# Detecting Opinion Polarisation on Twitter by Constructing Pseudo-Bimodal Networks of Mentions and Retweets

Igor Zakhlebin[1], Aleksandr Semenov[1],
Alexander Tolmach[2], and Sergey Nikolenko[3,4,5]

[1] International Laboratory for Applied Network Research,
National Research University Higher School of Economics, Russian Federation.
{izakhlebin, avsemenov}@hse.ru
[2] Institute of Sociology, Russian Academy of Sciences, Moscow, Russia
quatsch.ad@gmail.com
[3] Kazan (Volga Region) Federal University, Kazan, Russia
[4] Laboratory for Internet Studies,
National Research University – Higher School of Economics, St. Petersburg, Russia
[5] Steklov Institute of Mathematics at St. Petersburg, Russia
sergey@logic.pdmi.ras.ru

**Abstract.** We present a novel approach to analyze and visualize opinion polarisation on Twitter based on graph features of communication networks extracted from tweets. We show that opinion polarisation can be legibly observed on unimodal projections of artificially created bimodal networks, where the most popular users in retweet and mention networks are considered nodes of the second mode. For this purpose, we select a subset of top users based on their PageRank values and assign them to be the second mode in our networks, thus called pseudo-bimodal. After projecting them onto the set of "bottom" users and vice versa, we get unimodal networks with more distinct clusters and visually coherent community separation. We developed our approach on a dataset gathered during the Russian protest meetings on 24th of December, 2011 and tested it on another dataset by Conover [13] used to analyze political polarisation, showing that our approach not only works well on our data but also improves the results from previous research on that phenomena.

**Keywords:** Twitter, opinion polarisation, two-mode networks, community detection

## 1 Introduction

Twitter has become one of the most popular social networking services among researchers due to the open nature of its communication and relatively easy access to its data via the API (Application Programming Interface). The scope of previous Twitter-related research includes detection of the users' psychological features [14], spread of diseases [2] and response to natural disasters [24], analysis of financial markets [6], electoral predictions [15], and marketing campaigns [10]. One of the most scrutinised directions of study, however, concerns the protest movements in Twitter like the "#occupy" movement [4] or the so-called Twitter revolutions in the Middle East [12].

In this direction, researchers compared language use in Egypt and Libya [7], analysed types of actors [18] and measured the recruitment patterns and dynamics [19]. All these topics are related to a general question about political polarisation on Twitter because it can be used by all sides of the conflict in question to promote their point of view, strengthen their group identity and discriminate the opposite sides [25].

One of the most famous examples of political polarisation on the Internet was presented in [1]. The authors used a network approach to analyze hyperlink patterns among US political blogs during the presidential campaign of 2004 and demonstrated highly separated nature of pro-Republican and pro-Democrat parts of the blogosphere. Although there is other evidence that hyperlinks in blogs can serve as a signal of ideological affiliation [20], applying this approach to Twitter might not work well because hyperlinks are used sporadically, don't stay visible for long as the timeline fills with another updates, and can be used by all sides of the conflict in both positive and negative way. This suggests that analysis and visualisation of networks based on hyperlinks will not result in a clear picture of community structure. Some of these and other important differences between hyperlink usage in blogs and Twitter have been discussed in [9].

Another approach to detecting the stance of Twitter users on an issue of interest is based on the usage of keywords or hashtags related to that issue. In practice, searching for a hashtag is one of the most popular ways to get a sample of tweets [8]; however, it might introduce its own problems with the bias of that sample. For example, data gathered from trending hashtags during other protest meetings in Russia showed that these hashtags form two distinct clusters with pro-opposition and anti-opposition tweets [22]. However, both hashtags and clusters they represent contain words with clearly negative connotations and do not show more cautious or casual opinions on the events that use more neutral synonyms like "meeting", "march" etc.

We have not been able to find any kind of network analysis on the resulting dataset in previous work, although network analysis had proven to be useful in similar situations. For example, in [13] the authors gathered hashtags associated with Democratic and Republican parties for several weeks during the midterm elections to the US Congress in 2010. It goes without saying that political polarisation discovered in this paper was not a huge surprise; however, it was discovered only for the retweets while networks of mentions were more homogeneous with a low modularity score of 0.17.

Although these results look well-grounded and reasonable, there is one caveat in the general approach. Gathering trending or politically biased keywords may result in biased and polarised datasets. But when we need to cover the entire spectrum of opinions, we need to gather as much neutral keywords or hashtags as possible. In this work, we suggest an approach to this problem similar in spirit to the work [17]. In that paper, the authors classified ordinary users of Twitter and media outlets via the politicians whom these users follow on Twitter. The rationale behind this is simple – the number of prominent politicians and Congress members is limited and their position on the political spectrum is well-known. Therefore, it is reasonable to assume that their followers share that position and hence put the main media outlets on this continuum through their profiles.

Since Twitter had changed its API limits for gathering data on followers, making it almost impossible to build large graphs on that type of relationship, we decided to apply this logic to the networks built from retweets and mentions. Previous research demonstrated that users tend to retweet those whose ideas they share [9,13,25] and that there are very few popular, prominent, and central users [3] who can serve as such opinion leaders and whose influence does not depend on their followers count alone [11]. This allows us to assume that these "influencers" might be seen as a special type, or, in network terms, a "second mode" of users. Hence, we can analyze unimodal networks of user communication as if they were bimodal networks, artificially separating top (most popular) users into a second mode[6]. We analyze these networks as if they were "normal" bimodal networks by projecting sets of both "top" and "bottom" users on each other to obtain two separate unimodal networks for "top" and "bottom" users. The results show that this approach leads to more distinct clusters in the Twitter mention networks compared to standard analysis of bimodal networks constructed from hashtags and hyperlinks.

The rest of the paper is organised as follows. In the next section, we describe the network features of communication on Twitter with an emphasis on data acquisition and network extraction methods. In Section 3, we describe our approach to the polarisation discovery via unimodal projections of pseudo-bimodal networks. In Section 4, we describe our dataset and its background. Section 5 demonstrates the results of our analysis. Finally, in conclusion we discuss possible limitations of our results and how further work can help avoid them and improve our approach.

## 2 Communication networks on Twitter

On Twitter, users communicate by posting short public text statuses, often containing hyperlinks and pictures, called "tweets". To indicate that a tweet deals with a certain topic, users insert "hashtags" in their tweets which consist of the number sign ("#") followed by an alphanumeric combination denoting the topic (like "#tcot").

There are three basic ways how users can interact. Users can "mention" particular persons by including the recipients user name prefixed with "@" sign in the tweet (like "@navalny"). Such a combination is called a Twitter "handle".

[6] The term pseudo-bimodal networks is based on the previously introduced notion of pseudo-tricluster in two paired bimodal networks [16]

To show support, users can quote tweets to their timelines, thus sharing them with their subscribers; this is called a "retweet." Retweets start with "RT" and a handle. A tweet immediately starting with a handle is called "reply" and is considered to be a direct message from one user to another.

Although there are substantial differences between mentions, retweets, and replies, we define a mention as any occurrence of a user's handle in a tweet. We did it mostly because other authors rarely analyze reply networks on their own and the most common type of networks used in the analysis are retweets and mentions. Moreover, the dataset from [13], which we use as a test case for our approach, follows that classification too. Thus, in this paper mentions formally include both replies and retweets.

As any mention denotes a reference from one user to another, two types of directed networks were constructed from it: mentions and retweets. Nodes in these networks stand for the users and edges denote the chosen type of interaction; they are directed from the users who posted tweets to the users being mentioned in them.

The fact that any user can see any tweet lets any user freely gain popularity on Twitter. As a consequence, some individuals can gain influence comparable to and even surpassing that of organisations like news companies represented on the service. Such important influencers generally include politicians, individual bloggers, and celebrities.

To measure user influence in terms of network topology, let us consider simple measures of their centrality. A certain set of nodes has small out-degrees and large in-degrees. Those users produce mostly original tweets that get mentioned often; in what follows we call them "top users". For most other nodes, the out-degrees are larger than in-degrees, which is commonly interpreted as their activity score. These are the users that retweet others but do not get retweeted often; in what follows we call them "ordinary users". Another network metric, which demonstrates the intuition that prominence of a user is defined by centrality of his peers, is PageRank [23] and it is used here in further analysis.

Another observation on our data is that top users rarely mention each other. Most interactions happen between the ordinary users and the top ones. Thus, if we try to pull the network of mentions between top users, it will be too sparse to search for communities in it. On the other hand, the full network of interactions may be too strongly interconnected to effectively partition it as well [13].

These observations allow us to cluster ordinary users according to which top users they retweet and mention. To reiterate, our approach is based on the following assumptions:

- there exists a small group of users in the network with high PageRanks;
- they rarely interact with each other and with ordinary users;
- ordinary users tend to mention mostly top users with whose opinions they agree;
- users from both sets predominantly belong to one group each according to their opinions.

# 3 Method for detecting opinion polarisation

Our proposed algorithm receives a directed communication network $G = (V, E)$ as an input, where $V$ and $E$ are, respectively, its sets of nodes and edges. The algorithm consists of the following operations sequentially performed on $G$:

1. *Select a set of top users for some threshold $k$.*

   To separate top users from the ordinary ones, we sort the corresponding nodes by their PageRank values and simply select $k$ nodes with highest values. This splits them in two disjoint sets: top users $V_T$ and ordinary users $V_B$, $V = V_T \cup V_B$, $V_T \cap V_B = \emptyset$.

2. *Make the network bimodal.*

   To complete separation of top users into the second mode, all edges in $E$ between nodes of the same set ($V_T \times V_T \cup V_B \times V_B$) are removed. This produces a bipartite graph which we call the *pseudo-bimodal* network, $G^* = (V_T, V_B, E^*), E^* \subset E$. Its edges show how did the ordinary users mention top ones and vice versa. It is subsequently analysed as if it were a regular bimodal network.

3. *Project the pseudo-bimodal network onto one of its node sets.*

   Having constructed the pseudo-bimodal network $G^*$, we can either study ordinary users by their mutual connections to the top ones or study top users by intersecting their audiences, i.e., subsets of ordinary users mentioning them. For that purpose, we use Newman's two-mode projection method [21] to get a unimodal undirected weighted network built on a selected set of users (that is, a projection of the network on the set $V_T$ or $V_B$). We begin by defining this process for the projection of $G^*$ on the set of top users $V_T$. For a pair of nodes $i, j \in V_T$, $L_{i,j}$ denotes the set of nodes connected to both $i$ and $j$, $L_{i,j} = \{l \in V_B | (l, i) \in E^*, (l, j) \in E^*\}$. Both $i$ and $j$ will occur in the projected one-mode network, and they will be connected iff the set $L_{i,j}$ is nonempty. The edge between $i$ and $j$ is weighted as $w_{i,j} = \sum_{l \in L_{i,j}} \frac{1}{k_l - 1}$, where $k_l = |\{i \in V_T | (l, i) \in E^*\}|$. Projection on the set $V_B$ is done similarly.

4. *Perform community detection on the resulting one-mode network.*

   The one-mode network obtained on the previous step is expected to have a more expressed structure with more tight-linked communities and higher modularity. To partition the users into groups with presumably similar political biases we use the Louvain graph clustering method [5], one of the best known and widely used methods for community detection, on the resulting one-mode network. The Louvain method looks for a graph partitioning that maximizes *modularity*, i.e., density of links inside communities compared to links between communities. Modularity is defined as $Q = \frac{1}{2m} \sum_{i,j} (1 - \frac{k_i k_j}{2m}) \delta(c_i, c_j)$, where $m$ is the total number of edges in the graph, $k_i$ is a degree of node $i$, $c_i$ and $c_j$ are the communities of the nodes, and $\delta$ is a delta function ($\delta(c_i, c_j) = 1$ if $c_i = c_j$ and 0 otherwise). $Q$ varies between 0 and 1, with 1 corresponding to a perfect separation of nodes, i.e., no edges between different clusters.

   As a result, we find community structures among top and ordinary users. These community structures for ordinary users represent how often they retweet and/or mention the same top users (i.e., whether they follow the same issues); for

| Name | Data sources | Number of | | | |
|---|---|---|---|---|---|
| | | users | tweets | mentions | retweets |
| 24th December, Russia | Streaming and Firehose API | 3,485 | 24,378 | 12,725 | 6,529 |
| U.S. Elections | Firehose API | 45,000 | 250,000 | 77,920 | 61,157 |

Table 1: Descriptive statistics for the datasets

top users they show how often they are mentioned by the same ordinary users (i.e., how much their audiences overlap). This leads to a different and more pronounced community structure, as we will see in practical examples below.

# 4 Datasets

We have used the following datasets related to political polarisation.
1. *Meetings on December 24th, 2011 in Russia.*

As a main source of data we used tweets on pro-government and protest political events happened in Moscow during December 24th, 2011 on Poklonnaya Gora and Prospekt Sakharova. We have collected them using Twitter's Streaming API and Firehose. The first is Twitter's own free source of data, which contains a 1% sample of all the tweets. Firehose is a full stream of tweets with a cap of 500,000 tweets per hour, provided on a commercial basis by DataSift (www.datasift.com). To collect only tweets that refer to political events, we filtered them according to hashtag "24дек" ("#24dec", short of Russian "December 24th"), which was heavily used by both sides during that day and did not favor any particular position. Thus, we gathered 24,378 tweets from 3,485 unique users with 12,725 mentions, 6,529 of which were retweets.
2. *U.S. midterm elections to the Congress in 2010.*

This dataset was used in the work [13] and has been made public. We use it to test our approach on similar data from a similar context because it is one of the rare cases of publicly available datasets from Twitter. Descriptive statistics of both datasets are provided in Table 1.

# 5 Results

From these datasets, we have constructed:

- two standard bimodal networks: a) network of users and hashtags which they used in their tweets (hashtag network), and b) network of users and domains to which these users referred via hyperlinks (domain network);
- two unimodal networks with respect to retweets and mentions.

We analysed the latter two as bimodal networks. To transform communication networks into bimodal ones, we chose the top 100 users by PageRank as the second mode in the network and projected the result onto a unimodal network with Newman's method [21] to see how they are connected among each other via ordinary users who reply to and/or retweet them. Then we clustered each network with the Louvain method and used its modularity coefficient, which is one of the most common measures of clustering quality on networks.

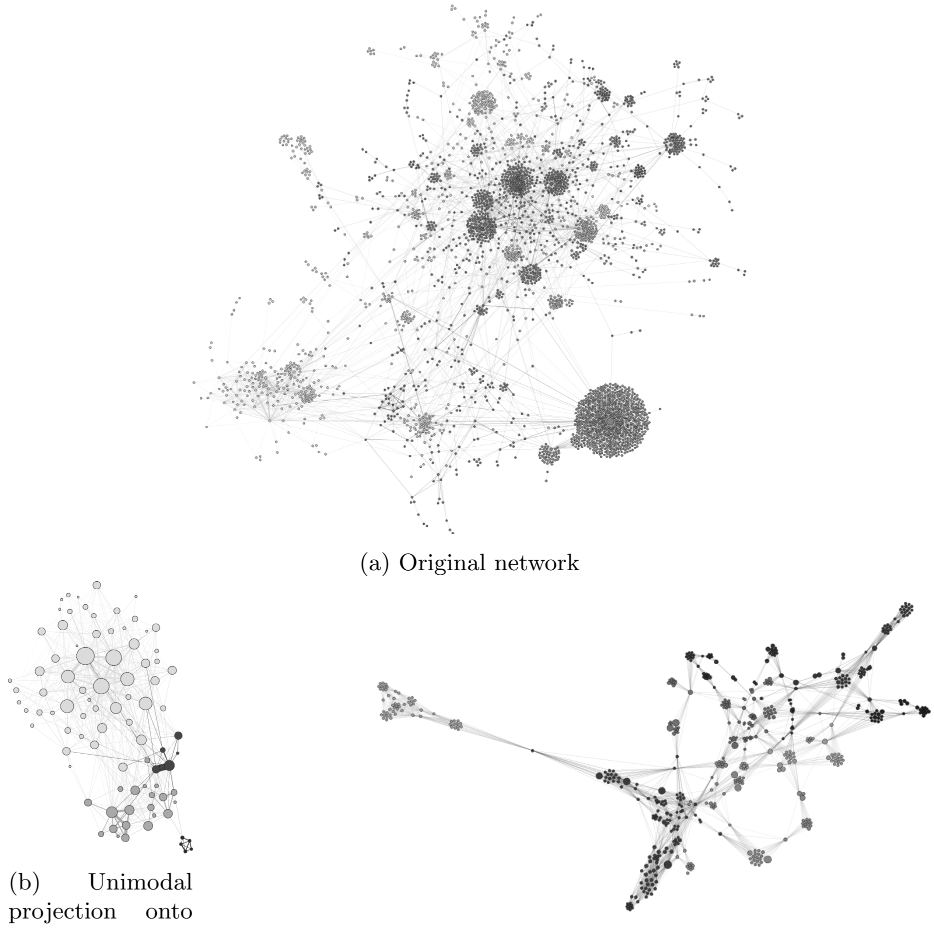
As expected, results from projections of bimodal networks constructed from hashtags returned the least readable clusters of users. With modularity score of 0.122, this network contained neutral hashtags with dates, names of cities, and other non-polarising keywords. The unimodal projection of the network constructed from URLs was a bit better in terms of modularity (0.485) but contained such hubs as youtube, livejournal, twitter, facebook, and vk.com, which once again were neutral in terms of possible content and usage. We view these results as evidence for the fact that unimodal networks from hashtags and URLs do not detect clusters particularly well in our case.

After the clustering, for our dataset collected during the events of 24 Dec. 2011, we get unimodal networks of retweets presented on Fig. 1 (for $k = 100$). It is clear that projected graphs are much better structured: connections are dense inside the clusters and sparse between them. What is even more surprising, this method also works for networks of mentions which are usually considered to be more homogeneous [9, 13]: both in our dataset and the test data from [13] the clusters obtained after projection are much better defined. Figure 2 shows how modularity of both unimodal networks changes with the percentage of top users; observe that even for a small number of top users (left part of the graph), which, naturally, do not form a modular graph, the modularity of the "ordinary" part of the graph increases.

## 6 Conclusion

We have proposed an approach to detect and explore opinion polarisation in Twitter communication networks, which leads to better defined clusters of users than methods employed in previous works. We have shown that our method works not only on our dataset, but also on a classical dataset previously used in literature, improving the quality of clustering.

However, our approach has some limitations. First, it would be good to have a mathematical proof that our results are not an artifact of bimodal networks and projection methods; this concern might also be solved via simulations or an analytic solution. Second, currently we have only analysed data with political origins, which are polarised by nature. Perhaps, in more homogeneous contexts such as tweets from scientific conferences or pop culture entertaining events this approach will not work so well. Hence, we need to test our approach on more datasets from different contexts on Twitter and maybe in other social media and domain areas. Third, although we have managed to improve upon the results of [13], we cannot verify all conclusions since not all information on the dataset has been provided by authors and also because of our lack of substantial knowledge of US political situation both offline and in Twitter.



(a) Original network

(b) Unimodal projection onto set of top users

(c) Unimodal projection onto set of ordinary users

Fig. 1: Variants of retweet network based on 24 Dec. 2011 dataset

Therefore, as further work we plan to test the approach analytically and via simulations, try different centrality measures (eigenvector, HITS), projection methods, and cut-off values, use semi-supervised annotation of tweet texts and labeling to attain a ground truth about existing opinions and clusters, add non-political and non-Twitter datasets into the analysis, and, finally, test how user preferences persist through time (across several datasets about Russian political events) and see if user separation remains stable.

**Acknowledgements** This paper was prepared within the framework of a subsidy granted to HSE by the Government of Russian Federation for implementation of the Global Competitiveness Program. The work of Sergey Nikolenko was supported by the Russian Science Foundation grant no. 15-11-10019.

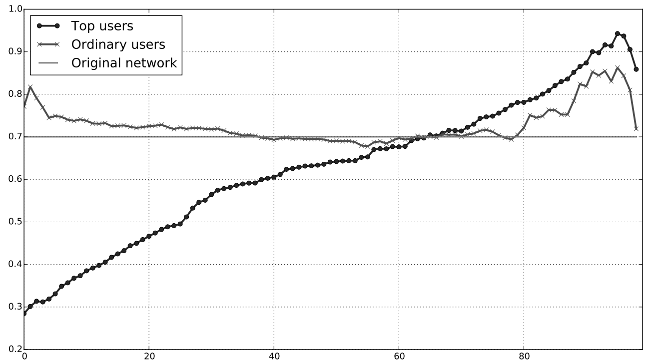



Fig. 2: Modularity as a function of the cutoff value for the Conover dataset.

## References


1. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 u.s. election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery. pp. 36–43. LinkKDD '05, ACM, New York, NY, USA (2005), `http://doi.acm.org/10.1145/1134271.1134277`
2. Aramaki, E., Maskawa, S., Morita, M.: Twitter catches the flu: detecting influenza epidemics using twitter. In: Proceedings of the conference on empirical methods in natural language processing. pp. 1568–1576. Association for Computational Linguistics (2011)
3. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer. In: Proceedings of the fourth ACM international conference on Web search and data mining - WSDM '11. p. 65. ACM Press, New York, New York, USA (2011), `http://portal.acm.org/citation.cfm?doid=1935826.1935845`
4. Bennett, W.L., Johnson, C.N.: A Model of Crowd-Enabled Organization : Theory and Methods for Understanding the Role of Twitter in the Occupy Protests. International Journal of Communication 8, 646–672 (2014)
5. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10008 (2008)
6. Bollen, J., Mao, H., Zeng, X.J.: Twitter mood predicts the stock market. arXiv preprint 1010.3003 (2010), `http://arxiv.org/abs/1010.3003`
7. Bruns, A., Highfield, T., Burgess, J.: The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks. American Behavioral Scientist 57(7), 871–898 (Jun 2013), `http://abs.sagepub.com/cgi/doi/10.1177/0002764213479374`
8. Bruns, A., Burgess, J.E.: The use of twitter hashtags in the formation of ad hoc publics. In: 6th European Consortium for Political Research General Conference (2011)
9. Bruns, A., Highfield, T.: Political Networks on Twitter. Information, Communication & Society 16(5), 667–691 (Jun 2013), `http://www.tandfonline.com/doi/abs/10.1080/1369118X.2013.782328`
10. Bulearca, M., Bulearca, S.: Twitter: a viable marketing tool for smes. Global Business and Management Research: An International Journal 2(4), 296–309 (2010)

11. Cha, M., Haddai, H., Benevenuto, F., Gummadi, K.P.: Measuring User Influence in Twitter : The Million Follower Fallacy. In: International AAAI Conference on Weblogs and Social Media (2010)
12. Comunello, F., Anzera, G.: Will the revolution be tweeted? A conceptual framework for understanding the social media and the Arab Spring. Islam and ChristianMuslim Relations 23(4), 453–470 (Oct 2012), `http://www.tandfonline.com/doi/abs/10.1080/09596410.2012.712435`
13. Conover, M., Ratkiewicz, J., Francisco, M., Gonalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. In: ICWSM (2011), `http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847%3Cbr/3275`
14. Coppersmith, G., Dredze, M., Harman, C.: Quantifying mental health signals in twitter. In: Proceedings of ACL Workshop on Computational Linguistics and Clinical Psychology. Association for Computational Linguistics (2014)
15. Gayo-Avello, D.: A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data. Social Science Computer Review 31(6), 649–679 (Aug 2013), `http://ssc.sagepub.com/cgi/doi/10.1177/0894439313493979`
16. Gnatyshak, D., Ignatov, D.I., Semenov, A., Poelmans, J.: Gaining insight in social networks with biclustering and triclustering. In: Perspectives in Business Informatics Research - 11th International Conference, BIR 2012, Nizhny Novgorod, Russia, September 24-26, 2012. Proceedings. pp. 162–171 (2012), `http://dx.doi.org/10.1007/978-3-642-33281-4_13`
17. Golbeck, J., Hansen, D.: A method for computing political preference among Twitter followers. Social Networks 36, 177–184 (Jan 2014), `http://linkinghub.elsevier.com/retrieve/pii/S0378873313000683`
18. Gonzalez-Bailon, S., Borge-Holthoefer, J., Moreno, Y.: Broadcasters and Hidden Influentials in Online Protest Diffusion. American Behavioral Scientist 57(7), 943–965 (Mar 2013), `http://abs.sagepub.com/cgi/doi/10.1177/0002764213479371`
19. González-Bailón, S., Borge-Holthoefer, J., Rivero, A., Moreno, Y.: The dynamics of protest recruitment through an online network. Scientific reports 1, 197 (Jan 2011), `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3240992\&tool=pmcentrez\&rendertype=abstract`
20. Highfield, T.: Mapping intermedia news flows: Topical discussions in the Australian and French political blogospheres. Ph.D. thesis, Queensland University of Technology (2011)
21. Newman, M.E.J.: Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Physical Review E 64(1) (Jun 2001)
22. Nikiporets-Takigawa, G.: Tweeting the Russian protests. Digital Icons: Studies in Russian, Eurasioan and Central European New Media 9(2013), 1–25 (2013)
23. Page, L., Brin, S., Rajeev, M., Terry, W.: The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University (1998)
24. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1079–1088. ACM (2010)
25. Yardi, S., Boyd, D.: Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter. Bulletin of Science, Technology & Society 30(5), 316–327 (Sep 2010), `http://bst.sagepub.com/cgi/doi/10.1177/0270467610380011`