Author Version

Mediatisation in Twitter:

An exploratory analysis of the

2015 Spanish General Election (*)

Authors:

Tomás Baviera
Polytechnic University of Valencia

Dafne Calvo
University of Valladolid

Germán Llorca-Abad
University of Valencia

(*) This manuscript was accepted for publication by *The Journal of International Communication* on June 18th, 2019. The text finally published has slightly changed with respect to the present document. The reference of the paper is:

Baviera, Tomás; Calvo, Dafne, & Llorca-Abad, Germán (2019). Mediatisation in Twitter: An exploratory analysis of the 2015 Spanish General Election, *The Journal of International Communication*, DOI: 10.1080/13216597.2019.1634619

Mediatisation in Twitter: An exploratory analysis of the 2015 Spanish General Election

The mediatisation model in politics assumes that media conveys political messages between parties and citizenship, with the risk of promoting issues that frame the electoral content in terms of competition. These dynamics could distract from the debate of ideas and political policies. However, digital media like Twitter provide direct communication channels between parties, candidates and users. The present research explores Twitter content during an electoral campaign focused on the four *issues* proposed by Patterson (1980) to assess mediatisation: *political, policy, campaign* and *personal* (regarding the candidate). The goal of this research study is to evaluate the degree of mediatisation on Twitter using this typology. The research also evaluates the influence of the issue on retweet volume. The study's basis was a 15.8 million-tweet corpus obtained during the 2015 Spanish General Election pre-campaign and campaign. This dataset was analysed using an automatic classification system. The results highlighted a predominance of *policy issues* during both the pre-campaign and campaign, except for the two televised debates, during which *campaign* issues were the most prevalent. On the election night, users commented much more on *political issues.* Finally, the kind of issue most likely to be retweeted was *policy issues.*

Keywords: political communication, mediatisation, Twitter, electoral campaign, computer-assisted content detection

The mediatized model of political communication establishes a public space heavily dependent on the mass media. According to this view, the relationship between the political system and citizens is embedded in the media (Mancini and Mazzoleni 1995, Mazzoleni 2004, Strömbäck and Esser 2009, Couldry and Hepp 2013). However, this setting introduces factors that altered the dynamics of the public sphere. This is especially evident in television media where an increasingly competitive environment tends to frame politics as a competition, to arouse more interest and attract more viewers (Patterson 1993). An obvious risk in this model is that issues that should be treated during political debates, such as questions about the economy or health system, might be underestimated during electoral campaigns. In contrast to this scenario, digital media provide channels for

direct interaction between the political actors and citizens that lessen a dependence on the mass media for political communication (Kreiss 2012, Bor 2014).

Chadwick (2013) features this public scenario as a hybrid media system. The vertical and hierarchical structures of the mass media incorporate participatory Internet practices, while social media continues to draw from press, radio and television news sources. However, this hybridisation is not easy to evaluate. The integration of new and old media raises relevant questions about the current validity of the mediatisation model, mainly due to the role played by the active users in the digital sphere. This study wishes to contribute to exploring the extent to which political communication on Twitter has become mediatized. The research focuses on the electoral pre-campaign and campaign for the 2015 Spanish General Election.

Researchers currently assess mediatisation processes through two kinds of indicator: *framings* and *journalistic narratives* (Martínez Nicolás *et al.* 2014). For this aim, the most used frames are strategic, conflict-oriented, and personalised. The main journalistic narratives that allow this assessment are descriptive, interpretative, and sceptic. These indicators require a careful qualitative assessment. As we focus on large volumes of short texts, as in the case of Twitter, we need to use computer-assisted tools. Nevertheless, implementing those qualitative methodologies with these kinds of tools is a challenge and this is why we turned to a previous methodology developed by Patterson (1980) to study the mediatisation effect on Twitter.

Patterson proposed a four-category-issue typology through which it was possible to quantify media influence on shaping political content disseminated by the media. This categorisation facilitates differentiation between the issues that are typical of the political sphere and those prioritising information about the campaign and candidates. Although topic detection is better developed in computer-assisted analysis than framing or journalistic narratives, the approach based on Patterson's typology was not free from difficulties. The short texts and the abstract concepts involved in the typology drove us to develop a system based on artificial intelligence methods in which several Spanish Natural Language Processing (NLP) research centres collaborated. However, this system limited our conclusions as we prioritised the reliability of the results. This is why our analysis should be considered as exploratory. In any case, the relevance of this research study must be framed as a step forward in assessing the mediatisation in the digital sphere of Twitter through computer-assisted methods.

This article is structured as follows. First is a review of the literature on electoral campaigns, political mediatisation, and computer-assisted content analysis on Twitter. Once the research questions have been formulated, the 2015 Spanish electoral context is outlined and the methodology described. Following that, the results obtained are presented and some threats to their validity are acknowledged to counterbalance the research significance. Finally, the results are discussed and the main conclusions presented.

Literature review

Mediatisation and Twitter

The relationship between media and politics has been an important topic of interest since

the Hypodermic Needle Theory. Researchers have been concerned about delimiting the scope of media influence on people. Strömbäck and Esser (2009) stand for a clear distinction between *politically mediated* and *politically mediatisated* messages. While *mediation* should be understood as the neutral act of transmitting a message through a technological mean, *mediatisation* should be seen as process by which media can influence the message beyond transmitting it in a particular format. "The media should rather be understood as an ever-present social and cultural system of production, broadcast, circulation, and dissemination of symbols, signs, messages, meanings, and values" (Strömbäck and Esser 2009, p. 209).

The mediatisation model proposed by Mazzoleni and Schulz (1999) also highlights the complexity of the concept and the encompassing position that the media holds in political communication. The interaction between parties and citizens takes place mainly in a public space created by the media given that they provide the channels for transmitting messages to a broad and general public. The media constitute the space in which power relations are decided (Castells 2009). Furthermore, it is broadly acknowledged that the interest of attracting an audience raises questions regarding the objective delivery of political content. The risks of this mediatisation of politics have been highlighted by various researchers (Brants and Neijens 1998, Blumler and Kavanagh 1999, Williams and Delli Caprini 2011, Couldry and Hepp 2013).

The study of mediatisation is the analysis of the performative capacity of the media in the social field, in particular on the media users' expectations (Strömbäck and Esser 2009, Finnemann 2014). The mediatisation is directly related to the increase of the presence of media and communication systems in the social field (Couldry and Hepp 2013). It is understood that its performative capacity is profound and diverse, and it affects very different levels of the social domain. Therefore, mediatisation processes are hard to delimit (Strömbäck and Esser 2009). This difficulty has become more complicated with the disruption of digital media, as they have enabled more trajectories for human communication. Message selection is not confined to the media organisations on the Internet, neither its dissemination; rather both depend on technical, sociocultural, institutional and, most importantly, individual factors (Finnemann 2014).

Strömbäck (2008) considers a close relationship between the Internet and mediation since the network is a supplement to traditional media as a source of information. Furthermore, most of the relevant political events have an impact on the online sphere, and social actors aim to disseminate their particular messages on the Net to tackle the absence of coverage in traditional news media. Richards (2010) describes this new context as *the emotional public sphere*. The Internet has become a dynamic and fertile ecosystem for citizen participation, as the driving force behind a new media regime (Delli Caprini and Williams 2001). Indeed, since its launch in 2006, Twitter has demonstrated its potential in playing the gatekeeper role of traditional media (Kalsnes *et al.* 2014).

Despite the growing number of Internet users, the media continue to play a significant role in determining the content, provided we do not give a disproportionate importance to the technological component by itself (Wright 2012). Blogs initially, along with social networks today, channel a significant part of the political discussion, with evident correlations with the media (Conway *et al.* 2015). The effects of agenda setting on Twitter have been extensively researched (Neuman *et al.* 2014, Vargo *et al.* 2014, Conway *et al.* 2015, Guo and Vargo 2015, Vergeer and Franses 2016). The attention paid to salient topics on Twitter tends to be correlated with the importance attributed by the media, although there are differences in the rhythms of that attention between both spaces

(Neuman *et al.* 2014). Jungherr (2014) confirmed this dependence during an electoral campaign by breaking down the volume of mentions of candidates by written media, television and Twitter. He found a component that was exclusively internal to Twitter, along with two others in which the references to candidates made in those three types of media were interrelated. Furthermore, the Twitter conversation reveals interest in topics overlooked by the media (Rogstad 2016). In any case, it seems plausible that the digitalisation process has posed a severe challenge to the conceptualisation of mediatisation, as the digital-based communication has provided more flexible forms of spreading content (Jensen 2013).

The increasingly complex landscape of technical-based communication processes makes the role of intermediation even more critical when elections approach. On one hand, the performative effects of the media on the social sphere gain interest during electoral campaigns (López-García *et al.* 2018). On the other hand, social networks are used mostly for non-political purposes and only during elections does the interest in politics increase these interactions (Nielsen and Vaccari 2013). One way to assess mediatisation is by analysing the political issues dealt with by the media, as long as the centrality of the media in the political debate is not questioned (Mancini and Mazzoleni 1995, Castells 2009). According to Mazzoleni (2004), the issue typology proposed by Patterson (1980) suffices to study the progressive mediatisation of political campaigns. It is assumed that mediated political discourse involves the elaboration of the form and content of the message by the interests of the medium (Meyen *et al.* 2014). This process creates a framework of thematic representation where certain considerations might be highlighted over others (Bartholomé *et al.* 2018). This is why Patterson's typology allows a precise distinction to be established between the topics of political interest and their representation frames, as it separates the issues of politics and the conflict inherent in the news representation (Martínez Nicolás *et al.* 2014, Bartholomé *et al.* 2018).

Patterson's typology proposes four categories: *political issues* correspond to issues concerning the more abstract issues of electoral confrontation, such as parties' ideologies, political alignment and relationships with groups governing at that time; *policy issues* include sector policy issues that affect the lives and interests of citizens, such as the health and education systems and the economy; *campaign issues* cover aspects related to the strategies and organisation of electoral campaigns, and *personal issues* concern the candidates' lives and activities. Several studies have shown how the growing popularity of *campaign issues* during electoral campaigns detracts from an interest in issues that are more specific to political interests, such as *political issues* and *policy issues* (Grossi *et al.* 1985, Patterson 1993, Mancini and Mazzoleni 1995).

Few studies have applied Patterson's analytical categories to Twitter. López-García (2016) used this typology to evaluate tweets posted by the leading candidates during the 2015 Spanish electoral campaign. An extended typology was used to assess the level of presence of policy issues during the electoral debates (López-García *et al.* 2018). However, an analysis of Twitter's global conversation based on these categories remains unexplored. The purpose of this research is to address this gap. Through this typology, we might distinguish between the issues expected to be aligned with citizens' interests (*political* and *policy)*, from those supposed to be more promoted by media's interests (*campaign* and *personal*). This would allow us to assess the mediatisation degree on Twitter. On this ground, the following research question is proposed:

RQ1: What kind of issues (*political*, *policy*, *campaign* or *personal*) predominates on Twitter during the electoral pre-campaign and campaign?

Message dissemination during electoral campaigns through Twitter

An increasing number of academic papers have addressed the role of Twitter during electoral campaigns. For such an aim, computer-assisted tools have been developed to collect and analyse large-scale data (Jungherr 2016, Campos-Domínguez 2017). Computational social science has researched the central topics present in the Twittersphere conversation (Kruikemeier 2014, Neuman *et al.* 2014, Rill *et al.* 2014, Xu *et al.* 2014, Antonakaki *et al.* 2017, Stier *et al.* 2018). The phenomenon of the *second screening* has attracted the attention of scholars specialised in electoral campaigns and social media (Gil de Zúñiga *et al.* 2015). Through this practice, the user shares their comments and opinions with other users, while watching a TV programme. Twitter is a tool particularly apt at broadening the experience of watching the political talk shows (Giglietto and Selva 2014, Ceron and Splendore 2018, Baviera *et al.* 2019) and the candidate debates (D'heer and Verdegem 2014, Freelon and Karpf 2015, Tremayne and Minooie 2015, Vaccari *et al.* 2015, Shah *et al.* 2016, Vergeer and Franses 2016).

Some investigations shed light on citizen support and mobilisation towards popular political accounts, such as voter-candidate and party engagement (Vaccari and Nielsen 2013, Kratzke 2017, Yang and Kim 2017). In the same vein, researches have discussed the influence of online debate and participation in the election outcome. Notably, sentiment analysis techniques have contributed to this examination of the Twitter landscape (DiGrazia *et al.* 2013, Guerrero-Solé *et al.* 2014, Jacobs and Spierings 2015, Murthy 2015, Burnap *et al.* 2016, Jungherr *et al.* 2017, Grover *et al.* 2019). Other studies have also captured datasets to detect political bots, fake accounts and their influence in polarising the discourse and the online dissemination of disinformation (Murthy *et al.* 2016, Filer and Fredheim 2017, Shao *et al.* 2018, Vosoughi *et al.* 2018).

Klinger and Svenson (2015) proposed the concept of *mass media logic* as opposed to *network media logic*. According to the first kind of logic, the information is conveyed vertically by media organisations, and is closely related to the mediatisation process; whereas in the second kind of logic, content distribution is performed through the users who share the message with their follower network in the digital environment. Social networking sites are clearly affected by the network media logic. In our case, we have to look closely how this logic operates on Twitter in disseminating messages.

There are four ways of interacting with the content in Twitter: liking, replying, retweeting and quoting a tweet. These interactions have a direct effect on content dissemination. By liking, the user expresses approval of the content; by replying, the user adds a comment to a tweet's thread; by retweeting, the user incorporates the tweet to her/his timeline so that her/his followers will read it; and by quoting, the user adds a comment to a retweeted tweet so that the new text is read before the original tweet. Each interaction pushes the original content through the network structure in a different way. The first one enables the system to show the liked tweet to the user's followers, but it does so randomly. The added content through the second and the fourth interaction can reinforce the original message, but can also change it by criticising it or by using irony.

Another way of assessing the mediatisation on Twitter would be to evaluate what effect the tweet's *issue* has on this network dissemination. For this aim, the most appropriate interaction would be retweeting, as the user contributes to the overall diffusion of the original tweet by sharing it with her/his followers. This interaction has been the research objective of several studies. Some of them pointed out that journalists

and news media elites drive the Twitter conversation during the televised candidate debates thanks to a more significant proportions of retweets than non-elite users (Hawthorne *et al.* 2013, Tremayne and Minooie 2015). Dang-Xuan et al. (2013) showed the influence of the presence of hashtags and links in the tweet, as well as of the number of the user's followers, on the number of retweets. In a smaller sample, they manually coded the subject of the tweet issue, and found dependency relationships in only some of the issues analysed.

The present research aims to better clarify the relationship between tweet content and dissemination by retweeting, in the context of the mediatisation. According to Patterson's framework, we wonder whether the original tweet's issue might influence its being retweeted. In this way, we might know what kind of issue most shapes the message dissemination through the network structure conformed by the Twitter users, as equivalent to the issue prevalence imposed by media vertically. Consequently, the following research question is proposed:

RQ2: How does the kind of issues (*political*, *policy*, *campaign* or *personal*) of a tweet influence the volume of its retweets during the electoral pre-campaign and campaign?

Materials and Methods

Background to the 2015 Spanish General Election

The chosen scenario for answering these RQs is the 2015 Spanish General Election. The two traditional parties, PP (conservative) and PSOE (socialist) had been decreasing in popularity as the General Election approached. The PP was hampered by corruption cases that predicted the loss of the absolute majority obtained in the previous elections. Its candidate, Mariano Rajoy, based a good part of the party's electoral strategy on the economic results obtained during its legislature. The PSOE's candidate was its secretary general, Pedro Sánchez, whom party militants had elected in June 2014. Opposing them was a liberal party of Catalan origins, Ciudadanos, and another recently created party, the progressive Podemos. Their respective candidates, Albert Rivera and Pablo Iglesias, exercised strong leadership and channelled much of the desire for political regeneration (Dader and Campos-Domínguez 2017, López-García and Valera-Ordaz 2017).

More than on other occasions, these two new candidates needed to attract the attention of the electorate, and to do this, television exposure was critical. A novelty in this pre-campaign was the participation of candidates in infotainment programmes (López-Rico and Peris Blanes 2017). During the campaign, the framing of strategy and competition dominated newspaper headlines and the information provided by the parties (Palau-Sampio *et al.* 2017). Moreover, the intense mediatisation of politics in these elections ran parallel to a considerable activity in social networks (Campos-Domínguez and Calvo 2017, Fenoll and Cano-Orón 2017). This intensification cannot be understood without considering the impact that the 15-M movement had on Spanish society in 2011, with a strong political disaffection unfolding amongst citizens (Díaz-Parra and Jover-Báez 2016). Not only were there numerous public demonstrations throughout the country, but social networks showed intense political debates and calls to action (Sampedro and Lobera 2014).

The political campaign for the 2015 Spanish General Election was framed around three main points: the general awareness that the two-party system paradigm was no longer relevant (Orriols and Cordero 2016), the need to capitalise media attention by the new candidates, and the vivid digital conversation by particularly active users. These features seemed to be appropriate to assess the mediatisation effect on digital media, and particularly on Twitter.

Data Collection

The General Election took place on December 20th, 2015. Three periods of analysis were established: pre-campaign, campaign and election day. They were chosen to better highlight the dynamics of the conversation on Twitter.

The pre-campaign period start was set one month before the electoral campaign began. The day of reflection was included as part of the campaign period because its dynamics were very different from the election day. Given that the intense conversation during the electoral night lasted until the early hours of the next morning, the day following the ballot was considered part of the election day itself. Thus, the analysis periods were as follows: *pre-campaign*, from November 2nd to December 3rd (32 days); *campaign*, from December 4th to December 19th (16 days), and *election day*, December 20th and 21st (2 days). The three periods totalled 50 days.

The tweets were obtained through the Twitter Streaming API using Python. We captured tweets that contained at least one of these terms: “#20D”, “20-D” (terms related to the election event); “Rajoy”, “@marianorajoy”, “Pedro Sanchez”, “Pedro Sánchez”, “@sanchezcastejon”, “Pablo Iglesias”, “@Pablo_Iglesias_”, “Rivera”, “@Albert_Rivera” (terms related to the four presidency candidates); “ PP ” (with blank spaces), “@PPopular”, “PSOE”, “@ahorapodemos”, “Ciudadanos”, “C’s”, “@CiudadanosCs” (terms related to the four main parties).

The term “Podemos” was not possible to be included as a searching term because it has a vast meaning in Spanish, and there was a risk of including tweets that were not related to the political content being studied. The inclusion of general terms associated with the elections was intended to address this inconvenience partially. An additional criterion was that the tweet had to be written in Spanish, given that the content was to be analysed. This supposes a limitation in our study, as several autonomous languages coexist in Spain. Lastly, as there are Spanish celebrities named “Rivera”, we performed a cleaning process to eliminate the tweets that mentioned these famous people. The final volume of the extracted corpus amounted to 15,806,057 tweets.

Computer-assisted classification according to Patterson's issues

Mazzoleni (2004) warned of the difficulty of using Patterson's typology (1980) to classify specific texts as *political* or *policy issues*, as there may be times when the content include both. In addition to this methodological difficulty, there are also those problems inherent to the computer-assisted detection of issues in social networks. Aspects such as a lack of context, the language informality, and the short length of the texts on Twitter impede the complete classification of issues (Hu and Liu 2012). Various procedures have been used to classify issues in political communication, such as those based on hashtags (Conover *et al.* 2011, Rill *et al.* 2014) and semantic network analysis (Lee *et al.* 2011), the most popular

being dictionary-based methods (Tumasjan *et al.* 2010, Conway *et al.* 2015, Guo and Vargo 2015, Vergeer and Franses 2016). However, Patterson's categorisation could not be carried out according to the mere presence of specific terms. Dictionary-based methods overlook issues in which exclusive, discriminating words are hard to find, as it would be the case of the more abstract issues or those related to candidates. We thought this problem could be appropriately addressed with classification algorithms based on supervised machine learning (Batrinca and Treleaven 2015).

The development of such a classification tool consisted of three phases. First, the necessary training and test datasets for supervised machine learning were manually coded. A three-coder team (two of the authors and a third researcher experienced in content analysis) conducted one pilot test with a 600-tweet corpus. The intercoder reliability was very low (Krippendorf's alpha = 0.541). The main problem was the difficulty mentioned earlier in classifying the tweets according to the Patterson typology. The presence of ironic expressions also posed difficulties in interpretation. This situation led to defining the codebook more accurately, and clarifying the criteria among the coders. After this, a subsample of 4,000 tweets was randomly extracted and coded separately by each of the three coders. The Krippendorf's alpha obtained was 0.7957. This value is very close to 0.8, considered the limit for acceptance (Krippendorf 2013).

Giving the inherent difficulty of the typology, we thought we should use a training set as clear as possible. This circumstance would favour the following phases of the process, where there would be several models operating on the dataset. In consequence, we decided to discard any tweet that all three coders did not agree on. In this way, the algorithm would be trained with a more accurate dataset. Thus, the resulting coded dataset amounted to 3,116 tweets. The distribution by issue was as follows: *political issues*, 738 (23.68%); *policy issues*, 1,102 (35.36%); *campaign issues*, 718 (23.04%); *personal issues*, 199 (6.38%), and *other issues*, 359 (11.52%).

The second phase comprised arranging an *evaluation task* where a variety of Spanish NLP research centres could develop classification systems based on the manually annotated dataset (Authors et al. 2017). Two tables were provided, extracted from the annotated corpus: one for training (80%) and another for testing (20%).

The result of each presented model was evaluated in the following manner: For each category, *precision* and *recall* were calculated. In this case, *precision* corresponds to the number of tweets correctly classified according to a specific category, divided by the number of tweets that the system classified according to that category. Meanwhile, *recall* divides the number of correctly classified tweets according to a category, by the number of real tweets in that category. *Precision* provides a measure of the accuracy of the classifier while *recall* assesses its completeness. Following that, the F_1 score was calculated for each category as the harmonic mean between *precision* and *recall*. Finally, the $F_{1\text{-macro}}$ served as a measure for the classification of each system, which was calculated as the average of the five F_1. The process can be seen in Equation 1.

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$
$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad (1)$$
$$F_1 = \frac{2 * precision * recall}{precision + recall}$$

$$F_{1-macro} = \frac{1}{L} * \sum_{l \in L} F_{1l},$$

$$where\ L\ represents\ the\ categories: political\ issues, policy\ issues, campaign\ issues, candidate\ issues\ and\ other\ issues$$

Seventeen teams presented a total of 39 models. There were two main difficulties for this competition task: firstly, *political issues* and *policy issues* shared common terms, and secondly the training sample was extremely unbalanced. Despite this, the results were acceptable. The maximum value of the $F_{1\text{-macro}}$ obtained was 64.82% (Authors et al., 2017).

The third phase consisted of labelling the 15.8 million-tweet corpus. To do this, we built a pooling system using five of the models presented. The best performing team submitted two models and other three teams submitted one model each. Table 1 describes the models used for the pooling system along with the baseline for the shared task (Authors et al., 2017).

Insert Table 1 about here

Since the levels of reliability of each system separately were moderate, we decided that valid labels would only be those that were agreed on by four or more systems (80% agreement). Tweets not meeting this condition were discarded in order to prioritise reliability. The final corpus was made up of 10,023,870 tweets, published by 986,565 different users. This sample represented 63.42% of the extracted corpus. Figure 1 shows the process for labelling the tweet corpus.

Insert Figure 1 about here

Variables for the regression analysis on retweets

RQ2 was evaluated using a multivariate regression analysis. The unit of analysis was the original tweets contained in the corpus, grouped into each of the three periods. The independent variables were divided into two blocks: the tweet's basic aspects and the *issue* type.

Number of retweets (n_RT). To obtain the *n_RT* value of a tweet, we looked for the highest value of retweets for each original tweet in the corpus. Table 2 shows the mean and the standard deviation in each of the three periods. Of interest were the maximum registered values as they were considerably far from the mean. This is due to the significant number of tweets that were not retweeted. The maximum value and the mean retweets for the pre-campaign period were 10,757 and 2.11; for the campaign they were 8,115 and 2.26 and for the day of the elections they were 4,011 and 1.72. For this dispersion of values, the logarithm of *n_RT* was taken as the dependent variable for the regression analysis (Dang-Xuan *et al.* 2013).

Basic features of each tweet. Two variables relative to the publication moment were taken: *Posting Hour*, which indicates the time the tweet was posted in 24-hour format, and *Days remaining until the election*. This variable was removed for the third period. Additionally, the number of *followers* the author had at the time the tweet was posted was taken. For the same reasons as with *n_RT* , the logarithm of *n_followers* was taken as the independent variable. Finally, the presence of formal elements in the tweet may somewhat influence whether the tweet is retweeted (Dang-Xuan *et al.* 2013). In this case, the presence of *images*, *hashtags* and *urls* in the tweet was evaluated as *dummy* variables.

Issue variables. The nominal variable of the tweet *issue* was divided into four fictitious *dummy* variables using the *Other issues* category as the reference.

Insert Table 2 about here

Results

Table 3 shows the basics statistics of the issues for each period, while Figure 2 shows its daily evolution. In the pre-campaign period, *policy issues* are predominant (57,228.78 on average daily). This category is followed by *campaign issues* (33,705.19), *political issues* (23,029.22) and *personal issues* (13,080.91). Figure 2 confirms this order of *issues* in the pre-campaign for almost every day. The only exception is November 26th and 27th, where *personal issues* notably stand out and surpass *political issues*. Three prominent days in the pre-campaign period were identified. The first one was November 14th. On the day before there were several terrorist attacks in Paris. The most relevant day was November 30th, when the newspaper *El País* hosted a digital debate among Pedro Sánchez, Pablo Iglesias and Albert Rivera. Figure 2 shows how all *issues* increased on that day, except for *personal issues*. There was another peak on December 2nd. On that day Mariano Rajoy attended a well-known TV show as the only protagonist. In general, *policy issues* were more abundant in the pre-campaign, and dominated the conversation on Twitter until approximately one week prior to the start of the campaign. From then onwards, *campaign issues* attracted more user attention, but did not exceed *policy issues*.

Insert Table 3 about here

Policy issues and *campaign issues* volumes during the electoral campaign were very similar, as Table 3 shows. On this occasion, the highest daily average corresponds to *campaign issues* (108,345.25), and with 5% less, *policy issues* (103,588.81). Lagging quite a bit behind were *policy issues* (52,573.88) and *personal issues* (19,438.38). The standard deviation of *campaign issues* (49,795.15) is much higher than that of *policy issues* (25,536.80). This critical difference is confirmed by Figure 2: *policy issues* predominate during most of the campaign, except for two peaks in which *campaign issues* strongly attracted almost all users' attention. These peaks happened on December 8th and 14th. Both dates are associated with the two televised candidate debates. The first took place on December 7th, and it brought together three of the leading candidates: Pedro Sánchez,

Pablo Iglesias and Albert Rivera. Mariano Rajoy declined to attend and so the Vice President of the Government, Soraya Sáenz de Santamaría, represented him. The conversation on Twitter about this debate extended along the night, and this explains the one-day delay. The second debate consisted of a face-to-face debate between the leaders of the two traditional parties: Mariano Rajoy and Pedro Sánchez. It was held on December 14th.

It should be noted that a third relevant peak occurred on December 16th and 17th: *personal issues* grew disproportionately, to the extent it matched *campaign issues*. On the 16th an unfortunate incident occurred: a young man physically attacked Mariano Rajoy. This resulted in an increase in *personal issues* that polarised the Twittersphere on the day it occurred and the day after.

Insert Figure 2 about here

However, the conversation on the election day did not follow the *policy issues* trend. The dominant issue during this third period was *political issues* (260,036.50 tweets on average during these two days), more than doubling the next issue type, *campaign issues* (114,875.50), and these in turn almost doubling the third category, *policy issues* (62,321). *Personal issues* attracted very little attention (9,053). Figure 2 reflects this pattern: the trend of *political issues* increases between December 20th and 21st, while *campaign issues* decrease from one day to the next.

The regression analysis for the three periods is shown in Table 4. All coefficients were significant, except for the variable *Days until the election* during the pre-campaign period. Since both samples were extensive, we expected the coefficients to be very significant. The only negative coefficients were those corresponding to *Presence of urls*: the tendency to retweet was more significant if the original tweet had no link.

Insert Table 4 about here

The importance of the standardised coefficients follows the same order throughout the three periods: *policy issues, campaign issues, political issues* and *personal issues.* During the pre-campaign and campaign, the difference between the coefficients of *policy issues* and of *campaign issues* is double. This difference reduces noticeably on election night: *policy issues* have a standardised coefficient of 0.084, *campaign issues* of 0.070 and *policy issues* of 0.063. The standardised coefficients of *personal issues* are extremely low for all three periods.

Threats on validity

This study has some limitations. The tweet extraction process along with the classification methods raised critical issues for the internal and external validity of the results.

Internal validity refers to whether the interpretation of the results is correct. The first difficulty lies in the corpus origin. Twitter Streaming API does not guarantee all the

tweets matching the searching words will be provided, especially when several are operating simultaneously (Felt 2016). Once the extraction was finished, we obtained a significant volume of tweets, enough to consider the volume appropriate for our research goals. The second difficulty was associated with the issue typology for classifying the tweets, as it proposes two items which can sometimes be hard to distinguish (Mazzoleni 2004). For this reason, the coder team had to be experienced both in the political field and in content analysis methodology. We tackled this threat by taking special care in the pilot test.

The tweet sample for the computer-assisted stage had to be as large and accurate as possible. The balance between these two features was achieved by deciding to code a large number of randomly selected tweets, and to retain *only* those tweets with full agreement. In this way, the algorithms could be trained with a better set. The essential threat behind this corpus was the unbalanced distribution among the categories.

As previously explained, the reasoning for avoiding dictionary-based methods at the computer-assisted stage was due to the lack of context in the tweets. This choice motivated us to develop a classification system, albeit with validity threats. First, we performed a tweet classification with a machine-learning system, adapted to our coded sample, and the results were actually poor: 46% accuracy. We countered this difficulty by inviting other research groups who were able to obtain a better performance with the dataset. Although the accuracy was improved, the F_1 indices obtained were only moderately acceptable for research. At this point, we decided to pool the systems with good performance and whose team agreed in collaborating with the project. The criterion of retaining tweets with 80% agreement among the systems (4 out of 5) assured a more reliable classification, but at the cost of reducing the internal validity of the results. We could only assess the issue in 63.42% of the collected tweets.

External validity refers to whether the findings are possible to generalise. This research focuses on a particular electoral campaign and as such limits the extent of our conclusions. In any case, the results could help understand the behaviour on Twitter in a situation similar to the one we have studied; an electoral campaign in a multi-party scenario.

Due to all these limitations, the conclusions of the present study should be taken with caution. Their contribution can only be taken as exploratory in the complex field of digital political communication, where there is a growing need for reliable measurement tools.

Discussion

According to the issue typology proposed by Patterson (1980), the present study shows certain evidence regarding the primacy of *policy issues* on Twitter during the pre-campaign and the electoral campaign for the 2015 Spanish General Election. On election night, the issues that were most dominant in the sample were those corresponding to *political issues*, i.e. issues more closely associated with ideologies. This evidence is consistent with the ballot results. The seat distribution among the four main parties was the following: PP, 123; PSOE, 90; Podemos, 42, and Ciudadanos, 40. No party reached a parliamentary majority, 176 seats. The new formed parliament favoured discussions that had greater ideological depth and focused on possible coalitions. What is relevant about this result is the enormous increase of these issues compared to the previous days.

Figure 2 served to confirm one of the characteristics indicated by Jungherr (2015) in his study of political conversation on Twitter: the detection of peaks in activity. Of particular relevance is the debate that took place on December 7th, which was promoted as the *deciding debate*. The sample registers that the main volume of tweets corresponded to *campaign issues* on that moment. This finding confirms TV's influence on the Twitter political conversation. Something similar happened in the following televised debate between the leaders of the two established parties: *campaign issues* once again surpassed *policy issues*, albeit less so this time. Conversely, it is worth highlighting the little weight that *personal issues* have had in the Twitter conversation throughout the pre-campaign and the campaign. This aspect, which could be associated with a more personalised policy because of the prominence given to candidates, is greatly reduced in the Twitter conversation. In response to RQ1, *policy issues* were more prevalent than any other issue during the pre-campaign and campaign. However, it should be noted that *campaign issues* were dominant during the televised candidate debates. This effect could be an indication of mediatisation in Twitter, because it is the moment at which TV most grabs the attention of the electoral campaign.

The effect of the topic in the dissemination through retweets confirms this view. Results from the regression analysis showed that the *issue* that had the highest standardised coefficient was *policy*, which evidenced that this type of issue is more likely to be retweeted than tweets containing *campaign* or *political issues*. Thus, users disseminate more tweets with strong political content than those who might perceive politics as a competition. Table 4 helps to interpret better what happened on election night. On those two days, the *political issues* coefficient is still less than *policy* and *campaign issues* coefficients. This is surprising considering the remarkable increase in the volume posted on the electoral night for each type of issue. These data suggest that there were probably few retweeted tweets containing *political issues*, since other issues had a more significant effect on the retweet activity.

The conclusions of this work are in line with other studies on Twitter users' profiles. In his research on Twitter users' ideological position, Barberá (2015) concluded that the political conversation on this social network is dominated by a small portion of users with strong political ideals. Based on a large corpus extracted for the 2017 German federal election campaign, Kratzke (2017) showed that right-wing and populist parties seem to have more active followers, so that the overall perception for those parties might be interpreted as having a "louder" voice in Twitter. In the same vein, the study by Vaccari et al. (2015) regarding *second screeners* during televised debates, pointed out these users' high level of online and offline political commitment.

If we assume that the Patterson's typology might be an indicator for mediatisation, these findings suggest that the topics discussed on Twitter during electoral campaigns are predominantly linked with issues that concern the electorate, such as *policy issues*, whereas the topics hypothetically interestedly promoted by the media have a secondary role in this online conversation. This might be interpreted as a sign of lack of mediatisation. Nevertheless, we should bear in mind the limitations of our results and, most importantly, that Patterson typology was conceived for content analysis published by media. Certainly, the intermediation operated by Twitter users, through the content generation and the message dissemination, shows a prevalence of issues different to those expected in a mediatized process. These findings do not imply that Twitter cannot conform to the intermediation process in an interested way. The concern about the automated activity on Twitter is increasing in the academic field (Filer and Fredheim

2017, Stella *et al.* 2018). This practice could be interpreted as a way of a performative influence, in the sense of prioritising interested messages artificially. Future research on mediatisation in the digital environment should try to identify operative bot networks to gauge the kind of messages spread by them. In any case, this would require a more accurate issue typology.

Conclusions

Political communication is shaped by the specific mechanisms of the media that operate in the public sphere. The mass media are essential in disseminating information about political news. However, they run the risk of sensationalising and personalising politics, driven by the need to attract viewer numbers (Mazzoleni 2004). An instrument particularly useful in evaluating this effect during electoral campaigns has been the typology proposed by Patterson (1980) as it allows issues that are more aligned with discussion of real policies to be distinguished from others that might frame the electoral context in an interesting way.

The present study applied this typology to the political conversation on Twitter during the 2015 Spanish General Election in order to assess the degree of mediatisation on this social network. The *issues* with the most political content (ideological and sector policy issues) predominate in conversations among users more so than those containing issues about the campaign and the candidates. This conclusion seems reasonable, considering that it is the citizens themselves who intervene in this digital public sphere, and will thus discuss what concerns them most directly. Another relevant finding has been the impact of the televised candidate debates on the topics discussed online. *Campaign issues* were very prevalent during these events. This provides evidence for the key role played by these debates in electoral campaigns and, most importantly for our research, it shows a possible indirect effect of the mediatisation on Twitter, as the conversation is closely linked to TV content and it fosters topics not strictly associated with sectorial policies. Finally, the Twitter conversation on election night was, by far, led by *political issues*. It could be considered a signal of a vibrant online political discussion, fed no doubt by the divided results that were obtained. The relevant finding is the overwhelming dominance of these issues during those final days, whilst in the remainder of the period studied they profiled low.

Our findings must be taken with a degree of caution. Despite the difficulty in assessing Patterson's typology for a large set of tweets, the results provide some evidence in most of the corpus and point out some trends in the Twitter political conversation. At least they do not contradict the hypothesis of Twitter being a democratic medium where parties, media and citizens communicate in a non-hierarchical structure. However, this does not mean that this medium is exempt of interests that might shape the network intermediation process. There are other mechanisms to influence the discussion, different to those used by the media taking advantage of their vertical position, as is the case for automatic activity driven by interested motivations. For our part, with this research study we have tried to contribute in order to clarifying the impact of the mediatisation process in an increasingly complicated context of media hybridisation.

References

Antonakaki, D., Spiliotopoulos, D., Samaras, C. V., Pratikakis, P., Ioannidis, S., and

Fragopoulou, P., 2017. Social media analysis during political turbulence. *PLoS ONE*, 12 (10), e0186836.

Barberá, P., 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political Analysis*, 23 (1), 76–91.

Bartholomé, G., Lecheler, S., and de Vreese, C., 2018. Towards A Typology of Conflict Frames. *Journalism Studies*, 19 (12), 1689–1711.

Batrinca, B. and Treleaven, P.C., 2015. Social media analytics: a survey of techniques, tools and platforms. *AI and Society*, 30 (1), 89–116.

Baviera, T., Peris, A., and Cano-Orón, L., 2019. Political candidates in infotainment programmes and their emotional effects on Twitter: An analysis of the 2015 Spanish general elections pre-campaign season. *Contemporary Social Science*, 14 (1), 144–156.

Blumler, J.G. and Kavanagh, D., 1999. The third age of political communication: Influences and features. *Political Communication*, 16 (3), 209–230.

Bor, S.E., 2014. Using Social Network Sites to Improve Communication Between Political Campaigns and Citizens in the 2012 Election. *American Behavioral Scientist*, 58 (9), 1195–1213.

Brants, K. and Neijens, P., 1998. The Infotainment of Politics. *Political Communication*, 15 (2), 149–164.

Burnap, P., Gibson, R., Sloan, L., Southern, R., and Williams, M., 2016. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies*, 41, 230–233.

Campos-Domínguez, E., 2017. Twitter and political communication. *El Profesional de la Información*, 26 (5), 785–794.

Campos-Domínguez, E. and Calvo, D., 2017. La campaña electoral en Internet: planificación, repercusión y viralización en Twitter durante las elecciones españolas de 2015. *Comunicación y Sociedad*, 29, 93–116.

Castells, M., 2009. *Communication Power*. New York, NY: Oxford University Press.

Ceron, A. and Splendore, S., 2018. From contents to comments: Social TV and perceived pluralism in political talk shows. *New Media & Society*, 20 (2), 659–675.

Chadwick, A., 2013. *The Hybrid Media System: Politics and Power*. Oxford: Oxford

University Press.

Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., and Menczer, F., 2011. Political polarization on twitter. *In*: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*. 89–96.

Conway, B.A., Kenski, K., and Wang, D., 2015. The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary. *Journal of Computer-Mediated Communication*, 20, 363–380.

Couldry, N. and Hepp, A., 2013. Conceptualizing Mediatization: Contexts, Traditions, Arguments. *Communication Theory*, 23 (3), 191–202.

D'heer, E. and Verdegem, P., 2014. Conversations about the elections on Twitter: Towards a structural understanding of Twitter's relation with the political and the media field. *European Journal of Communication*, 29 (6), 720–734.

Dader, J.L. and Campos-Domínguez, E., eds., 2017. *La Búsqueda Digital del Voto. Cibercampañas Electorales en España 2015-16*. Valencia: Tirant lo Blanch.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., and Neuberger, C., 2013. An Investigation of Influentials and the Role of Sentiment in Political Communication on Twitter During Election Periods. *Information, Communication & Society*, 16 (5), 795–825.

Delli Caprini, M.X. and Williams, B.A., 2001. Let Us Infotain You: Politics in the New Media Age. *In*: W.L. Bennet and R.M. Entman, eds. *Mediated politics: Communication in the future of democracy*. New York: Cambridge University Press, 160–181.

Díaz-Parra, I. and Jover-Báez, J., 2016. Social movements in crisis? From the 15-M movement to the electoral shift in Spain. *The International Journal of Sociology and Social Policy*, 36 (9/10), 680–694.

DiGrazia, J., McKelvey, K., Bollen, J., and Rojas, F., 2013. More Tweets, More Votes: Social Media as a Quantitative Indicator of Political Behavior. *PLOS ONE*, 8 (11), e79449.

Felt, M., 2016. Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3 (1), 1–15.

Fenoll, V. and Cano-Orón, L., 2017. Citizen engagement on Spanish political parties' Facebook pages: Analysis of the 2015 electoral campaign comments. *Comunicación y Sociedad*, 30 (4), 131–145.

Filer, T. and Fredheim, R., 2017. Popular with the Robots: Accusation and Automation in

the Argentine Presidential Elections, 2015. *International Journal of Politics, Culture and Society*, 30 (3), 259–274.

Finnemann, N.O., 2014. Digitization. *In*: K. Lundby, ed. *Mediatization of Communication*. Berlin, Germany: Mouton de Gruyter, 297–321.

Freelon, D. and Karpf, D., 2015. Of Big Birds and bayonets: Hybrid Twitter interactivity in the 2012 presidential debates. *Information, Communication & Society*, 18 (4), 390–406.

Giglietto, F. and Selva, D., 2014. Second Screen and Participation: A Content Analysis on a Full Season Dataset of Tweets. *Journal of Communication*, 64 (2), 260–277.

Gil de Zúñiga, H., Garcia-Perdomo, V., and McGregor, S.C., 2015. What Is Second Screening? Exploring Motivations of Second Screen Use and Its Effect on Online Political Participation. *Journal of Communication*, 65 (5), 793–815.

Grossi, G., Mancini, P., and Mazzoleni, G., 1985. *Giugno 1983: Una campagna elettorale*. Turín: ERI.

Grover, P., Kar, A.K., Dwivedi, Y.K., and Janssen, M., 2019. Polarization and acculturation in US Election 2016 outcomes – Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change*, 145, 438–460.

Guerrero-Solé, F., Corominas-Murtra, B., and Lopez-Gonzalez, H., 2014. Pacts with Twitter. Predicting voters' indecision and preferences for coalitions in multiparty systems. *Information, Communication & Society*, 17 (10), 1280–1297.

Guo, L. and Vargo, C.J., 2015. The Power of Message Networks: A Big-Data Analysis of the Network Agenda Setting Model and Issue Ownership. *Mass Communication and Society*, 18 (5), 557–576.

Hawthorne, J., Houston, J.B., and McKinney, M.S., 2013. Live-Tweeting a Presidential Primary Debate: Exploring New Political Conversations. *Social Science Computer Review*, 31 (5), 552–562.

Hu, X. and Liu, H., 2012. Text Analytics in Social Media. *In*: C.C. Aggarwal and C. Zhai, eds. *Mining Text Data*. New York: Springer-Verlag, 385–414.

Jacobs, K. and Spierings, N., 2015. Saturation or maturation? The diffusion of Twitter and its impact on preference voting in the Dutch general elections of 2010 and 2012. *Journal of Information Technology & Politics*, 13 (1), 1–21.

Jensen, K.B., 2013. Definitive and sensitizing conceptualizations of mediatization. *Communication Theory*, 23 (3), 203–222.

Jungherr, A., 2014. The Logic of Political Coverage on Twitter: Temporal Dynamics and Content. *Journal of Communication*, 64 (2), 239–259.

Jungherr, A., 2015. *Analyzing Political Communication with Digital Trace Data: The Role of Twitter Messages in Social Science Research*. Cham: Springer International Publishing Switzerland.

Jungherr, A., 2016. Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13 (1), 72–91.

Jungherr, A., Schoen, H., Posegga, O., and Jürgens, P., 2017. Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support. *Social Science Computer Review*, 35 (3), 336–356.

Kalsnes, B., Krumsvik, A.H., and Storsul, T., 2014. Social media as a political backchannel. *Aslib Journal of Information Management*, 66 (3), 313–328.

Klinger, U. and Svensson, J., 2015. The emergence of network media logic in political communication: A theoretical approach. *New Media & Society*, 17 (8), 1241–1257.

Kratzke, N., 2017. The #BTW17 Twitter Dataset–Recorded Tweets of the Federal Election Campaigns of 2017 for the 19th German Bundestag. *Data*, 2 (34), 19.

Kreiss, D., 2012. *Taking Our Country Back: The Crafting of Networked Politics from Howard Dean to Barack Obama*. New York: Oxford University Press.

Krippendorf, K., 2013. *Content Analysis: An Introduction to its Methodology*. 3rd ed. Thousand Oaks, CA: Sage Publications.

Kruikemeier, S., 2014. How political candidates use Twitter and the impact on votes. *Computers in Human Behavior*, 34, 131–139.

Lee, K., Palsetia, D., Narayanan, R., Patwary, M.M.A., Agrawal, A., and Choudhary, A., 2011. Twitter Trending Topic Classification. *In*: *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*. Washington, DC, USA: IEEE Computer Society, 251–258.

López-García, G., 2016. 'New' vs 'old' leaderships: the campaign of Spanish general elections 2015 on Twitter. *Communication & Society*, 29 (3), 149–167.

López-García, G., Llorca-Abad, G., Valera-Ordaz, L., and Peris-Blanes, À., 2018. Los debates

electorales, ¿el último reducto frente la mediatización? Un estudio de caso de las elecciones generales españolas de 2015. *Palabra Clave*, 21 (3), 772–797.

López-García, G. and Valera-Ordaz, L., eds., 2017. *Pantallas Electorales: El discurso de partidos, medios y ciudadanos en la campaña de 2015*. Barcelona: UOC.

López-Rico, C. and Peris Blanes, À., 2017. Agenda e Imagen de los Candidatos en las Elecciones Generales de 2015 en España en Programas Televisivos de Infoentretenimiento. *El Profesional de la Información*, 26 (4), 611–620.

Mancini, P. and Mazzoleni, G., 1995. Verso campagne sempre più mediatizzate. *In*: P. Mancini and G. Mazzoleni, eds. *I media scendono in campo. Le elezioni politiche 1994 in televisione*. Turín: ERI, 9–67.

Martínez Nicolás, M., Humanes, M.L., and Saperas Lapiedra, E., 2014. La mediatización de la política en el periodismo español. Análisis longitudinal de la información política en la prensa de referencia (1980-2010). *Trípodos*, 34, 41–59.

Mazzoleni, G., 2004. *La comunicazione politica*. 2nd ed. Bologna: Il Mulino.

Mazzoleni, G. and Schulz, W., 1999. 'Mediatization' of Politics: A Challenge for Democracy? *Political Communication*, 16 (3), 247–261.

Meyen, M., Thieroff, M., and Strenger, S., 2014. Mass Media Logic and The Mediatization of Politics. *Journalism Studies*, 15 (3), 271–288.

Murthy, D., 2015. Twitter and elections: are tweets, predictive, reactive, or a form of buzz? *Information, Communication & Society*, 18 (7), 816–831.

Murthy, D., Powell, A.B., Tinati, R., Anstead, N., Carr, L., Halford, S.J., and Weal, M., 2016. Bots and political influence: A sociotechnical investigation of social network capital. *International Journal of Communication*, 10, 4952–4971.

Neuman, W.R., Guggenheim, L., Jang, S.M., and Bae, S.Y., 2014. The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication*, 64 (2), 193–214.

Nielsen, R.K. and Vaccari, C., 2013. Do People "Like" Politicians on Facebook? Not really. Large-Scale Direct Candidate-to-Voter Online Communication as an Outlier Phenomenon. *International Journal of Communication*, 7, 2333–2356.

Orriols, L. and Cordero, G., 2016. The Breakdown of the Spanish Two-Party System: The Upsurge of Podemos and Ciudadanos in the 2015 General Election. *South European*

Society and Politics, 21 (4), 469–492.

Palau-Sampio, D., Carratalá, A., and Valera-Ordaz, L., 2017. Mediatización y encuadres de campaña. Análisis comparado de la información de partidos y medios en las elecciones generales de 2015. *El Profesional de la Información*, 26 (4), 602–610.

Patterson, T.E., 1980. *The Mass Media Election: How Americans Choose Their President*. New York: Praeger Special Studies.

Patterson, T.E., 1993. *Out of Order*. New York: Alfred A. Knopf Incorporated.

Richards, B., 2010. News and the emotional public sphere. *In*: S. Allan, ed. *The Routledge Companion to News and Journalism*. New York, NY: Routledge, 301–311.

Rill, S., Reinel, D., Scheidt, J., and Zicari, R. V., 2014. PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. *Knowledge-Based Systems*, 69, 24–33.

Rogstad, I., 2016. Is Twitter just rehashing? Intermedia agenda setting between Twitter and mainstream media. *Journal of Information Technology & Politics*, 13 (2), 142–158.

Sampedro, V. and Lobera, J., 2014. The Spanish 15-M Movement: a consensual dissent? *Journal of Spanish Cultural Studies*, 15 (1–2), 61.

Shah, D. V., Hanna, A., Bucy, E.P., Lassen, D.S., Van Thomme, J., Bialik, K., Yang, J., and Pevehouse, J.C.W., 2016. Dual Screening During Presidential Debates. *American Behavioral Scientist*, 60 (14), 1816–1843.

Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.-C., Flammini, A., and Menczer, F., 2018. The spread of low-credibility content by social bots. *Nature communications*, 9, 4787.

Stella, M., Ferrara, E., and De Domenico, M., 2018. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115 (49), 201803470.

Stier, S., Bleier, A., Lietz, H., and Strohmaier, M., 2018. Election Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter. *Political Communication*, 35, 50–74.

Strömbäck, J., 2008. Four Phases of Mediatization: An Analysis of the Mediatization of Politics. *The International Journal of Press/Politics*, 13 (3), 228–246.

Strömbäck, J. and Esser, F., 2009. Shaping Politics: Mediatization and Media Interventionism. *In*: K. Lundby, ed. *Mediatization: Concept, Changes, Consequences*.

New York, NY: Peter Lang Publishing Group, 205–223.

Tremayne, M. and Minooie, M., 2015. Using Social Media to Analyze Candidate Performance During Televised Political Debates. *Electronic News*, 9 (3), 143–159.

Tumasjan, A., Sprenger, T.O., Sandner, P.G., and Welpe, I.M., 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *In*: *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*. 178–185.

Vaccari, C., Chadwick, A., and O'Loughlin, B., 2015. Dual Screening the Political: Media Events, Social Media, and Citizen Engagement. *Journal of Communication*, 65 (6), 1041–1061.

Vaccari, C. and Nielsen, R.K., 2013. What Drives Politicians' Online Popularity? An Analysis of the 2010 U.S. Midterm Elections. *Journal of Information Technology & Politics*, 10 (2), 208–222.

Vargo, C.J., Guo, L., McCombs, M.E., and Shaw, D.L., 2014. Network Issue Agendas on Twitter During the 2012 U.S. Presidential Election. *Journal of Communication*, 64, 296–316.

Vergeer, M. and Franses, P.H., 2016. Live audience responses to live televised election debates: time series analysis of issue salience and party salience on audience behavior. *Information, Communication & Society*, 19 (10), 1390–1410.

Vosoughi, S., Roy, D., and Aral, S., 2018. The spread of true and false news online. *Science*, 359 (6380), 1146–1151.

Williams, B.A. and Delli Caprini, M.X., 2011. *After Broadcast News: Media Regimes, Democracy, and the New Information Environment*. Cambridge: Cambridge University Press.

Wright, S., 2012. Politics as usual? Revolution, normalization and a new agenda for online deliberation. *New Media & Society*, 14 (2), 244–261.

Xu, W.W., Sang, Y., Blasiola, S., and Park, H.W., 2014. Predicting Opinion Leaders in Twitter Activism Networks: The Case of the Wisconsin Recall Election. *American Behavioral Scientist*, 58 (10), 1278–1293.

Yang, J. and Kim, Y.M., 2017. Equalization or normalization? Voter–candidate engagement on Twitter in the 2010 U.S. midterm elections. *Journal of Information Technology & Politics*, 14 (3), 232–247.

Table 1. Classification models used in the pooling system and the $F_{1\text{-macro}}$ obtained in the shared task to classify the manual-coded sample.

Team	Model description	$F_{1\text{-macro}}$
Team 1 (2017)	Multilayer Perceptron (MLP) with bag-of-words representation, and with the scaling of the loss function adjusted on part of the training set.	64.82%
Team 1 (2017)	A majority voting scheme among three models: an MLP trained with part of the training set, an MLP trained with all the training set, and an MLP with the scaling of the loss function adjusted on part of the training set.	64.00%
Team 2 (2017)	Neural model with a Global Average Pooling layer followed by two dense layers, with word and n-gram embedding representation.	61.57%
Team 3 (2017)	Support Vector Machine (SVM) with character n-grams, reference tokens and word embedding representation.	60.54%
Team 4 (2017)	Combination of four models: Logistic Regression, an SVM, Naive Bayes, and a K-Nearest Neighbours classifier, with Term frequency–inverse document frequency (Tf-idf) representation.	58.59%
Baseline (Authors et al, 2017)	Random Forest with Tf-idf representation.	42.36%

Table 2. Descriptive statistics of the dependent and independent variables for the regression analysis.

	Pre-campaign		Campaign		Election Day	
	Mean	St. Dev.	Mean	St. Dev	Mean	St. Dev.
Dependent variable						
n_RT	2.11	23.026	2.26	26.842	1.72	24.375
Log n_RT	0.1098	0.32196	0.1103	0.32477	0.0753	0.27267
Basic aspects						
Posting hour	14.33	6.340	14.28	6.655	14.24	7.018
Days until the election	30.45	9.445	7.92	4.446	-	-
Followers	10,408.5	131,518.4	10,099.3	125,580.8	13,773.9	164,804.5
Log followers	2.7005	0.91911	2.6603	0.92828	2,6874	0.91045
Presence of images	0.12	0.330	0.14	0.0348	0.10	0.306
Presence of hashtags	0.29	0.453	0.33	0.471	0.30	0.459
Presence of urls	0.51	0.500	0.46	0.499	0.38	0.485
Issues						
Political Issues	0.20	0.399	0.22	0.412	0.60	0.490
Policy Issues	0.34	0.475	0.26	0.439	0.09	0.290
Campaign Issues	0.24	0.430	0.35	0.476	0.21	0.408
Personal Issues	0.05	0.218	0.07	0.261	0.02	0.143
N	1,262,140		1,332,512		288,563	

Notes: The analysis units correspond to the original tweets in the corpus.

Table 3. Descriptive statistics of the issues.

	Pre-campaign			Campaign			Election Day		
	Total	Mean	St. Dev.	Total	Mean	St. Dev.	Total	Mean	St. Dev.
Political Issues	736,935	23,029.22	10,399.98	841,182	52,573.88	19,716.40	520,073	260,036.50	34,524.49
Policy Issues	1,831,321	57,228.78	15,493.39	1,657,421	103,588.81	25,536.80	124,642	62,321.00	4,350.12
Campaign Issues	1,078,566	33,705.19	17,364.45	1,733,524	108,345.25	49,795.15	229,751	114,875.50	61,253.12
Personal Issues	203,538	6,360.56	9,301.00	311,014	19,438.38	30,979.92	18,106	9,053.00	1,909.19
Other Issues	418,589	13,080.91	6,940.74	283,673	17,729.56	7,375.32	35,535	17,767.50	2,069.70
Tweets	4,268,949	133,404.66	43,688.69	4,826,814	301,675.88	73,994.26	928,107	464,053.50	26,357.41
N		32			16			2	

Notes: The analysis units are the days in the study period. All tweets are considered: the original tweets and retweets.

Table 4. Standardised coefficients of the log n_RT regression models.

	Pre-campaign	Campaign	Election Day
Basic aspects			
Posting hour	0.015**	0.005**	0.011**
Days until the election	0.000	0.005**	-
Log followers	0.328**	0.334**	0.345**
Presence of images	0.166**	0.172**	0.124**
Presence of hashtags	0.077**	0.087**	0.033**
Presence of urls	-0.089**	-0.076**	-0.060**
Issues:			
the reference is Other Issues			
Political Issues	0.047**	0.033**	0.063**
Policy Issues	0.119**	0.101**	0.084**
Campaign Issues	0.068**	0.056**	0.070**
Personal Issues	0.014**	0.014**	0.011**
N	1,262,140	1,332,512	288,563
Adjusted R^2	0.189	0.196	0.157
F	29,378.83**	32,580.04**	5,977.40**

Notes: Multicollinearity between independent variables was not detected.
**p<0.01.

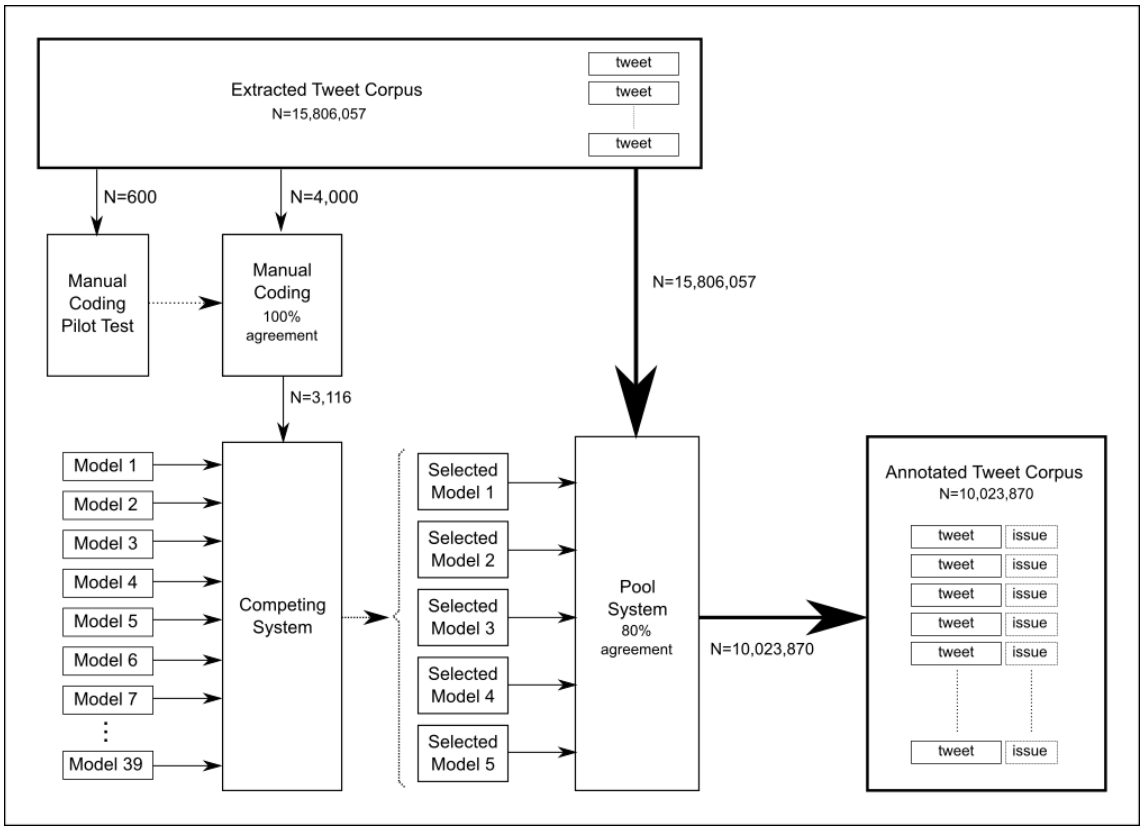

Figure 1. Process to annotate the tweet corpus with the Patterson's issues.

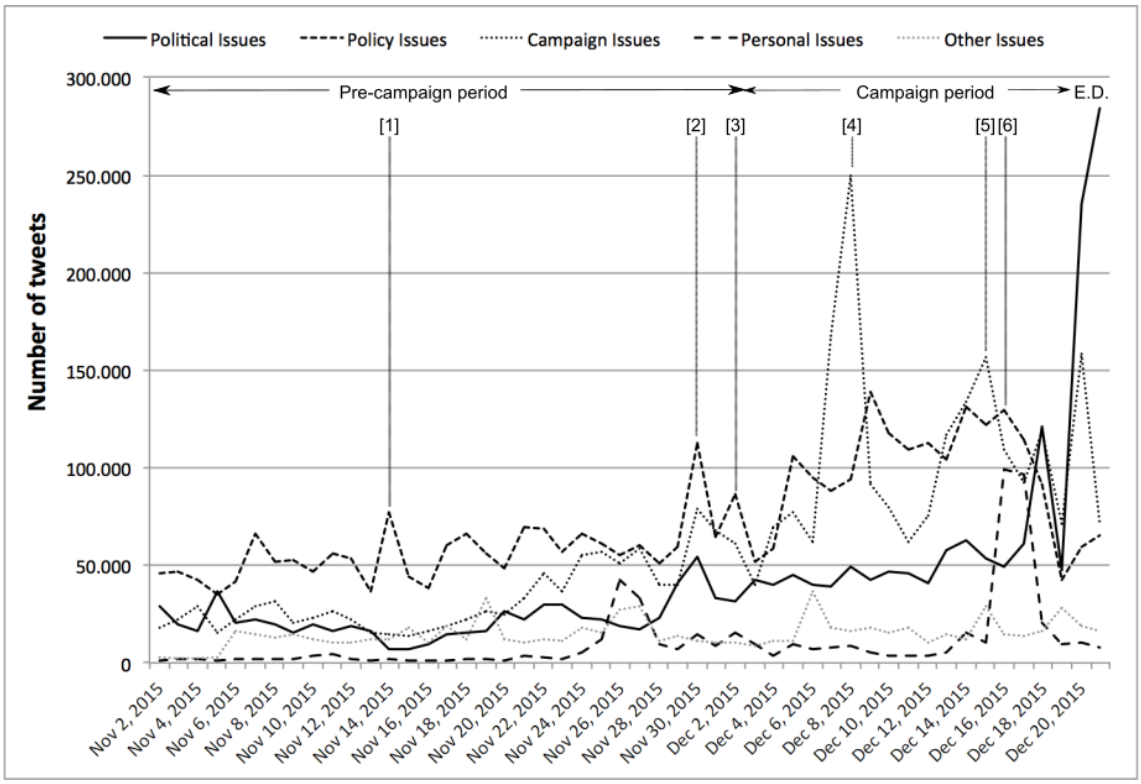

Figure 2. Daily evolution of issues discussed in Twitter during the pre-campaign, the campaign and election day (E.D) for the 2015 Spanish General Election. N=10,023,870. Main events: [1] Nov 14th, the day after the Paris attacks; [2] Nov 30th, digital debate hosted by *El País* among Pedro Sánchez, Pablo Iglesias and Albert Rivera; [3] Dec 2nd, Mariano Rajoy appeared on the TV Show "En la tuya o en la mía"; [4] Dec 8th, the day after the TV debate hosted by *Antena 3* among Pedro Sánchez, Pablo Iglesias, Albert Rivera and Soraya Sáenz de Santamaría; [5] Dec 15th, the day after the TV debate hosted by *La 1 de TVE* between Mariano Rajoy and Pedro Sánchez; [6] Dec 16th, Mariano Rajoy was punched by a citizen.

ANNEX

References to the IBEREVAL Workshop Proceedings:

Authors et al. (2017):

Giménez, M., Baviera, T., Llorca, G., Gámir, J., Calvo, D., Rosso, P., and Rangel, F., 2017. Overview of the 1st Classification of Spanish Election Tweets Task at IberEval 2017. *In*: R. Martínez, J. Gonzalo, P. Rosso, S. Montalvo, and J. Carrillo-de-Albornoz, eds. *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, September 19, 2017*. Aachen, Germany: CEUR Workshop Proceedings, 1–14.

Team 1 (2017):

González, J.-Á., Pla, F., and Hurtado, L.-F., 2017. ELiRF-UPV at IberEval 2017: Classification Of Spanish Election Tweets (COSET). *In*: R. Martínez, J. Gonzalo, P. Rosso, S. Montalvo, and J. Carrillo-de-Albornoz, eds. *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, September 19, 2017*. Aachen, Germany: CEUR Workshop Proceedings, 55–60.

Team 2 (2017):

Ambrosini, L. and Nicolò, G., 2017. Comparative study of neural models for the COSET shared task at IberEval 2017. *In*: R. Martínez, J. Gonzalo, P. Rosso, S. Montalvo, and J. Carrillo-de-Albornoz, eds. *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, September 19, 2017*. Aachen, Germany: CEUR Workshop Proceedings, 20–27.

Team 3 (2017):

Khandelwal, A., Swami, S., Akhtar, S.S., and Shrivastava, M., 2017. Classification Of Spanish Election Tweets (COSET) 2017 : Classifying Tweets using Character and Word Level Features. *In*: R. Martínez, J. Gonzalo, P. Rosso, S. Montalvo, and J. Carrillo-de-Albornoz, eds. *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th*

Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, September 19, 2017. Aachen, Germany: CEUR Workshop Proceedings, 49–54.

Team 4 (2017):

De la Peña Sarracén, G.L., 2017. Ensembles of methods for Tweet Topic Classification. *In*: R. Martínez, J. Gonzalo, P. Rosso, S. Montalvo, and J. Carrillo-de-Albornoz, eds. *Proceedings of the Second Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017) co-located with 33th Conference of the Spanish Society for Natural Language Processing (SEPLN 2017), Murcia, Spain, September 19, 2017*. Aachen, Germany: CEUR Workshop Proceedings, 15–19.