

## ARTÍCULOS

# A SAMPLE DESIGN PROPOSAL FOR THE ANALYSIS OF *TWITTER* IN POLITICAL COMMUNICATION

## Propuesta de diseño muestral para el análisis de *Twitter* en comunicación política

Salvador Percastre-Mendizábal, Carles Pont-Sorribes and Lluís Codina

**Nota:** Este artículo se puede leer en español en:

[http://www.elprofesionaldelainformacion.com/contenidos/2017/jul/02\\_esp.pdf](http://www.elprofesionaldelainformacion.com/contenidos/2017/jul/02_esp.pdf)



**Salvador Percastre-Mendizábal**, is a doctoral researcher in Political Communication, member of the *Political Communication, Media and Democracy Research Group (Polcom)* at *Pompeu Fabra University (UPF)*, Barcelona, and is attached as a doctoral researcher to *ReSIC* at the *Free University of Brussels*. Researcher on the *UPF* scientific project on *Communication 2.0 in Emergencies* (funded by *Fundación-BBVA*). Member of the *European Communication Research and Education Association* and of the *Latin-American Association of Electoral Campaign Researchers*. He is a *Concyt-Cocitbc* fellowship beneficiary.

<http://orcid.org/0000-0002-5947-5102>

[salvador.percastre@upf.edu](mailto:salvador.percastre@upf.edu)

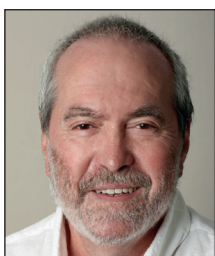


**Carles Pont-Sorribes** is doctor in Social Communication and lecturer-researcher in the *Department of Communication* at the *Pompeu Fabra University (UPF)*, Barcelona, where he is also post-graduate sub-director. His research interests include risk communication and crisis, political communication, public opinion, and journalism. Author of scientific articles published in journals indexed by the databases *Social Sciences Citation Index* and *Scopus*, among others, and of several books and book chapters. Coordinator of the *Political Communication, Media and Democracy Research Group (Polcom)*.

<https://upf.edu/es/web/polcom>

<http://orcid.org/0000-0003-1135-4245>

[carles.pont@upf.edu](mailto:carles.pont@upf.edu)



**Lluís Codina** is an associate professor of the *Journalism and Audiovisual Communication* undergraduate degree courses and coordinator of the *Master in Social Communication* in the *Department of Communication* at the *Pompeu Fabra University (UPF)*, Barcelona. He is a sponsoring member of the *Digital Documentation and Interactive Communication Research Group (DigiDoc)*, coordinator of the *DigiDoc Research Seminar* and, as a *DigiDoc member*, also a member of the *Journalism Research Unit*, a consolidated group recognized by *Generalitat of Catalonia*.

<http://lluiscodina.com>

<http://observatoriocibermedios.upf.edu>

<http://orcid.org/0000-0001-7020-1631>

[lluiscodina@upf.edu](mailto:lluiscodina@upf.edu)

*Universidad Pompeu Fabra, Departamento de Comunicación*  
Roc Boronat, 138. 08018 Barcelona, Spain

### Abstract

Based on an analysis of several sampled political discussions on *Twitter*, a new methodology is proposed that allows researchers to obtain a significant, replicable, and manageable data sample from a universe of *Twitter* metadata. The proposal is a new model called *Top discussion indicator (TDI)*. The aim of *TDI* is to assist researchers in obtaining a representative set of text from *Twitter* that includes the minimum amount of information needed to generalize the results.

Manuscript received on 20-02-2017

Accepted on 23-03-2017

## Keywords

Communication; Political communication; *Twitter*; Elections; Methodology; Metrics; Inputs; Interactions.

## Resumen

A partir del análisis de diversos ejemplos de discusiones en procesos de comunicación política en *Twitter*, se hace una propuesta metodológica para la obtención de una muestra de datos significativa, replicable y manejable de un universo de metadatos de *Twitter*, en investigaciones acerca de la comunicación política. La propuesta se basa en un nuevo modelo denominado *Top discussion indicator (TDI)* o Indicador de la máxima discusión. Su objetivo es ayudar a los investigadores a obtener un conjunto representativo de un universo que, aunque pueda superar cientos de miles de registros, incluya la mínima información que permita generalizar los resultados.

## Palabras clave

Comunicación; Comunicación política; *Twitter*; Elecciones; Metodología; Métricas; Inputs; Interacciones.

**Percastre-Mendizábal, Salvador; Pont-Sorribes, Carles; Codina, Lluís (2017).** "A sample design proposal for the analysis of *Twitter* in political communication". *El profesional de la información*, v. 26, n. 4, pp. 579-588.

<https://doi.org/10.3145/epi.2017.jul.02>

## 1. Introduction

The mission statement of the microblogging platform *Twitter* is

"to give everyone the power to create and share ideas and information, instantly, without barriers." (*Twitter*, 2016).

The company thereby defines itself as a global information empowerment platform, i.e. a medium for digital political communication (Percastre-Mendizábal; Dorantes-Aguilar, 2016).

*Twitter's* use as a global digital platform for social interaction is indisputable. According to data for March 2016, *Twitter*, 79% of whose users come from outside the USA, has 310 million monthly active users, records a billion unique monthly visits to websites with access to the platform and is available in over 40 languages (*Twitter*, 2016).

*Twitter* has not only become popular among social network users, but has also become a research subject for a growing number of academic and business researchers. In the academic sector, it has been particularly studied in the areas of (Bruns; Stieglitz, 2013):

- political communication
- crisis communication
- brand communication
- specific experiences of using *Twitter* as a backchannel
- *Twitter* use for interpersonal relationships.

Most research into *Twitter*, as used for political communication, is carried out from a quantitative and eminently positivist methodological perspective —using tools and models imported from the formal sciences- or from a statistical analysis perspective, due to the large volumes of immediate and objective data that can be gathered from this microblogging platform.

However, one of the most common problems for researchers is determining a valid sample, i.e., accurately identifying a set of data that will enable proper analysis of a research topic.

Our research focused on reducing *Twitter* data volumes for analysis by identifying the point in time at which the debate on a certain topic peaked and by examining topics through semantic nodes of discursive interconnection, such as keywords in the form of either hashtags or mentions.

The main objective of our research, therefore, undertaken from a multidisciplinary communication studies perspective, was to describe an indicator for marking particular points in time and highlighting key tweets in case studies pertaining to political communication via *Twitter*, so as to establish an optimal and meaningful set of data for analysis extracted from a much larger universe of data stored in a database.

“*Twitter* is defined as a global information enhancement platform, that is, as digital political communication media”

## 2. Theoretical framework. *Twitter* and political communication

Some of the more notable *Twitter* studies in the field of political communication concern the study and analysis of communication during election campaigns (Jungherr; Schoen; Jürgens, 2016; Conway; Kenski; Wang, 2015; Larsson; Ihlen, 2015). In addition to studies on *Twitter* use in elections in a global context, other noteworthy case studies have been carried out that address specific issues of political communication in elections, such as the following:

- vote predictions (Burckhardt; Duch; Matsuo, 2016; Guerrero-Solé; Corominas-Murtra; López-González, 2014);
- *Twitter* use by electoral candidates, parties and voters (Quevedo-Redondo; Portalés-Oliva; Berrocal-Gonzalo, 2016; Miller; Ko, 2015);
- electoral debates and information (López-Meri, 2015; López-García et al., 2015);
- simultaneous commenting during electoral debates (Park

- *et al.*, 2016; **Trilling**, 2015);
- impact on electoral outcomes and influence on/persuasion of voters (**Hosch-Dayican et al.**, 2016; **Vaccari et al.**, 2015);
- sentiment analysis in election-related discussions (**Himelboim et al.**, 2014; **Stieglitz**; **Dang-Xuan**, 2012);
- virality of messages during election campaigns (**Congosto**, 2015; **Penney**, 2014);
- negative campaigns (**Ceron**; **D'Adda**, 2015; **Just et al.**, 2012).

Political communication via the *Twitter* microblogging platform has also been studied by disciplines that include the computer sciences, sociology; and political science. In addition the classic tools of social communication, such as discourse analysis, content analysis, the theoretical perspective of framing and even formal linguistics and linguistic ethnography has been studied on *Twitter*.

As shown in a *Pew Research Center* study from the end of 2014, people —at least in the USA- who are interested in public affairs are more likely to view political content on *Twitter* than on any other digital social network (**Gottfried**, 2014). It can therefore be concluded that democratic access to digital social media has impacted the *status quo* of traditional information hierarchies.

Most studies of *Twitter* as used for political communication are made from a quantitative and positivist methodological perspective using models imported from the formal sciences

Mainstream media have had to adapt to new digital forms of information sharing —as evidenced by various empirical studies (**Pont-Sorribes**; **Codina**; **Pedraza-Jiménez**, 2010)- and even more so since the advent of digital social media, where what a person is or represents is less important than what they say in posts and share, tweet, or retweet to friends or followers.

Since *Twitter* emerged as a platform for political communication, it has fundamentally changed the dynamics of electoral campaigns and political processes, not only in terms of how candidates and political actors fulfil their electoral promises and governing objectives, but also in how information flows between the public, activists, and political elites (**Ammann**, 2010).

### 3. Method

Any study of political communication phenomena requires an analysis of the communication processes between three actors in a social system (**Wolton**, 1989). These actors, who may legitimately voice their opinions on matters of public interest, are as follows:

- 1) Public: public institutions, public officials, and politicians.
- 2) Media: the media, journalists, and communicators.
- 3) Citizens: who can be divided into two types:

- individuals belonging to interest or pressure groups, platforms or associations with clearly defined objectives and causes;
- individuals who express themselves using instruments which can be used to measure public opinion, including, in the era of hyper mediation, digital social communication platforms and most especially social networks.

Digital political communication processes —whether institutional, electoral, or relating to public management of a crisis or emergency- have three elements in common:

- participation of at least one actor from any of the three groups of actors described above.
- a formal or informal starting point, which can usually be determined according to the kind of event, e.g. the legal duration of an electoral campaign, the unfolding of a crisis, or the formal period dedicated by a government to a political event.
- a natural curve of developments, with a beginning, a period of development, a climax, and a conclusion.

We found that the above characteristics were shared by three cases of political communication via *Twitter* that we analyzed: an election campaign, an airline accident, and a public health crisis (described in Section 4). The data used to establish the study universe was based on a compilation of tweets and retweets using specific hashtags and mentions.

#### 3.1. Metrics

Determining a measurement system was fundamental to our research, but first we had to identify which variables to measure in order to be able to decide which measurement system to use. The identified variables not only reflected the nature of the studied phenomenon and the objectives of the research, but also the very nature of the *Twitter* ecosystem itself, i.e. the possibilities for exploiting quantitative elements of the platform.

Depending on the conceptual categories determined for a research study, independent variables, dependent variables, and control variables need to be established in accordance with the requirements of a particular analytical model.

Quantifiable variables included:

- followers;
- visits;
- number of tweets and retweets;
- the most influential tweeters;
- tweeters who assiduously followed other tweeters;
- keywords in the form of hashtags or mentions;
- tweeter biographies or descriptive information;
- tweeter, keywords and discussions geolocation;
- trending topics;
- positive, negative or neutral comments, and;
- tweets as responses to offline discussions (e.g. *Twitter* discussions about televised news or debates).

The variables that we considered crucial to an understanding of the political communication processes that we studied on *Twitter* and that ultimately formed the basis for our analytical model are described below in terms of core variables, inputs, and interactions.

## A. Core variables

Core variables are highly quantifiable elements that describe the importance, relevance, presence, and hierarchy of tweeters. They enabled an initial evaluation according to digital audience size, measured as the number of followers of public, media, and citizen actors (i.e. political and institutional tweeters, political news media, journalist and communicator tweeters, and grassroots/activist tweeters).

a) *Followers*. The number of tweeters following an account or forming part of the universe of study at the time of data collection, bearing in mind that the number of followers can vary considerably in a short period of time, especially for candidate and political party tweeters. In fact, one of the objectives of social network-based campaigning is to obtain the largest possible number of followers in the geographical area where the campaign, electoral process, or institutional crisis is unfolding.

b) *Following*. The number of tweeters following the studied *Twitter* accounts at the time of data collection.

c) *Retweets in*. Retweets received by a *Twitter* account as a proportion of all posts referring to a particular case study, whether delimited by time, hashtags, or mentions.

d) *Likes (formerly favorites)*. The number of likes by other tweeters of posts by a *Twitter* account included in the study universe, which yields the volume of data to ultimately be analyzed.

## B. Inputs

Inputs are the messages transmitted through *Twitter* accounts, posted by *Twitter* users who compose their own tweets, retweet posts by others; or copy-and-paste textual, audiovisual, or interactive content obtained elsewhere. These inputs are usually manual, although they can also be automated using a computer program called a bot.

a) *Original tweets*. Original messages posted by the tweeter, which can take the form of alphanumeric characters, videos, audios, podcasts, gifs, photographs, illustrations, graphics, digital posters, memes, hyperlinks, emoticons, mentions of other *Twitter* accounts, etc.

b) *Retweets out*. *Twitter* messages by the same or another tweeter that are retweeted, i.e. messages retweeted by the *Twitter* account owner, whether or not composed by them.

c) *Total inputs*. The sum of all tweets and retweets out.

## C. Interactions

These *Twitter* elements are the means by which tweeters interact with and establish relationships with other tweeters, either through specific topics identified by hashtags or directly through explicit mentions of *Twitter* handles (IDs).

a) *Hashtags*. Words or phrases without spaces preceded by a hash (#) symbol that express a concept or reflect a topic of particular interest in tweetosphere conversations.

b) *Mentions*. Direct references to other accounts, indicated by a username preceded by the at (@) symbol, whether as straightforward mentions or as part of the content of a message.

c) *Shared links*. Hyperlinks consisting of an electronic address that reroutes users to content elsewhere, usually (but not always) outside the platform (websites or other social networks).

d) *Total interactions*. The sum of all interactions, whether the interaction units themselves (mentions or hyperlinks) or messages reflecting interactions, provided that the parameter for computing the interactions has been established and justified.

## 3.2. Data selection using the *Top discussion indicator (TDI)*

Although numerous methodologies are available for analyzing political communication on *Twitter*, sampling is random and the criteria used tend not to be uniform. Some recent studies of political communication tweets are listed below.

**Jungherr, Schoen and Juergens** (2016) analyzed, in their study of the run-up to an election campaign, 6,677,795 tweets posted by 1,248,667 *Twitter* users. **Burckhardt, Duch, and Matsuo** (2016) analyzed a random sample of 8 million tweets from 25 million compiled tweets that referred by name to the six largest political parties in the UK and their leaders over the five and a half months prior to an election campaign.

**Hosch-Dayican et al.** (2016) analyzed 368,855 tweets collected over two and a half weeks, sorted them by hashtags and filtered them first by language (Dutch) and then by location. Using a proprietary selection algorithm, they excluded irrelevant tweets to only include those directly related to parliamentary elections, parties, and candidates.

**López-Meri** (2015) analyzed 500 randomly selected tweets about Catalan elections posted on voting day, the two days before, and the day after.

**Miller and Ko** (2015) analyzed tweets by 50 *Kuwaiti Parliament* electoral candidates during an election campaign.

**Ceron and D'Adda** (2015) analyzed 15,053 tweets by the eight largest Italian parties and their leaders during an electoral campaign.

**Congosto** (2015) compiled 1,552,282 tweets by Spanish electoral list leaders and political parties, mentions of those tweets and tweets related to the election campaign, although they only classified 536,588 tweets (35.18%) of the compiled sample.

Even though it may seem that there is no limit to the generation of information in digital social networks, imposing "natural" limits enable specific data universes of political communication via *Twitter* to be rationally delimited, for instance, by geographic areas, keywords, or time frames.

Our *Top discussion indicator (TDI)* is a proposal for identifying a particular chronological moment for data collection that considerably reduces data volumes, enhances overall data relevance and potentially avoids noise.

The idea is to pinpoint a set of *Twitter* conversations and users that reflect the greatest volume of debate on a po-

political communication by identifying the moment in time when a specific topic—as defined by keywords—attracts the greatest attention in the tweetsphere, as reflected in the number of both inputs and tweeters.

Regardless of the kind of case study being conducted, a plausible volume of data can be extracted from information compiled in a *Twitter* metadata database by using the *TDI* to locate and demarcate the specific moment in time—day, hour or other specific time bracket—that concentrates the largest volume of mentions and hashtags. The *TDI* thus enables peak moments of debate to be identified, when potential message impact is greatest, given the large number of active tweeters and of exchanged messages.

Using the *TDI* we could identify, for subsequent analysis, tweets and tweeters corresponding to the most significant part of the entire chronological curve referring to the political communication of interest.

Given the overload resulting from data usually counting in the hundreds of thousands of items and the need to minimize noise resulting from overly large and poorly representative samples, it is undoubtedly useful to identify peak moments when debates are most heated, since such moments reflect key developments in the political communication of interest.

#### 4. Case studies

We analyzed three scenarios that generated significant tweeting activity:

- the Mexican federal elections of 2015;
- the *Germanwings* flight 9524 crash of 2015, and
- the Ebola crisis in Spain in 2014.

##### 4.1. Mexican elections

Federal parliamentary elections, held on June 7, 2015 in the 32 federal entities of Mexico to elect the 500 members of the *Chamber of Deputies* of the *Congress of the Union*, were marked by the highest voter participation in the history of



Figure 1. Trend over time for tweets using the hashtag #elecciones2015 in the pre-electoral period (5 April-4 June 2015) leading up to federal elections in Mexico.

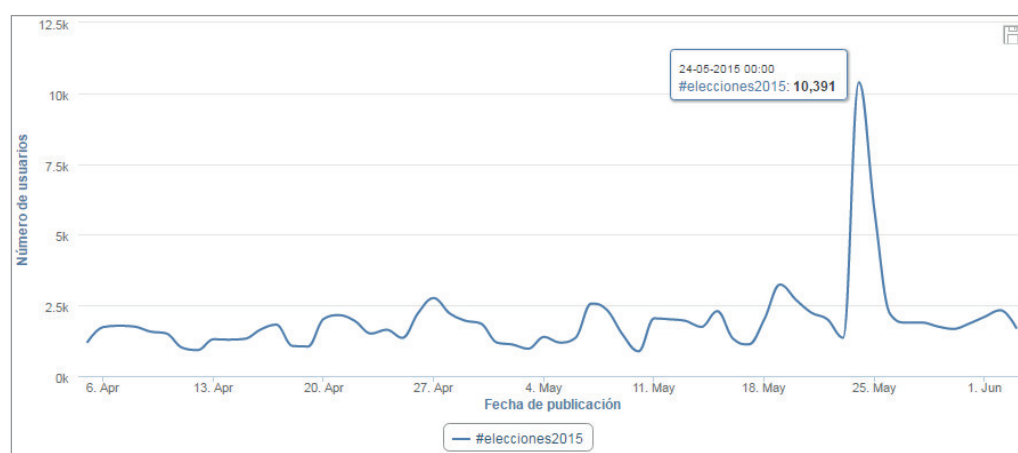


Figure 2. Trend over time for tweeters using the #elecciones2015 hashtag in the pre-electoral period (April 5 – June 4, 2015) leading up to federal elections in Mexico.

the country. The legal electoral campaign period lasted 61 days (April 5 to June 4).

During the campaign period, we collected data on all *Twitter* debates—original tweets, retweets, and modified tweets—that used the hashtag #elecciones2015 and reflecting what we refer to as digital semantic nodes of discursive interconnection.

Once filters were applied to distinguish between tweeters by geographical location (as indicated in bios) and by language, some 250,000 records resulted that were chronologically ordered and plotted as in Figure 1. The plot clearly points to May 25 as the day that generated the largest volume of debate regarding the analyzed hashtag; of the total universe of 250,000 records, 15,964 corresponded to this date.

Applying the *TDI*, of the 250,000 records, 684 were retweeted 50 times, and of these, 548 obtained 10 or more “likes” (0.21% of the total). Application of the *TDI*, therefore, resulted in a substantial reduction of the sample to be analyzed—from a quarter of a million records to just over 500.

Likewise, as shown in Figure 2, the date on which tweeting activity was greatest was also the date featuring the largest number of active tweeters (10,391) interacting around the hashtag #elecciones2015.

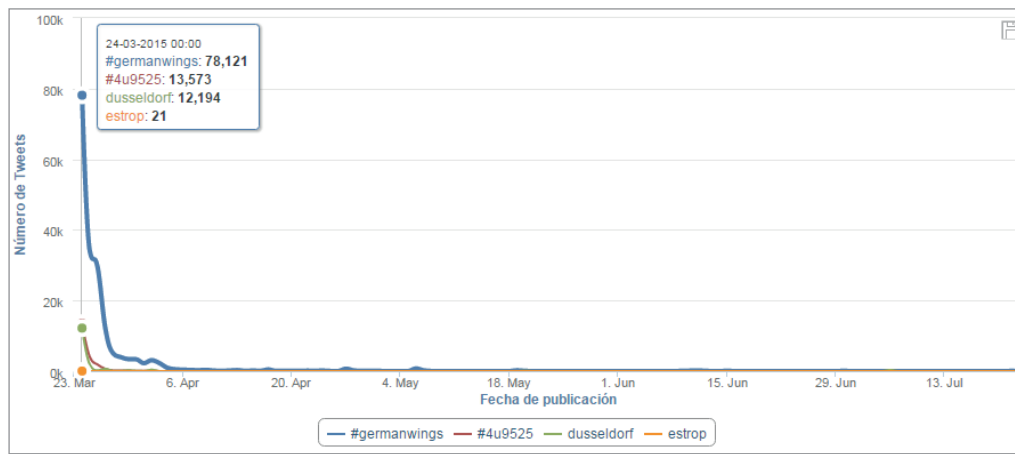


Figure 3. Trend over time for tweets using hashtags referring to the *Germanwings* flight 9525 crash of March 24, 2015.

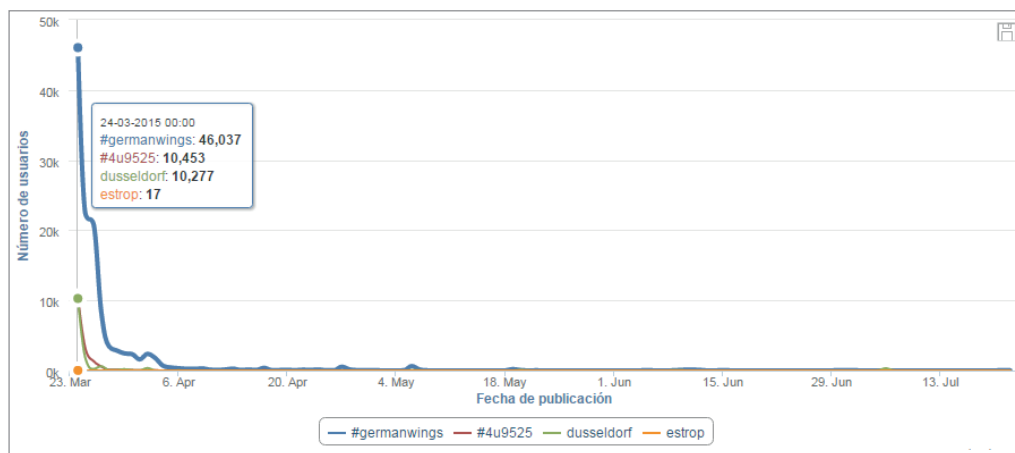


Figure 4. Trend over time for tweeters using hashtags referring to the *Germanwings* flight 9525 crash of March 24, 2015.

#### 4.2. Germanwings crash

*Germanwings* flight 9525, which took off from the Barcelona en route for Düsseldorf on March 24, 2015, crashed in the French Alps at around 10:30 am. All 150 people on board died on the spot. Investigations concluded that the accident was intentionally caused by the airplane’s co-pilot.

Figure 3 illustrates a *Twitter* debate that was hyper concentrated from when the earliest information on the flight’s disappearance from the radar and on the crash became available and that then tapered off until 22 July, when the airline officially announced details of compensation for the bereaved.

A universe of data was generated with 235,829 inputs from *Twitter* debates that used the hashtags:

- #Germanwings;
- #4U9525 (referring to the flight number);
- Dusseldorf;
- Estrop (alluding to Tête de l’Estrop, the Alpine peak near where the airplane crashed).

It can be observed that March 24, 2015 was when the greatest volume of hashtags and mentions of the accident were generated, amounting to 103,909 records for the most commonly used hashtags and mentions.

Applying the *TDI* to retrieve the most outstanding inputs from the total universe of 235,829 records, 1,273 received 50 or more retweets, and of these, 481 had 10 or more “likes” (0.20% of the initial total).

Figure 4 shows that the largest number of tweeters coincided with the date on which tweeting activity was at its highest. The greater the number of mentions, the greater the number of users interacting with these mentions, as was the case of the most used hashtag #Germanwings. The tweeters interacting with the four hashtags numbered 66,784 in total.

#### 4.3. Ebola crisis in Spain

This health crisis unfolded with the first case of Ebola infection reported in Spain (and in Europe) on October 6, 2014. Teresa Romero, a nursing assistant at *Hospital Carlos III of Madrid*, was infected after caring for a Spanish missionary, Miguel Pajares, who had been repatriated from Liberia after becoming infected following the outbreak of Ebola in West Africa in 2014.

A total of 1,880,750 records were collected from *Twitter* debates around the hashtags:

- #ebolaenespana
- #ebolaenespaña
- #salvemosalexcalibur (referring to Teresa Romero’s dog, eventually put down)
- #anamatodimision (referring to the Health Minister, Ana Mato)
- #javierrodriguezdimision (referring to the head of health-care for Madrid, Javier Rodríguez)
- #teresaromero
- #todossomosteresa, and
- #vamosamorirtodos (literally, “we’re all going to die”).

Records were collected from June 25, 2014, when the European Ebola emergency protocol was activated (in response to what turned out to be a false alarm in Valencia), and December 2, 2014, when the *World Health Organization* offi-

cially declared an end to the Ebola emergency in Spain.

The *TDI* analysis revealed that of the 1,880,750 tweets collected, 3,360 were retweeted 50 or more times, and of these, 1,113 received 10 or more “likes” (0.05%).

Figures 5 and 6 depict similar behavior as in the other two case studies, with a hyper concentration of tweets (358,290 in total) on a specific date, in this case, October 8, 2014, also the date on which the highest number of tweeters interacted around the same hashtags.

Two criteria are proposed for determining the *TDI*: the time frame and the viralization factor.

Determining a time frame requires establishing when tweeting activity is greatest (when debate peaks), whether a day, part of a day or even a specific hour. Adding all original tweets generated directly by *Twitter* users (excluding retweets), a relevancy criterion can be established to identify tweets with the greatest impact.

“ A typical problem faced by researchers is determining a valid sample or safely identifying a set of significant data for analysis as the object study ”

As for viralization, and bearing in mind how it functions in *Twitter*, considered as having significant content are original tweets with a high impact —i.e. retweeted at least 50 times (stated on the platform’s website to be the number that represents 50% of the latest users to have retweeted a public tweet). These tweets are further reduced by selecting those receiving at least 10 “likes”, a randomly selected cut-off point that serves as a second quality filter to identify messages of greatest qualitative interest to other tweeters.

These numerical cut-offs act as a discriminatory mechanism for determining the most significant tweets. The fact that a greater numerical value is awarded to retweets than to “likes” is justified in two ways:

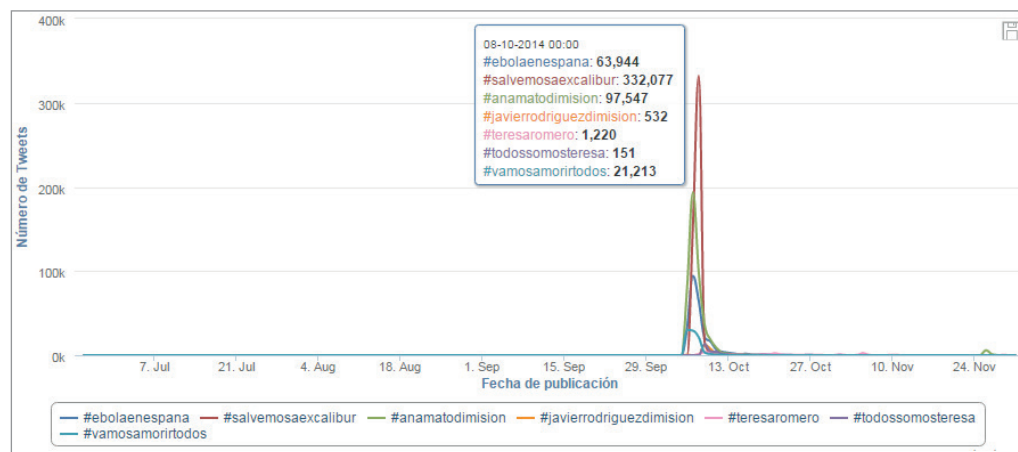


Figure 5. Trend over time for tweets using hashtags referring to the first Ebola case in Spain reported on October 6, 2014.

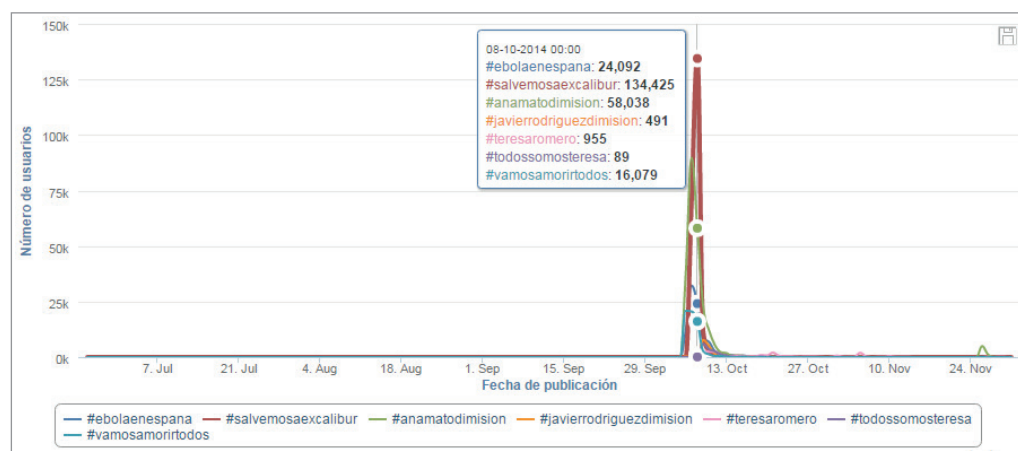


Figure 6. Trend over time for tweeters using hashtags referring to the first Ebola case in Spain reported on October 6, 2014.

- first, in the internal logic of the *Twitter* timeline, since a retweet gains greater visibility than a “like” it has a better probability of going viral; and
- second, as our analysis has shown, tweets tend to receive more retweets than they receive “likes”.

For our case studies, the *TDI* reflected content transmitted within a significantly delimited time period (of just a single day in each of our case studies) and messages that went viral. We were thus able to identify both when a debate peaked and which messages had the most impact (as indicated by retweets and “likes”).

The *TDI* proved useful in identifying a more manageable proportion of especially relevant *Twitter* data —as shown by the analyses of the three cases described above- with which to perform qualitative ethnographic, semantic, or discursive analyses.

### 5. Discussion and conclusions

The algebraic formula for identifying a key set of records from a universe of *Twitter* metadata can be expressed as follows:

$$cTDI = (U-tx)-(m)-(s)$$

where

- *cTDI* represents the *TDI*-selected set of data to be analyzed,
- *U* represents the universe of metadata, *tx* represents data from a specific time frame (established as days, a day, part of a day, hours, minutes, or seconds),
- *m* represents all low-impact inputs (tweets with fewer than 50 retweets) and, finally,
- *s* represents tweets receiving fewer than 10 “likes”.

In a big data or data intelligence contexts and considering the vast amounts of data available in *Twitter*, research into communication (which is mainly qualitative) encounters difficulties in dealing with and selecting data for analysis. This is due not only to the amount of data, but also to the very nature of data storage, which often leads to problems with management, classification, and analysis using traditional study tools, including the information technologies. Sample design using the *TDI* allowed us to obtain more manageable data sets for analysis using either a qualitative and quantitative approach, although, for obvious reasons, the *TDI* is more usefully applied in research adopting a qualitative approach.

“ We propose a valid indicator for determining a moment in time in which to extract an optimal set of significant data from a *Twitter* universe for analysis ”

As far as the exploitation of data is concerned, most studies are carried out using quantitative methods —e.g. content analysis, sentiment analysis, or network analysis— or implementing highly descriptive statistical analyses. This circumstance has led to problems not only in how we manage large volumes of data, but also in how we decide which models to use for analysis and which tools to use for data collection, and even to doubts as to the kind of specialist knowledge needed to understand these new ways of measuring data.

Depending on research objectives and approach, researchers need to be able to methodologically filter their universe of study so as to obtain a sample of data that can be easily managed and efficiently analyzed in a way that adapts to the methodological and computational tools available.

Although experience with the *TDI* is for the moment limited to political communication, it is likely to be applicable to data universes reflecting other communication or social science phenomena; however, it remains for other researchers to demonstrate its applicability to other disciplines.

The *TDI* is undoubtedly of use in a context in which no replicable and universal criteria have yet been developed for determining sample proportionality according to reasonable methodological principles and plausible criteria of representativeness. Researchers lack an efficient method for studying social media use —somewhat like what happened in the 1960s when television emerged—. Although a wide range of methodological approaches and analytical tools are available, in case studies on themes that are epistemologically very similar, researchers use a wide variety of —

sometimes diametrically opposed— methods for weighting, variable determination, and categorization, not to mention data collection, database configuration, and sample selection.

The *TDI* is proposed as a methodological tool that identifies a *temporis momentum* that allows the researcher to extract a reduced yet meaningful data set from a universe of data collected on one or more *Twitter* topics.

Although we found the *TDI* to be methodologically useful for our case studies, its approach to sample selection system has two shortcomings.

- The first is that, on the basis of the empirical evidence available, it cannot be assumed that the *TDI* can usefully be applied to other disciplinary fields.
- The second is that, even though a final sample may apparently be plausible and significant, there is no statistical correlation with the data universe that would vouch for the proportionality of the sample.

That said, however, even statistical models have been shown to be incapable of determining unequivocal correlation (as has been seen in the case of public opinion studies).

Our proposal of the *TDI* as a means for sampling from large data sets aims to contribute to and promote a debate—as well as encourage further empirical research—on methodological criteria aimed at standardizing academic inquiry into *Twitter* and similar platforms.

“ The *Top discussion indicator (TDI)* substantially reduces the volume of data to analyze, thereby avoiding possible “noise” and avoiding less significant data collection moments ”

## Note

This work is an output on the scientific project *Communicating in emergency situations, tools 2.0 and new protocols in the efficient management of emergency communication*, financed by the *BBVA Foundation*, and is part of the project *Creation and interactive content in the communication of audiovisual information: audiences, design, systems and formats*, CSO2015-64955-C4-2-R (*Mineco / Feder*), *Ministry of Economy and Competitiveness* (Spain).

## 6. References

- Ammann, Sky** (2010). *A political campaign message in 140 characters or less: The use of Twitter by U.S. Senate candidates in 2010*. SSRN Papers. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1725477](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1725477)
- Bruns, Axel; Stieglitz, Stefan** (2013). “Towards more systematic Twitter analysis: metrics for tweeting activities”. *International journal of social research methodology*, v. 16, n. 2, pp. 91-108. <https://eprints.qut.edu.au/58132> <https://doi.org/10.1080/13645579.2012.756095>



- Burckhardt, Philipp; Duch, Raymond; Matsuo, Akitaka** (2016). "Tweet as a tool for election forecast: UK 2015. General election as an example". En: *Third annual meeting of the Asian Political Methodology Society in Beijing*. [http://asiapolmeth.princeton.edu/sites/default/files/polmeth/files/uk\\_election\\_tweets\\_asia\\_polmeth.pdf](http://asiapolmeth.princeton.edu/sites/default/files/polmeth/files/uk_election_tweets_asia_polmeth.pdf)
- Ceron, Andrea; D'Adda, Giovanna** (2015). "E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election". *New media & society*, v. 18, n. 9, pp. 1935-1955. <https://goo.gl/xZEDnC>  
<https://doi.org/10.1177/1461444815571915>
- Congosto, María-Luz** (2015). "Elecciones europeas 2014: viralidad de los mensajes en Twitter". *Redes. Revista hispana para el análisis de redes sociales*, v. 26, n. 1, pp. 23-52. <https://doi.org/10.5565/rev/redes.529>
- Conway, Bethany; Kenski, Kate; Wang, Di** (2015). "The rise of Twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary". *Journal of computer-mediated communication*, v. 20, n. 4, pp. 363-380. <https://goo.gl/3j2DKG>  
<https://doi.org/10.1111/jcc4.12124>
- Gottfried, Jeffrey** (2014). "Facebook and Twitter as political forums: Two different dynamics". *Pew Research Center*, 12 November. <http://www.pewresearch.org/fact-tank/2014/11/12/facebook-and-twitter-as-political-forums-two-different-dynamics/>
- Guerrero-Solé, Frederic; Corominas-Murtra, Bernat; López-González, Hibai** (2014). "Pacts with Twitter. Predicting voters' indecision and preferences for coalitions in multiparty systems". *Information, communication & society*, v. 17, n. 10, pp. 1280-1297. <https://goo.gl/XvWq9G>  
<https://doi.org/10.1080/1369118X.2014.920040>
- Himmelboim, Itai; Sweetser, Kaye; Tinkham, Spencer; Cameron, Kristen; Danelo, Matthew; West, Kate** (2014). "Valence-based homophily on Twitter: Network analysis of emotions and political talk in the 2012 presidential election". *New media & society*, v. 18, n. 7, pp. 1382-1400. <https://doi.org/10.1177/1461444814555096>
- Hosch-Dayican, Bengu; Amrit, Chintan; Aarts, Kees; Dassen, Adrie** (2016). "How do online citizens persuade fellow voters? Using Twitter during the 2012 Dutch parliamentary election campaign". *Social science computer review*, v. 34, n. 2, pp. 135-152. <https://doi.org/10.1177/0894439314558200>
- Jungherr, Andreas; Schoen, Harald; Jürgens, Pascal** (2016). "The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013". *Journal of computer-mediated communication*, v. 21, n. 1, pp. 50-68. <https://goo.gl/zBw3dz>  
<https://doi.org/10.1111/jcc4.12143>
- Just, Marion; Metaxas, Panagiotis; Mustafaraj, Eni; Crigler, Ann** (2012). *It's trending on Twitter: An analysis of the Twitter manipulations in the Massachusetts 2010 special Senate election*. <http://www.researchgate.net/publication/256026592>
- Larsson, Anders; Ihlen, Øyvind** (2015). "Birds of a feather flock together? Party leaders on Twitter during the 2013 Norwegian elections". *European journal of communication*, v. 30, n. 6, pp. 666-681. <https://goo.gl/N9YpYF>  
<https://doi.org/10.1177/0267323115595525>
- López-García, Guillermo; Gámir-Ríos, José-Vicente; García-Ull, Francisco-José; Llorca-Abad, Germán; Cano-Orón, Lorena; González-Esteban, José-Luis** (2015). "El debate sobre Europa en Twitter. Discursos y estrategias de los candidatos de las elecciones al Parlamento Europeo de 2014 en España". *Revista de estudios políticos* (nueva época), n. 170, pp. 213-246. <https://doi.org/10.18042/cepc/rep.170.07>
- López-Meri, Amparo** (2015). "Redes sociales y campañas electorales: Twitter como fuente informativa en las elecciones catalanas del #25N". *Comunicació: Revista de recerca i d'anàlisi*, v. 32, n. 2, pp. 115-137. <https://goo.gl/duUNdN>
- Miller, Noah; Ko, Rosa** (2015). "Studying political microblogging: Parliamentary candidates on Twitter during the February 2012 election in Kuwait". *International journal of communication*, v. 9, pp. 2933-2953. <http://ijoc.org/index.php/ijoc/article/view/1910/1464>
- Park, Se-Jung; Park, Ji-Young; Lim, Yon-Soo; Park, Han-Woo** (2016). "Expanding the presidential debate by tweeting: The 2012 presidential election debate in South Korea". *Telematics and informatics*, v. 33, n. 2, pp. 557-569. <https://goo.gl/nZHJCP>  
<https://doi.org/10.1016/j.tele.2015.08.004>
- Penney, Joel** (2014). "Motivations for participating in 'viral politics': A qualitative case study of Twitter users and the 2012 US presidential election". *Convergence: The international journal of research into new media technologies*, v. 22, n. 1, pp. 71-87. <https://goo.gl/BiaWgU>  
<https://doi.org/10.1177/1354856514532074>
- Percastre-Mendizábal, Salvador; Dorantes-Aguilar, Gerardo-Luis** (2016). "Comunicación política electoral en la era digital. La campaña electoral de 2015 en la ciudad de México". *Vivat academia. Revista de comunicación*, diciembre, n. 137, pp. 36-56. <http://www.vivatacademia.net/index.php/vivat/article/view/976/1083>
- Pont-Sorribes, Carles; Codina, Lluís; Pedraza-Jiménez, Rafael** (2009). "Comunicación de riesgo y sistemas de información en la Web: cinco modelos". *El profesional de la información*, v. 18, n. 4, pp. 389-397. <https://doi.org/10.3145/epi.2009.jul.05>
- Quevedo-Redondo, Raquel; Portalés-Oliva, Marta; Berrocal-Gonzalo, Salomé** (2016). "The image use on Twitter during the 2015 municipal election campaign in Spain". *Revista latina de comunicación social*, v. 71, pp. 85-107. <https://doi.org/10.4185/RLCS-2016-1085en>
- Stieglitz, Stefan; Dang-Xuan, Linh** (2012). "Political communication and influence through microblogging – An empirical analysis of sentiment in Twitter messages and retweet

behavior". En: *IEEE computer society*, pp. 3500-3509.

<https://goo.gl/PswMJK>

<https://doi.org/10.1109/HICSS.2012.476>

**Trilling, Damian** (2015). "Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter". *Social science computer review*, v. 33, n. 3, pp. 259-276.

<https://goo.gl/ve2AMe>

<https://doi.org/10.1177/0894439314537886>

Twitter (2016). About.

<https://about.Twitter.com/company>

**Vaccari, Cristian; Valeriani, Augusto; Barberá, Pablo; Bonneau, Rich; Jost, John; Nagler, Jonathan; Tucker, Joshua** (2015). "Political expression and action on social media: Exploring the relationship between lower- and higher-threshold political activities among Twitter users in Italy". *Journal of computer-mediated communication*, v. 20, n. 2, pp. 221-239.

<https://doi.org/10.1111/jcc4.12108>

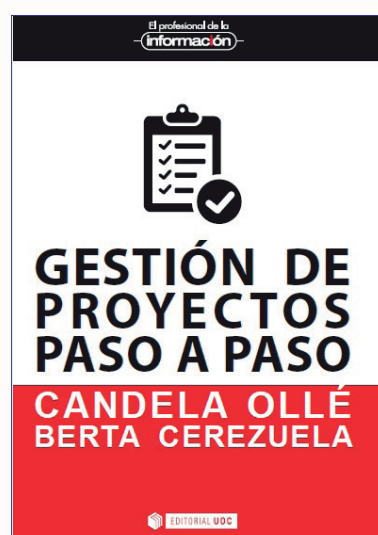
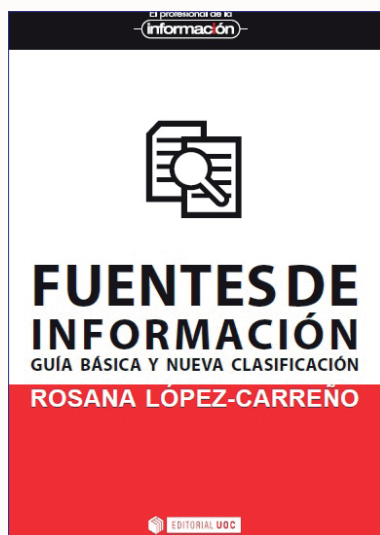
**Wolton, Dominique** (1989). "La communication politique: construction d'un modèle". *Hermès*, n. 4, pp. 27-42.

<https://doi.org/10.4267/2042/15353>

## Colección de libros de bolsillo

### *El profesional de la información (Editorial UOC)*

#### Últimos títulos publicados



Media

Publishing

Editorial UOC

## NOVEDADES

Más información:

<http://www.elprofesionaldelainformacion.com/libros.html>