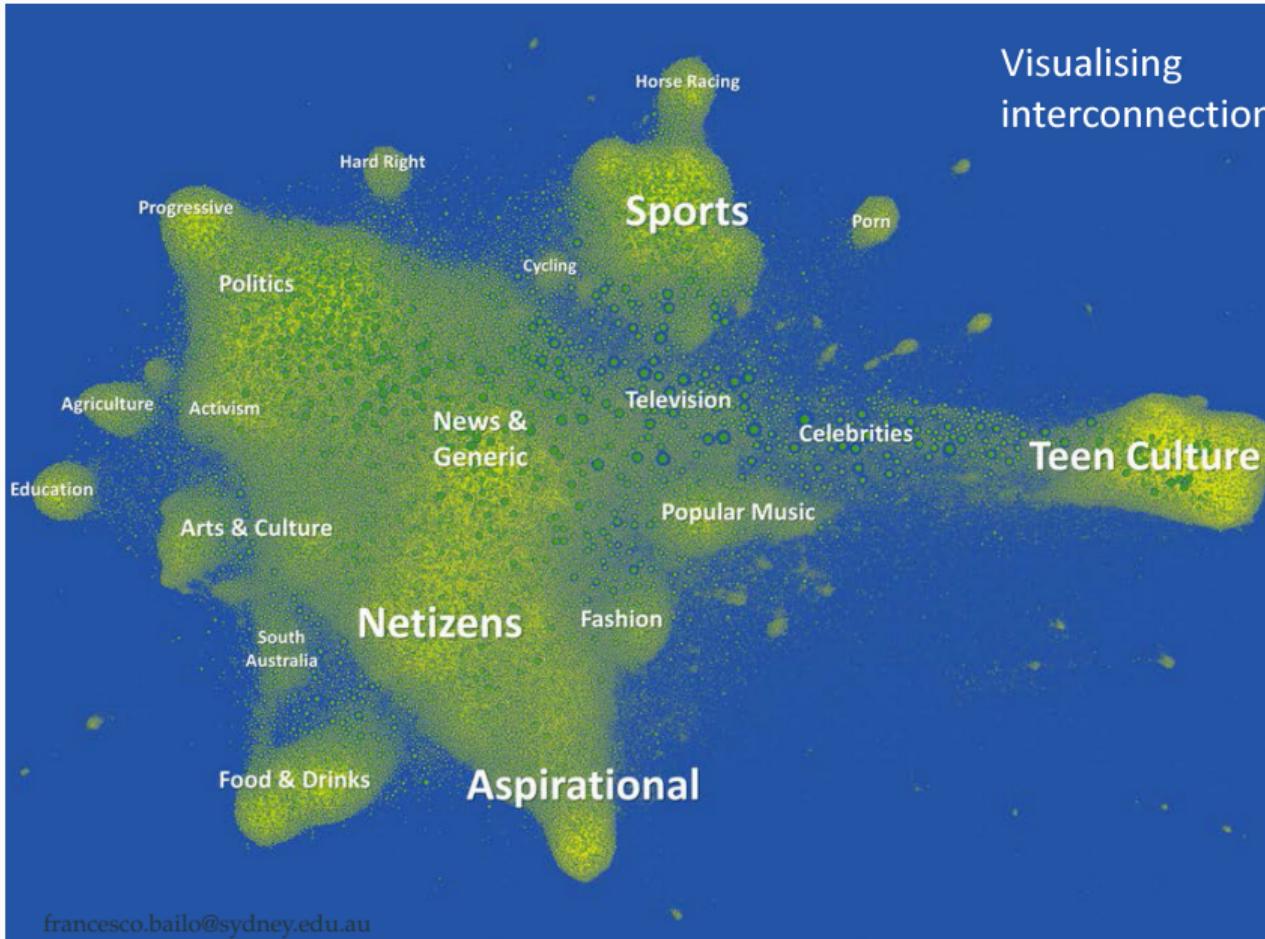


Figure 2: The Australian Twitter sphere where each node represents a user and links their interconnections
(Credit: Axel Bruns / QUT Digital Media Research Centre)

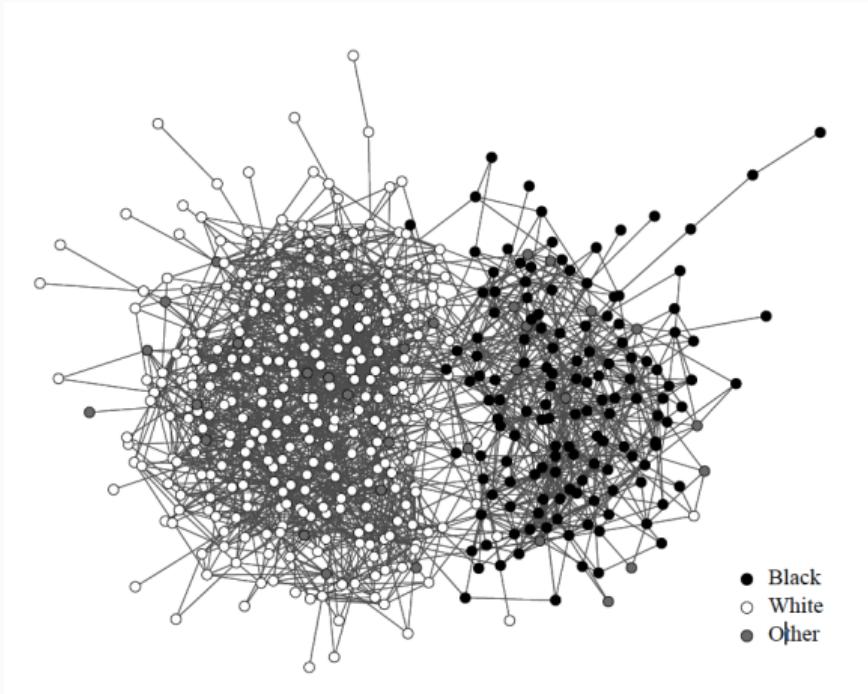


Figure 3: Friendship network at a US high school (Newman, 2010, p. 221).

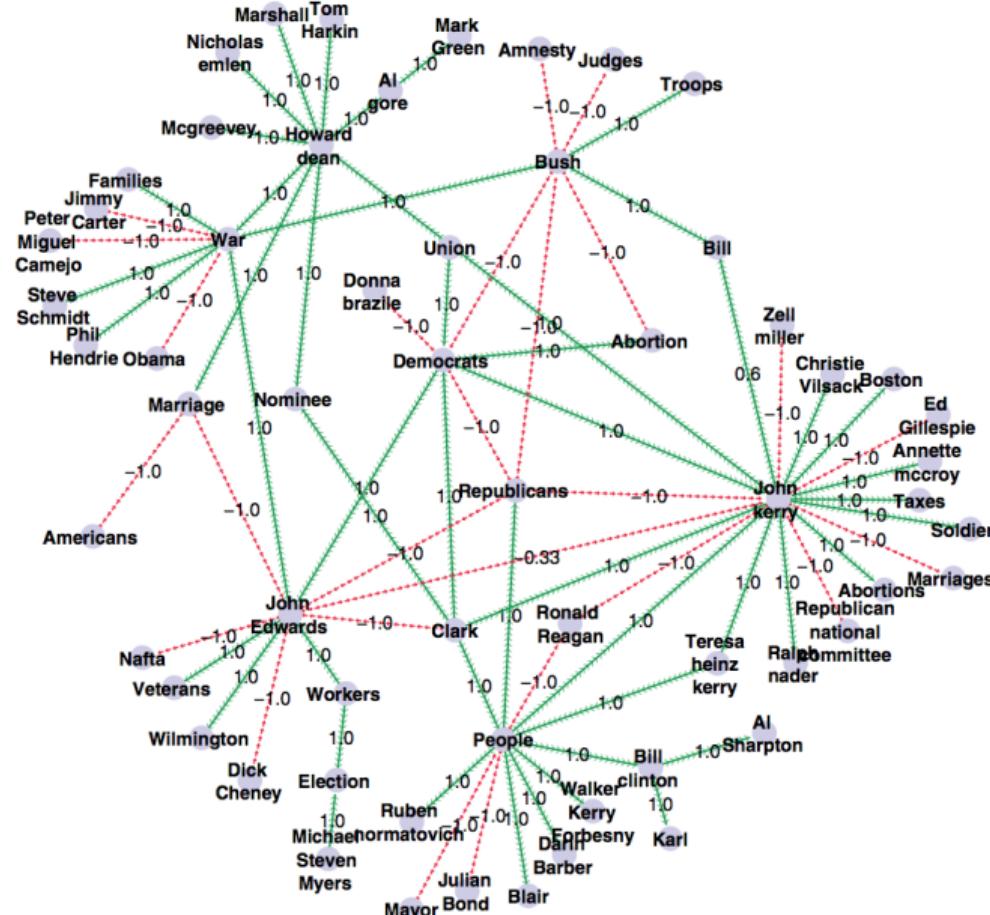


Figure 4: Network with positive and negative edges between “entities” mentioned in the NYT (US presidential election data: January to August 2004) (Sudhahar et al., 2015).

How do we formally represent networks?

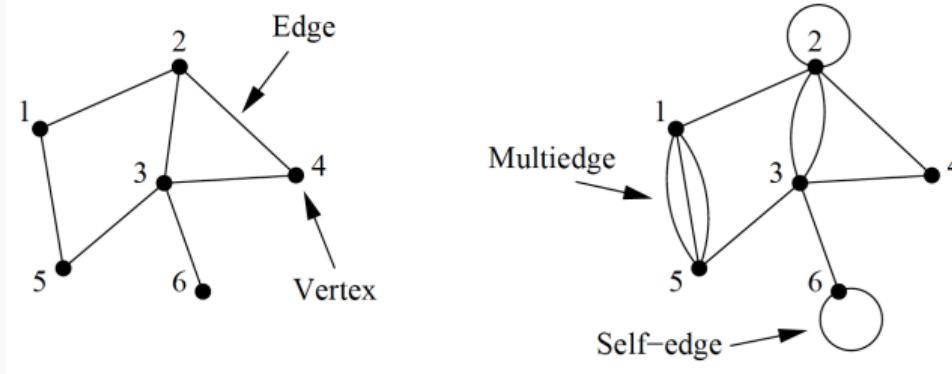


Figure 5: Traditional visualisation of two small networks...

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 6: ... and the adjacency matrix of the left-hand network (Newman, 2010, p. 111).

How do we formally represent networks?

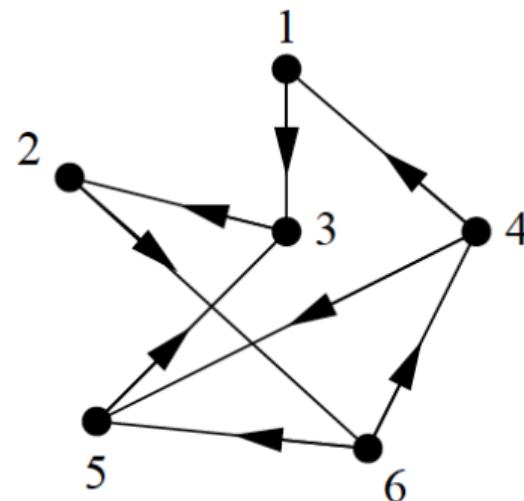


Figure 7: A directed network...

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 8: ... and its adjacency matrix (not symmetric!) (Newman, 2010, p. 112).

What can I use network analysis for?

- Understand the **role** that nodes play within the network: for example a node can be central, peripheral, a broker connecting different parts of the network or relatively “influential” if it is connected with many other nodes.
- Understand the **structure** of the overall network: for example, we would expect a social network to be structurally very different from a network of followers/followees on a social media website.

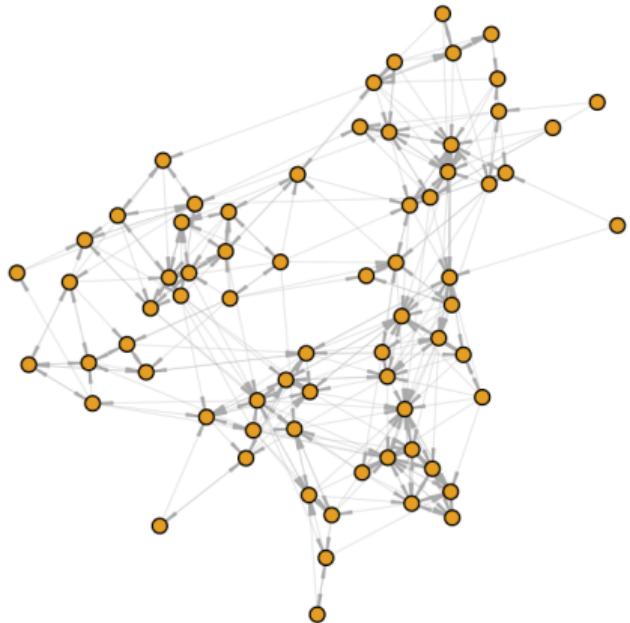


Figure 9: A friendship network of highschool students

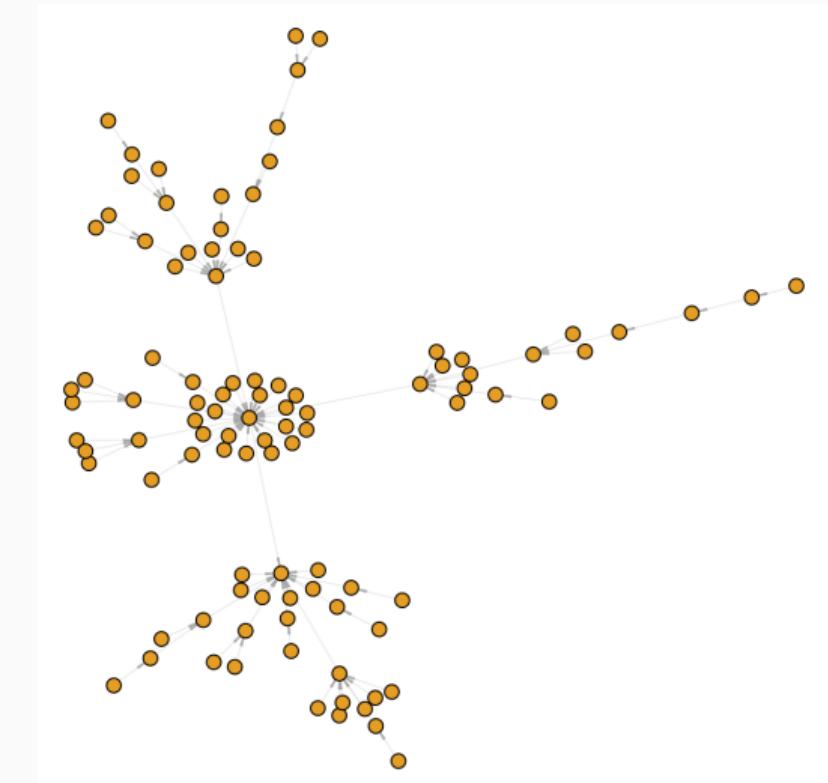
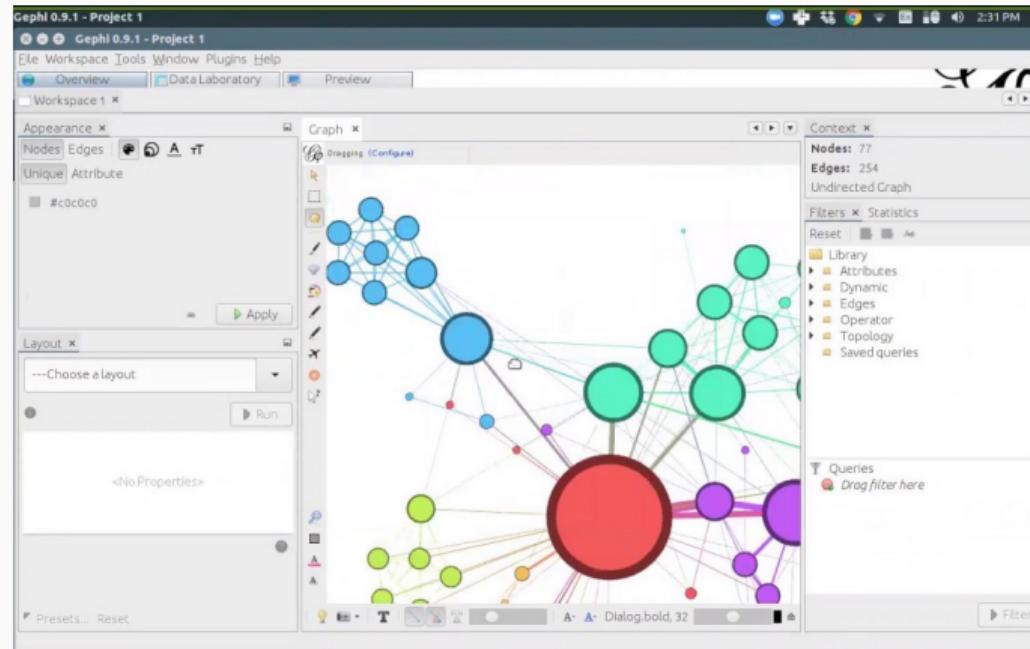


Figure 10: Followers and followees on Twitter

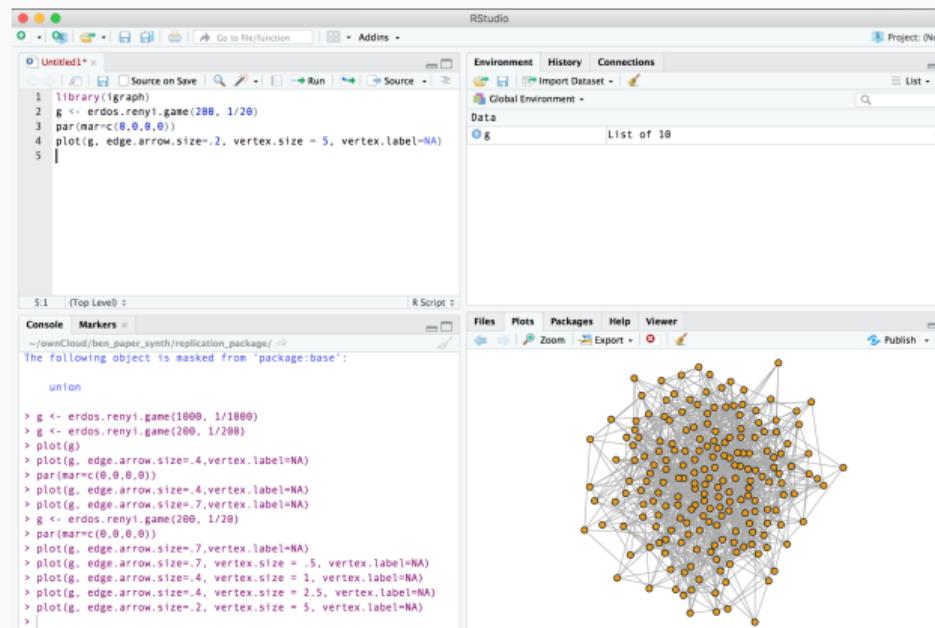
Tools

Easy, small n: Gephi (gephi.org)



Tools

Hard, big n igraph package (igraph.org) in R (www.r-project.org) or Python (www.python.org)



Bibliographic resources

Getting started bibliography:

Easy Scott, *What is social network analysis?*, 2012

Important Marin and Wellman, 'Social network analysis: An introduction', 2011

Hard Newman, *Networks*, 2010

Online resources

Tutorials for beginners by Katherine Ognyanova (Rutgers University):



- Network visualisation with Gephi (kateto.net/sunbelt2016)
- Network visualization with R (kateto.net/network-visualization)
- Network Analysis and Visualization with R and igraph
(kateto.net/networks-r-igraph)

Online deliberative discourse

The Five Star Movement



(a) (Credit: Enrico Lo Storto)



(b) (Source: 'Logo MoVimento 5 Stelle', 2013)

Timeline

- January 2005: www.beppegrillo.it launched

“What’s politics? Nobody knows it anymore. Does it still make sense to talk of Right, and Left and centre?
Maybe it makes more sense to talk of above and below. [...] In politics we don’t need a leader, we are grown up
people. We need a vision of the world [...].”

(28 January 2005)

- June 2005: First M5S Meetup

“I thought on how to do to give all who follow my blog the opportunity to meet to discuss, take the initiative,
see each other in person. To transform a virtual discussion into an opportunity to change. I discussed with my
collaborators and I decided to use MeetUp. MeetUp is a site that allows to organise in a simple way meetings
among people interested in a topic.”

(16 July 2005)

Timeline

- 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 25.50 percent popular vote (1st party with 8,689,168 votes)
- May 2014: M5S won 21.16 percent popular vote (2nd party with 5,792,865 votes)
- March 2018: M5S won 32.70 percent populat vote (1st party with 10,732,066 votes) and now is in the government coalition.

Online deliberative discourse on the GMI

I present now a discourse analysis based on natural language processing of the deliberation process for the institution of a guaranteed minimum income.

The analysis explores how the discourse on the GMI took shape in the online discussion of the online community of citizen-users, and in the broader national public sphere.

I argue that

1. The online debate contributed significantly in pushing the GMI onto the public agenda and to parliamentary deliberation
2. As the idea was discussed in the commenting sections, Grillo progressively legitimised it by presenting it as *economically* sound.
3. The *label* of the policy idea was appropriated by the Movement in the run-up to the general election of 2013 and it trickled down to the discourse of candidates of other parties.
4. After the general election, at least three parties, including the M5S, discussed the ‘citizen’s income’ in Parliament and bills to introduce nation-wide some sort of GMI were presented.

Mapping documents in the concept space

The problem How to make sense of thousands of comments posted in an online forum over the years? Specifically:

1. How comments relate one with the other, and
2. How comments relate to meaningful **concepts**.

Solution 1 We could attempt to use a **topic model**, so to relate each **document** to a **topic** (and consequentially each comment to the other comments). But this doesn't solve the problem of topics being abstract representation and not necessarily meaningful.

Solution 2 Instead, we could measure the “distance” separating each comments and a limited number of documents, each representing a well defined *semantic concept*.

Documents position within the concept space

The **concept space** can be thought of *spatially*. Individual concept documents have a position within the concept space relatively to all the other concept documents, with their distance being measured based on the relative frequency of their terms.

But what can be used as concept document?

Gabrilovich and Markovitch, 'Computing semantic relatedness using Wikipedia-based explicit semantic analysis.', 2007 proses to use a selection of *Wikipedia articles*.

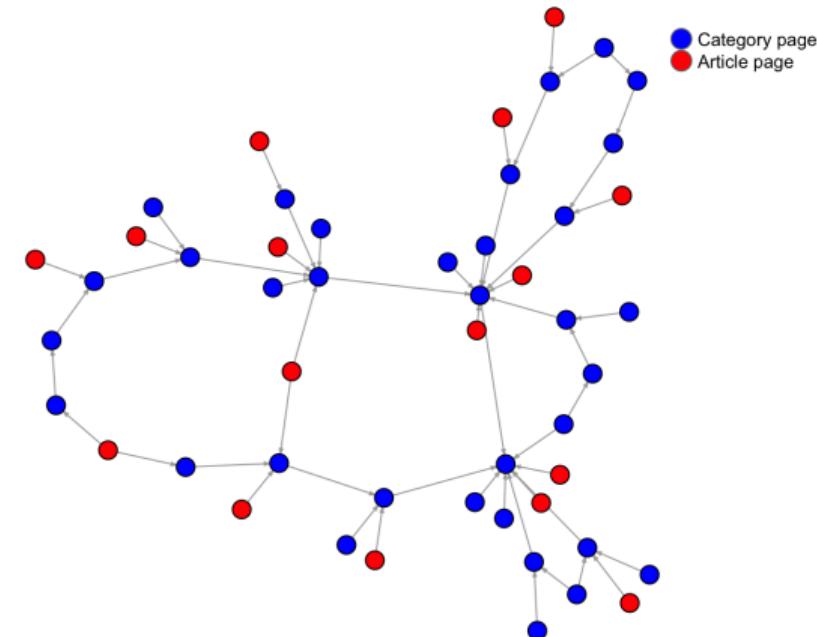
Documents position within the concept space

Wikipedia articles have two important characteristics:

1. They are precisely, concisely and meaningfully identified by their title (e.g. “Car”, “Leonardo da Vinci”, “Poverty”)
2. The scope of the article itself is limited to the subject of the article’s title; that is, the terms used in the article are all related to the title.

Define the concept space: the problem

Yet, there are more than 5 million articles in the English Wikipedia. How do we only select articles that can help our analysis? We can leverage the networked structure of Wikipedia: with Categories hierarchically linked to Subcategories and Articles linked to Categories.

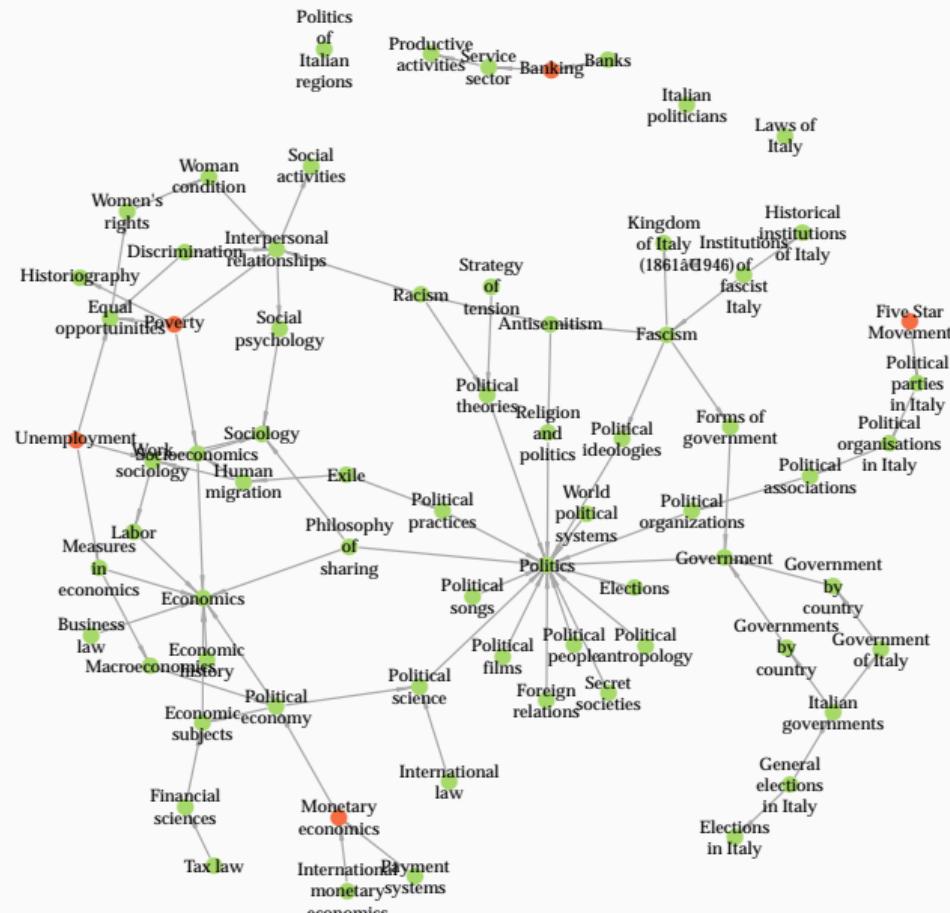


Define the concept space: a solution

I set nine *macro*-categories: Poverty, Unemployment, Banking, Income, Politics, Politics of Italy, Monetary economics, Constitution of Italy and Five Star Movement.

From these nine *macro*-categories I queried the network for all categories with a maximum of two-degree (two hops) separation from them, obtaining a set of 540 categories.

All articles not belonging to these 540 categories were removed, bringing the final number of Wikipedia articles for the concept space to 4827.



Define the concept space: a solution

Finally, the distance between each comment and the 4827 concept documents (the Wikipedia articles) forming the concept space was measured based on the frequency of terms contained by the comments and the frequency of terms contained by the Wikipedia articles.

In practice, I obtained two results:

- I could determine the importance of each concept in relation to each comment, and
- I could determine the closeness of comments.

These two results allowed me to infer what were the main concepts framing the debate and how these concepts changed overtime.

Example of concepts assigned to a post

1) we are talking about basic income, non-guaranteed minimum income 2) and in any case with the various cuts you don't reach 30 billion per year 3) money to subsidise idleness and unproductivity? absurd and who should create jobs since companies are closing because Italy is a tax hell? the state? Another fake welfarism. 4) with both citizen's income and with GMI the general Italian mentality is fraudulent so it would cost even more than any normally budgeted calculation.

Concepts (Wikipedia articles)

Productivity; Tax wedge; Fiscal cliff; Rebirth of Christian Democracy; Profit; Government agency; Tax haven; Social stigma; Italian fiscal code card; Declaration of taxable income

The GMI public debate in Italy

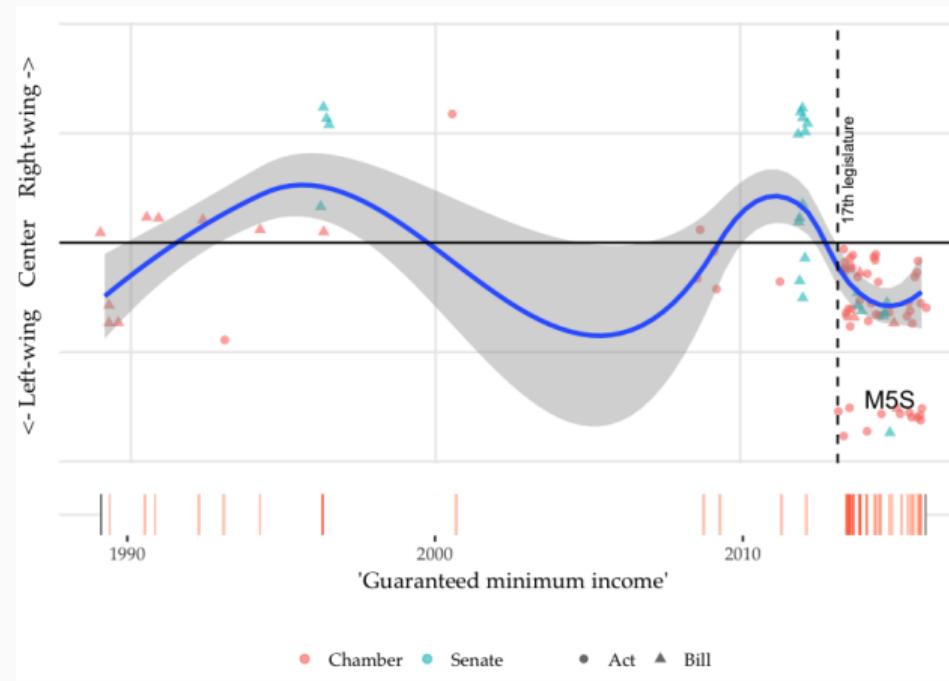


Figure 12: Number of acts and bills presented mentioning 'guaranteed minimum income'

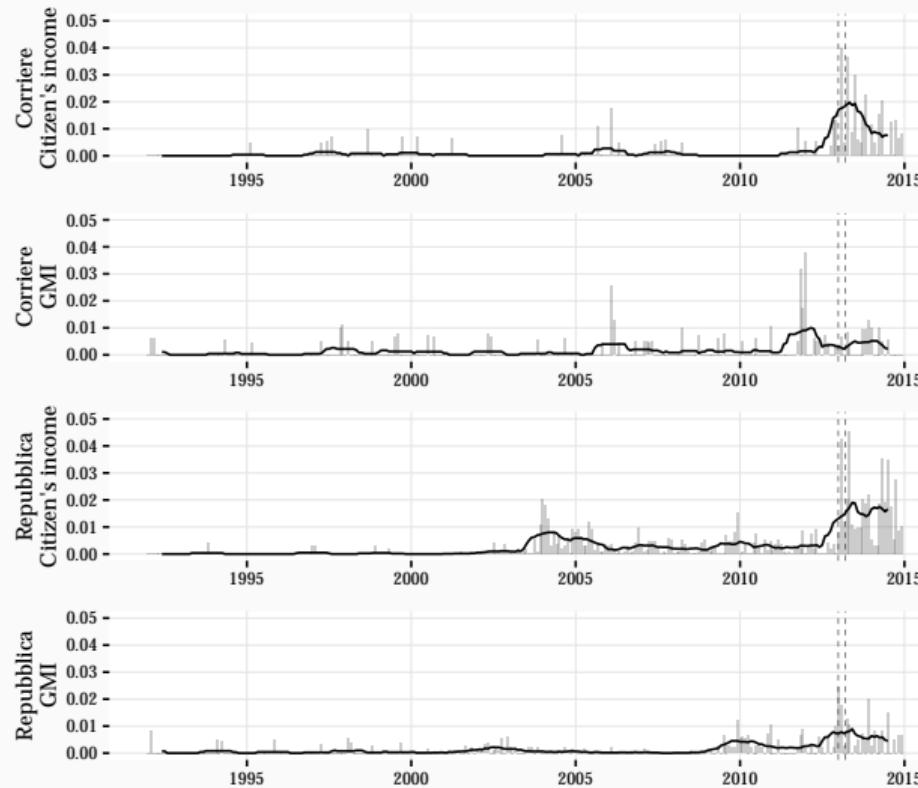


Figure 13: Number of articles mentioning 'citizen's income' as fraction of articles mentioning 'politics' published in the same period.

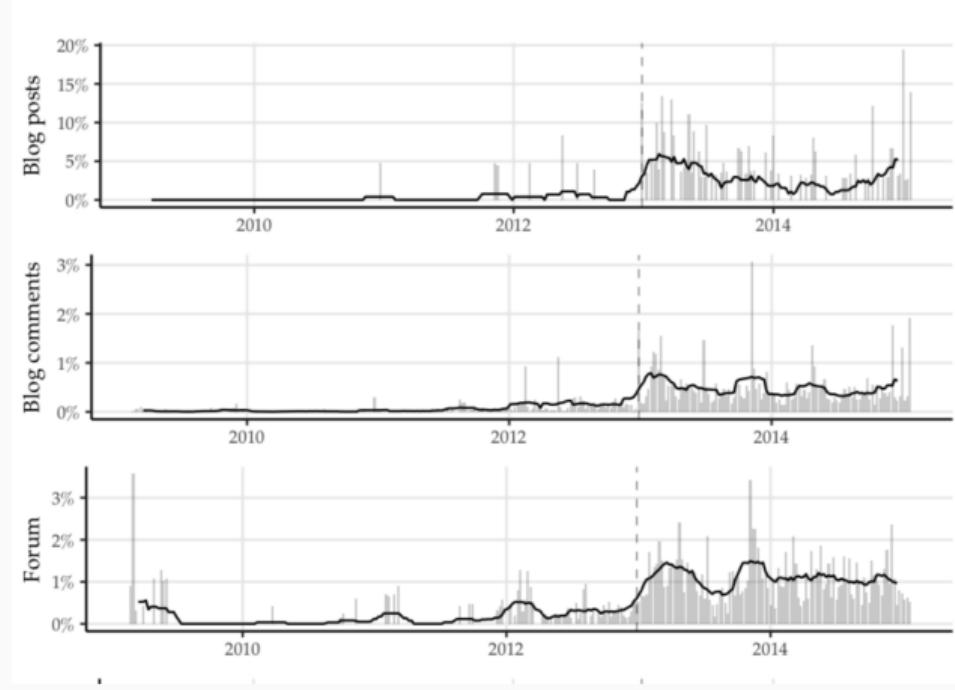


Figure 14: Number of postings mentioning 'citizen's income' or 'GMI'

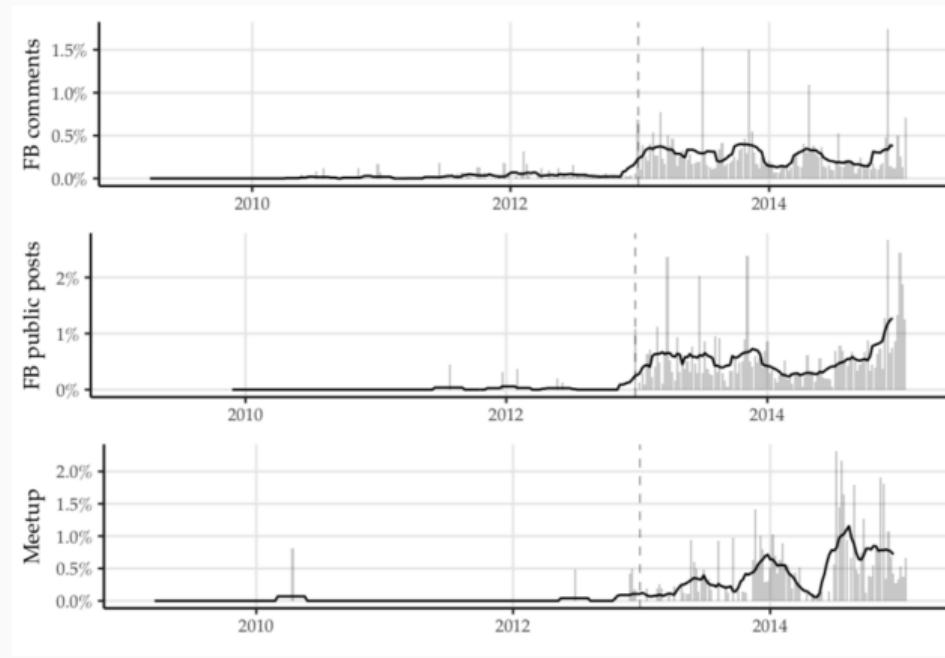
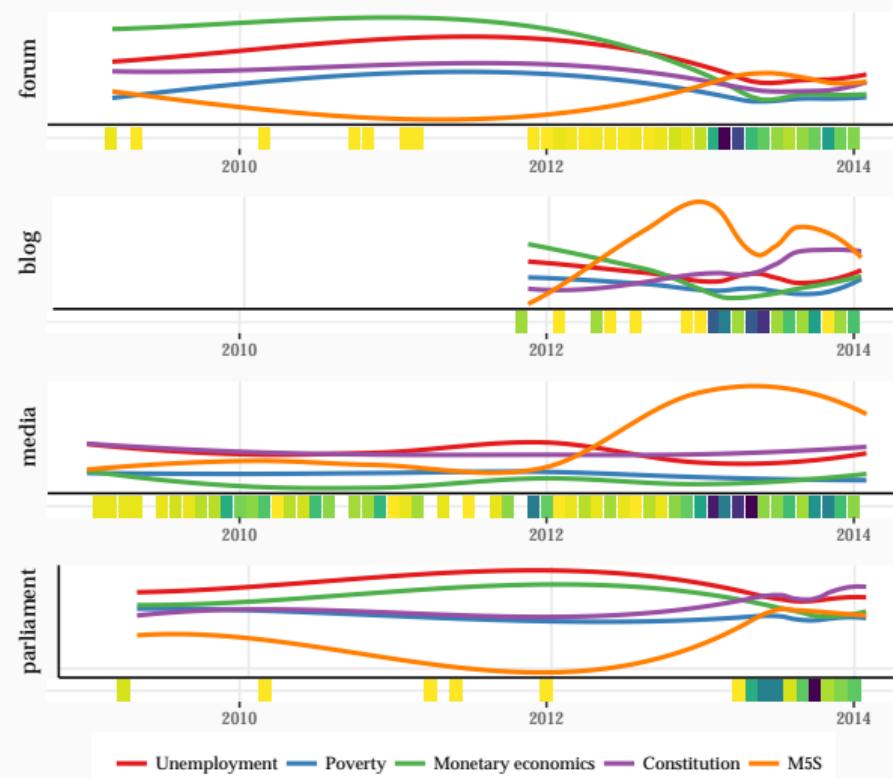


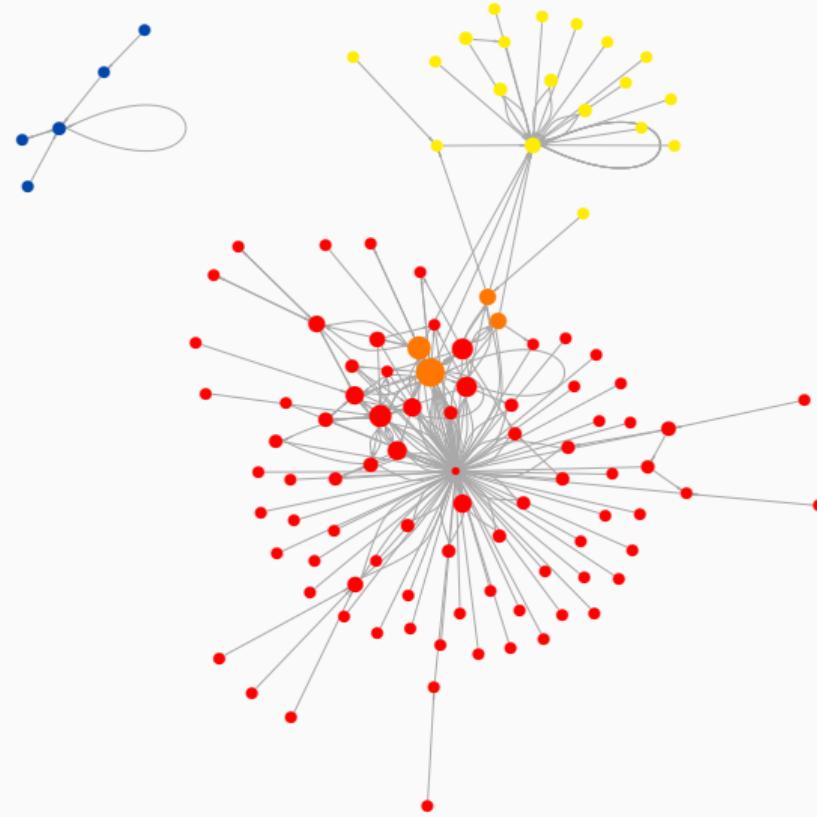
Figure 15: Number of postings mentioning 'citizen's income' or 'GMI'

Comparing the GMI online debate with the public debate based

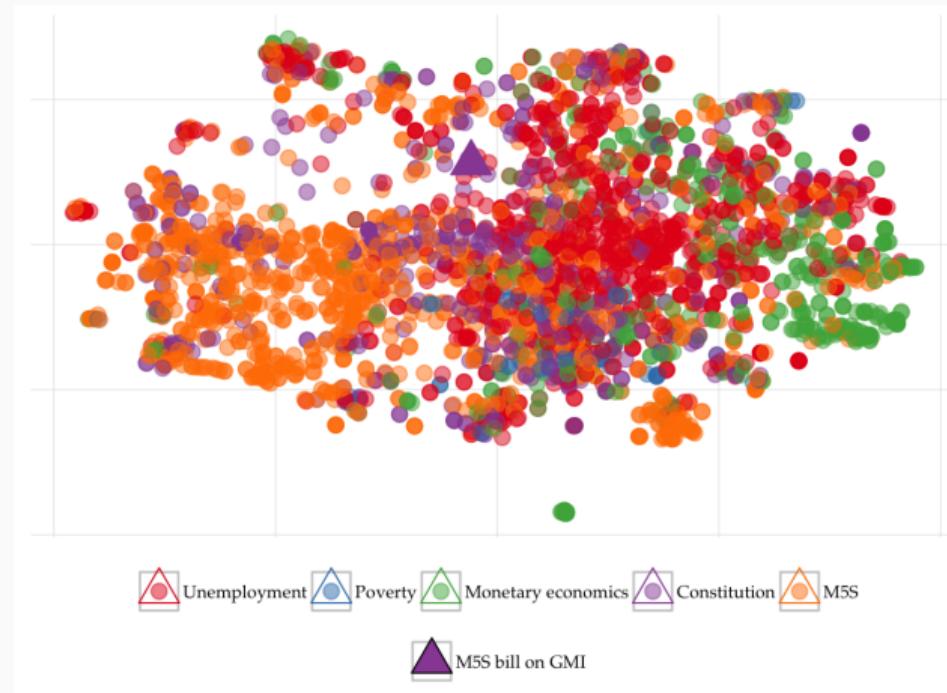
Weight of concepts in the discussions (lines) and frequency of interventions (heatmap)



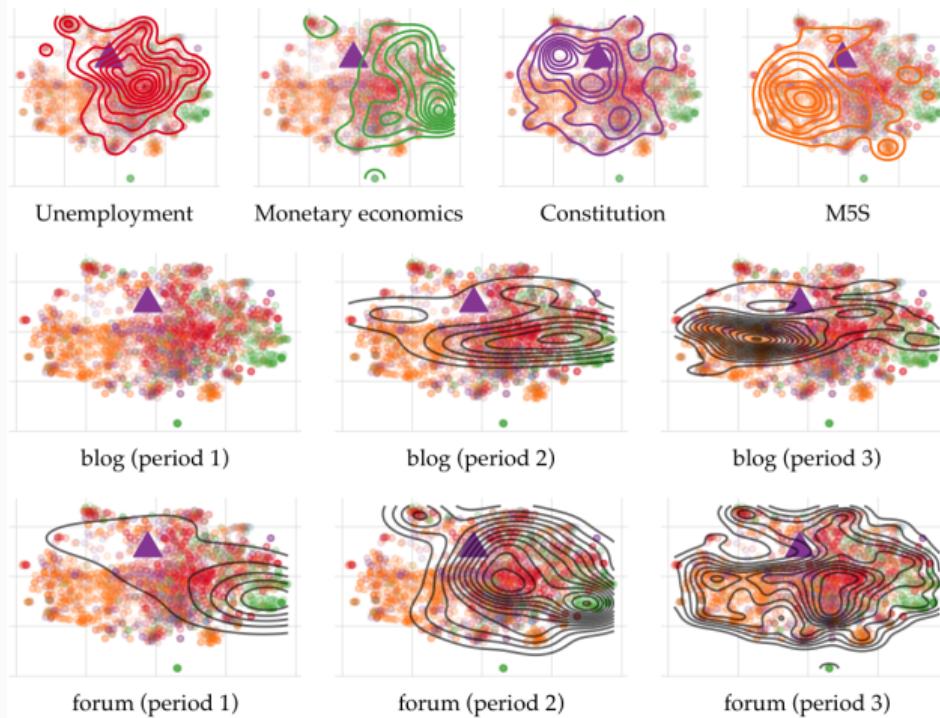
Users active in three selected threads: on bank seigniorage (red nodes), GMI (yellow nodes), on value added tax (blue nodes) and on both seigniorage and GMI (orange nodes). Direct links represent a comment from user to user and nodes' size are proportional to the level of activity of the corresponding user.



2D visualisation of comments within the concept space



2D visualisation of comments within the concept space



Conclusions

1. The online fora were able to push an item onto the agenda not against a set of strong alternatives but in the absence of any.
2. Online deliberation by itself had a significant impact only when its discourse was endorsed by Grillo; the impact of online deliberation was mediated by the leadership of the Movement.

References

-  Gabrilovich, E. & Markovitch, S. (2007). Computing semantic relatedness using Wikipedia-based explicit semantic analysis.. In *Ijcai* (Vol. 7, pp. 1606–1611). Retrieved February 21, 2016, from
<http://www.aaai.org/Papers/IJCAI/2007/IJCAI07-259.pdf>
-  Logo MoVimento 5 Stelle. (2013). Retrieved May 2, 2014, from
http://it.wikipedia.org/wiki/File:MoVimento_5_Stelle_logo.png
-  Marin, A. & Wellman, B. (2011). Social network analysis: An introduction. In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 11–25). London: SAGE Publications.
-  Newman, M. E. J. (2010). *Networks: An introduction*. Oxford: Oxford University Press.

-  Scott, J. (2012). *What is social network analysis?* London: Bloomsbury Publishing.
-  Sudhahar, S., Fazio, G. D., Franzosi, R. & Cristianini, N. (2015). Network analysis of narrative content in large corpora. *Natural Language Engineering*, 21(1), 81–112.