

Detecting sociosemantic communities by applying social network analysis in tweets

Rocio Abascal-Mena, Rose Lema, Florence Sèdes

▶ **To cite this version:**

Rocio Abascal-Mena, Rose Lema, Florence Sèdes. Detecting sociosemantic communities by applying social network analysis in tweets. Social Network Analysis and Mining, 2015, vol. 5 (n° 1), pp. 1-17. 10.1007/s13278-015-0280-2 . hal-01283863

HAL Id: hal-01283863
https://hal.science/hal-01283863

Submitted on 7 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 15373

To link to this article : DOI :10.1007/s13278-015-0280-2
URL : http://dx.doi.org/10.1007/s13278-015-0280-2

To cite this version : Abascal-Mena, Rocio and Lema, Rose and Sèdes, Florence *Detecting sociosemantic communities by applying social network analysis in tweets.* (2015) Social Network Analysis and Mining, vol. 5 (n° 1). pp. 1-17. ISSN 1869-5450

Any correspondance concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Detecting sociosemantic communities by applying social network analysis in tweets

Rocío Abascal-Mena[1] · **Rose Lema**[1] · **Florence Sèdes**[2]

Abstract Virtual social networks have led to a new way of communication that is different from the oral one, where the restriction of time and space generates new linguistic practices. Twitter, a medium for political and social discussion, can be analyzed to understand new ways of communication and to explore *sociosemiotics* aspects that come with the use of the hashtags and their relationship with other elements. This paper presents a quantitative study of tweets, around a fixed hashtag, in relation with other contents that users bring to connection. By calculating the frequency of terms, a table of nodes and edges is created to visualize tweets like graphs. Our study applies social network analysis that, going beyond mere topology, reveals relevant *sociosemantic communities* providing insights for the comparison of social and political movements.

Keywords Information mining · Social Web · Social Network Analysis · Graph theory · Community detection · Sociosemantics · Twitter

✉ Rocío Abascal-Mena
mabascal@correo.cua.uam.mx

Rose Lema
rose@xanum.uam.mx

Florence Sèdes
sedes@irit.fr

[1] Information Tecnhology Department, Universidad Autónoma Metropolitana-Cuajimalpa, Mexico City, Mexico

[2] Institut de Recherche en Informatique de Toulouse, Toulouse, France

1 Introduction

The new information technologies are now deeply transforming the research in linguistics. Indeed, the availability of large electronic resources (very large corpus, dictionaries of any kind) enables to develop methods that were still unthinkable a decade ago, to study the language. In particular, statistical methods, which were relegated to relatively marginal treatments (analysis of lexical frequency in studies of texts, for example), have gained momentum. Their role in the syntactic and semantic analysis of texts plays a more important status (see for example Habert et al. 1997).

Actually, the existence of digital social networks has changed the way people interact in social and political movements. They are composed by nodes (persons, groups or organizations linked to others according to their interests and/or points of view in common). Virtual social networks have led to a new way of communication that is different from the oral one, where the restriction of time and space generates new linguistic practices.

Twitter is a platform registered under the category of social networking and microblogging. It allows users to write brief texts, a maximum length of 140 characters, and publish them. These messages are called *tweets*. Keywords or topics are marked using the symbol '#' which is called a *hashtag*. Using common hashtags it is possible to follow, search and find conversations related to a given theme. The tweets take a nonlinear way by allowing the capacity to jump from one idea to another thanks to the use of links and hashtags. Twitter is also composed by usernames (using the symbol '@' before the name of the user) and links to images and websites.

For researchers from various disciplines, Twitter is a key element to be studied and analyzed in order to understand

and learn new ways of communication and language through the use of *techno-discursive elements* that exist in the tweets. New elements such as the shortening of words or the assumption of knowledge on a given topic lead to the encryption and the encoding of the message that creates a new digital culture. For example, Honeycutt and Herring (2009) showed that Twitter is not only used in one direction of communication but generally as a media for conversation between two persons. In their study, they found that randomly selecting a set of tweets, 31 % will contain the '@' symbol and 91 % will be sent to a specific user. Cunha et al. (2011) studied the propagation of hashtags, containing the symbol #, within speech communities which are groups of people whose members linguistically and culturally influence each other. In this way, they analyzed in which cases the hashtags are adopted and reused in future messages. In their research, they found a relationship between the distribution of hashtags within the rankings of frequency and the length of the hashtag to be adopted by the users. According to Cunha et al. (2011), the hashtags that are more used correspond to those containing more characters and that are more explicit or that are already known by the speech community.

In Cunha et al. (2012), the gender is included as a social determinant that influences the user in selecting hashtags about a particular topic. In the same way, Romero et al. (2011) studied how information spreads differently online by analyzing the ways in which tokens, known as hashtags, spread on a network defined by the interactions among Twitter users.

Chew and Eysenbach (2010) found the existence of variability in the use and context of hashtags when they studied the use of *#H1N1* and *#swineflu* in the tweets generated during 2009.

The hashtags have also been studied to determine the feelings issued within tweets. Rodrigues et al. (2012) presented a study of a corpus of tweets recovered during the 2010 Brazilian presidential elections where about ten hashtags were categorized into positive, negative or neutral. Hence, they count the number of tweets that are included to analyze if the hashtags can be used to determine feelings issued in a tweet. They found that the hashtag has already an intention that reflects a sentiment.

Twitter has also been used to create communities through the use of hashtags (Zappavigna 2011; Bruns and Burgess 2011) that are applied to help the interweaving of ad hoc public around specific issues. The hashtags are used to quickly expand events and emerging issues. Such is the case of, for example, *#yosoy132* (Mexican motion generated during the presidential campaign of 2012) or *#5-M* (protests in Spain from 2011 to 2013). In this context, we find similarities in the labels used among various movements around the world, as a way to show the same ideology as concrete connections and similarities despite their geographic situation or cultural context.

The previously presented works give us an idea of the way in which Twitter is analyzed to understand the new way of communication. But, no study explores more *sociosemiotics* aspects of the use of the hashtags and their relationship with other elements. These *sociosemiotics* aspects refers to signs, symbols and semantics induced through users by the use of hashtags to be able to better reflect the meaning of the tweet. These aspects reflect the dynamics of the vocabulary of the users and evolve with the users as compared to formal static knowledge structures (Singh 2000). The hashtags evolve, over the time, to insert new information, to clarify certain facts that cannot be explained in the tweet because of the restriction of characters.

In this paper, we present the results of our research around the use of hashtags as basic elements in the composition of the message trying to understand and interpret the meaning of the relations between concepts, terms that are already categorized by the use of hashtags, over time. In this case, the hashtags contains *sociosemiotics* aspects to insert meta-information to the tweet but also to be able to arrive to a pertinent public. Meanings are constructed through the creation and interpretation of *signs*. According to Peirce (1931) people think only in signs which take the form of words, images, sounds, odors, flavors, acts or objects, but such things have no intrinsic meaning and become signs only when we invest them with meaning. Anything can be a sign as long as someone interprets it as *signifying* something—referring to or *standing for* something other than itself. Things are interpreted as signs by relating them to familiar systems of conventions. It is this meaningful use of signs that is at the heart of the concerns of semiotics.

Our research analyses tweets tagged with *#noalospluris* that were recovered in mid-December 2013. This hashtag was used to express an upset by the existence of multimember in the Chamber of Deputies and the Senate. The purpose of our research is to find visualizations of the information as *networks* (nodes and links between information, in this case between the hashtags) to understand the meaning of the use of different elements in the tweet. Is it possible to visualize the central problematic of the movement *#noalospluris* using a network? Is it possible to compare different movements by analyzing their main communities? We refer to communities as located groups that share similar characteristics. These are groups connected around hashtags and composed by symbols that are known and used only by the community.

Our work aims to help the analysis of political and social movements in Twitter through a quantitative analysis by the application of the graph theory to model the

tweets like a network and the use of basics of Social Network Analysis (SNA) to extract *sociosemantic* information. We have carried out, also, a qualitative analysis of the tweets using the results of our quantitative analysis to make an interpretation of our results which are compared with a new political movement *#menospluris*. This movement is promoted by the Institutional Revolutionary Party (PRI) that held the power for 71 years and of which current president of Mexico, Enrique Peña Nieto, is a member.

The rest of this article is organized as follows. In the next section, we briefly describe the main characteristics of a tweet. In Sect. 3, we present some concepts that we used in Graph Theory and Social Network Analysis (SNA). Section 4 presents some works that have for purpose the analysis of tweets to understand social behavior. Section 5 presents the dataset used. Section 6 is devoted to the presentation of the method used to extract and analyze our corpus. This methodology is based on techniques provided by the SNA field. The application of these techniques and the results are presented in Sect. 7. A comparison between the movements *#noalospluris* and *#menospluris* is provided in Sect. 8. Finally, we provide insights for future work and our conclusions.

2 Characteristics of a tweet: the case of the *#hashtag*

At the launch of Twitter in 2006, the main objective was that users answer the question "*What are you doing?*" being a kind of "*life streaming*" to make public the day a day of the people in Internet and opening the doors of the individuality and the privacy. However, it is from the daily use and trends of Web 2.0 (where the tagging of information is encouraged to make easier to recover information not only from its syntax but also from its meaning) that Chris Messina (technologist located in San Francisco) proposed, in 2007, from his blog and his Twitter space the use of the symbol '#' to improve filtering and contextualization in Twitter. Overall, the proposal aimed to create interest groups.

However, using the pad as a means of labeling can be considered as a *linguistic innovation* that allows the creation of jumps within a sentence and it is a reference, always present, supplementing the information (meta-information) that is being read. Also, Twitter enables the construction of *metadiscursive* comments where *metadiscourse* is the language used when the writer refers to his own thinking and writing as he thinks and writes—to summarize.

The use of labeling extends the ability of the message making the hashtag an identifier to (a) document the tweet using hypertextualization, (b) categorize the tweet in ad hoc groups and subject areas, (c) generate *metadiscursive* comments, (d) create expressive marks, (e) create personal interpretations and (f) identify a thematic as generic.

For the generation of metadiscursive comments, we use the definition of metadiscourse provided by Hyland and Tse (2004): *metadiscourse refers to all the strategies used by the writer to organize the text and allow the connection of different ideas*. However, the metadiscourse is a crucial element in the argumentative writing (found on Twitter) that facilitates persuasion in messages (Crismore et al. 1993). There is no more research on the use of metadiscourse in Twitter except the work of Russell (2011) and Poell et al. (2014) that confirm the use of metadiscourse processes to persuade tweeters around raised movements of the Arab Spring. In the case of Egypt, the tweets that were sent intended to build the history instead of being only transmitters of information in which the criticism and the scrutiny were present. Russell (2011) argues that the awareness and the metadiscourse that is inspired by these tweets have their precedents: *it is a recurring phenomenon in the history of professional journalism, which originates, particularly in times of political crisis or polarization* (Gitlin 1980). According to Georgakopoulou (2013), what stands out in these new forms of digital storytelling is the potential for greater interactivity with the user and the range of forms of involvement of the reader, who may even decisively change the course and interpretation of the story told.

Currently, the use of the generic label (hashtag) is not given at all because we still find that the messages, in Twitter, contain multiple hashtags to reach a wider audience. Bruns and Burgess (2011) argue that the use of a generic label would categorize tweets within a variety of themes making much easier the disambiguation of tweets that have nothing in common. In the same way, the generic tag allows an alternative explanation and emphasis without the use of visual elements (the use of bold, italic, etc.) or the use of emoticons.

3 From graph theory to the study of graphs using social network analysis

In this article, we present the constitution of graphs from the analysis performed to tweets to find how the labels evolve over time and what position they take within the subject. Thus, we apply graph theory to complement the linguistic analysis from the interpretation, over a period of time, of the existing relationships between the concepts used in the tweets. Our analysis corresponds to the need of studying visual elements that could help us to understand the formation and variation of language, as the composition of communities around a specific theme, when it is mediated by the use of current technologies.

The application of graphs in textual analysis does not imply to have a graph that models the whole meaning of the text; rather it focuses on capturing the dependence that exists between the concepts referred to in the text. In this case, the graph is seen as a social network of words where the meeting points or communities are the grammatical contexts in which these words appear.

When performing a text analysis using graphs, modeled as a social network, it is possible to use techniques that come from the Social Network Analysis (SNA). In our case, we use three techniques that allow us to identify key concepts and grammatical contexts: the *degree*, the *centrality* and the generation of *communities*.

The *degree* is the number of points adjacent to a given point. That is, the number that indicates how many connections (relationships) exist between a node and others nodes. In 1948, Bavelas introduced the idea of centrality applied to human communication in small groups and hypothesized about the relationship that exists between structural centrality and the influence in group processes (Freeman 1979). One point is central if it has a high degree, which corresponds to the intuitive idea of centrality in which a point is central if it is well connected to other points of its environment (Herrero 2000).

In our analysis, we also use the idea of *betweenness* (algorithm used: Brandes 2001) that determines to what extent a point is an intermediary between other points because it is located on the path between them (Herrero 2000). The centrality index based on *betweenness* is often used, with more success, to extract the words that are relevant in unstructured texts. This index has shown superiority over the use of word frequency to extract those words that are relevant to a set of documents (Hotho et al. 2005).

Communities, commonly called subgroups, are associated with an area of the graph that has a relatively high density. In some cases, these clusters allow to derive a family of related concepts.

4 The use of Twitter to analyze social behavior

The rise of social web, particularly Twitter, provides new opportunities to collect real-time data in large quantities directly from users. Big data can be analyzed in various ways to examine patterns in a wide range of subjects. In this way, tweets can be analyzed to track reactions to events, since Twitter enables to extract tweets and compose actual corpus.

There has been a lot of linguistic research applied in tweets. An increasing number of empirical analyses of sentiment and mood are based on textual collections of data generated on Twitter. In fact, they use sophisticated algorithms to preprocess, apply grammatical rules and classify tweets in mood categories. In this way, we found, for example, the use of a lexicon-based classifier as a data set that is also classified using Support Vector Machine and/or Naïve Bayes (Wijaya et al. 2013; Martínez and González 2013). A classifier is specifically tuned for tweets, using keywords, sentences and emoticons to determine the mood of each tweet (Bertrand et al. 2013). Several methods have been already proposed for exploring tweets to detect people's mood changes throughout the day (Hotho et al. 2005; Lampos et al. 2013).

In general, studies analyzing tweets by combining different sentiment analysis algorithms have been able to give new insights about patterns and motivations behind human social behavior, preferences, and interactions (e.g., Chung and Mustafaraj 2011; Dodds and Danforth 2010; Gruzd et al. 2011; Kramer 2010).

In our previous work (Abascal-Mena et al. 2013), we showed some results about analyzing tweets to study users' behavior before, during and after the presidential elections of 2012, in Mexico by applying semantic and ethnosemantic (on humor) techniques.

Following mood and sentiment analysis, several approaches to analyze, extract and automatically detect meaning in tweets have been proposed in the last years. They leave evidence of the limitations that already exist in the analysis of messages when applying only semantic approaches. The problem here is that mood and sentiment analysis form part of semantic interpretation in linguistic sciences. So, semantic interpretation involves removing features that are specific to a particular linguistic and cultural context.

An emergent social network, like Twitter, is a good source of (big) data to study social interactions as human social behavior. Thereby, an analysis of tweets should not only be based on the linguistics aspects but should also apply SNA, model tweets like a graph, to differentiate it from traditional sociological studies, which assume that it is the attributes of individual actors that matter. SNA produces an alternate view, where the attributes of individuals are less important than their relationships and ties with other actors within the network. The initial corpus is composed by 4509 tweets that were extracted from different days (see Table 1). For example, for December 6 we have extracted 1000 tweets but some of them correspond to previous days and in the corpus presented in Table 2 is shown the corpus used after a clean process.

A research establishes that the "exposure" to a social structure of friendship patterns influences individual behavior. This approach has turned out to be useful for explaining many real-world phenomena as we attempt to prove that it could be used to contextualize and to reveal relevant linguistic and social categories and their dynamics.

Table 1 Information about the corpus extracted each day with tweets corresponding to different days

Date	Number of tweets	Unique users	Words	Main words	Maximal term length
Dec 6	1000	672	15,159	1711	30
Dec 10	412	321	6854	807	30
Dec 11	599	437	8632	1203	26
Dec 12	899	596	12,975	1589	33
Dec 13	1000	572	14,544	1904	35
Dec 16	599	402	7945	1034	26

Table 2 Information about the corpus extracted each day with tweets corresponding to a unique date

Date	Number of tweets	Unique users	Words	Main words	Maximal term length
Dec 6	1000	672	15,159	1711	30
Dec 10	423	311	5564	843	24
Dec 11	399	286	6231	990	26
Dec 12	219	180	3052	652	33
Dec 13	296	224	4165	988	35
Dec 16	144	104	2138	420	25

Corpus composed by 2481 tweets

5 The *#noalospluris* dataset

To compose our corpus, we selected tweets about a citizen movement in Mexico that was carried out between 2009 and 2013 and that collected signatures to remove from the Chamber of Senators and Representatives all the multi-members who did not have their status by the legitimate vote of citizens. Thus, to gain a wider audience the movement used the tag *#noalospluris* in Twitter to get the attention of politicians and to show them the repudiation and discomfort of people about the existence of multi-members. In late 2013, this movement had its peak in Twitter, the tweets showed not only the labeling of *#noalospluris* but the main problems afflicting the people. In our corpus, we try to show how the movement of *#noalospluris* became a forum that provided a platform for the expression of the problem of the existence of deputies and senators multimember but also showed various problems that exist in Mexico.

Since January 2014, the movement has replaced the tag *#noalospluris* by *#intelecto* (meaning that the only way to gain against political imposition is the use of intellectual and constructive ideas). As confirmed by Heverin and Zach (2012) and Bastos et al. (2012) labels evolve over time, especially in the context of economic crisis and political uncertainty.

Our corpus containing the label *#noalospluris* is composed by 2481 tweets sent between 6 and 16 December 2013. In our analysis, we do not include the weekend of 14 and 15 December that have a radical decrease in the number of tweets because the radio program that promotes the use of the hashtag *#noalospluris* only has its broadcast from Monday to Friday.

To be able to obtain only the tweets that were sent exactly in the analyzed date we have decomposed each corpus by dates. In this way, we have obtained corpus that are composed with tweets of a unique date. In Table 2 is shown the corpus used after a process for eliminating stop words, punctuation, website addresses, actors and special characters excepting the hashtag symbol.

In recent years, we have seen the relationship that has Twitter with other traditional media as we find topics on Twitter that occupy the main headlines in the news (Kwak et al. 2010). This is an essential feature of Twitter: cover and discuss important events in society. The tweet is valid for a couple of days, what is important is what you read here and now. Yesterday is no longer news. The central mechanism used to highlight a tweet is the use of hashtags, as we mentioned before, that place a tweet in a specific subject making it immediately accessible to millions of users. The hashtags, because of brevity of Twitter, are keywords that show the current reality. Before making an automatic analysis of the recollected tweets, we were interested in analyzing the meaning of hashtags cited in the corpus to be able to understand visualizations of the information.

In this way, the use of hashtags corresponds to the need to present and discuss, quickly, important events and happenings. Thus, we find in the analyzed corpus some hashtags that disappear within the days as some others appear only for a couple of days but are specific to an event that happened in those days. Such is the case of *#diputadoencueratriz* (naked deputy) and *#posmeencuero* (so I'm going to undress me) that appeared, on 12 and 13 December 2014, to discuss and criticize a Mexican congressman that in a full podium has naked himself to protest against the energy reform.[1]

Despite that *#noalospluris* is used for a specific cause we find the inclusion of other hashtags which are pronounced against government actions. However, *#noalospluris* is a forum that is widely read and the inclusion of hashtags that are out of the context has a special objective: reach a larger audience, a strategy to increase the visibility. Thus, other labels that make their appearance between *#noalospluris* are *#noaumentocamionjalisco* and

[1] http://www.dailymail.co.uk/news/article-2522868/Mexican-congressman-takes-clothes-angry-protest-historic-energy-privatization-scuffles-break-doors-barricaded.html.

#*noalcamionazo*. These tags were used, since December 13, to report an increasing dissatisfaction with the rate of public transport in the State of Jalisco, Mexico. The introduction of these tags is done by the use of sentences from Mexican political celebrities (José López Portillo, Vicente Fox, Manuel López Obrador, Enrique Peña Nieto). These are used to criticize politicians who have not done anything about the country. Here, some examples of these comments, traduced to English:

"*In #Mexico* \" *I will defend the peso like a dog*\" *José López Portillo. #NoALosPluris y #NoAlCamionazo in #Jalisco*"

"*The 75 % of the #mexicanos have 1 washer, and not the one of 2 legs, but the one of two feet*\" *Vicente Fox #NoALosPluris #NoAlCamionazo in #Jalisco*"

"*In #Mexico* \" *To the hell with their institutions* \" *Andrés Manuel López Obrador #NoALosPluris y #NoAlCamionazo in #Jalisco*"

"*#Mexico* \"*Eat and then you go*\" *Vicente Fox say #NoALosPluris y #NoAlCamionazo in #Jalisco*"

"*#Mexico* \"*I'm not the woman of the house*\" *Enrique Peña Nieto #NoALosPluris y #NoAlCamionazo in #Jalisco*"

The introduction of new hashtags in a certain theme, even if there are not appropriated, is used to reach a larger public but also they correspond to new problems that need to be widespread. The analysis of this hashtags will let us gain a better comprehension of the network since we know, now, that all the information contained has a certain importance.

6 Methodology: extraction and composition of the corpus

In the proposed method, we analyze tweets using three steps. In the first step, we extract tweets from Twitter containing a given hashtag. In the second step, we parse them to clean them and to create two different tables: one with all the labels of the important words contained in the corpus (nodes) and another containing the relationships between each couple of words (edges). These two steps are automatically performed using R which is an interpreted computer language designed for statistical data analysis. The third step concerns the creation of the network using Gephi that is an open source software for graph and network analysis (Bastian et al. 2009). Gephi is a visualization and manipulation tool that allows users to interactively browse and explore graphs.

We have selected the language R for our approach since it contains some packages like TwitteR (allows to connect to Twitter and retrieve tweets) and *tm* (for text mining) that are useful for data import, corpus handling, preprocessing, data management and for the creation of term-document matrices.

Then, using *R* we have developed a computer program that is able to:

1. Extract tweets containing a given tag. In our case, we extracted tweets containing the hashtag *#noalospluris*.
2. Eliminate stopwords and punctuation.
3. Eliminate website addresses (anything that starts with *http*).
4. Eliminate actors (that is, all sequences that contains the symbol @ as the first character).
5. Delete numbers and special characters (except the symbol # to left all the hashtags contained in the tweets) and letting the numbers and special characters in tags.
6. Change the corpus to lowercase.
7. Eliminate accents.
8. Get the frequency of each word.
9. Generate the node for words that are above the third quintile (Q3). A quintile represents 20 % (or one-fifth) of the total number of words of a given corpus.
10. Determinate the edges for each node.

Once we have captured the tweets, we make a preprocessing, manipulation, cleaning and formatting to compose the corpus. The main structure for managing documents, using *tm* package in *R*, is the corpus that represents a collection of text documents (in our case a collection of tweets). With *tm* package *R*, we are able to create a term-document matrix from our corpus. The matrix contains the list of words extracted from the tweets (list of nodes) and a list with the relations between words or co-occurrence of words in a tweet (list of edges). The matrix is exported as a .CSV file to be used in Gephi.

In the next section, we present the analysis we have carried out on our corpus by applying techniques of SNA.

7 Applying social network analysis

Tables that were generated by *R* (table of nodes and edges) are used in Gephi to:

- Generate a network for the corpus.
- Calculate the degree for each node.
- Calculate the centrality for each node.
- Detect the main communities.

7.1 Generation of the network of the corpus

Some elements, such as the use of the hashtag, can be seen with a naked eye by reading a part or the whole corpus. However, the use of techniques from the SNA will give us greater certainty about the use and positioning that take some words and hashtags over the time.

In Figs. 1, 2, 3, 4, 5 and 6, we show the network of concepts generated for the each day of our analyzed corpus. As we explained before, a network is composed by nodes and edges (also called links or connections). In our case, nodes are the concepts (main words) within the network, and edges are the relationships between the concepts.

The first network, Fig. 1 shows at the middle (the green node) the hashtag *#noalospluris* and it is composed by 18 communities. In this network, there is no node with the same degree as *#noalopluris* so the others nodes have less connections with other nodes.

Figure 2 shows a network that is composed by 15 communities. In this case, *#noalospluris* is not the central node and even more there are other nodes like Mexico, society and citizens that are more important by relying different communities and with a higher degree. Also, in the network, it is possible to see groups of nodes that are not connected with the main theme. This means that they use the hashtag *#noalospluris* but they do not have any relation with it.

In Fig. 3, there are some subgroups that are represented by nodes where their degree and centrality is higher than *#noalospluris*. That is the case of Mexicans (mexicanos), Mexico or people (gente). It is interesting to see that around the main thematic of *#noalospluris* what matters in this network is people. As the precedent network, some subgroups are disconnected and others begin to move far away. The network is composed by 20 communities that contain terms that are related and that correspond to similar contexts.

The network of Fig. 4 is composed by 20 communities in which some of them are composed only by a few terms. That is the case of, for example, the community 18 that has only the terms *effective* and *suffrage* which are often used to reclaim real democracy.

The network corresponding to December 13 has 18 communities as *#noalospluris* is again placed at the center like in December 6 and 12. In this network, there are no subgroups behind the main theme. This means that all the

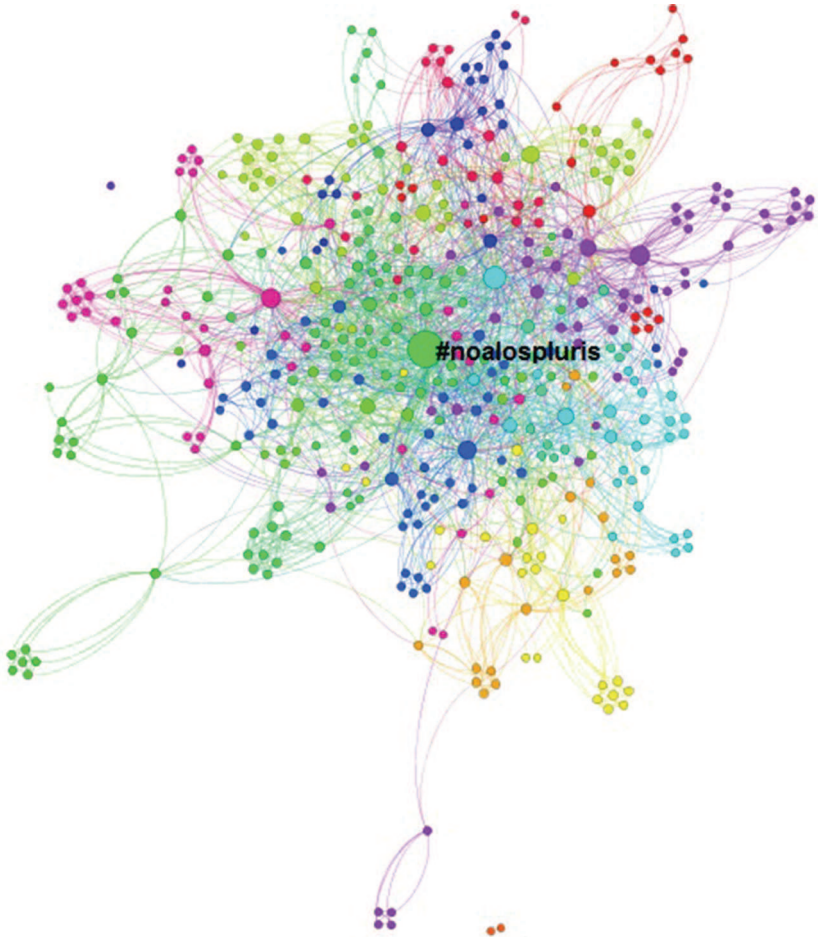


Fig. 1 Network of concepts generated for December 6, 2013 containing the hashtag: *#noalospluris*

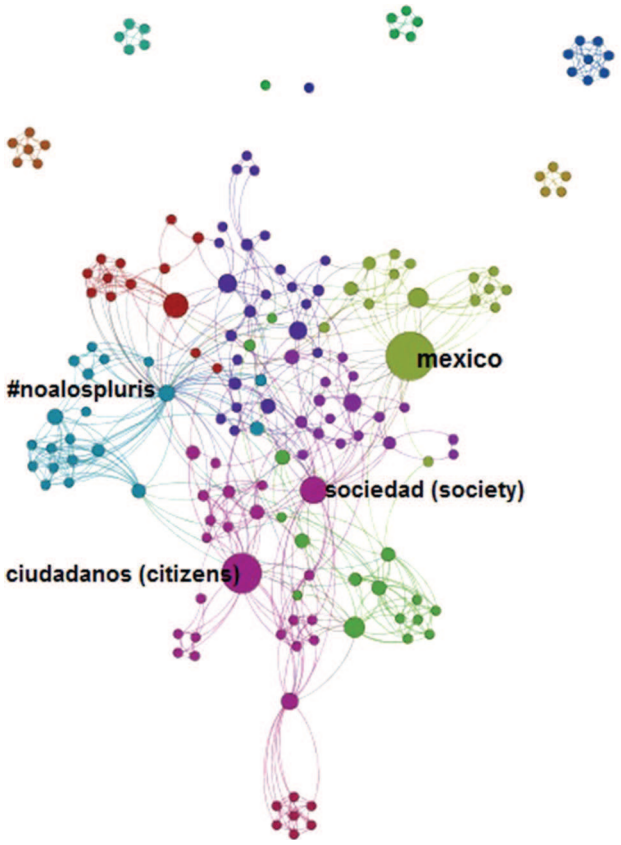


Fig. 2 Network of concepts generated for December 10, 2013 containing the hashtag: *#noalospluris*

subgroups have a relation between *#noalospluris* in contrast with the subgroup of congratulation of December 13.

On December 16, the network is composed by 10 communities because only the subgroups that have some relation within the main theme remain after a few days. And subgroups that unify big problems in Mexico like is the case of PEMEX (Petroleum of Mexico) or the new Tax Reform. However, *#noalospluris* is in the center and other problems are far away.

Visual representation of the network is important to understand the evolution of concepts used over the time as they tend to concentrate in communities that share the same thematic. The concepts can evolve to use meta-information, in the tweets, by the use of more precise hashtags as the time goes on.

7.2 Degree and centrality

The networks that are obtained have directed edges so their edges have a direction associated with them. So, using the degree, we have obtained the number that indicates how many nodes are connected to a respective node. For each hashtag, we identify the degree to compare the use of the hashtag in a period of time. Is this degree indicative of the permanence of the hashtag all along the entire movement? The often use of a hashtag in the tweets means that it is precise, pertinent and it can remain all the movement?

The degree shows the number of nodes that a focal node is connected to, and measures the involvement of the node in the network. The measure of the degree has its limitations: the measure does not take into consideration the global structure of the network. However, the identification of the centrality of nodes or focal points takes into account the whole network.

The centrality based on the intermediation (*betweenness*) measures the frequency with which a node appears on the shortest path between two other nodes, and is able to funnel the flow in the network. In so doing, a node can assert control over the flow. Analyzing the results, it is interesting to show the eccentricity of the nodes: the longest path of a node to reach any other in the network. The degree and the eccentricity are shown for the hashtags that appear in at least 2 days of our corpus even if they are not consecutive days (see Tables 3, 4).

In Table 5, we present the hashtags that have a centrality greater than zero (number indicating how many nodes must go through to travel a given network node) and that are used repeatedly in more than 1 day.

From these results, we found an important relationship between the degree and the centrality: the centrality is greater as is higher the degree and there is big chance of having a centrality (bigger than zero) if the hashtag is used repeatedly over time. If we apply this theory to concepts that are not used as hashtags but have a significant centrality and they are used in at least two different days, we can find concepts that appear as important nodes and that are part of clusters (subgroups) for a precise thematic. This means that there are hashtags that remains all over the time because they are known by the community as they are keywords that characterize a movement. However, some terms (known as concepts because they are a generalization from particular terms) also remain by the pass of the days because they are crucial to contextualize the main objective and the problem of the social movement.

In Fig. 7, we show these concepts as we are going to present the clusters in the next section. Some of these concepts are also used as hashtags as in the case of *#mexico.*

7.3 Subgroups

When the tweets of a particular topic are analyzed they might appear like if they are disorganized without a specific center where only spontaneity and anarchy prevail. But, while the tweets are set in our mind like a swarm of birds or insects attacking irrationally, as the meaning of Twitter, we find that the interior of the network is



Fig. 3 Network of concepts generated for December 11, 2013 containing the hashtag: *#noalospluris*

organized, rational and creative, a "*swarm intelligence*". This means that although there is not a centralized structure that dictates the users in what they have to contribute, the users seems organized to add facts and information to complement the main theme and sometimes persuade the readers. In this logic, according to Arquilla and Ronfeldt (2003), the swarm reaches its maximum effectiveness and its greater attack power when the network members are not together to combat their forces, but they concentrate on the dispersion, sharing important information so the swarm can reach its final destination. In this way, in social movements, it is possible to find groups of users with different points of view, positions and objectives. These groups of users are clearly identified by organizing them in clusters where the same kind of discourse can be found. So, there is a global behavior and for the case of terms or concepts if they are grouped according to their interactions, it is possible to determine located subthemes and positions about a certain movement. By studying networks, it becomes necessary to know the community structure, set of nodes that are grouped and in each set nodes are densely connected internally, to be able to extract and study only the pertinent information. Merely, in the case of tweets, the finding of subgroups can convey useful information where the groups introduce different themes and have some degree of independence. Also, it is possible to isolate subgroups in which information obeys to different objectives, for example the case of bots.

The detection and characterization of communities in the study of networks presents some problems like trying to find polarization where it can be key information for tasks such as opinion analysis. One highly effective approach is the optimization of the quality function known as "modularity" over the possible divisions of a network. Network segmentation with high modularity indicates that the social graph may be divided into clusters having many internal connections among nodes and few connections to the other group, what is widely accepted as an indication of polarization (Conover et al. 2011). In this polarization, it is possible to find subgroups composed by *sociosemantic* concepts in which subthemes are exposed.

So, to analyze the composition of our corpus we have decided to apply, in Gephi, the modularity algorithm that allows the visualization of subgroups. The subgroups

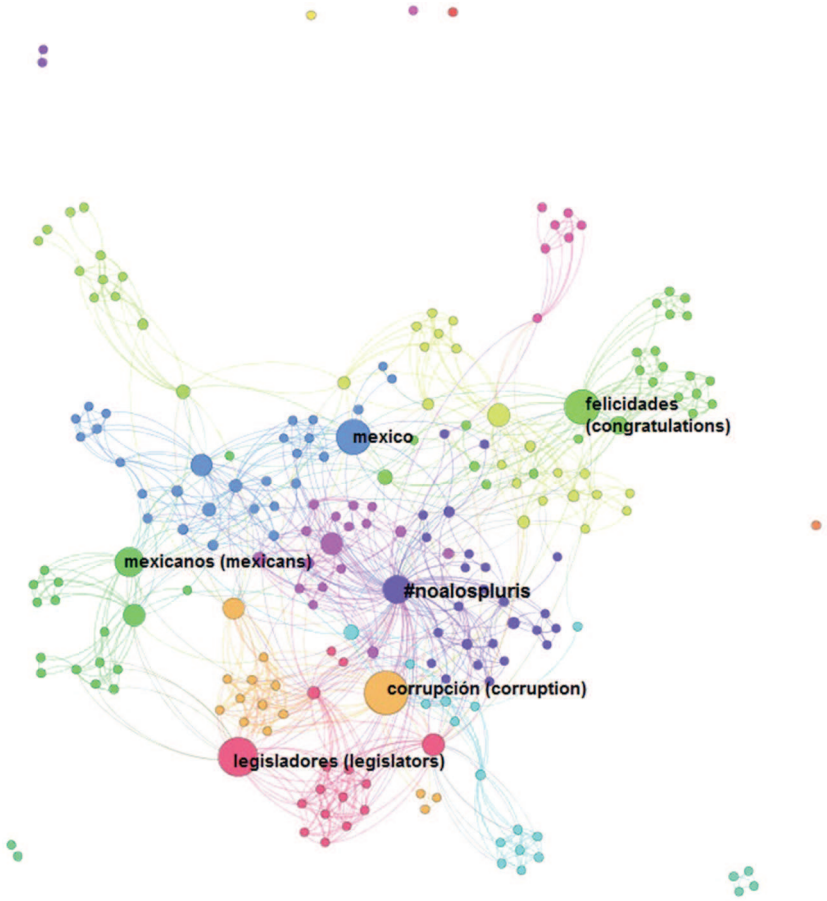


Fig. 4 Network of concepts generated for December 12, 2013 containing the hashtag: *#noalospluris*

concentrate terms that are preceded by a concept, a generalized term, and they work like keywords that characterize subthemes.

In this way, we have:

- 18 communities for December 6 and December 13,
- 15 communities for December 10,
- 20 communities for December 11 and December 12,
- 12 communities for December 16.

The number of communities in each corpus is not very interesting without reviewing the percentage of words that compose the daily communities. So, we find that the biggest agglomeration of words in a community is about 24.83 % found on December 16, the last day. The minimum percentage is found on the first day. Our results show that at the last day there is a crucial diminution of communities (Table 6).

By analyzing the communities that have a greater concentration of words, we find that the hashtag *#noalospluris* shows them. In Table 6, we show the minimum and maximum percentage of words obtained for the communities daily analyzed. Also, we show in bold the percentage obtained for the community containing *#noalospluris* which in some days (December 6, 11, 13 and 16) corresponds to the community with maximum percentage of words. See "Appendix" for a more detailed presentation of the percentage of words obtained by each community.

The communities containing *#noalospluris* are in bold in Table 6 to show that *it's not necessary to analyze the entire corpus but only the communities with high percentage in order to be able to know more about the corpus*. In this way, we have found the subgroups (communities) that provide concepts or words that help to disambiguate (with semantic information) the entire corpus. Also, the community with the tag *#noalospluris* has almost all the concepts shown in Table 6. This way, we find that concepts with biggest centralities are concentrated in the same community.

Our results show that it is not necessary to read the entire tweets to know about the theme because in some ways we can have tweets that are included with the main hashtag without having something in common. Instead, it is important to concentrate our attention in biggest centralities to retrieve the main concepts of the theme. However, communities with a percentage above 10 % concentrate important concepts as they provide more semantic information about the tweets.

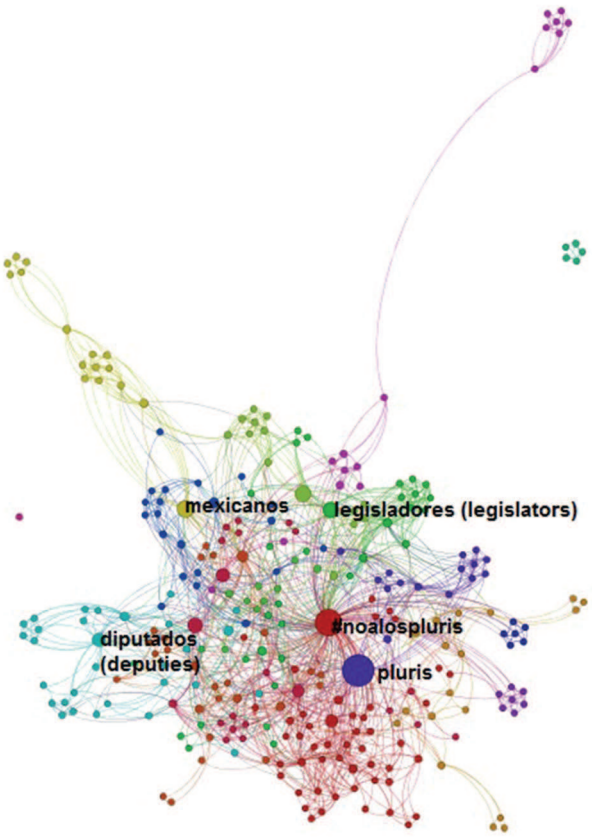


Fig. 5 Network of concepts generated for December 13, 2013 containing the hashtag: *#noalospluris*

8 Comparison between similar movements: *#noalospluris* and *#menospluris*

In this section, we present a comparison between similar movements but also, at the same time, we validate the process to extract and visualize *sociosemantic* communities.

The detection of *sociosemantic* communities can give us insights of what are the main issues in social and political movements. Even more, it is possible to compare movements by their communities to retrieve similarities and differences between them. Thus, two movements were selected to be compared: *#noalospluris* and *#menospluris*.

The old Institutional Revolutionary Party (PRI), of Mexico, returned to power in 2012, and is determined not to lose it again. After the historic constitutional reforms—in which highlights the energetic—and the upcoming elections of 2015 the strategy of the PRI does not admit mistakes. However, it requires strength to be reflected in concrete results. To generate sympathy among voters, since 2014, the PRI promotes the use of the hashtag *#menospluris*, similarly to the movement *#noalospluris*, and adopts it like his flag. *#menospluris* collect signatures via a website[2] which uses the slogan "+ *de signatures—pluris. Major support citizen in the shortest time.*"

In both movements, the discomfort of citizenship about the existence of deputies and senators multimember is shown, but their intention is not the same as we show in the next analysis of communities.

The network generated for *#noalospluris* uses the tweets extracted for December 16, 2013 (Fig. 8). The most relevant terms extracted from our analysis are (in this order): interests, represent, public, politics, understand, excellent, pri, campaign, corruption, mediocre, enough, laws, candidatures, Mexico, sign, join, taxes, module, deputies, disappearance, class, supports, remove, citizens, elected, excessive and action.

Table 7 shows the communities that were obtained for *#noalospluris* and it is possible to extract (encapsulate) in a community terms corresponding to other language. That is the case of the community 7 where the terms were already in English.

The network generated for *#menospluris* uses the 231 tweets that were extracted for October 9, 2014 (Fig. 9). By analyzing the tweets the most relevant terms are: menospluris, multimember, proposal, consultation, pri, parties, more, signatures, representation, goal, politics, debate, citizen, power, t-shirt, congress, money, deputies, Mexico, opposed, resignation, find, and always.

Comparing Table 7 with Table 8, it is possible to visualize some similarities: the community with the highest percentage of concepts contains the main hashtag. In the case of *#noalospluris,* it corresponds to the community with 24.83 % and for *#menospluris* is 25 %. In both cases, we find that four communities contain more than 10 % of concepts and that the word *Mexico* is contained in the main community.

By analyzing the main community, it is possible to extract the main differences with the two movements even if they are apparently similar (see Table 9):

- the movement *#noalospluris* is used, also, to report or promote other themes (*#coahuila*, *#jalisco*, *#mexico*, *#noalospluris*, *#noaumentocamionjalisco*, *#saltillo*, etc.) that are affecting the social and economic situation of Mexicans,
- more negative words are included in *#noalospluris* than in *#menospluris*. For example, the main community of *#noalospluris* contains: lower, corruption, cost, bribery and tapped.

With these differences it is evident that *#noalospluris* was a movement carried out by the society. In the case of *#menospluris,* it is actually carried out by the government.

[2] http://pri.org.mx/masconmenospluris/ consulted 10 October 2014.

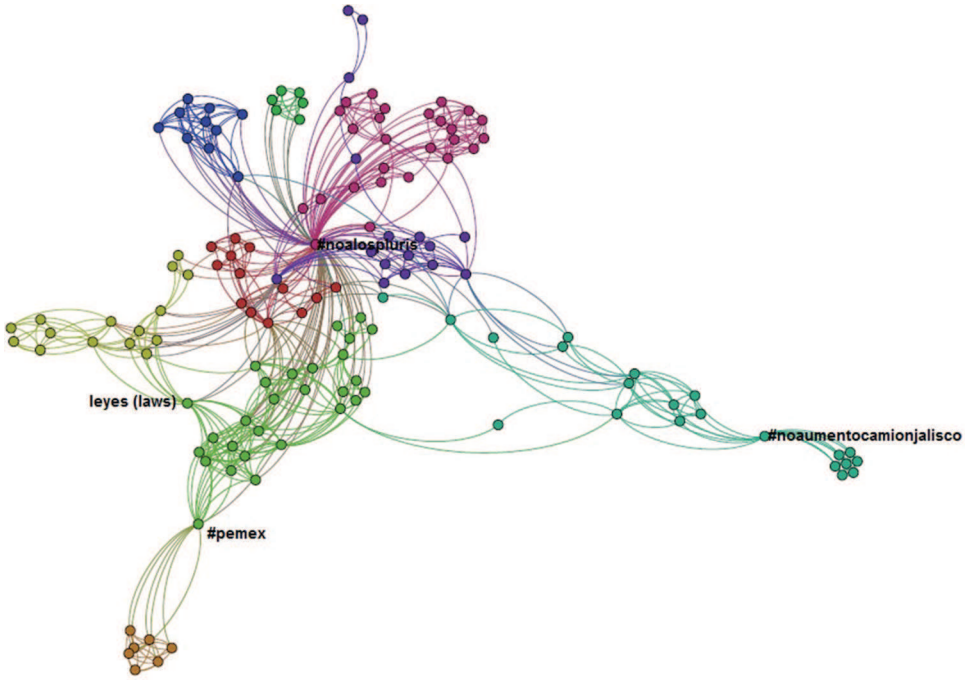


Fig. 6 Network of concepts generated for December 16, 2013 containing the hashtag: *#noalospluris*

Table 3 Degree for the principals hashtags contained in the corpus of *#noalospluris*

Hashtags	Degree					
	Dec 6	Dec 10	Dec 11	Dec 12	Dec 13	Dec 16
#intelecto	7	8	16	8	18	14
#jalisco	–	–	–	–	20	4
#mexico	20	6	62	21	45	27
#noalospluris	165	51	153	97	161	103
#pemex	–	–	–	13	–	10
#piedrasnegras	–	5	5	–	–	–
#posmeduermo	–	6	7	–	–	–
#posmeencuero	–	–	–	5	2	–
#reformaenergetica	–	–	9	14	–	–
#reformapolitica	12	–	–	–	5	–
#revoluciondelintelecto	7	–	8	–	–	–
#saltillo	–	7	9	–	–	7
#teregalo	3	–	4	–	–	–
#torreon	–	9	7	–	–	–
#yoquieroun2014sin	3	–	8	–	7	–

On one hand, *#noalospluris* could be considered a negative movement where there is an abundance of words like corruption, bribery, hustlers and inefficiency. Meanwhile, *#menospluris* is a movement of agreements, construction and consulting. In both movements, Mexico appears in the most important community. The movement led by the PRI, *#menospluris* is clearly controlled by not including issues unrelated to the movement. Instead, *#noalospluris* becomes a forum in which issues affecting different states of Mexico (Coahuila, Saltillo, Jalisco), reforms and laws to be discussed are addressed.

In this comparison, it is also possible to have the same results as in Sect. 6 where *it's not necessary to analyze the entire corpus but only the communities with high percentage in order to be able to know more about the corpus*. Communities reveal *sociosemantics* information that allows understanding the main movement.

Table 4 Eccentricity for the principals hashtags contained in the corpus of *#noalospluris*

Hashtags	Eccentricity					
	Dec 6	Dec 10	Dec 11	Dec 12	Dec 13	Dec 16
#intelecto	8	5	6	4	7	4
#jalisco	–	–	–	–	5	3
#mexico	5	4	4	4	5	3
#noalospluris	4	3	4	3	4	2
#pemex	–	–	–	4	–	4
#piedrasnegras	–	6	5	–	–	–
#posmeduermo	–	4	6	–	–	–
#posmeencuero	–	–	–	7	1	–
#reformaenergetica	–	–	6	6	–	–
#reformapolitica	3	–	–	–	4	–
#revoluciondelintelecto	5	–	5	–	–	–
#saltillo	–	4	6	–	–	1
#teregalo	6	–	1	–	–	–
#torreon	–	4	6	–	–	–
#yoquieroun2014sin	6	–	7	–	5	–

Table 5 Hashtags that have a centrality greater than zero

December 6, 2013		December 10, 2013		December 11, 2013	
#noalospluris	462	#noalospluris	66	#intelecto	27
#teregalo	49	#saltillo	1	#mexico	48
		#torreon	27	#noalospluris	306
				#revoluciondelintelecto	262
				#saltillo	26

December 12, 2013		December 13, 2013		December 16, 2013	
#mexico	10	#jalisco	9	#mexico	24
#noalospluris	603	#mexico	95	#noalospluris	366
#reformaenergetica	90	#noalospluris	1724		

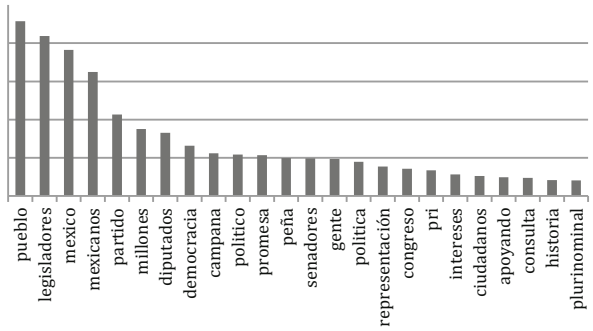


Fig. 7 Average of centralities obtained for each concept

By selecting corpus in Spanish, we were able to identify in the communities the sense of the terms and study *sociosemiotics* aspects like: signs (by the use of abbreviations, symbols) where the relation between the meanings is purely conventional and culturally specific. Also, the connotation of tweets is not described here but it is other aspect that can be appreciated by studying the resulting communities.

9 Conclusion

Social networking sites have provided alternative mediums for citizen communication and participatory journalism. The global civil society and public sphere now have the means to exist independently from political institutions (Castells 2008). Thus, social media used as organizing and mobilizing tools, as a medium for debate, dialog and collective decision-making. That is the case of Twitter: a new medium for unfiltered and decentralized participation.

Table 6 Percentage (minimum and maximum) of words obtained for the group of communities analyzed for each day as the percentage obtained for the community containing *#noalospluris*

	Dec 6	Dec 10	Dec 11	Dec 12	Dec 13	Dec 16
Minimum percentage	0.22	0.6	0.27	0.45	0.34	1.34
Percentage for community containing *#noalospluris*	**15.5**	**11.98**	**15.85**	**11.71**	**17.12**	**24.83**
Maximum percentage	**15.5**	14.37	**15.85**	13.06	**17.12**	**24.83**

Bold values show the communities containing the hashtag *#noalospluris*

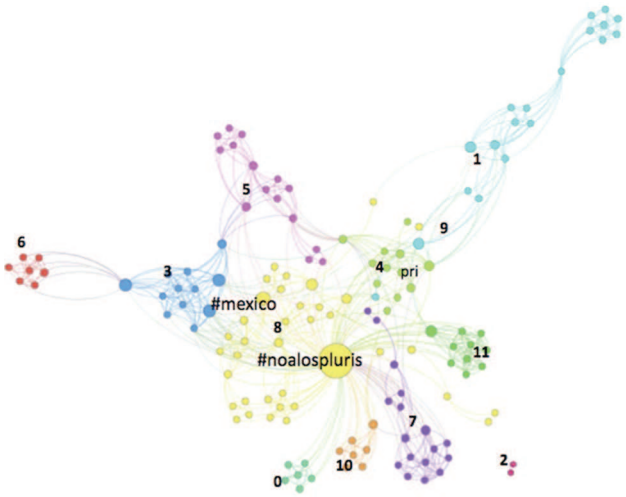


Fig. 8 Network of concepts generated for the movement: *#noalospluris*

McLuhan suggests that the message is the medium and to this extent there is a factor of cultural transformation that each medium causes when it gets implemented. Such is the case of Twitter, with 231.7 million active users around the World; it is a medium that develops its own language as it transforms the culture (Orihuela 2006).

The study of messages that are sent via Twitter are crucial to understand, for example, how non-traditional actors set the political agenda or how activists have devised social media for mobilizing and shaping the language in which movements are discussed.

This paper has presented a study along the cutting edge of the current analysis of online social network in relation with the contents communicated among users. Twitter data are carefully selected around the hashtag *#noalospluris* to study the specified content in relation with other contents that users bring to connection. A separate network of hashtags related (in tweets) is constructed for different days; the networks are analyzed within advanced Gephi package, providing several measures—degree, betweenness centrality, communities, as well as the longest path, by which the evolution of communication around specified concepts is quantified. The results are compared with other corpus containing the hashtag *#menospluris* to demonstrate the applicability of the proposed methodology.

Table 7 Percentage and concepts obtained for each community of *#noalospluris* corresponding to December 16, 2013

Community	% of concepts	Relevant terms
0	3.36	*#posmesalto*, *#todoesculpademina*, lost gift
1	4.7	Captives, defended, defending, taxes, paid
2	1.34	Jaded, keep
3	**12.75**	*#intelecto*, class, democracy, desire, scruples, heard, excellent, mediocre, goal, moral, political, *#politicarevolucion*, people, respected, third world country, will
4	**11.41**	Friends, street campaign, install, market, module, profile, standing, can, join, victory
5	**10.07**	America, fanned, starting, missing, treasury, inefficiency, laws, need, word, political, reform, secondary
6	8.05	Enough, elected, excessive interest, popular, representing, salaries, have, vote, voice
7	4.03	Add, help, now, please, support
8	**24.83**	*#coahuila*, *#jalisco*, *#mexico*, *#noalospluris*, *#noaumentocamionjalisco*, *#saltillo*, action, high, appears, arranged, lower, corruption, cost, elimination, obviously, Mexico, participate, multimember, proposal, public, bribery, tapped
9	8.05	Supports, citizens disappearance, deputies, eliminate, signature, ask, pri, visit
10	4.7	Candidates, require, independent, marches, resources, transparency
11	6.71	*#pemex*, learn, authoritarianism, understand, sample, assholes

Bold values show the communities that have a percentage above 10 %

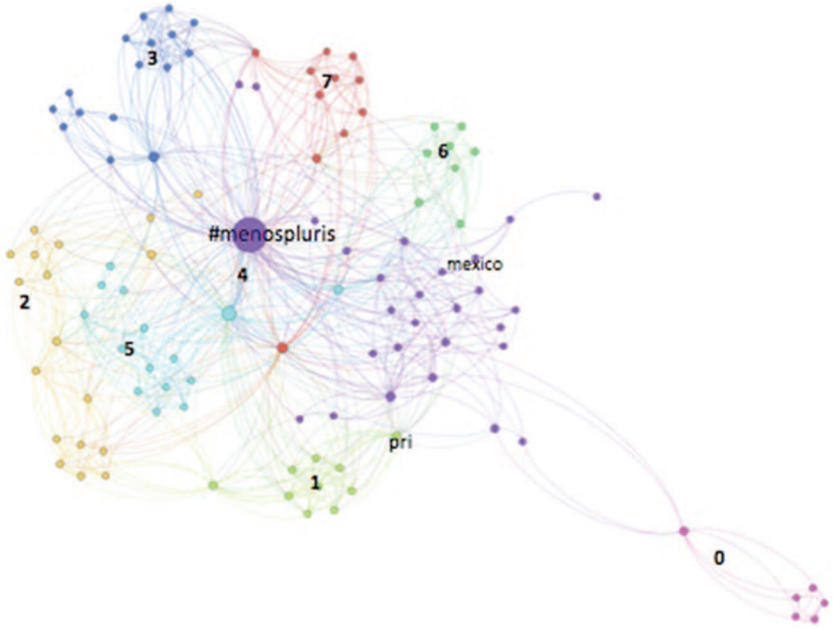


Fig. 9 Network of concepts generated for the movement: *#menospluris*

Table 8 Percentage and concepts obtained for each community of *#menospluris* corresponding to October 9, 2014

Community	% of concepts	Relevant terms
0	5.36	Activism, recognition, work
1	8.93	Raise, trampling, must, denunciation, hegemony, pri, proposal, retake, voice
2	**16.07**	Known, *#consultaenergetica*, consultation, money, drastically, expenses, boost, majority, parties, political, popular, trim, decrease, representativeness, *#salariodigno*
3	**14.29**	Slap, listen, extreme, reach, municipalities, multi, poverty, power, results, always, alive
4	**25**	Agreements, citizen, congress, construction, consultation, deputies, signing, Mexico, *#menospluris*, multimember, exceeds, all, lets go, senatorship
5	**13.39**	Article, t-shirt, campaigns, tailoring, contempt, expect, may, politics
6	7.14	Support, *#elentrego*, firms, overcome
7	9.82	Now, search, discussion, understand, today, opposed, renunciation

Bold values show the communities that have a percentage above 10 %

Table 9 Communities containing the main hashtag *#noalospluris* and *#menospluris*

#noalospluris	**24.83 %**	*#coahuila*, *#jalisco*, *#mexico*, *#noalospluris*, *#noaumentocamionjalisco*, *#saltillo*, action, high, appears, arranged, lower, corruption, cost, elimination, obviously, Mexico, participate, multimember, proposal, public, bribery, tapped
#menospluris	**25 %**	agreements, citizen, congress, construction, consultation, deputies, signing, Mexico, *#menospluris*, multimember, exceeds, all, lets go, senatorship

Our study is absolutely in the current trend of analysis of online social networks that, going beyond mere topology, reveals relevant linguistic and sociosemantic communities in which the interaction between the words is particularly intense, allowing to categorize them in certain subjects and to extract *semantic* interpretation from the network. The communities containing more than 10 % of concepts, over a period of time, provide insights to be able to understand the entire corpus.

Future work should apply techniques to study the tweets based on the gender or key actors to provide mechanisms to enhance the probability of obtaining relevant information according to the role of users (from individual to group behavior) that is relevant for social dynamics. Also, the

study of user groups it is crucial to analyze tweets according to their social interactions.

Appendix

In this appendix is shown in detail the percentage of words obtained daily by each community. For example, for December 11 we have 20 communities where from the total number of concepts extracted for this day we got that the smallest community has 0.27 % of concepts while the biggest has 15.85 %.

Dec 6	Dec 10	Dec 11	Dec 12	Dec 13	Dec16
0.22	0.6	0.27	0.45	0.34	1.34
0.44	0.6	0.82	0.45	0.68	3.36
0.66	2.99	0.82	0.45	1.71	4.03
1.31	2.99	1.09	0.45	2.05	4.7
2.62	2.99	1.09	0.9	2.4	4.7
4.15	3.59	1.64	0.9	3.08	6.71
4.15	4.19	2.19	1.8	4.79	8.05
4.8	4.79	2.46	2.25	4.79	8.05
5.02	7.78	3.28	2.25	5.48	10.07
5.02	9.58	3.83	3.15	5.48	11.41
5.24	9.58	4.1	5.41	5.48	12.75
5.46	10.18	4.64	6.76	6.51	**24.83**
6.11	**11.98**	6.28	6.76	6.51	
8.52	13.77	6.83	7.21	7.19	
9.39	14.37	7.38	7.66	7.19	
9.61		8.2	7.66	8.9	
11.79		8.74	9.46	10.27	
15.5		8.74	11.26	**17.12**	
		11.75	**11.71**		
		15.85	13.06		

Bold values show the communities that contain the hashtag *#noalospluris*

References

Abascal-Mena R, López-Ornelas E, Zepeda-Hernández JS (2013) User generated content: an analysis of user behavior by mining political tweets. In: Ant Ozok A, Zaphiris P (eds) Online communities and social computing. Lecture Notes in Computer Science Volume 8029, Springer, Heidelberg, pp 3–12

Arquilla J, Ronfeldt D (2003) Networks and netwars: the future of terror, crime, and militancy. RAND, Santa Monica, p 380

Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. International AAAI conference on weblogs and social media

Bastos MT, Puschmann C, Travitzki R (2012) Tweeting political dissent: retweets as pamphlets in #FreeIran, #FreeVenezuela, #Jan25, #SpanishRevolution and #OccupyWallSt. Internet, politics, policy 2012: big data, big challenges? Oxford. Panel 6A: the Arab spring and political dissent, vol 2. Oxford Internet Institute, Oxford, pp 1–20

Bertrand KZ, Bialik M, Virdee K, Gros A, Bar-Yam Y (2013) Sentiment in New York City: a high resolution spatial and temporal view. arXiv:1308.5010

Brandes U (2001) A faster algorithm for betweenness centrality. J Math Sociol 25(2):163–177

Bruns A, Burgess JE (2011) The use of Twitter hashtags in the formation of ad hoc publics. 6th European Consortium for political research general conference, University of Iceland, Reykjavik

Castells M (2008) The new public sphere: Global civil society, communication networks, and global governance. Ann Am Acad Political Social Sci 616(1):78–93

Chew C, Eysenbach G (2010) Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 5(11):e14118. doi:10.1371/journal.pone.0014118

Chung JE, Mustafaraj E (2011) Can collective sentiment expressed on Twitter predict political elections?. In: Burgard W, Roth D (eds) Proceedings of the 25th AAAI conference on artificial intelligence (AAAI 2011). AAAI Press, Menlo Park, pp 1768–1769

Conover M, Ratkiewicz J et al (2011) Political polarization on twitter. In: Proceedings 5th International AAAI conference on weblogs and social media (ICWSM)

Crismore A, Markannen R, Steffensen M (1993) Metadiscourse in persuasive writing: a study of texts written by American and Finnish University students. Writ Commun 10(1):39–71

Cunha E, Magno G, Comarela G, Almeida V, Gonçalves MA, Benevenuto F (2011) Analyzing the dynamic evolution of hashtags on twitter: a language-based approach. In: Proceedings of the workshop on language in social media (LSM 2011), pp 58–65

Cunha E, Magno G, Almeida V, Gonçalves MA, Benevenuto F (2012) A gender based study of tagging behavior in twitter. In: Proceedings of the 23rd ACM conference on hypertext and social media. ACM, pp 323–324

Dodds PS, Danforth CM (2010) Measuring the happiness of large-scale written expression: songs, blogs, and presidents. J Happiness Stud 11(4):441–456

Freeman LC (1979) Centrality in social networks: conceptual clarifications. Soc Netw 1(3):215–239

Georgakopoulou A (2013) Narrative analysis and computer-mediated communication. In: Herring S, Stein D, Virtanen T (eds) Pragmatics of computer-mediated communication. Mouton, Berlín, pp 695–715

Gitlin T (1980) The whole world is watching: mass media in the making and unmaking of the new left. University of California Press, Berkeley

Gruzd A, Doiron S, Mai P (2011) Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics. In: Proceedings of the 44th Hawaii international conference on system sciences. IEEE Computer Society, Washington, DC

Habert B, Nazarenko A, Salem A (1997) Les linguistiques de corpus. Armand colin, Paris

Herrero R (2000) La terminología del análisis de redes: problemas de definición y de traducción. Política y sociedad 33:199–206

Heverin T, Zach L (2012) Use of microblogging for collective sense making during violent crises: a study of three campus shootings. J Am Soc Inform Sci Technol 63(1):34–47

Honeycutt C, Herring SC (2009) Beyond microblogging: conversation and collaboration via Twitter. 42nd Hawaii international conference on system sciences, Hawaii, pp 1–10

Hotho A, Nürnberger A, Paaß G (2005) A brief survey of text mining. Ldv Forum 20(1):19–62

Hyland K, Tse P (2004) Metadiscourse in academic writing. Appl Linguist 25(2):156–177

Kramer ADI (2010) An unobtrusive behavioral model of "Gross National Happiness". In: Proceedings of CHI 2010. ACM Press, New York, pp 287–290

Kwak H, Lee C, Park H, Moon S (2010) What is Twitter, a social network or a news media? In: Proceedings of the 19th international conference on World Wide Web. ACM, pp 591–600

Lampos V, Lansdall-Welfare T, Araya R, Cristianini N (2013) Analysing mood patterns in the United Kingdom through Twitter content. arXiv preprint arXiv:1304.5507

Martínez V, González VM (2013) Sentiment characterization of an urban environment via Twitter. In: Urzaiz G, Ochoa SF, Bravo J, Chen LL, Oliveira J (eds) Ubiquitous computing and ambient intelligence. Context-awareness and context-driven interaction. Springer International Publishing, Cham, pp 394–397

Orihuela JL (2006) La revolución de los blogs. Cuando las bitácoras se convirtieron en el medio de comunicación de la gente. La Esfera de los Libros, Madrid, 20fbb06, ISBN: 84-9734-498-7, p 283

Peirce CS (1931) Collected writings (8 Vols.). Ed. Charles Hartshorne, Paul Weiss & Arthur W Burks. Harvard University Press, Cambridge

Poell T, de Kloet J, Zeng G (2014) Will the real Weibo please stand up? Chinese online contention and actor-network theory. Chin J Commun 7(1):1–18

Rodrigues BGA, Silva IS, Zaki M, Meira Jr W, Prates RO, Veloso A (2012) Characterizing the effectiveness of twitter hashtags to detect and track online population sentiment. In: CHI'12 Extended Abstracts on Human Factors in Computing Systems, ACM, pp 2621–2626

Romero D, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. International world wide web conference (WWW 2011). Hyderabad, India

Russell A (2011) The Arab spring extra-national information flows, social media and the 2011 Egyptian uprising. Int J Commun 5:10

Singh MP (2000) A social semantics for agent communication languages. In: Dignum F, Chaib-draa B, Weigand H (eds) Issues in agent communication. Proceedings of the IJCAI-99 Workshop on Agent Communication Languages, Springer, Berlin, pp 31–45

Wijaya V, Erwin A, Galinium M, Muliady W (2013) Automatic mood classification of Indonesian tweets using linguistic approach. 2013 international conference on information technology and electrical engineering (ICITEE), IEEE, pp 41–46

Zappavigna M (2011) Ambient affiliation: a linguistic perspective on Twitter. New Media Soc 13(5):788–806