Digital sources: a case study of the analysis of the Recovery of Historical Memory in Spain on the social network Twitter

Mariluz Congosto

Universidad Carlos III. Escuela Politécnica Superior, Avd. de la Universidad, 30. 28911 Leganés (Madrid)
email: mcongosto@inv.it.uc3m.es
ORCID iD: http://orcid.org/0000-0002-8826-729X

ABSTRACT: The incorporation of digital sources from online social media into historical research brings great opportunities, although it is not without technological challenges. The huge amount of information that can be obtained from these platforms obliges us to resort to the use of quantitative methodologies in which algorithms have special relevance, especially regarding network analysis and data mining. The Recovery of Historical Memory in Spain on the social network Twitter will be analysed in this article. An open-code tool called T-Hoarder was used; it is based on objectivity, transparency and knowledge-sharing. It has been in use since 2012.

KEYWORDS: Methodology; Data mining; Networks analysis; History; Twitter

Citation / Cómo citar este artículo: Congosto, Mariluz (2018) "Digital sources: a case study of the analysis of the Recovery of Historical Memory in Spain on the social network Tw". *Culture & History Digital Journal*, 7 (2): e015. https://doi.org/10.3989/chdj.2018.15

RESUMEN: *Fuentes digitales: un estudio de caso sobre la recuperación de la Memoria Histórica en España en Twitter*.- La incorporación de fuentes digitales procedentes de las redes sociales on-line a la investigación histórica aporta grandes oportunidades aunque no está exenta de retos tecnológicos. La ingente información que se puede obtener de estas plataformas aboca sin remedio al uso de metodologías cuantitativas en las que los algoritmos adquieren especial relevancia, especialmente en el análisis de redes y la minería de datos. En este artículo se analizará Recuperación de la Memoria Histórica en España en la red social Twitter. Se aplicará una metodología denominada T-Hoarder_kit, de código abierto, usada desde el año 2012, que cumple con los requisitos de objetividad, transparencia y compartición de conocimientos.

PALABRAS CLAVE: Metodología; Minería de datos; Análisis de red; Historia; Twitter

INTRODUCTION

The digital world is becoming so omnipresent that society is growing increasingly unaware of how immersed in it it actually is. Most real-world activities have their equivalent in the digital universe: shopping, entertainment, administrative formalities, conversations with friends and family, etc. There is little that does not have a digital counterpart. This immersion, which has intensified this decade, is bringing about social changes whose impact has not been felt yet.

Researchers need to extend their activity into the digital dimension – but the foundations are yet to be laid. Newspaper libraries are already just a small portion of the secondary sources. The role of media in shaping public opinion is being overtaken by the new digital environment. According to the Estudio General de Medios [General Media Study] conducted from February to November

2017,[1] the share of newspaper readers stands at 24.3% whilst Internet is accessed by 75.7% of the population, greater than the share of radio listeners (59.3%) and approaching that of television viewers (85.2%). The growth of the Internet as the preferred place for getting information and debating topics is unstoppable. Much of this growth has come from virtual social media, which have revolutionised the way in which content is delivered. We may not be witnessing a phenomenon of mass self-communication (Castells, 2009) but we are experiencing society's power to make the media agenda more or less relevant.

Many of the conversations and discussions that used to take place on the analogue plane are now being incessantly recorded in the digital world in the form of text, images or video thanks to social media. New, direct channels of communication are opening up between politics and the public that fall outside traditional media. Everything happens faster and more directly and leaves an indelible trace.

Online social media are in the hands of a few companies, such as Facebook (Facebook, Instagram and WhatsApp), Microsoft (LinkedIn), Alphabet (G+ and YouTube) and Twitter. These organisations use the information they obtain from people's profiles and interactions within their medium for commercial purposes. On the other hand, access by researchers to this information is very limited and is strictly controlled by these companies. In the case of Twitter, the information generated by most of its users is in the public domain and can be accessed through its API; however, the full volume of generated messages cannot be accessed for free – only a portion of it. Even so, Twitter is today the most widely used source of social data.

This new digital environment – where researchers will have to dip their toes in – offers great opportunities, but is not without technological challenges. The huge amount of information outputted by social media requires the application of quantitative methods and other data analysis techniques in order to study it. The other side of the coin is the low reliability of the identity of the users who publish or spread content, as is the case of Twitter, where fictitious, false or automatic profiles (bots) proliferate (Ferrara et al., 2016).

This paper lays down a cyclical, three-phase (data capture, data processing and data display) methodology to help qualify the profiles that publish information on Twitter. How to access the information in this social network is explained, the types of data that can be obtained and their limitations. Procedures for analysing Twitter user profiles are defined and different types of display for detecting behavioural patterns listed.

A tool called t-hoarder_kit[2], in use since 2012, was used as technological support; since it is an open-source tool it meets the requirements of transparency and knowledge sharing. This methodology was applied to a case study on the profiles of users who write about the Recovery of Historical Memory in Spain on the Twitter social network to determine the degree of reliability thereof.

THE ENVIRONMENT AND CHARACTERISTICS OF SOCIAL MEDIA

Virtual social media do not involve the whole of society, only a percentage of the people that have an Internet connection. According to the data obtained by the Spanish Centre for Sociological Research (CIS) in a survey it conducted right after the 2016 general elections in Spain,[3] 67.8% of Spaniards accessed the Internet during the last three months leading up to the elections, 74.6% of which belonged to the Facebook social network, the most widespread of all social media in Spain.

But not only is there a part of society that is not connected via these platforms, there are also access gaps by gender (Fig 1) and by age (Fig 2). This lopsided presence of different social groups greatly hinders opinion prediction and analysis methodologies (Gayo-Avello, 2011). Any statistical study based on social media data would have a large social bias. However, the amount of information that can be obtained from virtual social interactions opens the way to new types of studies based on network analysis and data mining.

Communication on social media is fragmented and consists in short messages that are shared or commented on. In the case of the Twitter platform, the size of the messages – known as tweets – has been limited from its inception to 140 characters until it was doubled in November 2017. In other social media where this restriction does not exist, the length of the posts tends to be short. These messages are often accompanied by multimedia content, which in some platforms is more important than text.

Sticking to two very popular platforms in Spain (Facebook and Twitter), the differences in the number of users who participate in them, their degree of privacy and their restrictions to retrieving information can be seen in the following table.

The Facebook platform seems at first more suitable for retrieving information because of both the number of profiles it has and the segmentation of its users. However,

Percentage of profiles by gender

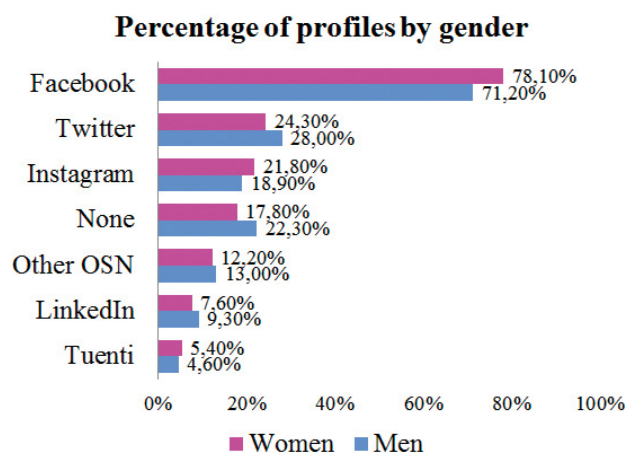|	Women	Men
Facebook	78.10%	71.20%
Twitter	24.30%	28.00%
Instagram	21.80%	18.90%
None	17.80%	22.30%
Other OSN	12.20%	13.00%
LinkedIn	7.60%	9.30%
Tuenti	5.40%	4.60%

FIGURE 1. Percentage of profiles on online social media by gender. Source: CIS: Survey conducted after the 2016 Spanish general elections – Socio-demographic variables
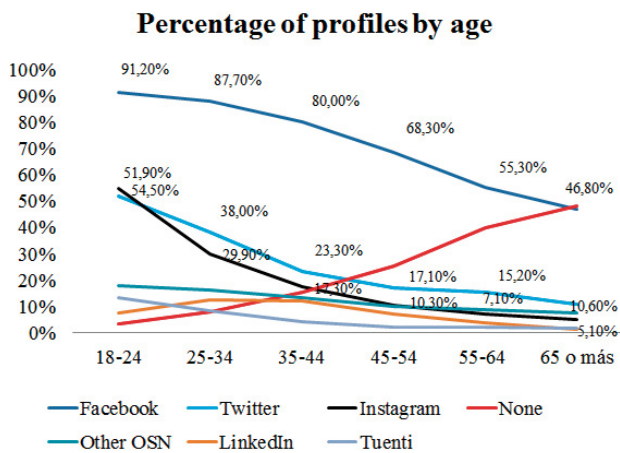
Percentage of profiles by age

FIGURE 2. Percentage of profiles on online social media by gender. Source: CIS: Survey conducted after the 2016 general elections – Socio-demographic variables.

the privacy restriction on personal profiles leaves only the profiles of entities visible, thereby drastically reducing its scope. This is why the possibility of obtaining data from personal profiles on Twitter has made this platform the main source of data for researching online social media. On the other hand, Twitter beats Facebook in immediacy; it is a platform where messages are shared faster and conversations are more agile. It is a place where current events are discussed and communication campaigns organised.

Twitter's strength as a public source of opinion becomes a weakness as far as certain hot topics are concerned owing to noise and overreaction. This causes the most extremist positions to prevail, creating what is known as the "spiral of silence" (Noelle-Neumann, 1995). Thus, the analyses should take this into account in order for their results not to become distorted. Although Twitter is not a perfect source of digital information for researchers, it has served as the basis for multi-

ple studies that have allowed the social pulse in limited environments to be analysed. The birth and development of social movements that have lead to social transformations, such as 15M (Toret, J., et al., 2013) (González-Bailón et al., 2013) (Peña-López et al., 2014), or the networking of new citizen platforms (Aragón et al., 2017) has been researched. Likewise, it has been used to research crisis situations that lead to political polarisation (Morales et al. 2015). It has also been used globally to research electoral campaigns in Spain (Barberá & Rivero, 2012) (Congosto, 2015), the United States (Livne et al., 2010) (Hanna et al., 2011) (Bessi & Ferrara, 2016) (Wang et al., 2016) and Europe (Jungherr et al., 2011) (Ferrara, 2017). Lately, the social alarm generated by fake news is being analysed from Twitter as it is one of the channels over which they are being spread (Fletcher et al., 2018) (Stella et al., 2018).

METHODOLOGY

There is no consolidated global methodology for analysing online social media. Researchers apply specific methods to obtain results from their experiments. Independently of the method that is ultimately applied, it must take into account the types of entities that communicate with each other on social media, the manner in which this communication takes place and the restrictions to collecting this information. Only the definition of a methodological framework that uses open, transparent tools can ensure that experiments can be repeated and checked by third parties.

Platforms provide Application Programming Interfaces (APIs) for obtaining their data. These mechanisms allow the data to be downloaded via a very efficient protocol, but under the conditions set by the platform, which may vary over time. The restrictions affect aspects such as data privacy, the age of the messages and the amount of information that is provided per unit of time (rate limit). An alternative consists in using web scraping tech-

TABLE 1. Comparison of the Facebook and Twitter platforms.

Characteristic	Facebook	Twitter
Number of profiles	~1,800m	~350m
User segmentation	Yes	No
Difference between person and entity	Yes	No
Privacy of personal profiles	Yes	Most No
Privacy of entity profiles	No	No
Fake profiles	Yes	Yes
Message size limitation	No	Yes
Restriction to retrieving messages on privacy grounds	Yes (personal profiles)	Most No
Restriction to retrieving old messages	No	Yes
Restriction on the number of messages retrieved in a time window	Yes	Yes

niques (Iacus, 2015) to obtain data directly from the platforms' websites. This option allows sidestepping the message age limitation, but the retrieval of information is less efficient; in some cases, some kinds of data cannot even be downloaded. Its use is only advisable when the APIs find it impossible to retrieve information of a certain age.

Twitter is the social network that is the most agreeable to data downloading because its messages are mostly public. Even if it is not the best data source, as it is not the most widely used and has gender and age gaps, it is the most readily available to researchers. This is why the methodology set out herein is going to focus on Twitter.

Access to Twitter's API

Twitter has several platform access APIs.[4] The ones which are related to data downloading are listed below:

- The REST API: it gives access to multiple types of Twitter data. It allows downloading all the data that can be displayed by the graphic interface, such as the users' profiles, their followers and the people they follow (following), their tweets, trending topics, and so on. It is the most suitable API for analysing user profiles and the relations among users.
- The Search API: it allows tweets to be retrieved from the tweet history. Queries may be done by entering sets of keywords separated by logical connectors or through advanced searches. In addition, results can be filtered by language and by location.
- The Streaming API: it provides a real-time flow of tweets by establishing a permanent connection with Twitter's servers. Data can be filtered by ten different types of parameters. The most common ones are: keywords, users, and locations. This is the API that is most suited to gathering information about a subject matter on a continual manner over a long period of time (months or years).

APIs limit the amount of information that can be downloaded over a certain period of time.

- The REST API has a variable limitation according to the requested method. The restriction is measured in the number of requests that can be made in a 15-minute period. The values range from 15 to 900. In turn, an operation can include a multiple answer (pagination), which speeds up downloading. The most restrictive methods are those that provide lists of users' followers or of the people users follow, which are kept down to 15 requests. The less restricted methods are the requests for user tweets, which allow up to 900 queries.
- In the Search API, the limitation stands at 180 requests every 15 minutes.

- Since the Streaming API provides a constant flow of tweets, the restriction applies to the flow that is received, which can never exceed 50 tweets per second.

There is also a time limitation for retrieving information from different levels:

- The Search API can only retrieve information that is seven days old at the most. On the other hand, the REST API can download a user's last 3,200 tweets. In this latter case, the age of the data will depend on the tweeting frequency, which can range from years to months.

Additionally, neither the Search API nor the Streaming API (in free mode) provide all tweets after just one query. The percentage of tweets that are outputted ranges from 85% to 95% – the criterion according to which the tweets are filtered is not known. Nevertheless, this is an acceptable percentage for analysis purposes.

There are several open-code tools that provide access to Twitter's APIs, such as TwapperKeeper[5], Twitter-Tap[6] twitterstream-to-mongodb[7], dmi-tcat[8] and t-hoarder_kit. The last tool was the one that was used to conduct this study.

The t-hoarder kit tool

T-hoarder-kit is an evolution of the T-Hoarder platform (Congosto et al., 2017). It consists of a collection of open-code software that allows Twitter information to be both downloaded and processed so as to make it easier to use in network analysis and information display tools. Since the analysis of online social media involves working with massive amounts of information, it is essential to display it to let patterns or singularities emerge to guide the analysis in its next phase.

t-hoarder_kit uses Twitter's REST, Search and Streaming APIs and allows the following kinds of information to be retrieved:

- All information that is associated with a user's profile, followers, following, tweets posted and the lists to which they belong (REST API).
- The existing follower-following relations for a set of users (REST API).
- The most recently posted 3,200 tweets from a set of users (REST API).
- Consulting the Twitter history for a set of tweets that match a search pattern (Search API).
- Consulting in real time a set of tweets that match a search pattern (Streaming API).

Tweet processing is aimed at information aggregation and inference. Aggregation will allow the degree of repetition of some tweet components to be quantified, and inference will let the underlying characteristics of the infor-

mation emerge. The different types of processing are listed below:

- Entities. A tweet is just a short text message (280 characters) but is made up of multiple entities that are included in it or in the metadata provided by Twitter APIs. The entities t-hoarder-kit takes into consideration in each tweet are references to other users, the most frequently used words, hashtags for tweet classification, URLs, images and the application from which the tweet was posted (metadatum). By quantifying the frequency of appearance of these entities, the tool produces a summary of the collection of information it captured.
- Diffusion. One of the actions that characterises Twitter is the re-posting of information, or retweeting (RT). Retweeting is a convention that Twitter users settled on at the very beginning for sharing tweets with their followers. Initially it was done by posting other users' tweets by preceding them with the letters RT and the name of the author. In 2009 Twitter included an RT button that did the same thing but automatically, which greatly facilitated the propagation of tweets. Users normally spread tweets with which they agree, so each retweet can be considered to be a positive vote for the original tweet (Conover et al., 2010). This feature causes the information that circulates through Twitter to be very redundant. Therefore, quantifying the diffusion of original tweets results in a ranking of the discourse that has been spread.
- Location. The location of tweets can be known in two ways. The first one consists in obtaining the location that appears on a user's profile. This piece of data might not have been entered or consist in the name of a fictional location, so not all tweets can be pinpointed geographically. However, a large percentage of tweets (60-70%) can still be geolocated. The second option consists in collecting the geolocated tweets of those users who have activated the geolocation feature in Twitter. In this case, the percentage is much smaller (1.5% in the Spanish case). Standardising the location of tweets allows presenting them using a mapping tool and thus having an overall view of the geographical location of tweets.
- User characterisation. When users use Twitter they leave a trace from which their characteristics, such as their personality (role), impact (h-index), network ratio, propagation ratios, and link and image usage frequency, can be inferred.
  – Role: users have been categorised into the following role types based on González-Bailón et al's definition of influence (González-Bailón et al. 2013) (Fig 3):
    - Speaker: their tweets are retweeted. According to the mean value of diffusion of their tweets, they can be classified as either low (at least

three times), medium (at least ten times) or high (at least one hundred times).
    - Networker: their tweeting frequency is high and they make and receive pretty much the same number of retweets.
    - Retweeter: they mostly spread information (60% of RTs).
    - Replicator: their most usual activity is to respond to tweets (60% of replies).
    - Monologist: the tweets they post are hardly retweeted (less than 70% are not retweeted).
    - Isolated: they have an insular attitude; they neither retweet nor are retweeted.
    - Common: their level of activity is very low and they hardly interact with other users.
  – h-index: this indicator is employed to calculate a user's impact (Hirsch, 2005) by simultaneously measuring the quality and the quantity of their scientific output. The calculation of the indicator is adapted to the Twitter environment by replacing tweets with publications and RTs with citations (Fig 4). This algorithm sorts tweets by number of RTs received and looks for the point where the number of tweets and the number of retweets match. For example, a user with an h-index of 40 has made at least 40 tweets that have been retweeted 40 times. This metric rewards continued success instead of one-offs.

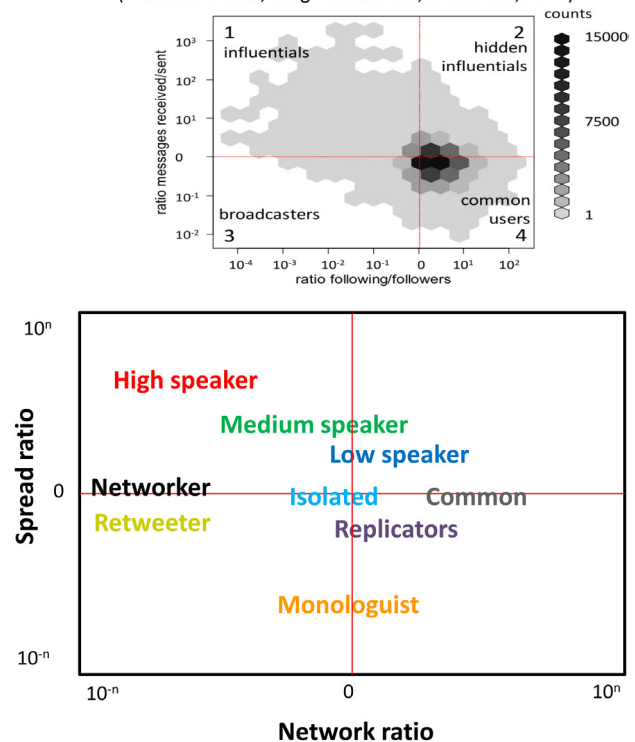(González-Bailón, Borge-Holthoefer, & Moreno, 2012).

FIGURE 3. Classification of user roles according to their level of activity and impact.
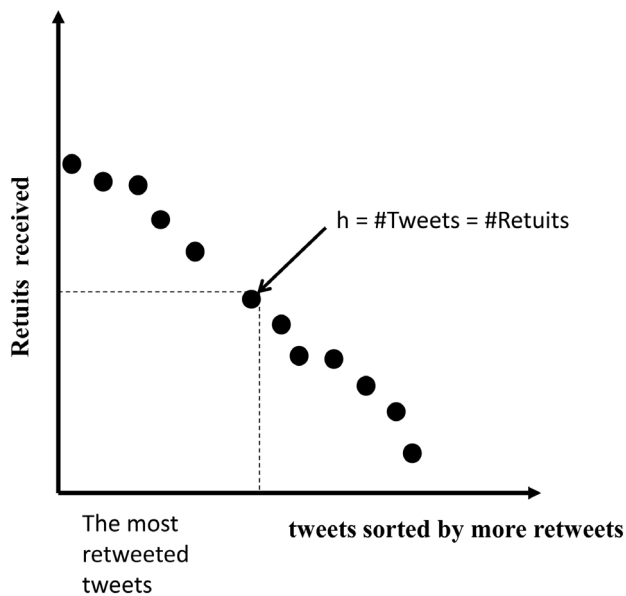
FIGURE 4. Way of calculating the h-index.

– Network ratio: it identifies the asymmetry of a user's declared network.

$$Network\ ratio = \frac{Number\ followers}{Number\ following}$$

– RT_in ratio: it is used to calculate a user's retweeting capacity.

$$RT\_in\ ratio = \frac{Number\ RTs\ received}{Number\ original\ tweets}$$

– RT_out ratio: it measures a user's tendency to be retweeted.

$$RT\_out\ ratio = \frac{Number\ RTs\ sent}{Number\ Tweets}$$

– Hashtag ratio: it shows the frequency of tweet hashtagging.

$$Hashtag\ ratio = \frac{Number\ tweets\ with\ hashtag}{Number\ tweets}$$

– Link ratio: it calculates the frequency of tweets with URLs.

$$Link\ ratio = \frac{Number\ Tweets\ with\ links}{Nuber\ tweets}$$

– Media ratio: it indicates the use of multimedia.

$$Media\ ratio = \frac{Number\ tweets\ with\ media}{Number\ Tweets}$$

• Relations. A Twitter user is connected either because they express an interest in what other user says (by following them) or because other users become interested in them (their followers). This declaration of interest of some users for others creates the network of declared relations through which their

tweets will flow, the dynamic network being the network that will emerge from the interactions among them. The declared network and the dynamic network might not be the same. One of the first analyses of Twitter (Huberman et al., 2009) shows how both networks differ. A user does not interact with all the nodes in their declared network, and sometimes they interact with nodes from other networks. This is down to the fact that tweets are public; thus, they can be accessed in other ways. A user can access tweets from another user with whom they do not have any declared relation either because they get a retweet from a user in their network or because they check the tweets associated with some word or hashtag. Both types of relation – declared and dynamic – can be extracted using the t-hoarder-kit tool, which will generate a graph that can then be imported into network analysis and graph display tools.

Irrespective of the type of relation, declared or dynamic, according to which the graph is generated, the mathematical model for determining network parameters is the same, as the graph is a mathematical abstraction: a graph G is an ordered pair $G = (V, E)$, where V is a set of vertices or nodes and E is a set of links or arcs that connect these nodes.

Of the multiple associated network parameters, the following have been selected:

– Degree centrality: the number of links in a node, ie the number of nodes to which the former is connected.
– Indegree centrality: the number of incoming links to a node; in other words, how many nodes link to it.
– Outdegree centrality: the number of outgoing links from a node, or the number of nodes it links to.
– Closeness centrality: as defined by Hanneman & Riddle (Hanneman & Riddle 2005):

> *Degree centrality measures might be criticized because they only take into account the immediate ties that an actor has, or the ties of the actor's neighbors, rather than indirect ties to all others. One actor might be tied to a large number of others, but those others might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighborhood. Closeness centrality approaches emphasize the distance of an actor to all others in the network by focusing on the distance from each actor to all others. Depending on how one wants to think of what it means to be "close" to others, a number of slightly different measures can be defined,*

– Betweenness centrality: it measures a node' capacity to intermediate in other nodes' connections. It is the percentage of times a node is in between

the different paths connecting other nodes. This metric identifies the nodes that hold the network together and tie different subgroups together.

– PageRank: it is based on Google's PageRank algorithm (Page et al., 1998), adapted to network analysis by the Gephi graph platform.[9]

> *An iterative algorithm that measures the importance of each node within the network. The metric assigns each node a probability that is the probability of being at that page after many clicks. The page rank values are the values in the eigenvector that has the highest corresponding eigenvalue of a normalized adjacency matrix A'. The standard adjacency matrix is normalized so that the columns of the matrix sum to 1.*

– Modularity: it is a network metric that lets the different communities in a graph emerge. It groups those nodes that have the most solid ties among them within a group and hardly any relations with members of other groups. There are many modularity algorithms (Newman & Girvan, 2004) (Leskovec et al., 2010) (Grabowicz et al., 2012). Blondel et al's algorithm (Blondel et al., 2008) is the one that has been used in this study.

t-hoarder-kit's role in the display of data is basically to transform the data so that display tools can import it. There are four different ways in which to display data:

• *Timeline:* it is a display that shows the evolution of one or several variables in chronological order. The information is represented along the X and Y axes: the X axis shows the units of time and the Y axis the value of the variables. The data is prepared by linking to the variables their time information.
• Variable comparison: tabulated information of a set of variables in order for them to be visually represented as histograms, bar charts, correlations, and so forth.
• Map: it is used to represent geolocated information. The data is structured by associating a geolocation with a set of variables.
• Graph: it is used for network analysis. The data is modelled as a set of nodes having attributes and a number of links among them.

CASE STUDY

The case study focused on those Twitter profiles that tweet about the 2nd Spanish Republic, the Spanish Civil War and the Franco Regime. Behind those profiles are both associations and private individuals that spread all kinds of contents – some of them new.

The group that was the subject of this study comprised 70 profiles whose common denominator is the historical memory topic above any other. The group was identified in two phases. During the first phase, 61 profiles were manually catalogued based on the available information on historical memory associations. During the second, 9 other profiles were discovered by analysing the interactions on Twitter among the members of the initial group.

The Twitter profile of each of the users under study, their declared relations (follower-following) and the last 3,200 tweets they had posted were downloaded using t-hoarder-kit. This data provided an overview of the contents generated by this group, whilst it also enabled these users to be characterised based on their behaviour, their acceptance by other group members and their overall impact on Twitter. This characterisation allowed these users to be ranked according to several indicators and a reliability index to be calculated.

Group posting timeline

A timeline of the tweets posted by the members of this group every month starting from 2011 was generated (Fig 5). This timeline shows two variables: the number of tweets made (excluding RTs) and the number of RTs received. The variables had different scales (1-10), so they were proportionally represented in order to be able to see their correlation. The months of July were marked to see whether the Anniversary of the 18th of July led to an increase in tweets during that month; an increase, however, was only seen in 2015.

Can be seen that there is a growing trend to tweet and an increase in virality over time. Retweeting was very low in 2013 but it increased considerably in 2014 and 2015, when it was proportionally higher than in 2017.

This timeline allowed the peak moments of both tweeting and retweeting to be identified and the information limited in order to be able to analyse it in greater detail. This study does not dwell on this analysis because it revolves around analysing the sources. Nevertheless, the possibilities of this kind of display are duly noted.
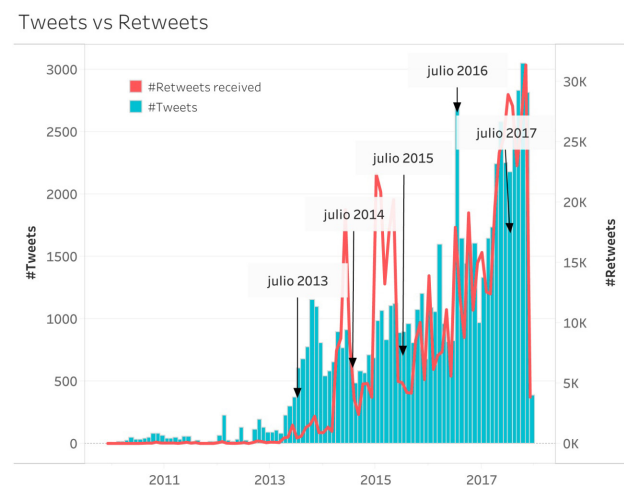

FIGURE 5. Timeline of tweets vs retweets of the set of Twitter profiles that talk about the 2nd Spanish Republic, the Spanish Civil War and the Franco Regime.

User behaviour

The first approach to the analysis of users focused on analysing their activity and its impact. The following indicators were calculated for each user, as set out in the Methodology section: role, h- index, RT_in ratio, RT_out ratio, Media ratio, URL ratio and HT ratio. The results are shown in Table 2 below, the users having been sorted from higher to lower h-index.

Users were sorted according to the way they interact, and a role was assigned to them. This sorting into roles has been applied in 18 study cases in different fields, such

TABLE 2. User characterisation parameters.

User	Role	h-index	Network ratio	RT_in ratio	RT_out ratio	Media ratio	URL ratio	HT ratio
deportado4443	M speaker	146	8,242.80	88.92	0.31	0.17	0.33	0.11
ARMH_Memoria	M speaker	143	5.19	76.93	0.56	0.17	0.50	0.28
DefensaDeMadrid	M speaker	116	18.96	35.49	0.25	0.35	0.11	0.35
Memoria_Publica	M speaker	113	87.10	88.10	0.33	0.16	0.67	0.06
CaosHistorico	M speaker	82	6.13	25.88	0.46	0.42	0.23	0.21
demiguelch	M speaker	61	4.73	20.82	0.64	0.06	0.49	0.20
amauthausen	M speaker	55	21.50	33.68	0.79	0.16	0.28	0.39
foromemoria	M speaker	49	9.72	23.52	0.73	0.14	0.36	0.46
amigosbrigadas	M speaker	46	31.75	12.60	0.60	0.17	0.39	0.57
recupmemoria	M speaker	44	1.72	15.08	0.75	0.19	0.25	0.86
19391936	L speaker	41	169.75	6.64	0.00	0.19	0.03	0.00
Dia_Como_Hoy	M speaker	41	0.00	38.39	0.05	0.84	0.01	1.05
IBMT_SCW	M speaker	38	7.43	17.33	0.73	0.23	0.43	0.52
Aledelafuent7	M speaker	35	1.35	12.28	0.42	0.18	0.58	0.67
jmgarretas	L speaker	31	2.22	4.13	0.28	0.34	0.20	0.21
corunamemoria	L speaker	30	1.04	3.56	0.27	0.10	0.80	0.16
inesgce	L speaker	29	1.83	9.14	0.38	0.22	0.37	0.48
foroporlamemori	M speaker	26	1.41	11.31	0.13	0.52	0.27	0.06
ARMHEXMemoria	L speaker	24	1.84	7.02	0.42	0.27	0.49	0.20
LincolnBrigade	L speaker	24	4.47	4.77	0.27	0.12	0.67	0.55
Valdenoceda	L speaker	23	1.08	4.61	0.24	0.07	0.68	0.90
muyfandel36	L speaker	22	6.15	4.06	0.41	0.16	0.54	0.55
MemoriaMallorca	L speaker	20	5.27	4.31	0.48	0.19	0.30	0.33
richardbaxell	L speaker	20	3.15	3.68	0.34	0.08	0.29	0.15
largocaballerof	L speaker	19	0.81	4.82	0.73	0.11	0.43	0.75
FAMYR_Asturias	Networker	19	2.10	2.82	0.33	0.12	0.61	0.30
ComisionVerdad_	Networker	19	3.59	2.91	0.55	0.07	0.59	0.61
ColumnaUruguaya	L speaker	19	1.94	4.57	0.31	0.23	0.47	0.21
RDignidad	L speaker	18	2.66	6.23	0.54	0.08	0.45	0.37
MaiMes_info	L speaker	18	1.18	3.02	0.40	0.19	0.32	0.59
spanje3639	L speaker	18	9.36	4.03	0.39	0.24	0.56	0.93
Openwatermelon	L speaker	17	1.53	3.82	0.62	0.26	0.27	0.54
SidBrint	L speaker	17	2.35	3.37	0.30	0.25	0.74	0.91
Buscameblog	L speaker	17	1.00	3.37	0.71	0.29	0.48	0.41
Toledo_GCE	Networker	17	5.02	2.54	0.17	0.35	0.39	0.67
bibrepublica	Networker	17	1.05	2.37	0.19	0.57	0.82	1.29

User	Role	h-index	Network ratio	RT_in ratio	RT_out ratio	Media ratio	URL ratio	HT ratio
garrielies	L speaker	17	2.62	4.53	0.79	0.23	0.23	0.55
armh_adh	L speaker	16	0.99	5.66	0.26	0.10	0.68	1.53
BunkerCapricho	L speaker	16	7.78	4.66	0.67	0.21	0.47	0.91
ASMJ_Salamanca	L speaker	14	3.97	5.44	0.33	0.19	0.48	0.33
investigando36	Networker	14	4.76	1.88	0.14	0.05	0.88	0.22
SOSCarabanchel	L speaker	13	1.07	3.05	0.78	0.11	0.37	0.55
memoristorica	Networker	13	1.03	2.44	0.07	0.01	0.25	0.67
DiarideGuerra	Networker	13	1.23	1.40	0.04	0.16	0.76	1.25
Ce_AQUA	L speaker	12	0.84	4.11	0.92	0.09	0.34	0.48
BATALLAEBRE	Networker	12	30.48	1.45	0.04	0.07	0.59	0.23
guerraenmadrid	Networker	11	1.50	0.70	0.05	0.02	0.67	0.04
MemoriaNuestra	L speaker	10	1.27	6.66	0.52	0.14	0.49	0.38
Gusen_Memorial	Networker	10	3.85	1.75	0.06	0.02	0.79	0.55
MyLMadrid	Networker	10	19.20	0.92	0.00	0.00	1.00	0.00
angelvinashist	Networker	10	23.47	2.68	0.00	0.00	0.89	0.05
FemMemoriaPV	L speaker	9	1.10	3.11	0.58	0.16	0.46	0.60
GuerraCivil1936	L speaker	9	0.99	3.03	0.86	0.46	0.23	0.44
AsocTajar	L speaker	8	0.78	5.21	0.81	0.29	0.32	0.41
MemoriadeHuelva	Networker	7	2.13	1.46	0.34	0.03	0.77	0.28
matilde_landa_	M speaker	7	1.21	20.63	0.97	0.15	0.45	0.35
GuerraCivilLeon	Networker	7	35.33	1.22	0.14	0.11	0.25	0.92
ateneodelaisla	Networker	6	0.57	1.49	0.11	0.12	0.73	0.19
F_Areneros	Networker	6	1.41	1.52	0.38	0.37	0.52	0.23
basquechildren	L speaker	5	1.58	3.60	0.34	0.24	0.36	0.32
Gerion74	Networker	5	2.15	1.83	0.02	0.86	0.95	0.00
AMHCIUDADREAL	Networker	4	0.83	1.41	0.55	0.09	0.51	0.44
exiliadas	Monologist	4	3.04	0.23	0.00	0.00	1.00	0.00
TLNAndalucia	Monologist	4	7.83	0.30	0.01	0.00	0.99	0.00
mhtorrejon	Networker	4	2.22	1.04	0.52	0.16	0.54	0.16
AsociacionArmha	Retweeter	3	0.83	1.96	0.75	0.10	0.35	0.29
FundacionNegrin	Networker	3	0.86	1.73	0.48	0.08	0.16	0.72
MemoriaDipCadiz	Monologist	2	2.17	0.11	0.00	0.00	1.00	0.00
GuerraCivil3639	Monologist	2	0.00	0.04	0.00	0.07	0.33	0.01
laguerracivil	No tweets	1	26.75	7.00	0.00	0.00	0.00	0.00

as the press, elections, trending topics and international events (Congosto, 2016). These roles' distributions vary depending on the field but fall within a delimited range of values: speakers are fewer than 3%; networkers amount to less than 2%; retweeters range from 6% to 15%; replicators from 3% to 17%; monologists from 1 to 4%; isolated users from 7% to 20%; and, finally, common users add up to more than 50%.

Taking these percentages of reference into account, and as can be seen in Figure 6, the study group is way above the percentages of speakers and networkers (low speaker: 42.03%; medium speaker: 21.74%; and networker: 28.9%) and way below the percentage of retweeters (1.45%). No replicator, common or isolated profiles were detected. The predominant profile is that of low speaker followed by those of networker and medium speaker. Therefore, it could be said that the members of this group generate contents that are spread by others.

h-index measures the impact of the tweets, that is, the echo they have in Twitter thanks to their being retweeted.

The higher this indicator for a user, the greater this user's capacity to grab the attention of other users in a continued manner and cause them to retweet their tweets. This is the result of their having a stable, motivated audience to the contents they post. The h-index distribution on Twitter follows a power law distribution (Newman 2005), where a few users have high values for this indicator and most users have a low value – usually below 4.

As can be seen in the histogram of the h-indexes of the users in the group (Fig 6), only eight have a value below 5. This means that most of these profiles have a relatively high h-index value compared to most Twitter users. Within the group, the most frequent value falls between 10 and 20, the highest being 146.

The network ratio is very irregular in this group; it ranges from 0 to 8,242.80. This indicator gives an idea of the asymmetry of a user's declared network. A value greater than 1 for a profile means that there are more people interested in the user than people in which the user is interested. The higher the value, the more popular the user is. This metric has to be taken into account, although it must always be qualified with the degree of acceptance of the tweets of a user since many of their followers may be passive and not interact with them (Huberman et al., 2009; Romero & Huberman, 2011).

The RT-in ratio for this set of profiles ranges from 0.04 to 88.92. This indicator is the result of calculating the average number of received RTs per tweet. It is a measure of the propagation capacity and is sensitive to diffusion peaks. It is calculated differently than the h-index because, in this case, the continued retweeting capacity is not calculated but an overall retweeting capacity. For example, a tweet that has been retweeted more than 1,000 times would cause the value of this indicator to go up dramatically, but not the h-index.

As can be seen in Figure 7, there is no correlation between the network ratio and the RT_in ratio; users with similar values for the RT_in ratio have different network ratios. The same is the case with the correlation between the network ratio and the h-index (Fig 8). This goes to confirm that the declared network not always matches the dynamic network of interactions and that the indicators of the latter provide a metric that is more in tune with reality. Therefore, the h-index and the RT_in ratio appear to be more realistic metrics of reference for ranking users.

As indicated above, even though the h-index and the RT_in ratio measure user interactions, they do not do it in the same manner. The h-index measures the continued retweeting capacity (x messages retweeted over x times), whereas the RT_in ratio provides the average numbers of retweets per message. Figure 9 shows that both variables have a rather high coefficient of determination. Of the two metrics the h-index was chosen as the indicator because it is considered to measure not only the number of retweets but also the retweeting success.

Acceptance among group members

Another way to rank the members of this group is to analyse how they value each other. This appreciation can
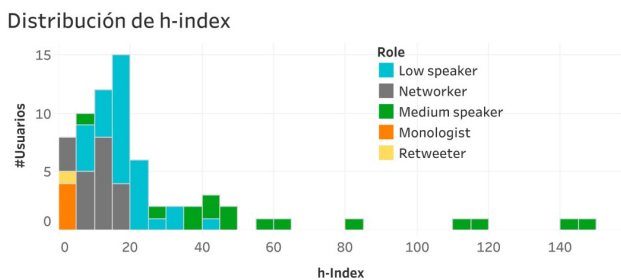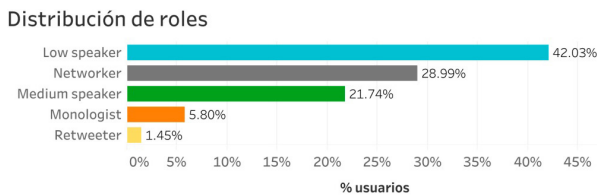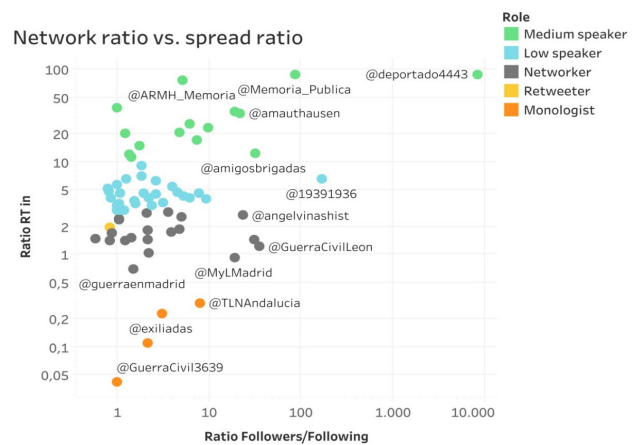

FIGURE 7. Distribution of roles according to the network ratio vs the RT_in ratio.

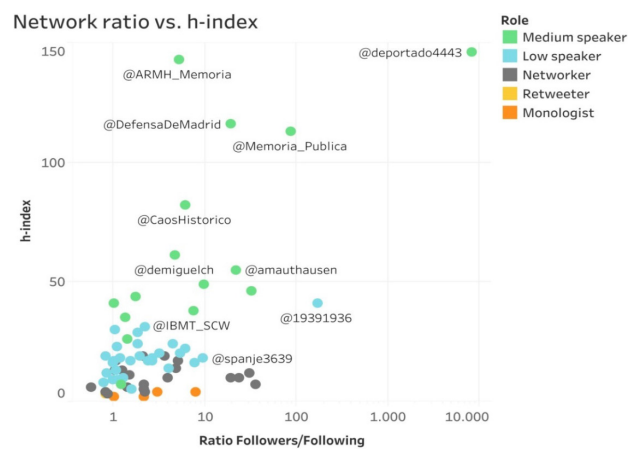FIGURE 6. Distribution of roles and h-indexes.

FIGURE 8. Distribution of roles according to the network ratio vs the h-index.
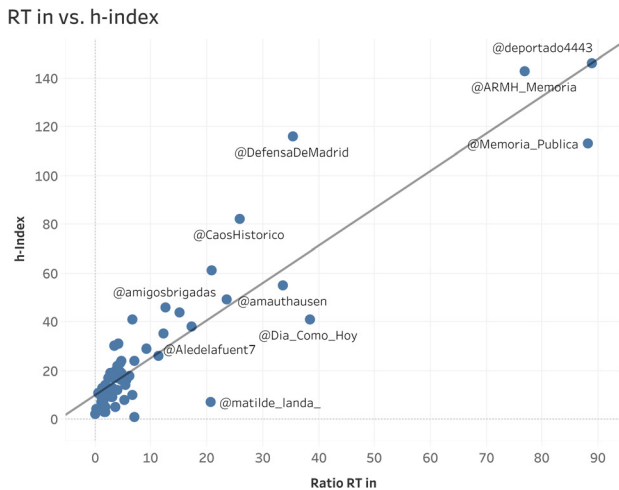
FIGURE 9. Correlation between the RT_in ratio and the h-index.

be measured in two ways: the way in which they follow each other and the way in which they interact with each other. In both cases network analysis was used to determine their degree of connection (degree centrality, indegree centrality, outdegree centrality) and their position in the network as regards their proximity to other nodes (closeness centrality) and their intermediation (betweenness centrality). The PageRank algorithm was also used for evaluation purposes. Additionally, users were grouped into communities according to their connections (modularity). Network analysis and graph display were carried out using the Gephi tool.[10]

The way in which the different group profiles follow each other determines the declared follower-following network. However, as previously mentioned, this network may be a declaration of intentions rather than an active relation. In the case of the analysis of a group that is associated with a specific goal, the types of relations among users are important since the act of following a certain user invests the latter with credibility.

In order to establish how group members follow each other, the declared connections were obtained with t-hoarder-kit and graphed. In this graph the nodes correspond to the members of the group and the links to follower-following relations. Relations are asymmetric, that is, it is not necessary for two users to follow each other; it is enough for a member to follow another for there to be a connection. This gives rise to a directed graph where relations have a direction from one node to another. If two users were to follow each other, there would be two relations, each one starting at a respective node and ending in the other.

The graph has been visually represented in Figure 10. This figure shows how the nodes are connected by follower-following relations. The size of the nodes is directly proportional to indegree centrality, those profiles that are more followed within the group standing out. The colour of the nodes corresponds to the three communities formed by those users. The red community represents

42.86% of the nodes; the most followed users are foromemoria and ARMH_Memoria. The blue community, which is the same size as the red one, has Buscameblog and AmigosBrigadas as its most followed profiles. The green community encompasses 14.28% of group members and SOSCarabanchel and amauthausen stand out for the number of followers they have.

Table 3 lists the network parameters for each of the members of the analysed group sorted by PageRank. It can be seen at first glance that there is not a very strong correlation among indegree centrality, closeness centrality, betweenness centrality and PageRank This is due to the fact that they measure different node characteristics that some take the quality of the connections into account and others do not. (Connection quality is understood to mean the weight that is assigned to links coming from important nodes in the network. It is not the same thing to be linked to peripheral, hardly connected nodes than to central, highly connected nodes.) The metric that takes the quality of connections more into account is PageRank, so it was the one that was used to rank the sources.

In order to find out how real the declared declarations are, the interactions within the group as to the way of being mentioned or cited were analysed. To this end, the last 3,200 tweets posted by each of the members of the analysed group were downloaded with t-hoarder-kit, and a graph was generated which only included the mentions among them – those made about users outside the group being discarded.

In this case, the graph is also directed, that is, mentions go from one user to another – which may be returned or not. Unlike the follower-following graph above, where there was only one relation, in this graph a user might have mentioned another user several times. For the purposes of this graph, a user's multiple mentions of an-
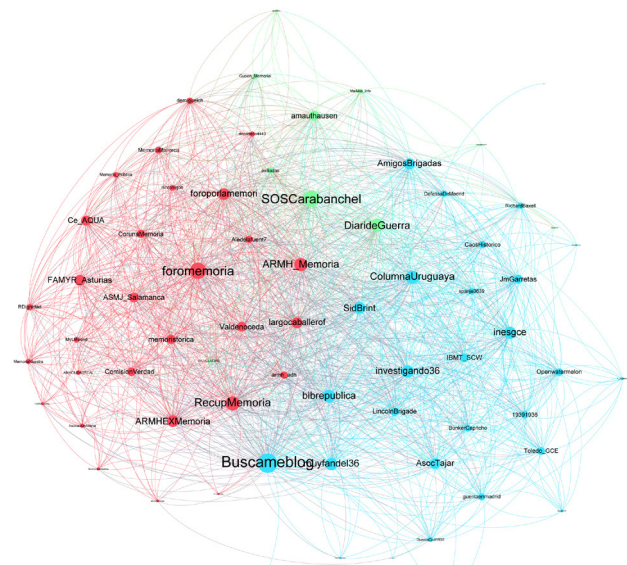

FIGURE 10. Network of follower-following connections of those Twitter profiles that talk about the 2nd Spanish Republic, the Spanish Civil War and the Franco Regime.

TABLE 3. Parameters of the declared network among group members sorted by PageRank.

Member	Modularity_	In degree	Out degree	Degree	Closeness	Betweenness	PageRank
foromemoria	0	49	46	95	0.7556	0.0705	0.0368
Buscameblog	1	51	63	114	0.9315	0.1206	0.0341
ARMH_Memoria	0	48	28	76	0.6296	0.0263	0.0320
AmigosBrigadas	1	43	21	64	0.5913	0.0136	0.0307
Gusen_Memorial	2	19	10	29	0.5113	0.0162	0.0293
SOSCarabanchel	2	41	54	95	0.8293	0.0531	0.0283
amauthausen	2	38	18	56	0.5763	0.0115	0.0268
DiarideGuerra	2	33	44	77	0.7391	0.0327	0.0259
foroporlamemori	0	34	33	67	0.6602	0.0161	0.0251
RecupMemoria	0	38	45	83	0.7473	0.0285	0.0232
inesgce	1	30	40	70	0.7083	0.0323	0.0207
memoristorica	0	29	24	53	0.6018	0.0102	0.0202
investigando36	1	31	40	71	0.7083	0.0145	0.0200
ARMHEXMemoria	0	33	34	67	0.6602	0.0135	0.0198
LincolnBrigade	1	30	18	48	0.5763	0.0036	0.0193
SidBrint	1	30	38	68	0.6869	0.0116	0.0188
deportado4443	0	27	1	28	0.3400	0.0003	0.0187
IBMT_SCW	1	28	20	48	0.5812	0.0073	0.0187
ColumnaUruguaya	1	26	49	75	0.7816	0.0263	0.0179
Valdenoceda	0	28	28	56	0.6296	0.0055	0.0174
AsocTajar	1	26	35	61	0.6667	0.0103	0.0173
muyfandel36	1	27	45	72	0.7391	0.0234	0.0172
19391936	1	26	16	42	0.5620	0.0051	0.0168
ComisionVerdad_	0	30	18	48	0.5763	0.0046	0.0164
Ce_AQUA	0	27	30	57	0.6355	0.0090	0.0162
FAMYR_Asturias	0	27	34	61	0.6667	0.0096	0.0162
JmGarretas	1	22	29	51	0.6355	0.0054	0.0162
demiguelch	0	19	18	37	0.5763	0.0058	0.0161
MemoriaMallorca	0	22	16	38	0.5620	0.0031	0.0161
ASMJ_Salamanca	0	27	27	54	0.6182	0.0044	0.0155
DefensaDeMadrid	1	21	15	36	0.5620	0.0153	0.0152
bibrepublica	1	24	50	74	0.7907	0.0184	0.0150
BunkerCapricho	1	24	14	38	0.5528	0.0016	0.0148
RichardBaxell	1	20	15	35	0.5313	0.0042	0.0148
Aledelafuent7	0	21	20	41	0.5812	0.0016	0.0143
largocaballerof	0	24	39	63	0.6939	0.0095	0.0140
MyLMadrid	0	24	7	31	0.5191	0.0014	0.0140
spanje3639	1	22	14	36	0.5528	0.0009	0.0137
Toledo_GCE	1	21	22	43	0.5913	0.0057	0.0135
guerraenmadrid	1	20	18	38	0.5714	0.0065	0.0135
CaosHistorico	1	21	20	41	0.5862	0.0026	0.0133
Openwatermelon	1	17	26	43	0.6126	0.0052	0.0115

Member	Modularity_	In degree	Out degree	Degree	Closeness	Betweenness	PageRank
Memoria_Publica	0	19	11	30	0.5231	0.0013	0.0110
BATALLAEBRE	2	18	3	21	0.4474	0.0002	0.0109
exiliadas	2	17	15	32	0.5528	0.0016	0.0100
CorunaMemoria	0	15	31	46	0.6476	0.0038	0.0091
mhtorrejon	0	13	19	32	0.5763	0.0021	0.0091
garrielies	2	7	5	12	0.4595	0.0001	0.0090
GuerraCivil1936	1	13	14	27	0.5440	0.0020	0.0090
MaiMes_info	2	10	15	25	0.5484	0.0023	0.0087
RDignidad	0	13	22	35	0.5862	0.0025	0.0080
matilde_landa_	0	10	5	15	0.4690	0.0002	0.0080
armh_adh	0	13	27	40	0.6182	0.0020	0.0078
AMHCIUDADREAL	0	9	15	24	0.5574	0.0003	0.0073
MemoriaNuestra	0	10	18	28	0.5714	0.0006	0.0065
Gerion74	1	9	5	14	0.4857	0.0004	0.0064
TLNAndalucia	0	7	2	9	0.4533	0.0000	0.0063
MemoriadeHuelva	0	10	6	16	0.5075	0.0006	0.0058
AsociacionArmha	0	8	16	24	0.5620	0.0004	0.0057
F_Areneros	1	6	8	14	0.5113	0.0006	0.0056
GuerraCivilLeon	1	6	3	9	0.5000	0.0000	0.0055
angelvinashist	2	7	0	7	0.0000	0.0000	0.0053
basquechildren	1	5	3	8	0.4503	0.0000	0.0052
FemMemoriaPV	2	5	4	9	0.5113	0.0001	0.0046
ateneodelaisla	0	3	10	13	0.5313	0.0002	0.0041
laguerracivil	1	3	0	3	0.0000	0.0000	0.0039
FundacionNegrin	1	3	0	3	0.0000	0.0000	0.0036
Dia_Como_Hoy	1	1	0	1	0.0000	0.0000	0.0033
GuerraCivil3639	1	1	0	1	0.0000	0.0000	0.0029
MemoriaDipCadiz	0	0	0	0	0.0000	0.0000	0.0024

other user count as just one relation, although it has been assigned a weight that corresponds to the number of times it was mentioned.

The graph that represents the mentions among group members is shown in Figure 11. Node size is directly proportional to indegree centrality, whereby the most mentioned nodes stand out. The colour of the nodes corresponds to the communities into which they have been grouped. All group members have been mentioned by others save for GuerraCivil3639, which has been classed as a monologist.

The manner in which members mention each other gave rise to four communities, the red one being the one with the most members (47.06%) – where the ARMH_memoria, Memoria_Publica, Buscameblog, foromemoria and SOSCarabanchel profiles stand out. AmigosBrigadas, LincolnBrigade and jmgarretas stand out in the purple community (33.82%). DiarideGuerra and bibrepublica

are prominent in the yellow community (8.82%). deportado4443 and demiguelch are noticeable in the green community (8.82%). These communities do not perfectly overlap the communities detected in the follower-following graph, but they nevertheless have certain elements in common.

Table 4 lists the parameters of the dynamic network formed by users' mentions to others. In this case indegree, outdegree and degree were included with the weight of each node. Users were sorted by PageRank.

Reliability ratio

Group members were ranked in two environments: the exogenous and the endogenous. The exogenous environment corresponds to the positions of each of the members of the analysed group in relation to all Twitter users, and the endogenous to how each member is perceived
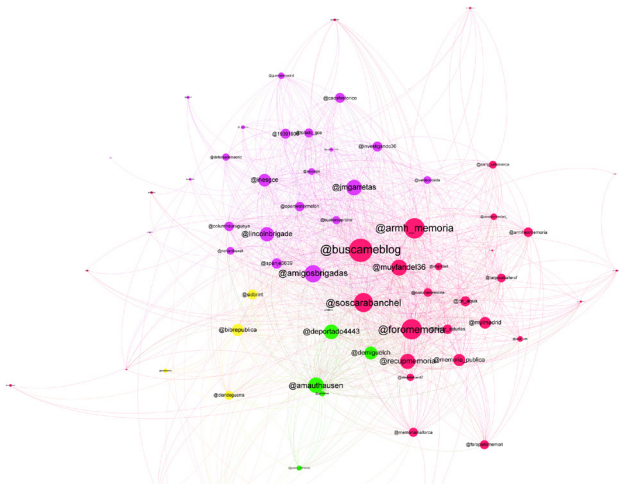
FIGURE 11. Dynamic network of mentions among the Twitter profiles that talk about the 2nd Spanish Republic, the Spanish Civil War and the Franco Regime.

within the group. Endogenous acceptance had more weight in the determination of the reliability ratio than exogenous acceptance because the perception of the members of this group by Twitter profiles that specialise on the same topic was considered more important than that of more generalist users.

The declared and dynamic networks were taken into account in endogenous assessment (the latter having more weight), whereas only the dynamic part of the network was considered in the case of the exogenous assessment (Fig 12). The following formula was used:

$$Reliabiliti = h_{index} \cdot \left(PR_{declared\ network} + \left(2 \cdot PR_{dynamic\ network} \right) \right)$$

Table 5 ranks every group member according to their final reliability ratio. It can be seen here that the exogenous appreciation is qualified by the endogenous, whereby group members with a high h-index were relegated to less prominent positions because they had a low endogenous score. For instance, DefensaDeMadrid, which has

TABLE 4. Parameters of the dynamic network among group members sorted by PageRank.

Member	Modularity	W. In-Degree	W. Out-Degree	W. Degree	Closeness	Betweenness	PageRank
ARMH_Memoria	1	1,950	1,068	3,018	0.5946	0.0210	0.1014
deportado4443	0	935	474	1,409	0.4818	0.0028	0.0535
Memoria_Publica	1	583	241	824	0.5116	0.0047	0.0514
Buscameblog	1	2,085	1,530	3,615	0.7500	0.1184	0.0503
demiguelch	0	1,041	1,029	2,070	0.5455	0.0055	0.0485
IBMT_SCW	3	1,093	730	1,823	0.5690	0.0152	0.0369
foromemoria	1	1,127	967	2,094	0.7097	0.0649	0.0316
amauthausen	0	2,180	1,886	4,066	0.6600	0.0277	0.0313
SOSCarabanchel	1	1,156	1,092	2,248	0.7021	0.0421	0.0288
AmigosBrigadas	3	1,211	1,397	2,608	0.7333	0.0325	0.0285
LincolnBrigade	3	692	257	949	0.5593	0.0115	0.0274
recupmemoria	1	1,270	322	1,592	0.5789	0.0075	0.0241
DiarideGuerra	3	640	681	1,321	0.5841	0.0274	0.0230
muyfandel36	1	827	1,110	1,937	0.7857	0.0754	0.0201
FundacionNegrin	3	20	16	36	0.0000	0.0000	0.0170
inesgce	2	1,298	1,402	2,700	0.7097	0.0400	0.0170
jmgarretas	2	720	206	926	0.5546	0.0094	0.0166
richardbaxell	3	453	266	719	0.5500	0.0082	0.0155
corunamemoria	1	481	464	945	0.5789	0.0088	0.0152
CaosHistorico	2	542	494	1,036	0.5280	0.0011	0.0145
spanje3639	3	480	1,044	1,524	0.6947	0.0140	0.0140
SidBrint	3	483	340	823	0.5946	0.0040	0.0137
foroporlamemori	1	92	27	119	0.4286	0.0004	0.0131
RDignidad	1	386	399	785	0.5841	0.0034	0.0130
bibrepublica	3	392	399	791	0.6168	0.0149	0.0126
FAMYR_Asturias	1	414	516	930	0.6168	0.0163	0.0124

Member	Modularity	W. In-Degree	W. Out-Degree	W. Degree	Closeness	Betweenness	PageRank
MemoriaMallorca	1	576	663	1,239	0.5789	0.0104	0.0122
BATALLAEBRE	3	50	12	62	0.4521	0.0147	0.0114
guerraenmadrid	2	238	166	404	0.5410	0.0030	0.0113
FemMemoriaPV	3	56	52	108	0.4783	0.0001	0.0112
basquechildren	3	30	20	50	0.3860	0.0000	0.0110
garrielies	0	614	1,211	1,825	0.6346	0.0031	0.0108
Valdenoceda	2	563	692	1,255	0.5893	0.0052	0.0108
ComisionVerdad_	1	474	548	1,022	0.5893	0.0072	0.0103
Gusen_Memorial	0	86	38	124	0.4748	0.0004	0.0101
Openwatermelon	2	684	1,602	2,286	0.7253	0.0188	0.0095
largocaballerof	1	383	503	886	0.6168	0.0076	0.0091
Toledo_GCE	2	423	249	672	0.5739	0.0065	0.0087
MaiMes_info	3	154	193	347	0.5238	0.0027	0.0081
19391936	2	402	65	467	0.5238	0.0036	0.0079
angelvinashist	2	20	4	24	1.0000	0.0003	0.0078
DefensaDeMadrid	2	143	76	219	0.4783	0.0011	0.0073
MyLMadrid	1	197	0	197	0.0000	0.0000	0.0071
Ce_AQUA	1	216	383	599	0.5739	0.0056	0.0070
ASMJ_Salamanca	1	119	141	260	0.5197	0.0021	0.0066
investigando36	2	305	270	575	0.6226	0.0090	0.0066
ARMHEXMemoria	1	175	158	333	0.5593	0.0037	0.0062
ColumnaUruguaya	2	214	326	540	0.6535	0.0062	0.0060
Aledelafuent7	1	444	1,212	1,656	0.5739	0.0053	0.0057
armh_adh	1	124	131	255	0.5323	0.0004	0.0057
BunkerCapricho	2	283	736	1,019	0.6535	0.0059	0.0052
AMHCIUDADREAL	1	9	16	25	0.4783	0.0000	0.0044
GuerraCivilLeon	1	27	47	74	0.5238	0.0003	0.0043
F_Areneros	2	88	38	126	0.5000	0.0001	0.0042
TLNAndalucia	3	5	0	5	0.0000	0.0000	0.0042
AsocTajar	2	179	963	1,142	0.7021	0.0202	0.0040
MemoriaNuestra	1	21	40	61	0.5116	0.0008	0.0036
GuerraCivil1936	2	274	970	1,244	0.5641	0.0004	0.0036
AsociacionArmha	1	17	72	89	0.5323	0.0022	0.0033
mhtorrejon	1	6	28	34	0.5116	0.0000	0.0029
MemoriadeHuelva	1	5	18	23	0.4648	0.0004	0.0028
memoristorica	1	9	1	10	0.3158	0.0001	0.0026
Gerion74	2	19	0	19	0.0000	0.0000	0.0026
exiliadas	1	6	1	7	0.4342	0.0000	0.0026
MemoriaDipCadiz	1	3	0	3	0.0000	0.0000	0.0026
matilde_landa_	1	6	197	203	0.4783	0.0001	0.0026
laguerracivil	2	1	0	1	0.0000	0.0000	0.0025
Dia_Como_Hoy	4	0	0	0	0.0000	0.0000	0.0025
ateneodelaisla	0	0	0	0	0.0000	0.0000	0.0000

Exógena	———	h-index
Endógena	Page Rank	Page Rank
	Red declarada	Red Dinámica

FIGURE 12. Indicators for profile evaluation according to the endogenous and the exogenous environment.

an h-index of 116, went from third according to the h-index to tenth according to overall score. After analysing the tweets made by this profile, it was found that it posts current political news rather than tweets about historical memory, so it is to be expected that the rest of the group would echo its tweets less. This correction is fitting because it means that this profile is not a source of specialised content only.

CONCLUSIONS AND FUTURE WORK

There are no consolidated methodologies to help humanities and social sciences researchers handle large amounts of information. In view of the difficulty to address this issue in a generic manner, there is always the possibility to provide partial solutions to identified needs. This paper aims to contribute to this by enriching the in-

TABLE 5. Ranking of the members of the group according to their overall reliability ratio.

Member	h-index	Declared network PageRank	Dynamic network Page Rank	Reliability ratio
ARMH_Memoria	143	0.0320	0.1014	33.60
deportado4443	146	0.0187	0.0535	18.36
Memoria_Publica	113	0.0110	0.0514	12.86
demiguelch	61	0.0161	0.0485	6.90
amauthausen	55	0.0268	0.0313	4.91
foromemoria	49	0.0368	0.0316	4.90
AmigosBrigadas	46	0.0307	0.0285	4.03
IBMT_SCW	38	0.0187	0.0369	3.51
CaosHistorico	82	0.0133	0.0145	3.47
DefensaDeMadrid	116	0.0152	0.0073	3.45
RecupMemoria	44	0.0232	0.0241	3.14
Buscameblog	17	0.0341	0.0503	2.29
LincolnBrigade	24	0.0193	0.0274	1.78
inesgce	29	0.0207	0.0170	1.58
JmGarretas	31	0.0162	0.0166	1.53
19391936	41	0.0168	0.0079	1.34
foroporlamemori	26	0.0251	0.0131	1.33
muyfandel36	22	0.0172	0.0201	1.26
CorunaMemoria	30	0.0091	0.0152	1.18
SOSCarabanchel	13	0.0283	0.0288	1.12
DiarideGuerra	13	0.0259	0.0230	0.93
RichardBaxell	20	0.0148	0.0155	0.91
Aledelafuent7	35	0.0143	0.0057	0.90
Valdenoceda	23	0.0174	0.0108	0.90
MemoriaMallorca	20	0.0161	0.0122	0.81
SidBrint	17	0.0188	0.0137	0.79
FAMYR_Asturias	19	0.0162	0.0124	0.78
ARMHEXMemoria	24	0.0198	0.0062	0.77

Member	h-index	Declared network PageRank	Dynamic network Page Rank	Reliability ratio
spanje3639	18	0.0137	0.0140	0.75
ComisionVerdad_	19	0.0164	0.0103	0.70
bibrepublica	17	0.0150	0.0126	0.68
RDignidad	18	0.0080	0.0130	0.61
largocaballerof	19	0.0140	0.0091	0.61
ColumnaUruguaya	19	0.0179	0.0060	0.57
Toledo_GCE	17	0.0135	0.0087	0.52
garrielies	17	0.0090	0.0108	0.52
Openwatermelon	17	0.0115	0.0095	0.52
Gusen_Memorial	10	0.0293	0.0101	0.49
investigando36	14	0.0200	0.0066	0.46
MaiMes_info	18	0.0087	0.0081	0.45
BATALLAEBRE	12	0.0109	0.0114	0.41
BunkerCapricho	16	0.0148	0.0052	0.40
ASMJ_Salamanca	14	0.0155	0.0066	0.40
guerraenmadrid	11	0.0135	0.0113	0.40
Ce_AQUA	12	0.0162	0.0070	0.36
Dia_Como_Hoy	41	0.0033	0.0025	0.34
memoristorica	13	0.0202	0.0026	0.33
armh_adh	16	0.0078	0.0057	0.31
MyLMadrid	10	0.0140	0.0071	0.28
FemMemoriaPV	9	0.0046	0.0112	0.24
angelvinashist	10	0.0053	0.0078	0.21
AsocTajar	8	0.0173	0.0040	0.20
GuerraCivil1936	9	0.0090	0.0036	0.14
basquechildren	5	0.0052	0.0110	0.14
MemoriaNuestra	10	0.0065	0.0036	0.14
FundacionNegrin	3	0.0036	0.0170	0.11
GuerraCivilLeon	7	0.0055	0.0043	0.10
matilde_landa_	7	0.0080	0.0026	0.09
F_Areneros	6	0.0056	0.0042	0.08
MemoriadeHuelva	7	0.0058	0.0028	0.08
AMHCIUDADREAL	4	0.0073	0.0044	0.06
exiliadas	4	0.0100	0.0026	0.06
mhtorrejon	4	0.0091	0.0029	0.06
TLNAndalucia	4	0.0063	0.0042	0.06
Gerion74	5	0.0064	0.0026	0.06
AsociacionArmha	3	0.0057	0.0033	0.04
ateneodelaisla	6	0.0041	0.0000	0.02
MemoriaDipCadiz	2	0.0024	0.0026	0.02
laguerracivil	1	0.0039	0.0025	0.01
GuerraCivil3639	2	0.0029		0.01

formation provided by the Twitter platform about its users, a lack that is always present when analysing data from this social network.

This paper has put forward a methodology for ranking groups of Twitter profiles in a specific field. This methodology makes it easier to aggregate and infer information on the activity of this platform's users of by using quantitative methods and other techniques that are typical of data analysis.

A user profile reliability ratio was devised based on the acceptance of user profiles within the social network. This acceptance is won unknowingly by users when they interact, which leaves useful clues about how other users perceive them. This perception is given more weight when it comes from users who specialise on one topic than when it comes from more generalist users.

To illustrate this methodology, a case study of Twitter profiles that tweet about the 2nd Spanish Republic, the Spanish Civil War and the Franco Regime was used. After successive steps, the result was a ranking of users according to their perceived reliability as providers of contents about historical memory.

In addition to the reliability ranking, this methodology provides other information of interest, such as the role these profiles play on Twitter, the type of content they post (use of URLs or multimedia), the subgroups they form within a group and their position in the network of declared and dynamic relations. This data can also shed light on the research into Twitter communities.

In the future it would be interesting to add network analysis algorithms to t-hoarder-kit in order for it to be possible to automatically calculate the reliability ratio. Part of this analysis required the use of external network analysis tools.

ACKNOWLEDGEMENTS

NOTES

1   Available at http://www.aimc.es/a1mc-c0nt3nt/uploads/2017/05/resumegm317.pdf
2   Available at https://github.com/congosto/t-hoarder_kit
3   Postelectoral 2016 Spanish General Elections, socio-demographic variables. Question 20a http://datos.cis.es/pdf/Es3145sd_A.pdf
4   Documentation of the Twitter API https://developer.twitter.com/en/docs
5   Available at github https://github.com/540co/yourTwapperKeeper
6   Available at https://github.com/janezkranjc/twitter-tap
7   Available at https://github.com/gdelfresno/twitterstream-to-mongodb
8   Available at https://github.com/digitalmethodsinitiative/dmi-tcat
9   Available at https://github.com/gephi/gephi/wiki/PageRank
10 Open Graph Plataform Gephi https://gephi.org/

REFERENCES

Aragón, P. et al. (2017) "Online network organization of Barcelona en Comú, an emergent movement-party". *Computational Social Networks*, 4(1), p.8. Available at: http://computationalsocialnetworks.springeropen.com/articles/10.1186/s40649-017-0044-4.

Barberá, P. & Rivero, G. (2012) "Desigualdad en la discusión política en Twitter". *Congreso ALICE*.

Bessi, A. & Ferrara, E. (2016) "Social Bots Distort the 2016 US Presidential Election Online Discussion". *First Monday,* 21(11), pp.1–15.

Blondel, V.D. et al. (2008) "Fast unfolding of communities in large networks". *Journal of Statistical Mechanics: Theory and Experiment,* p.6. Available at: http://arxiv.org/abs/0803.0476 [Accessed July 10, 2014].

Castells, M. (2009) *Comunicación y Poder* Alianza Editorial, ed., Alianza Editorial.

Congosto, M. (2015) "Elecciones Europeas 2014 : Viralidad de los mensajes en Twitter". *Revista redes*, 26, pp.23–52.

Congosto, M., Basanta-Val, P. & Sanchez-Fernandez, L. (2017) "T-Hoarder: A framework to process Twitter data streams". *Journal of Network and Computer Applications*, 83(August 2016), pp.28–39. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1084804517300486.

Congosto, M.L. (2016) *Caracterización de usuarios y propagación de mensajes en twitter en el entorno de temas sociales*. Universidad Carlos III. Available at: http://e-archivo.uc3m.es/bitstream/handle/10016/22826/tesis_maria-luz_congosto_martinez_2016.pdf?sequence=1.

Conover, M.D. et al. (2010) Political "Polarization on Twitter". *Networks*, pp.89–96.

Ferrara, E. (2017) "Disinformation and social bot operations in the run up to the 2017 french presidential election". *First Monday,* 22(8).

Ferrara, E. et al. (2016) "The Rise of Social Bots". *Communications of the ACM,* 59(7), pp. 96-104. Available at: http://arxiv.org/abs/1407.5225%0A http://dx.doi.org/10.1145/2818717.

Fletcher, R. et al. (2018) "Measuring the reach of fake news and online disinformation in Europe". *Factsheets Reuters Institute (February),* pp.1–10. Available at: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/Measuring%20the%20reach%20of%20fake%20news%20and%20online%20distribution%20in%20Europe%20CORRECT%20FLAG.pdf.

Gayo-Avello, D. (2011) "Don't turn social media into another "Literary Digest" poll". *Communications of the ACM*, 54(10), pp.121–128. Available at: http://dl.acm.org/citation.cfm?doid=2001269.2001297 [Accessed March 1, 2012].

González-Bailón, S., Borge-Holthoefer, J. & Moreno, Y. (2013) "Broadcasters and Hidden Influentials in Online Protest Diffusion". *American Behavioral Scientist*, (0).

Grabowicz, P. a et al. (2012) "Social features of online networks: the strength of intermediary ties in online social media". *PloS one*, 7(1), p.e29358. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3256152&tool=pmcentrez&rendertype=abstract [Accessed March 2, 2012].

Hanna, A. et al. (2011) "Mapping the Political Twitterverse : Candidates and Their Followers in the Midterms". *Artificial Intelligence*, pp.510–513.

Hanneman, R. a. & Riddle, M. (2005) "Introduction to Social Network Methods: Table of Contents". *Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/)*, 13(October). Available at: http://www.faculty.ucr.edu/~hanneman/nettext/.

Hirsch, J.E. (2005) "An index to quantify an individual's scientific research output". *Proc Natl Acad Sci U S A*, 102(46), pp.16569–16572. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16275915.

Huberman, B.A., Romero, D.M. & Wu, F. (2009) "Social networks that matter : Twitter under the microscope". *First Monday 14(1)*. Available at SSRN: http://ssrn.com/abstract=1313405.

Iacus, S.M. (2015) "Automated Data Collection with *R* - A Practical Guide to Web Scraping and Text Mining". *Journal of Statistical Software*, 68(Book Review 3). Available at: http://www.jstatsoft.org/v68/b03/.

Jungherr, A., Jurgens, P. & Schoen, H. (2011) "Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. 'Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment'." *Social Science Computer Review*. Available at: http://ssc.sagepub.com/cgi/doi/10.1177/0894439311404119 [Accessed April 11, 2012].

Leskovec, J., Lang, K.J. & Mahoney, M. (2010) "Empirical comparison of algorithms for network community detection". *Proceedings of the 19th international conference on World wide web - WWW '10*, p.631. Available at: http://portal.acm.org/citation.cfm?doid=1772690.1772755.

Livne, A. et al. (2010) "The Party is Over Here : Structure and Content in the 2010 Election". *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media,* pp.201–208.

Morales, A. J. et al. (2015) "Measuring political polarization: Twitter shows the two sides of Venezuela". *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25, p.33114. Available at: http://scitation.aip.org/content/aip/journal/chaos/25/3/10.1063/1.4913758.

Newman, M.E.J. (2005) "Power laws, Pareto distributions and Zipf's law". *Contemporary physics,* 46(5), pp.323–351. Available at: http://arxiv.org/abs/cond-mat/0412004.

Newman, M.E.J. & Girvan, M. (2004) "Finding and evaluating community structure in networks". *Physical Review E,* 69(2), pp.1–16. Available at: http://arxiv.org/abs/cond-mat/0308217%0A http://dx.doi.org/10.1103/PhysRevE.69.026113.

Noelle-Neumann, E. (1995) "La Espiral del silencio: opinión pública: nuestra piel social". *Paidós comunicación*, 62, p.331. Available at: http://biblioteca.uoc.edu/llibres/19198.htm.

Page, L. et al. (1998) "The PageRank Citation Ranking: Bringing Order to the Web". *World Wide Web Internet And Web Information Systems*, 54(1999–66), pp.1–17. Available at: http://il-pubs.stanford.edu:8090/422.

Peña-López, I., Congosto, M. & Aragón, P. (2014) "Spanish Indignados and the evolution of the 15M movement on Twitter: towards networked para-institutions". *Journal of Spanish Cultural Studies*, 15(1–2), pp.189–216.

Romero, D.M. & Huberman, B.A. (2011) "Influence and Passivity in Social Media". In *Machine learning and knowledge discovery in databases*. Springer Berlin Heidelberg, pp. 18–33.

Stella, M., Ferrara, E. & De Domenico, M. (2018) *Bots sustain and inflate striking opposition in online social systems,* pp.1–10. Available at: http://arxiv.org/abs/1802.07292.

Toret, J., Calleja, A., Miró, Ó. M., Aragón, P., Aguilera, M., & Lumbreras, A. (2013) *Tecnopolítica: la potencia de las multitudes conectadas. El sistema red 15M , un nuevo paradigma de la política distribuida.* Universitat Oberta de Catalunya, Internet Interdisciplinary Institute, Working Paper Series RR13-001.

Wang, Y., Li, Y. & Luo, J. (2016) "Deciphering the 2016 U.S. Presidential Campaign in the Twitter Sphere: A Comparison of the Trumpists and Clintonists". *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM),* (Icwsm), pp.723–726. Available at: http://arxiv.org/abs/1603.03097.