Physica A 414 (2014) 403–413

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Mapping the online communication patterns of political conversations

J. Borondo, A.J. Morales, R.M. Benito *, J.C. Losada

Grupo de Sistemas Complejos and Departamento de Física y Mecánica, Universidad Politécnica de Madrid, ETSI Agrónomos, 28040 Madrid, Spain

HIGHLIGHTS

- The political conversation is constrained by both ideology and language.
- Politicians, the main characters; Traditional media, still the main source of information.
- The political communication is still driven by a minority.
- There is a relation between the political alignment of users and the language in which they tweet.

ARTICLE INFO

Article history:
Received 14 February 2014
Received in revised form 13 May 2014
Available online 27 July 2014

Keywords:
Social network analysis
Political conversations
Languages
Twitter
Political elections
Online communication patterns

ABSTRACT

The structure of the social networks in which individuals are embedded influences their political choices and therefore their voting behavior. Nowadays, social media represent a new channel for individuals to communicate, what together with the availability of the data, makes it possible to analyze the online social network resulting from political conversations. Here, by taking advantage of the recently developed techniques to analyze complex systems, we map the communication patterns resulting from Spanish political conversations. We identify the different existing communities, building networks of communities, and finding that users cluster themselves in politically homogeneous networks. We found that while most of the collective attention was monopolized by politicians, traditional media accounts were still the preferred sources from which to propagate information. Finally, we propose methods to analyze the use of different languages, finding a clear trend from sympathizers of several political parties to overuse or infra-use each language. We conclude that, on the light of a social media analysis perspective, the political conversation is constrained by both ideology and language.

1. Introduction

Opinions are not formed in a social vacuum, but on a social network where those around us affect what we think. A classic topic in political science is how social interactions shape individuals political views, and whether the social network in which they are enmeshed has an influence on their voting behavior [1–4]. In fact, back more than 60 years ago, Lazarsfeld et al. [5] pointed out in "The People's Choice" that the impact social contacts have on voting decision is bigger than that from mass media or politicians. This resulted on the Two-step flow of communication [6–8]. According to this theory, ideas flow from mass media (or politicians in an electoral context) to opinion leaders, and from them to the population. More

* Corresponding author. Tel.: +34 913365646; fax: +34 913365726.
E-mail address: rosamaria.benito@upm.es (R.M. Benito).

http://dx.doi.org/10.1016/j.physa.2014.06.089

recently, studies have shown that political participation is affected by the social environment; as friends, family members, or co-workers exhibit similar behavior [9,1,10]. The explanation for this influence is a prevalent theme of research. Burt [11], distinguished between contagion by cohesion and contagion by equivalence. According to the first one, people's political preferences are directly influenced by their social network. This theory sees social influence as a result of intimacy within primary social grouping, and is referred in literature as direct political influence, political assimilation, or socialization [1,12,13]. Alternatively contagion by equivalence proposes a structural equivalence model to explain this influence, where people base their behavior on what they observe from others that occupy a similar position to them [14–16].

In this paper, we intend to explore the structure of the online social networks in which individuals are embedded when discussing politics. This is an increasingly relevant topic since online social networks and social media platforms, such as Twitter, are the latest new medium being exploited by politicians for decisive competitive advantage. Thus, today's culture is changing, Internet and social media represent a new channel through which information and ideas can quickly flow [17], bringing people a wider (and cheaper) variety of information. Today's new culture sees value in sharing information, and relies on collective wisdom [18]; just take Wikipedia [19] as an example. Social media fit perfectly this new context as they are about listening and being heard, about sharing information with those you trust, and about having a variety of sources of information at hand from where to choose. So, nowadays, when trying to understand the opinion formation process of individuals, we have to take into account not only their face to face relations or the propaganda coming from traditional mass media, but also the online communications that are increasingly taking place through social media platforms such as Twitter. In fact, recent research has brought evidence to show that political mobilizations in an online social network can influence real world voting behavior [20]. Moreover, the availability of the data represents a big opportunity to study social phenomena, such as politics [21,22], viral marketing [23], information diffusion [24], or social influence [25].

The target of this paper is to map the communication patterns behind the political conversations taking place on social media to uncover possible constrains in the online political discussion, and to answer questions such as: Do really social media platforms represent a channel through which more voices can be heard and encourage political discussion? To this end we have analyzed the mention and retweet networks within the framework of complex networks [26,27] and performed a community structure [28] analysis to understand how users of social media have grouped themselves around the different ideologies. We discuss the different role played by politicians and traditional mass media, when propagating news, or capturing the collective attention. Next, we identified the language in which each tweet was posted and explored to which extent the diversity of languages limits the communication among the different sides. Furthermore, we propose a method to determine the proximity between the different languages and political alignments. Finally, we discuss the major implications of our results.

2. System & methods

2.1. Twitter

The number of users engaged to online social networks or social media is rapidly growing all around the globe. More specifically in Spain the percentage of population using online social networks has reached over 49% of the population [29].

Twitter is one of the most popular social media platforms and its main feature consists in allowing people to post and exchange text messages limited by 140 characters. This platform features several interaction mechanisms to facilitate communication among users. The first of these interaction mechanisms is the ability of people to follow and be followed by the rest of users. This is a passive mechanism that allows users to receive the messages written by their followees at real time. The Twitter followers network is a directed graph where non reciprocal relations are admitted and it states the social substratum through which information will flow. Although having a large number of followers increases the visibility of the tweets posted by users, it not necessarily makes them influential [30].

Another important mechanism to interact is the message retransmission or retweet. This mechanism allows individual messages to propagate and travel throughout the network [31], and also, it serves as a way for people to endorse their point of view over specific subjects [32]. In addition to this, another relevant way for direct interaction is the mention mechanism. By mentioning someone's username in the message text, people is able to send directed messages to the mentioned user's inbox. This mechanism is often used to establish conversations between users, through the exchange of messages, or just to refer somebody in the messages text [33]. At last, all messages on Twitter, may be identified using keywords called hashtags [34]. This mechanism generates the trending topics, and people use it to discuss and exchange ideas without the necessity of having any explicit relation.

2.2. Dataset

Our datasets are constructed from public access messages posted on Twitter regarding political conversations. The first dataset, labeled 20*N*, relates to the last 20th of November 2011 Spanish general elections. The second one, labeled as 25*N*, regards the last Catalan elections that took place on the 25th of November of 2012. We downloaded all the tweets using the Twitter API interface and searching for a specific keyword that identified each conversation. Regarding the general elections we downloaded all the messages that included the keyword 20*N* posted in a three week period including the official electoral

Table 1
General information about the networks built from the two studied datasets, including the number of nodes, number of edges and assortativity by language (r_L) of each network. Note that the 25N networks do not include nodes whose language could not be analyzed since they did not post a message.

		Retweet network			Mention network		
Dataset ID	Region	Nodes	Edges	r_L	Nodes	Edges	r_L
20*N*	Spain	075546	153549	0.43	39631	86029	0.38
25*N*	Catalonia	102959	289486	0.70	29088	72129	0.37

campaign and voting day. This dataset has already been used to show that the activity taking place on Twitter was correlated to the election outcomes; and to characterize politicians behavior, observing a lack of debate among political parties [21]. Analogously, to build the dataset corresponding to the Catalan elections, we downloaded all the messages that included the keyword 25*N* posted in a seven week period including the official electoral campaign and voting day. We chose these tags, 20*N* and 25*N*, for being ideologically neutral identifiers, used all around Spain and by all the political parties when referring to these elections.

To better understand these datasets we have included, in Appendix A, a brief description about the Spanish political situation.

2.3. Twitter networks

The networks presented on this paper are constructed from the datasets described in the preceding section. To build the mention networks we filtered all the data remaining only messages that contained a mention to another user. These networks are directed and weighted graphs. A new edge appears when a new message posted by user, *A*, contains a mention to another user, *B*, (with direction $A \rightarrow B$). *B* is notified about the new tweet, as it appears in his private in-box, what dramatically increases his chances of reading it. In a similar way we build the retweet networks that are also directed and weighted graphs. However, in the retweet graph an edge is created whenever user, *A*, retweets a message originally posted by user, *B*, (direction $A \rightarrow B$). Therefore the tweet not only appears on *B*'s public panel but also on *A*'s, increasing the chances to draw, both *A*'s and *B*'s followers attention. General details about each network can be found in Table 1.

2.4. Pagerank as a measure of popularity

In the scope of network analysis, the centrality of a node determines the relative importance of a node within the graph. Therefore in a social context it represents popularity or how influential a person is in the social network. Despite there are numerous widely used measures of centrality, here we will use the pagerank centrality.

The pagerank centrality, a network-based diffusion algorithm initially proposed to rank web pages [35,36], has arisen as one of the preferred methods to rank a vast amount of data in different types of networks. For instance, it has been used to effectively determine the relevance of scientists in the citation network [37,38], as well as to rank streets [38], ecological species [39] or leadership groups on social networks [40]. We could think about it as if a 'random surfer' surfs the net by following links between nodes, eventually the surfer decides to jump to a randomly chosen node and continues the process. The probability of the surfer visiting each node is determined by its pagerank. Hence, a node has a high pagerank when highly connected or while attached to leading ones.

2.5. Language detection

To automatically determine the language in which each user preferentially tweeted we have taken advantage of guess-language [41]. This is a Python library that determines the natural language of a given text. We determined the language in which each tweet was posted and assigned each user the language in which most of his tweets were written.

3. Mapping the communication patterns

3.1. Community structure

Online social networks provide us information on how people interact with each other through the Internet. The best approach to analyze the upper hierarchical level of such networks is community structure [28,42]. Within complex networks, a community can be defined as a subgroup of densely connected nodes inside a larger graph. Thus, in social networks a community would correspond to a group of strongly related friends. Similarly, in a politicized context, such as the present study, one expects people to group themselves by ideology. Following this idea communities correspond to the different political parties, as already noted for blogs [43]. Analyzing the 20*N* dataset by means of community structure we intend to map the political communication strategies, identifying large groups of users among which information flows quickly and

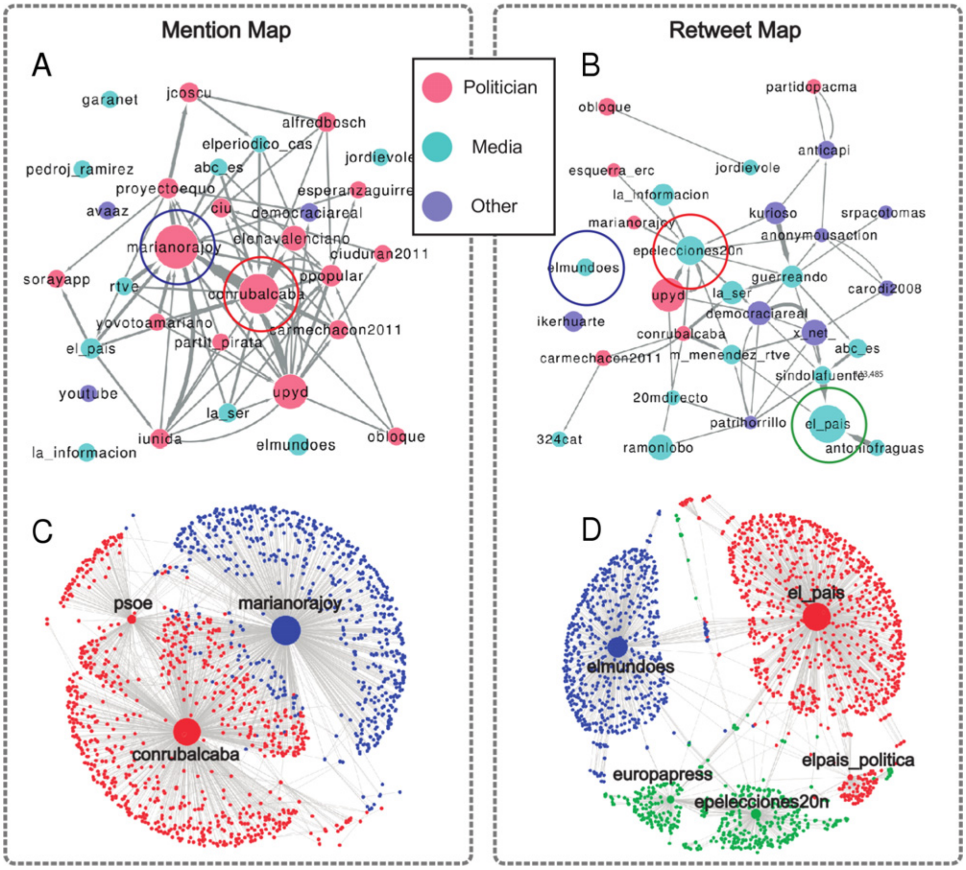

Fig. 1. Map of the 20*N* mention (A) and retweet (B) c-networks, showing the 30 most important communities. The color of communities indicate their class (politician, media, others), and the size represent their pagerank centrality. For a clearer representation we only show the links representing at least 0.005% of the total inter-module flow. A detail of the inside structure of some important communities (marked with circles) is displayed in (C) for the mention c-network and (D) for the retweet c-network. Nodes size represent their centrality and their color indicate the community to which they belong. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

understand the user behavioral patterns engaged among them. In order to perform the community structure analysis we used the map equation algorithm [44], based on random walks, and the intuitive idea that if a community exists the random walker tends to get trapped inside it. We decided to use this method for considering it specially appropriate to study systems where links represent information exchange among nodes, such as the mention and retweet networks built from Twitter data. In this way, we build networks of communities (c-networks) [45], where each community is represented by a c-node and the flow of information going on across communities constitute the links.

In the top panels of Fig. 1 we present, for the 20*N* conversation, both c-networks: the mention c-network (A) and retweet c-network (B). In order to obtain a clear and simplified map from where to extract useful information, we only display the 30 most representative (higher aggregated pagerank) c-nodes, and the links representing at least 0.005% of the total inter-community flow for each c-network. Despite the small number of communities displayed, the map captures the most relevant information, as it represents over 30% of the total activity. Moreover, these most popular communities attracted the attention of users belonging to many other communities without necessarily directing tweets at them. In the figure the name of each c-node corresponds to the username with highest popularity (pagerank) inside it, and the size to the pagerank of the community (aggregated pagerank of all nodes within the community). Communities are colored according to their classification: politicians (red), media (blue) and others (purple). The thickness of the links represent the communication flow going from one community to another. For more details about the statistical properties of the c-networks see Appendix B.

3.2. Politicians, the main characters; traditional media, still the main source of information

Politicians have drawn most of the attention, as their accounts were the most mentioned. This means that users addressed their messages to them by referring to their accounts when tweeting. This is illustrated in Fig. 1A (mention c-network map), that show how the majority and more significant communities centered their attention at them. For example, the map is centered around the two most voted candidates c-nodes (*marianorajoy* and *conrubalcaba*, belonging to the two dominant parties: PP and PSOE respectively), as their accounts were the most popular.

Table 2
Characterization of the inside structure of the top 5 communities (c-nodes) for each network of the 20*N* dataset. We present them ordered, firstly by their network type (NW): mention (M) or retweet (RT) and secondly by their total pagerank (PR). In the table each community is identified by its c-node name, that corresponds to the leading account of the community, and classified according to its class (politician or media). Following we show the size of each c-node in terms of number of nodes (N) and PR; the number of official accounts inside it together with their affiliation; the dominant affiliation; the relative PR of the official accounts belonging to it (PR affiliation).

NW	Class	C-node	Size(N/PR)	Official accounts	Affiliation/media	PR affiliation
M	Politician	marianorajoy	612/5,90%	7(PP)	PP	44,94%
M	Politician	conrubalcaba	804/5,20%	11(PSOE)	PSOE	54,72%
M	Politician	upyd	533/4,00%	22(UPyD)	UPyD	47,09%
M	Politician	elenavalenciano	2/2,00%	1(PSOE)	PSOE	55,00%
M	Politician	ppopular	303/1,20%	8(PP)	PP	53,09%
RT	Media	el_pais	986/2,50% %	5(El Pais)	el Pais	74,52%
RT	Politician	upyd	863/2,20%	30(UPyD)	UPyD	41,72%
RT	Media	epelecciones20n	461/1,80%	3(EUR)	Europapress (EUR)	52,78%
RT	Media	laser	774/1,30%	5(Ser)	Cadena Ser (Ser)	74,15%
RT	Media	la info	140/1,22%	2(LI)	La Informacion (LI)	87,50%

To further understand the mention c-network and how the collective attention has been directed to politicians, we have explored the inside structure of the politicized communities. We found that communities grow around a single political alignment official accounts, as all the politicians inside a same community belong to the same party (as can be seen in Table 2). Politicians accounts, although representing a very small fraction of the total members; have attracted most of the inside collective attention, concentrating a very high fraction of the total pagerank (Table 2). To illustrate this phenomena, Fig. 1C visualizes the inside structure of the two most central communities corresponding to the dominant parties PP and PSOE. Both of them are large communities (over 600 nodes), with a high pagerank (5.9% and 5.2%), and with a small number of politicians (7 and 11). As it can be seen, the vast majority of users only mentioned a single political party accounts. However, a minority of intermediaries linked the two sides of the conversation mentioning both candidates.

Despite being the main characters regarding mentions, politicians loose importance when considering retweets in favor of media and anonymous people such as bloggers. This resulted in traditional media accounts prevailing in the retweet map, where they were the top retweeted in the most important communities (can be checked in Fig. 1B and Table 2). This means that media accounts were the preferred source of information from where users propagated information regarding the elections. For example, the media account *epelecciones*20*n* has not only attracted many retweets within its community, but also from other communities. We must note that this account was created by the national news agency (Europapress) to report about the 20*N* elections.

When exploring the inside structure of the retweet communities, (Fig. 1D) we found that users spread information from a minority of official accounts belonging to a same media. This fact together with the limited number of retweets going on across communities reflects the fidelity of people to their preferred media, as they do not compare different sources of information. Or if they do, they just propagate news from their favorite one. Whether each community hosts a single hub or various, depends on each media online communication strategy. For example, media using various accounts to publish news on Twitter (El Pais or Europa Press), present various hubs inside their community; whereas those using just a single global account (El Mundo), present an unique hub. All this is illustrated in Fig. 1D, where the internal structure of the mentioned communities together with their inter-links are displayed.

4. Language & politics

4.1. Language polarization

Here, we explore the impact that the existence of several co-official languages has on the Spanish politics and its polarized landscape. For this we complement the 20*N* dataset with a more local political conversation: the 25th of November Catalan elections (25*N*). We classified each user according to their most written language (see Section 2.5 for details) and analyzed the communication patterns among the different languages.

Twitter reflects the Spanish linguistic diversity, as Tweets were posted in all the different official and co-official languages. This is illustrated in Fig. 2 where we have visualized the 20*N* (A) and 25*N* (B) retweet networks coloring each node according to their language. In Fig. 3 we present the percentage of users tweeting in each language for both datasets, and compare these results with those observed when considering all the raw Twitter data coming from Spain and Catalonia [46]. Regarding the general elections, Spanish was the most used language (78%), followed by Catalan (18%). However, when comparing this data to the statistics presented in Ref. [46] (see Fig. 3), we found that the share of users tweeting in Catalan is above its expected value (3.11%). We also observed this same behavior during the Catalan elections, where Catalan presents a share of 54%, quantify far above its expected value (28%). This suggests that the use of Catalan on Twitter increases in politicized contexts such as electoral campaigns.

Next, we explore the communication patterns among users tweeting in different languages. For this we measure the language polarization by calculating the assortativity [47] by language. This measure quantifies the tendency for users on

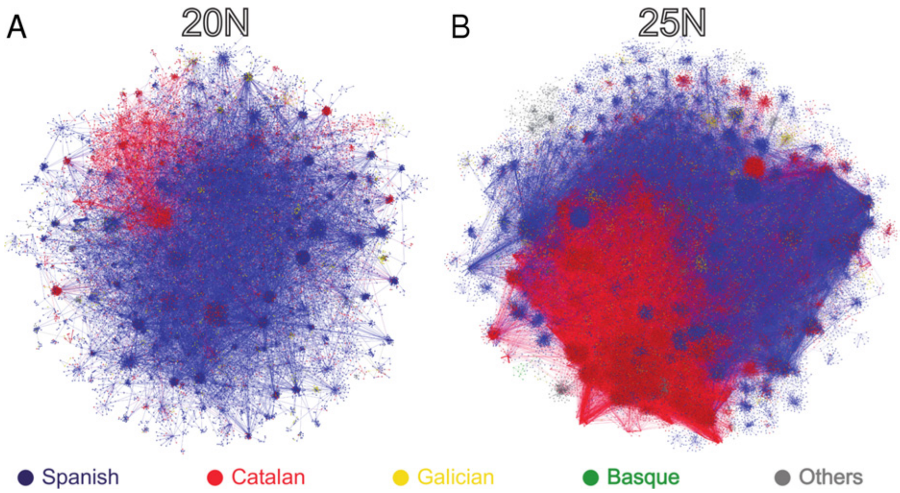

Fig. 2. This figure visualizes the 20N (A) and 25N (B) retweet networks. Each node represents a user and has been colored according to his most used language. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

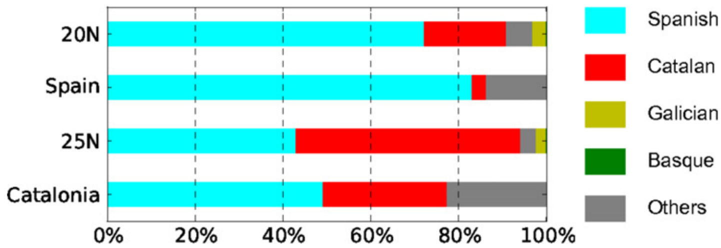

Fig. 3. Language share for the two analyzed datasets, 20N and 25N, together to the overall share found for Spain and Catalonia in general Twitter conversations [46].

Twitter to communicate to other users that tweet in their same (or not the same) language and is given by the following expression:

$$r_L = \frac{\mathrm{Tr}\,\mathbf{e} - \|\mathbf{e}^2\|}{1 - \|\mathbf{e}^2\|} \tag{1}$$

where $\mathbf{e}$ is the language mixing matrix, whose elements e_{ij} quantify the number of retweets going from users posting on language i to users posting on language j; Tr stands for trace; and $\|\mathbf{e}\|$ represents the sum of all elements in the matrix $\mathbf{e}$. This measure ranges between -1 and 1. A value of 1 means that users only retweet those tweeting on their same language. On the contrary, -1 means that users only retweet users who tweet in a different language. In between, $r_L = 0$, indicates users retweet others regardless their language.

Our results show the existence of a clear trend to retweet only those who tweet in the same language (see Table 1) and hence, in Spain, language represents a constrain for political communication. This tendency becomes more extreme for the Catalan conversation, meaning that for more local conversations the segregation by language becomes higher.

4.2. Language preferences

Nationalist currents are known to have a big impact in the Spanish politics. In fact, in some regions, such as Catalonia, the nationalist component of several parties can be much more influential than their ideology (liberal/conservative). Here, we intend to answer the following question: what is the relation between the language in which users tweet and the political party they support?

As a proxy to determine the party that each user supports we use the retweet network, since retweet is the Twitter interaction mechanism that exhibits the highest segregated partisan structure [21,22]. We define the Proximity between a language (L_i) and a political alignment (A_j), $PROX(L_i, A_j)$, as the relative frequency of retweets going on from users posting in language, L_i, to official accounts of party, A_j, observed in the empirical network compared to its expected value in the randomized version. Hence, it can be expressed as:

$$PROX(L_i, A_j) = \frac{P(L_i, A_j)}{\langle P_r(L_i, A_j)\rangle} \tag{2}$$

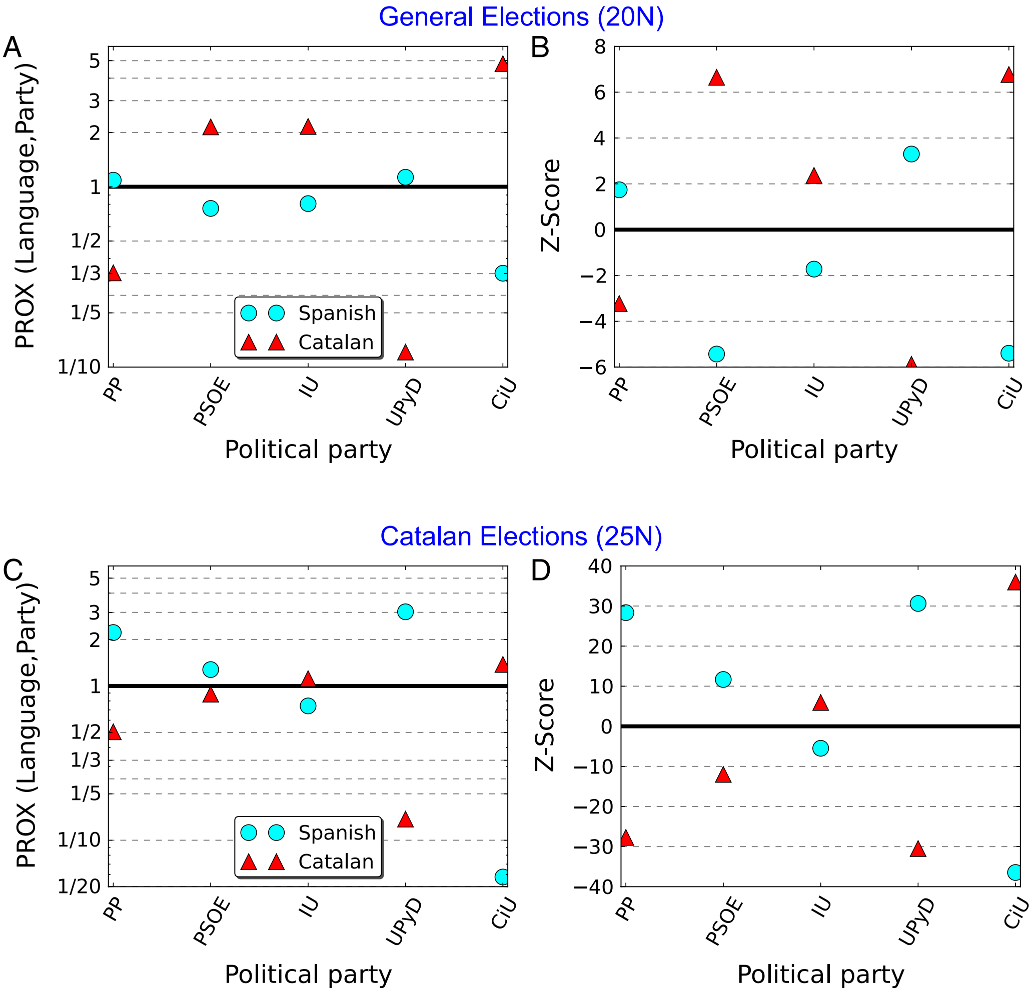

Fig. 4. Measuring proximity between languages and political parties. Circles (cyan) correspond to Spanish, while triangles (red) correspond to Catalan. Panel (A) shows the Proximity measures for the 20N dataset. Their corresponding Z-scores are plotted in panel (B). Panel (C) shows the Proximity measures for the 25N dataset. Their corresponding Z-scores are plotted in panel (D).

where $P(L_i, A_j)$ is the probability that a user tweeting in language, L_i, retweets a message originally posted by a politician affiliated to party, A_j. Analogously $\langle P_r(L_i, A_j)\rangle$ represents the expected value (averaged from 100 realizations) of this same probability but for the randomized networks. When randomizing the networks we have preserved all the individual characteristics of each user, meaning that each node preserves his preferred language, and number of outgoing and ingoing retweets. In order to randomize the networks we used the Markov-chain algorithm that repeatedly exchanges randomly chosen pairs of connected nodes. By comparing to randomized networks we compensate for the effects of differences in party popularity and in number of users tweeting on each language. Finally, to determine the significance of each Proximity measure we can calculate its corresponding Z-score, that is given by the following expression:

$$Z(L_i, A_j) = \frac{P(L_i, A_j) - \langle P_r(L_i, A_j)\rangle}{\sigma_r(L_i, A_j)} \qquad (3)$$

where $\sigma_r(L_i, A_j)$ is the standard deviation of $P_r(L_i, A_j)$, computed out of 100 randomized networks.

In Fig. 4 we present the Proximity between the most voted political parties (PP, PSOE, IU, UPyD, CiU) and the two most used languages (Spanish and Catalan), together with its corresponding Z-score, for both datasets. The results corresponding to the general elections (20N) are displayed on panels A and B, while the Catalan elections (25N) are displayed on panels C and D. Deviations from the standard preference line ($PROX = 1$) imply that users preferentially ($PROX > 1$) or not preferentially ($PROX < 1$) retweeted specific parties. The deviations found for both datasets measure the closeness in ideology of each party to the Catalan nationalist current. For example, CiU is a Catalan party with a remarkable nationalist current, thus, its sympathizers preferentially tweeted in Catalan and hardly did it in Spanish. Conversely users tweeting in Catalan hardly retweeted PP or UPyD, as these parties strongly defend the unity of Spain.

5. Discussion

In this paper we have analyzed the communication patterns of two Spanish political conversations taking place on Twitter; one of national scope, the 2011 general elections; and a more local one, the 2012 Catalan elections. We found that the Spanish political conversation is centralized around a small fraction of influential accounts. Politicians are the main characters, since their accounts were the most mentioned, and captured most of the collective attention. However, despite their accounts are still influential in the retweet network, users tend to propagate information from the same sources as in the offline world. Thus, traditional media accounts were the most retweeted. Therefore we can affirm that, in the light of our results, despite social media should allow more voices to be heard, the political communication is still driven by a minority of political parties and elite media.

Analyzing the mention network by means of community structure, we have been able to map the flow of political information going through Twitter during the campaigns. We show it to be considerably polarized by political ideology, as users crowded around a single political party accounts, and preferentially communicated with those of their same political stance. However, there were a small fraction of users, more exposed to political disagreement, who sustained the exchange of information among these polarized communities. Similar conclusions can be made from the retweet analysis, where we found that despite the countless sources of information available, users do not take full advantage of it, tending to just rely on their preferred ones. In this regard users mostly retweet from just one traditional media official accounts with whose editorial line they feel identified. Hence, we can affirm that the social network in which individuals are enmeshed is ideologically homogeneous. Such networks might be too insular, limiting the opportunities to learn about politics and contrast information. However, they represent effective information shortcuts to access information [48–50].

The results of this paper also speak about the importance that nationalist currents have in the Spanish political landscape, where at some regions the nationalist component of the parties becomes much more influential than the ideology. It is widely known that the strong feeling of membership to the autonomous community, that exists in several regions of Spain, is frequently used as a political argument. This is reflected on Twitter in several ways: (i) Catalan tends to be overused in political conversations. (ii) Despite the vast majority of speakers of a co-official language are bilingual, the conversation is highly segregated by language. (iii) There is an obvious relationship between the political alignment of users and the language in which they tweet.

However, the utility of the proposed method to determine the proximity between languages and political parties is not particular of Spain, but generalizable to other countries. For example, let us think about the United States. Although English is the main language, there is a diversity of cultures co-existing inside the country that speak different languages. Hence, the proposed methods could be used to estimate the proximity between the Chinese or Hispanic communities and the liberal or conservative parties. Thus, being useful to estimate electoral outcomes.

Of course, one cannot directly extrapolate from the Twitter information diffusion process to the overall opinion formation of the population, as users of the social network may not be the representative of the whole society. Even if they were, there are other channels through which individuals receive information, such as mainstream media (e.g. TV, radio) or personal relations. However, as the percentage of the population engaged to these platforms is rapidly growing, they have become the latest new medium being exploited for decisive competitive advantage. This makes the understanding of the communication patterns behind these platforms an increasingly important topic to further uncover opinion formation process of individuals.

Acknowledgment

This research was supported by the Ministry of Economy and Competitiveness—Spain under Grant No. MTM2012-39101.

Appendix A. Information on the Spanish political situation

Spain is a parliamentary monarchy since the approval of the last constitution in 1978, and belongs to the European Community since 1986. The country is organized into 17 autonomous communities (AC) and two autonomous cities. In Spain there is a strong feeling of membership to the AC. In fact despite Castilian being the official language for the whole territory, several ACs (Galicia, Basque Country, Catalonia, Valencia and Balearic Islands) have their own co-official language. Moreover in some of them there is strong nationalist current, with the existence of nationalist parties represented in parliament, above all at Catalonia, Basque Country, and Galicia.

In a general elections context we can distinguish two kinds of political parties, those of national scope, and others of regional character also standing in election. The most important parties can be found in Table A.1, where they are briefly characterized.

Regarding the national scope, we can distinguish two main ideologies: the conservative represented by Partido Popular (PP), and the liberal represented mainly by Partido Socialista Obrero Español (PSOE), and the smaller and far more left-wing party Izquierda Unida (IU). Recently a new party, Union Progreso y Democracia (UPyD), whose political posturing is difficult to classify in the classic conservative/liberal framework has broken in the parliamentary spectrum. The Financial Times and The Economist classified it as of center ideology, as it combines social liberalism with a defense of 'the unity of Spain'.

Table A.1
Information about the most important Spanish political parties that took part on the 20th of November of 2011 general elections. Name of the party; acronym; level at which the party stood to the elections, nationally (N) or just in one Autonomous Community (R); important official Twitter accounts.

			Official Accounts	
Political Party	Acronym	Level	Candidate	Others
Partido Popular	PP	N	marianorajoy	ppopular, esperanzaguirre, sorayapp
Partido Socialista Obrero Español	PSOE	N	conrubalcaba	psoe, carmechacon2011, elenavalenciano
Izquierda Unida	IU	N	cayo_lara	iunida, gllamazares
Union Progreso y Democracia	UPyD	N		upyd, tonicanto1
EQUO	EQUO	N	isabanes	proyectoequo
Convergencia i Unio	CiU	R (Catalonia)	ciuduran2011	ciu
AMAIUR	AMAIUR	R (Basque Country)		bildueh, aralarnafarroa
Partido Nacionalista Vasco	PNV	R (Basque Country)	jerkoreka	eaj_pnv
Ezquerra Republicana de Catalunya	ERC	R (Catalonia)	bosch	esquerra_erc
Bloque Nacionalista Gallego	BNG	R (Galicia)		bng, obloque
Iniciativa per Catalunya	ICV	R (Catalonia)	jcoscu	
Partido Andalucista	PA	R (Andalusia)		pandalucista, p_andalucista

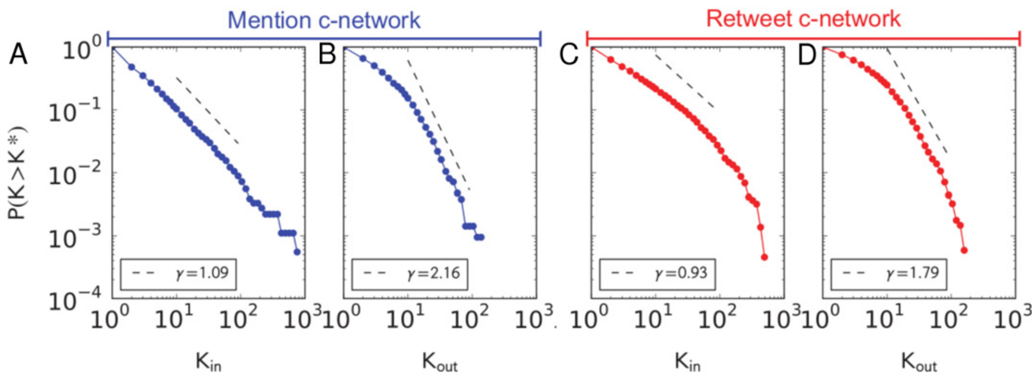

Fig. B.1. Indegree (A, C) and outdegree (B, D) complementary cumulative distribution functions for the mention (A, B) and retweet (C, D) c-networks of the 20*N* conversation. Dashed lines and their corresponding γ values are included for comparison. They do not represent fitted curves.

Within the autonomic plane we can also distinguish liberal and conservative parties, though their nationalist component is much more influential.

Appendix B. Statistical properties of the c-networks

In this appendix we characterize the statistical properties of the two constructed c-networks: the mention c-network and the retweet c-network. In Fig. B.1 we compare the complementary cumulative distributions of the in and out degree for both c-networks. The in and out degree measures for a community define the number of inter-community edges pointing to (indegree) and from (outdegree) the community. In other words the indegree quantifies the attention a community received from other communities, and the out degree the attention that the community paid to other communities. The indegree distribution for the mention c-network can be well fitted to a power law, in the form $P(k > k^*) = k^{-\gamma}$ where $\gamma = 1.09$, meaning that the growth process of this c-network follows the preferential attachment rules: popular communities become more popular [51]. However the retweet c-network presents a slightly less heterogeneous distribution not reaching so extreme values. Outdegree distributions follow similar decrements for both of the c-networks. These distributions are less heterogeneous than those observed for the indegree, as they present an initial smooth decadent to promptly decrease for values larger than 10.

To further understand the structure of these c-networks and how the collective attention has been distributed, we have calculated the popularity, measured as the aggregated pagerank, of each community and compared it to its indegree and outdegree. Note that the aggregated pagerank of a community is calculated as the sum of the pagerank (in the mention or retweet networks) of all the nodes inside it. Thus, it is not a measure resulting from the c-network structure, but from the network at the lowest hierarchical level. On the contrary the indegree and outdegree of a community are calculated from the c-network. In Fig. B.2 we present each community's indegree versus its popularity in a logarithmic scale for the mention (A) and the retweet (C) c-networks. The results reflect that the popularity of each community is well correlated to its indegree in the corresponding c-network. Meaning that the small fraction of very popular communities have been frequently targeted from users belonging to many others. This is usually explained by the presence of a famous account inside it, as discussed in Section 3.2. This result differs to what we observe when plotting the outdegree versus the popularity of each community, where as we can see in panels B and D of Fig. B.2, correlation is lost. Overall, we can affirm that the most

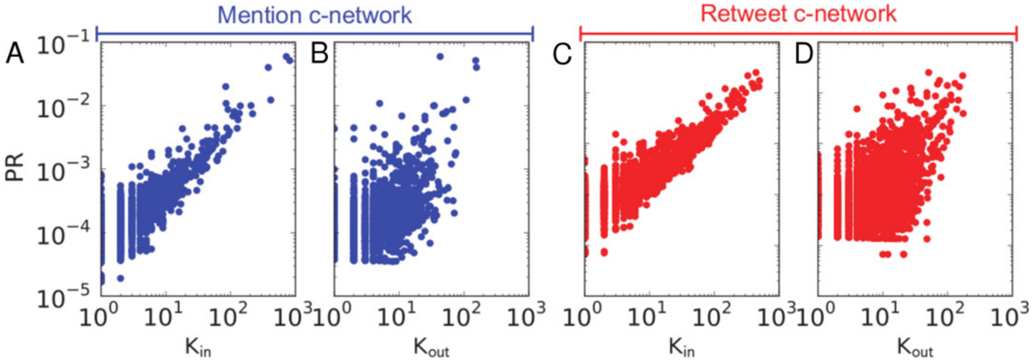

Fig. B.2. Comparison of the correlation between the aggregated pagerank (PR) and degree (measured from the c-network) of communities resulting from mentions (A, B) and retweets (C, D) for the 20*N* conversation. Aggregated pagerank plotted versus indegree (A, C), and versus outdegree (B, D).

popular communities have been preferentially mentioned or retweeted by members belonging to many other communities. However, such important communities have not necessarily directed tweets to other communities in order to attract this attention.

References

[1] R. Huckfeldt, J. Sprague, Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign, in: Cambridge Studies in Political Psychology and Public Opinion, Vol. 60, Cambridge University Press, 1995, http://psycnet.apa.org/psycinfo/1995-98606-000.

[2] D. Knoke, Political Networks: The Structural Perspective, in: Structural Analysis in the Social Sciences, 4, Cambridge University Press, 1990, http://books.google.com/books?id=9djTlBLaOccC&pg=PA159.

[3] B.R. Berelson, P.F. Lazarsfeld, W.N. McPhee, Voting: A Study of Opinion Formation in a Presidential Campaign, University of Chicago Press, 1954, http://www.press.uchicago.edu/ucp/books/book/chicago/V/bo3616092.html.

[4] R. Huckfeldt, Interdependence, density dependence, and networks in politics, Amer. Politics Res. 37 (5) (2009) 921–950. http://apr.sagepub.com/cgi/doi/10.1177/1532673X09337462.

[5] P.F. Lazarsfeld, B. Berelson, H. Gaudet, The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign, Columbia University Press, 1944, http://www.mendeley.com/catalog/peoples-choice-voter-makes-up-mind-presidential-campaign.

[6] J.P. Robinson, Interpersonal influence in election campaigns: two step-flow hypotheses, Public Opinion Quarterly 40 (3) (1976) 304–319. http://www.jstor.org/stable/2748249.

[7] E. Katz, The two-step flow of communication: an up-to-date report on an hypothesis, Public Opin. Quart. 21 (1) (1957) 61–78. http://poq.oupjournals.org/cgi/doi/10.1086/266687.

[8] M.S. Weatherford, Interpersonal networks and political behavior, Amer. J. Political Sci. 26 (1) (1982) 117–143. http://www.jstor.org/stable/2110843.

[9] C.B. Kenny, Political participation and effects from the social environment, Amer. J. Political Sci. 36 (1) (1992) 259–267. http://www.jstor.org/stable/2111432.

[10] P. Nieuwbeerta, H. Flap, Crosscutting social circles and political choice Effects of personal network composition on voting behavior in The Netherlands, Social Networks 22 (4) (2000) 313–335. http://linkinghub.elsevier.com/retrieve/pii/S0378873300000290.

[11] R.S. Burt, Social contagion and innovation: cohesion versus structural equivalence, Amer. J. Sociol. 92 (6) (1987) 1287–1335. http://www.journals.uchicago.edu/doi/abs/10.1086/228667.

[12] M.K. Jennings, L. Stoker, J. Bowers, Politics across generations: family transmission reexamined, J. Politics 71 (03) (2009) 782. http://www.journals.cambridge.org/abstract_S0022381609090719.

[13] M.K. Jennings, R.G. Niemi, The transmission of political values from parent to child, Amer. Political Sci. Rev. 62 (1) (1968) 169–184. http://www.jstor.org/stable/1953332.

[14] A. Downs, An Economic Theory of Democracy, Vol. 13, Harper, 1957, http://www.amazon.com/Economic-Theory-Democracy-Anthony-Downs/dp/0060417501.

[15] S.L. Popkin, The Reasoning Voter: Communication and Persuasion in Presidential Campaigns, second ed., University of Chicago Press, 1991, http://www2.lib.ncsu.edu/catalog/record/NCSU774992.

[16] R. Huckfeldt, J. Sprague, Discussant effects on vote choice: intimacy, structure, and interdependence, The Journal of Politics 53 (01) (1991) 122–158. http://journals.cambridge.org/abstract_S0022381600047940.

[17] D. Lazer, A.S. Pentland, L. Adamic, S. Aral, A.L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, et al., Life in the network: the coming age of computational social science, Science (New York, NY) 323 (5915) (2009) 721.

[18] L.A. Adamic, J. Zhang, E. Bakshy, M.S. Ackerman, Knowledge sharing and yahoo answers: everyone knows something, in: Proceedings of the 17th International Conference on World Wide Web, ACM, 2008, pp. 665–674.

[19] Kathryn Zickuhr, L. Rainie, Wikipedia, past and present, 2011 http://www.pewinternet.org/Reports/2011/Wikipedia.aspx.

[20] R.M. Bond, C.J. Fariss, J.J. Jones, A.D.I. Kramer, C. Marlow, J.E. Settle, J.H. Fowler, A 61-million-person experiment in social influence and political mobilization, Nature 489 (7415) (2012) 295–298. http://dx.doi.org/10.1038/nature11421.

[21] J. Borondo, A.J. Morales, J.C. Losada, R.M. Benito, Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish presidential election as a case study, Chaos 22 (2) (2012) 023138 ST. http://www.ncbi.nlm.nih.gov/pubmed/22757545.

[22] M.D. Conover, B. Gonc, A. Flammini, F. Menczer, Partisan asymmetries in online political activity, EPJ Data Science 1 (6) (2012) 1–17. http://dx.doi.org/10.1140/epjds6.

[23] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, ACM Transactions on the Web (TWEB) 1 (1) (2007) 5.

[24] P.A. Grabowicz, J.J. Ramasco, E. Moro, J.M. Pujol, V.M. Eguiluz, Social features of online networks: the strength of intermediary ties in online social media, PLoS ONE 7 (1) (2012) e29358.

[25] N.A. Christakis, J.H. Fowler, Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives, in: Little, vol. 3, Brown and Company, Little, Brown and Company, 2009, http://www.amazon.com/dp/0316036145.

[26] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang, Complex networks: structure and dynamics, Phys. Rep. 424 (4-5) (2006) 175–308. http://linkinghub.elsevier.com/retrieve/pii/S037015730500462X.

[27] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 58. http://dx.doi.org/10.1137/S003614450342480.

[28] M.E.J. Newman, Communities, modules and large-scale structure in networks, Nat. Phys. 8 (1) (2011) 25–31. http://dx.doi.org/10.1038/nphys2162.

[29] P. Research, Social Networking Popular Across Globe (2012). http://www.pewglobal.org/2012/12/12/social-networking-popular-across-globe/.
[30] M. Cha, H. Haddadi, F. Benevenuto, P.K. Gummadi, Measuring user influence in twitter: the million follower fallacy, ICWSM 10 (2010) 10–17.
[31] A.J. Morales, J.C. Losada, R.M. Benito, Users structure and behavior on an online social network during a political protest, Physica A 391 (21) (2012) 5244–5253. http://www.sciencedirect.com/science/article/pii/S037843711200386X.
[32] D. Boyd, S. Golder, G. Lotan, Tweet, tweet, retweet: conversational aspects of retweeting on twitter, in: HICSS, IEEE Computer Society, 2010, pp. 1–10. http://doi.ieeecomputersociety.org/10.1109/HICSS.2010.412.
[33] C. Honeycutt, S.C. Herring, Beyond Microblogging: Conversationand Collaboration via Twitter, in: HICSS-42nd Hawaii InternationalConference on System Sciences, vol. 0, IEEE Computer Society, Los Alamitos, CA, USA, 2009, pp. 1–10.
[34] J. Pöschko, Exploring Twitter Hashtags, CoRR abs/1111.6. http://arxiv.org/abs/1111.6553.
[35] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, in: Seventh International World-Wide Web Conference (WWW 1998), 1998, http://ilpubs.stanford.edu:8090/361/.
[36] A.N. Langville, C.D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Rankings, University Press, Princeton, 2006, http://www.worldcat.org/isbn/0691122024.
[37] P. Chen, H. Xie, S. Maslov, S. Redner, Finding Scientific Gems with Google, http://dx.doi.org/10.1016/j.joi.2006.06.001.
[38] D. Walker, H. Xie, K.-K. Yan, S. Maslov, Ranking scientific publications using a simple model of network traffic, J. Stat. Mech. (2007) P06010. http://dx.doi.org/10.1088/1742-5468/2007/06/P06010.
[39] S. Allesina, M. Pascual, Googling food webs: can an eigenvector measure species' importance for coextinctions? PLoS Comput Biol 5 (9) (2009) e1000494+. http://dx.doi.org/10.1371/journal.pcbi.1000494.
[40] F. Pedroche, F. Moreno, A. Gonzlez, A. Valencia, Leadership groups on social network sites based on personalized pagerank, Math. Comput. Modelling 57 (7–8) (2013) 1891–1896. http://www.sciencedirect.com/science/article/pii/S0895717711007898.
[41] guess-language. https://pypi.python.org/pypi/guess-language.
[42] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113. http://dx.doi.org/10.1103/PhysRevE.69.026113.
[43] L.A. Adamic, N. Glance, The political blogosphere and the 2004 {U.S.} election: divided they blog, in: Proceedings of the 3rd International Workshop on Link Discovery, LinkKDD'05, New York, NY, USA, 2005, pp. 36–43.
[44] M. Rosvall, C.T. Bergstrom, Maps of random walks on complex networks reveal community structure, Proc. Natl. Acad. Sci. 105 (4) (2008) 1118–1123. http://www.pnas.org/content/105/4/1118.abstract.
[45] M.C. González, H.J. Herrmann, J. Kertész, T. Vicsek, Community structure and ethnic preferences in school friendship networks, Physica A 379 (1) (2007) 307–316.
[46] D. Mocanu, A. Baronchelli, N. Perra, B. Gonalves, Q. Zhang, A. Vespignani, The twitter of babel: mapping world languages through microblogging platforms, PLoS ONE 8 (4) (2013) e61981.
[47] M.E.J. Newman, Mixing patterns in networks, Phys. Rev. E 67 (2) (2003) 026126.
[48] P.V. Marsden, Core discussion networks of Americans, Amer. Sociol. Rev. 52 (1) (1987) 122–131. http://www.jstor.org/stable/2095397.
[49] D.C. Mutz, Hearing the Other Side: Deliberative Versus Participatory Democracy, vol. 57, Cambridge University Press, 2006, http://www.amazon.com/dp/0521612284.
[50] A. Lupia, Shortcuts versus encyclopedias: information and voting behavior in California insurance reform elections, American Political Science Review 88 (1) (1994) 63–76. http://www.jstor.org/stable/2944882?origin=crossref.
[51] A.L. Barabási, R. Albert, Emergence of scaling in random networks, science 286 (5439) (1999) 509–512.