A review on political analysis and social media*

Una revisión del análisis político mediante la web social

David Vilares y **Miguel A. Alonso**
Grupo LyS, Departamento de Computación, Universidade da Coruña
Campus de A Coruña, 15071 A Coruña, Spain
{david.vilares, miguel.alonso}@udc.es

Resumen: En los países democráticos, conocer la intención de voto de los ciudadanos y las valoraciones de los principales partidos y líderes políticos es de gran interés tanto para los propios partidos como para los medios de comunicación y el público en general. Para ello se han utilizado tradicionalmente costosas encuestas personales. El auge de las redes sociales, principalmente Twitter, permite pensar en ellas como una alternativa barata a las encuestas. En este trabajo, revisamos la bibliografía científica más relevante en este ámbito, poniendo especial énfasis en el caso español.
Palabras clave: Análisis político, Análisis del sentimiento, Twitter

Abstract: In democratic countries, forecasting the voting intentions of citizens and knowing their opinions on major political parties and leaders is of great interest to the parties themselves, to the media, and to the general public. Traditionally, expensive polls based on personal interviews have been used for this purpose. The rise of social networks, particularly Twitter, allows us to consider them as a cheap alternative. In this paper, we review the relevant scientific bibliographic references in this area, with special emphasis on the Spanish case.
Keywords: Political Analysis, Sentiment Analysis, Twitter

1 *Introduction*

The adoption of social media and its use for widespread dissemination of political information and sentiment is so remarkable that it has impacted traditional media. Nowadays, Twitter is a convenient tool for journalists in search of quotes from prominent news sources, e.g., politicians (Lassen and Brown, 2011), as they can add direct quotes to stories without having the source in front of a microphone or camera (Broersma and Graham, 2012).

Current computational techniques (Mohammad et al., 2015) make possible to automatically determine the sentiment (positive or negative) and the emotion (joy, sadness, etc.) expressed in a tweet, the purpose behind it (to point out a mistake, to support, to ridicule, etc.) and the style of writing (statement, sarcasm, hyperboles, etc.). As a result, a lot of research activity has been devoted to analyze social media. In the field of political analysis on Twitter, most research has focused on predicting electoral outcomes, although Twitter is also a valuable tool for tasks such as identifying the political preferences of the followers of an account (Golbeck and Hansen, 2011) and monitoring day-to-day change and continuity in the state of an electoral campaign (Jensen and Anstead, 2013; Wang et al., 2012).

In this article, we review the relevant scientific literature dealing with Twitter as a source for political analysis, with special emphasis on the Spanish case. In Section 2 we consider work focused on predicting electoral outcomes, while in Section 3 we consider that work dealing with the political preferences of individual users. In Section 4 we consider the use of Twitter as a forecasting tool in the Spanish political arena. Conclusions are presented in Section 5.

2 *Predicting electoral outcomes*

One of the first studies on the prediction of electoral outcomes was performed by Tumasjan et al., (2010). They analyze 104 003 Twitter messages mentioning the name of at

* This research is partially supported by Ministerio de Economía y Competitividad (FFI2014-51978-C2). David Vilares is partially funded by the Ministerio de Educación, Cultura y Deporte (FPU13/01180)

least one of the six parties represented in the German parliament or prominent politicians of these parties, that were published in the weeks leading up to the 2009 federal election of the German parliament. The messages were downloaded in German and automatically translated into English to be processed by LIWC (Pennebaker, Francis, and Booth, 2001). They show that the mere number of tweets mentioning parties or politicians reflect voters preferences and comes close to traditional election polls. They find surprising that, although only 4% of users wrote more than 40% of messages, these heavy users were unable to impose their political opinion on the discussion, a fact they attribute to the large number of participants on Twitter who make the information stream as a whole more representative of the electorate. Therefore, the main conclusion of Tumasjan et al., (2010) is that Twitter may complement traditional polls as political forecasting tool, although they also point out several limitations of their approach: a Twitter sample may not be representative of the electorate, replies to messages in the sample that do not mention any party or politician may be relevant but they are missed, the dictionary may be not well-tailored for the task, and the results may be not generalizable to other specific political issues.

Bermingham and Smeaton (2011) use the 2011 Irish General Election as a case of study, collecting 32 578 tweets relevant to the five main Irish parties, where relevance is defined by the presence of the party names and their abbreviations, along with the election hashtag #ge11, with tweets reporting poll results being discarded. They apply a volume-based measure defined as the proportional share of mentions for each party, and sentiment analysis measures that represent the share of tweets with positive and negative opinion and, for each party, the log-ratio of sentiment for the tweets mentioning it. They find that the best method for predicting election results is the share of volume of tweets that a given party receives in total, followed closely by the share of positive volume. However, the mean absolute error of 5.85% is significantly higher than that achieved by traditional polls. Examining the errors, they find that they forecasted a higher result for the Green party, whose supporters tend to be more tech-savvy and have a disproportionately large presence in social media, and a lower result for Fianna Fáil, a party that attracted a low volume of tweets and plenty of negativity, however it is traditionally the largest Irish party and thus it enjoys a degree of brand loyalty.

In a similar line, Effing, van Hillegersberg, and Huibers (2011) test whether there exists a correlation between the use that Dutch politicians made of social media and the individual votes. Their study concludes that the results of national elections where correlated with the compromise of politicians with social media, but the same was not true for local elections. One of the novelties of the study is the introduction of an standarized framework, Social Media Indicator (SMI), for measuring the participation of politicians and how they interact with the public.

O'Connor et al., (2010) try to determine if the opinions extracted from Twitter messages about the US presidential approval and the 2008 US presidential elections, correlate the opinions obtained by means of classical polls. They collect messages over the years 2008 and 2009 and derive day-to-day sentiment scores by counting positive and negative messages: a message is defined as positive if it contains any positive word, and negative if it contains a negative one (a message can be both positive and negative). With this simple sentiment analysis technique, they find many examples of falsely detected sentiment, but they consider that, with a fairly large number of measurements, these errors will cancel out relative to the aggregate public opinion. They also find that recall is very low due to the lexicon, designed for standard English. To make predictions, day-to-day sentiment is volatile, so smoothing is a critical issue in order to force a consistent behavior to appear over longer periods of time. Finally, they find the sentiment rate correlated the presidential approval polls, but it does not correlate to the elections polls. Unlike Tumasjan et al., (2010), they find that message volume has not a straightforward relationship to public opinion. For the same 2008 US presidential elections, Gayo-Avello (2011) collects 250 000 Twitter messages published by 20 000 users in seven states, finding that the correlation between population and number of tweets and users was almost perfect. He applies four simple sentiment analysis techniques to that collection that also fail to predict the election outcomes, concluding that the prediction error is due to younger people is overrepresented in Twitter, and that Republican supporters had tweeted much less than Democratic voters.

DiGrazia et al., (2013) analyze 542 969 tweets mentioning candidates as well as data on elections outcomes from 795 competitive races in the 2010 and 2012 US Congressional Elections and socio-demographic and control variables such as incumbency, district partisanship, median age, percent white, percent college educated, median household income, percent female and media coverage. They show that there is a statistically significant association between tweets that mention a candidate for the US House of Representatives and the subsequent electoral performance. They also find that districts where their models under-perform tend to be relatively noncompetitive and that a few districts have idiosyncratic features difficult to model, such as a rural district that had voted for a Democratic congressman while voting strongly for the Republican presidential candidate. They conclude that (1) social media are a better indicator of political behavior than traditional TV media; (2) they can serve as an important supplement to traditional voter surveys; and (3) they are less likely to be affected by social desirability bias than polling data, i.e., a person who participates in a poll may not express opinions perceived to be embarrassing or offensive but socially undesirable sentiments are captured in social media.

Contractor and Faruquie (2013) try to use Twitter to predict the daily approval rating of the two candidates for the 2012 US presidential elections. They formulate the issue as a time series regression problem where the approval rate for each candidate is dependent on the bigrams (two consecutive words) mentioned in messages written by his supporters. They find that 227 bigrams were causal for the Democratic candidate and 183 bigrams for the Republican candidate. Nooralahzadeh, Arunachalam, and Chiru (2013) compare the sentiment that prevailed before and after the presidential elections taking place in 2012 in USA and France. In the case of the US Presidential election, they find that there are more tweets relating Obama (incumbent candidate) with positive and neutral opinions and less tweets with negative opinions than Rommey. On the contrary, in the case of French Presidential election, the elected President Holland has less tweets than Sarkozy (incumbent candidate) with positive and neutral opinions but also much less tweets with negative opinions.

Caldarelli et al., (2014) monitor 3 million tweets during the 2013 General election in Italy in order to measure the volume of tweets supporting each party. In this election, the three major parties got a similar number of votes but few traditional polls were able to predict the final outcomes. Although the tweet volume and time evolution do not precisely predicted the election outcome, they provided a good proxy of the final results, detecting a strong presence in Twitter of the (unexpected) winner party and the (also unexpected) relative weakness of the party finally occupying the fourth position. They find that predicting results for small parties is difficult, receiving a too large volume of tweets when compared to their electoral results. Moreover, a relevant 7.6% of votes went to very small parties which were not considered in their study.

Lampos, Preoţiuc-Pietro, and Cohn (2013) propose an approach for filtering irrelevant tweets from the stream in order to accurately model the polls in their prediction in voting intentions for the three major parties in the United Kingdom and for the four major parties in Austria. Gaurav et al., (2013) predict with a low error margin the winners of the Venezuelan, Parguayan and Ecuatorian Presidential elections of 2013. The best results are attained with a volume-based approach consisting of measuring the number of tweets mentioning the full names of candidates or mentioning the aliases of candidates jointly with a electoral keywork.

2.1 Controversy

As a consequence of the mixed results obtained in these studies, some authors are skeptics about the feasibility of using Twitter to predict the outcomes of electoral processes.

Jungherr, Jürgens, and Schoen (2012) argue that, taking into account all of the parties running for the elections, and not only the six ones with seats in the German parliament, the approach of Tumasjan et al., (2010) would actually have predicted a victory of the Pirate Party, which received a 2% of the votes but no seats in the parliament.

Gayo-Avello (2012) indicates that sentiment analysis methods based on simplistic assumptions should be avoided, devoting more resources to the study of sentiment analysis in politics before trying to predict elections. Moreover, Metaxas, Mustafaraj, and Gayo-Avello (2011) find that electoral predictions on Twitter data using the published research methods at that time are not better than chance and that even when the predictions are better than chance, as when they were applied to a corpus of messages during the 2010 US Congressional elections (Gayo-Avello, Metaxas, and Mustafaraj, 2011), they were not competent compared to the trivial method of predicting through incumbency given that current holders of a political office tends to maintain the position in an electoral process. To corroborate their statement, they apply a lexicon-based sentiment analysis technique to a dataset of Twitter data compiled during the 2010 US Senate special election in Massachusetts (Chung and Mustafaraj, 2011) and they find that, when compared against manually labeled tweets, its accuracy is only slightly better than a classifier randomly assigning the three labels of positive, negative and neutral to Twitter messages.

On the other hand, Huberty (2013) points out that US elections pose a very high bar, since forecasts must beat the simple heuristic of incumbency that reliably predicts future winners with high accuracy, even in ostensible competitive races. He also finds that algorithms trained on one election for the U.S. House of Representatives perform poorly on a subsequent election, despite having performed well in out-of-sample tests on the original election.

Prasetyo and Hauff (2015) point out that traditional polls in developing countries are less likely to be reliable than in developed countries, therefore they often result in a high forecasting error. Taking the 2014 Indonesian Presidential Election as a case study, they show that a Twitter prediction based on sentiment analysis outperformed all available traditional polls on national level.

3 Predicting political preferences

The aim of the work of Makazhanov and Rafiei (2013) is not to forecast election outcomes, but to predict the vote of individual users, arguing that political preference can be predicted from the interaction with political parties. For this purpose, they build an interaction profile for each party as a language model from the content of the tweets by the party candidates, and the preference of a user is assessed according to the alignment of user tweets with the language models of the parties. Their method is evaluated on a set of users whose political preferences are known based on explicit statements made on election day or soon after, in the context of Alberta 2012 general election. They find that, although less precise than humans, for some parties their method outperforms human annotators in recall, and revealed that politically active users are less prone to change their preferences than the rest of users. Pennacchiotti and Popescu (2011) try to classify 10 338 Twitter users as being either Democrats or Republicans, finding that the linguistic content of the user's tweets is highy valuable for this task, while social graph information has a negligible impact on the overall performance.

Monti et al., (2013) analyze the phenomenon of political disaffection in Italy, i.e., negative sentiment towards the political system in general, rather than towards a particular politician, policy or issue. For this purpose, they apply sentiment analysis techniques on political tweets to extract those with negative sentiment, to then select the tweets that refer to politics in general rather than specific political events of personalties. They find a strong correlation between their results and political disaffection as measured in public opinion surveys. They also show that important political news of Italian newspapers are often correlated with the highest peaks of disaffection.

There are great difference in how electoral processes are driven in developed and developing countries. In this respect, Razzaq, Qamar, and Bilal (2014) use sentiment analysis to study the Twitter messages related to the 2013 Pakistan general election, finding there are two groups of users, one formed by people living outside Pakistan and that only could participate in political discussion in social media, and a second group of users living in Pakistan. In this latter group, they also observed differences, both in volume and sentiment, among users living in large cities and in rural areas with low literacy rates. Fink et al., (2013) analyze the 2011 Nigerian Presidential election and find that volume counts of the mentions of the two major candidates correlates strongly with polling and election outcomes, but that other candidates are overrepresented. However, the particular ethnic divide of Nigerian population makes religion the best predictor of electoral support, with place of living and education as significant predictors as well.

4 Twitter as a tool for political analysis in Spain

With respect to the analysis of messages regarding the political situation in Spain, Peña-López, Congosto, and Aragón (2011) study networked citizen politics, in particular the relations among the Spanish *indignados* movement, traditional political parties and mass media. Criado, Martínez-Fuentes, and Silván (2013) note the high degree of use of Twitter as a channel of political communication during electoral periods. Congosto, Fernández, and Moro Egido (2011) corroborate results obtained for other countries (Livne et al., 2011; Conover et al., 2012; Conover et al., 2011) that observed how Twitter users are grouped by political affinity when transmitting information. A similar grouping by ideological reasons is found by Romero-Frías and Vaughan (2012) among Spanish political parties and traditional mass media when analyzing the linking practices in the websites of both kinds of organization, with left-wing media closer to PSOE (socialist party) and right-wing media closer to PP (conservative party). In the same line, Romero-Frías and Vaughan (2010) find that ideology was the main factor in the clustering of European political parties belonging to the, at that time, 27 member states of the European Union, followed by country or regional affiliation.

Borondo et al., (2012) also find that politicians mentioned and retweeted mostly their own partisans, after analyzing 370 000 Twitter messages written by over 100 000 users during the 2011 Spanish general Election, where half of the messages were posted by only 7% of participants, just 1% of users were the target for half of the mentions, 78% of mentions were for politicians, 2% of the users causes half of the retweets and the source of 63% of the retweeted messages were created by mass media accounts. Aragón et al. (2013) analyze 3 074 312 Twitter messages published by 380 164 distinct users during the same election, concluding again that members of political parties tend to almost exclusively propagate content created by other members of their own party. They also observe that politicians conceive Twitter more as a place to diffuse their political messages than to enage in conversations with citizens, although minor and new parties are more prone to exploit the communication mechanisms offered by Twitter; and that messages corresponding to the winner party become more and more positive until election day.

Barberá and Rivero (2012) find that in the political debate in Twitter in Spain, 65% of participants are men compared to 35% of women and that the geographical distribution of users corresponds to the distribution of population in the country, except that Madrid is overrepresented, with no significant differences between the behavior of those living in large cities and in the rest of Spain. They also find a strong polarization of the political debate, since those citizens with a stronger party identification monopolize much of the conversation, with the communication related to PP being highly structured and hierarchical, while the communication concerning the PSOE is much more horizontal and interactive.

In this context, assuming that individuals prefer to follow Twitter users whose ideology is close to theirs, Barberá (2012) considers the ideology or party identification of a Twitter user as a latent variable that cannot be observed directly, proposing the use of Bayesian inference to derive it from the list of accounts that each user follows. He takes as seeds the accounts of the top 50 politicians from PP and PSOE with the highest number of followers. Then, he applies his approach on a random sample of 12 000 active users during the 2011 Spanish elections. He tries to validate the technique by considering as additional seeds the official accounts of other two minority parties (IU and UPyD), obtaining inconclusive results that seems to support the idea that the latent variable is not measuring ideology but rather a combination of both policy preferences and party support. In order to determine whether the approach places both politicians and citizens on the same scale, he applies a lexicon-based sentiment analysis technique, observing that socialist candidates attain a better average evaluation among left-wing Twitter users and that conservative candidates attain a better evaluation among right-wing users, as expected. A similar correlation is found between the value of the latent variable for each user and the support of hashtags promoted by the socialist and conservative parties.

Borondo et al., (2012) confirm the finding of Tumasjan et al., (2010) that there exists a correlation between the number of times a political party is mentioned on Twitter and the electoral outcomes, but they only consider parties that obtained more than 1% of votes. Deltell (2012) analyzes the presence of one of these minor parties, eQuo, on social media during the 2011 Spanish General Election to question the efficiency and effectiveness of social networks in the modification of the vote and in predicting election results. At that time, eQuo was a newly created green party, without enough budget for a conventional electoral campaign, thus, no TV, radio or newspapers ads were possible. In addition, as a completely new party, no free airtime was granted on public TV and radios, and any privately-owned media offered significant coverage. As a result, the electoral strategy of eQuo was based mainly on social media: for several days its proposals were trending topics on Twitter, its Facebook page was the most-visited and had more "likes" than the page of any other political party, and it was the party with more presence on YouTube. However, this apparently successful campaign on social media was not reflected in its electoral outcome, as the number of votes was so small than no representative was assigned to eQuo in the parliament. Surprisingly, the best results for eQuo were obtained in those districts in which this party was present physically by means of traditional activities such as meetings, pasting campaign posters, and recreational activities. The interesting point here is that, disregarding eQuo, simple methods relying on the number of Twitter followers and Facebooks "likes" seems to be reliable indicators of outcomes for the 2011 Spanish elections.

Deltell, Claes, and Osteso (2013) study the political campaign on Twitter during the Andalusian regional elections of 2012. They focus on monitoring the Twitter profiles of the six most-voted political parties in Andalusia in the Spanish elections of 2011 and their leaders for Andalusian regional elections in 2012. They compute the support of each party by the number of followers of the Twitter accounts of political parties and their leaders. For the two major parties, PP and PSOE, the results computed by Deltell, Claes, and Osteso (2013) are closer to the final election outcomes than traditional polls. So, for the PP they predict a 40.48% of votes for a 40.66% final result, while for PSOE they predict a 36.31% of votes for a final score of 39.52%. We must point out that traditional polls failed in most predictions: although they predicted rightly that PP would have more votes than PSOE, they predicted a 10% difference between them when in the end it was less than 2%. This situation allowed the leader of PSOE to be elected as regional president with the support of the elected parliamentarians of IU. The authors confirm that their method is not accurate for small or newly created political parties, in particular, they were completely wrong in predicting the votes for IU, which they attributed to the low activity on Twitter of IU's leader.

Cotelo et al., (2015) use the follower-folowee relationship to cluster politically active users in Twitter. This information is combined with the textual content of tweets in order to determine the sentiment (positive, negative or nautral) expressed in a given tweet with respect to PP and PSOE, attaining a 75% accuracy.

Vilares, Thelwall, and Alonso (2015) analyze the sentiment of 2 704 523 tweets referring to Spanish politicians and parties from 3 December 2014 to 12 January 2015. They describe the Spanish version of SentiStrength[1], an algorithm designed originally for analyzing the sentiment of English texts in social media (Thelwall, Buckley, and Paltoglou, 2012), and how their sentiment scores are used to build ranks for the politicians and their parties, giving popularity ratings that are comparable with those provided by the classic polls, although tweet volume was a much better predictor of voting intentions. A deeper analysis of politicians that had sentiment scores and that did not match those of their parties, suggested that these had attracted negative media publicity that had been amplified in Twitter. Thus, Twitter results may be useful to analyze the trajecto-

[1] http://sentistrength.wlv.ac.uk/#Non-English

ries of individual politicians and to evaluate the impact of negative press coverage on their popular perception.

The task 4 of the TASS 2013 competition (Villena-Román and García-Morera, 2013) consisted of classifying the political tendency of public figures (not necessarily politicians) into one of four wings: left, center, right or neutral. A training set was not provided, so participant teams need to define their own strategies to categorise the authors. This was a controversial issue since the same party might belong to a different wing depending on their statutes or the polls made to citizenship. The best performing system (Pla and Hurtado, 2013) considered a number of entities related with the main political parties, which were classified into one of the four proposed classes. If the messages of a user containing one of those entities tend to be negative the user is prone to be against that political orientation, and vice versa. The task 3 of this same workshop was related with politics too: given a tweet where a representation of an entity (one of the four main national parties in 2013) occurs, participants where intended to classify the polarity of that entity. In this case, the best performing system (Gamallo, García, and Fernández Lanza, 2013) assumed that the polarity of the whole tweets corresponded to the polarity of the entity.

For the TASS 2015 competition (Villena-Román et al., 2015), the STOMPOL training and test corpora were developed, formed by 784 and 500 tweets, respectively, related to the six major Spanish political parties gathered from 23rd to 24th of April 2015. For each tweet, the polarity of the aspects involved (economy, health, education, political parties, other political issues) were identified. Only three teams participated at TASS 2015 task 2 (aspect-based sentiment analysis) on STOMPOL. The best-performing system applied a regular sentiment analysis system on the context of each aspect under consideration, defining context as a fixed-size window around the aspect instance. Best results were attained by the approach by Park (2015), which clusters tweets according to party-aspect pairs, on the assumptions that people who share similar preference or status show similarity in the expression of sentiment and that people evaluate a political party in multiple ways regarding different aspects. Then, some clusters are grouped together attending to the left vs. right political dimension. The deep learning approach tried by Vilares et al., (2015), based on LSTM recurrent neural networks, did not outperformed well-established machine learning approches, probably due to unsupervised pretraining and sentiment-specific were not considered.

5 Conclusion and future work

Over the last five years a lot of studies have been conducted on the use of Twitter as a cheap replacement for expensive polls involving personal interviews. Some initial satisfactory results were followed by disappointing ones, sparking controversy over the methodology used and the management of biases introduced by the demographics of active users on Twitter. However, we can see how in recent years Twitter has been accepted as a valid tool of political analysis, although there are still problems to be solved, such as the handling of very small parties with very active Twitter users; the management of small constituencies; the detection of spam produced by robots or users engaged in propaganda; and the treatment of countries with multilingual population.

References

Aragón, P., K.E. Kappler, A. Kaltenbrunner, D. Laniado, and Y. Volkovich. 2013. Communication dynamics in Twitter during political campaigns: The case of the 2011 Spanish national election. *Policy and Internet*, 5(2):183–206, June.

Barberá, P. 2012. A new measure of party identification in Twitter. Evidence from Spain. In *2nd Annual General Conference of the European Political Science Association(EPSA)*, Berlin, June. EPSA.

Barberá, P. and G. Rivero. 2012. ¿Un tweet, un voto? Desigualdad en la discusión política en Twitter. In *I Congreso Internacional en Comunicación Política y Estrategias de Campaña*, Madrid, July. ALICE.

Bermingham, A. and A.F. Smeaton. 2011. On using Twitter to monitor political sentiment and predict election results. In S. Bandyopadhyay and M. Okumurra, editors, *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychol-*

ogy (SAAIP 2011), pages 2–10, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.

Borondo, J., A.J. Morales, J.C. Losada, and R.M. Benito. 2012. Characterizing and modeling an electoral campaign in the context of Twitter: 2011 Spanish presidential election as a case study. *AIP Chaos*, 22(2):023138, June.

Broersma, M. and T. Graham. 2012. Social media as beat. tweets as a news source during the 2010 British and Dutch elections. *Journalism Practice*, 6(3):403–419.

Caldarelli, G., A. Chessa, F. Pammolli, G. Pompa, M. Puliga, M. Riccaboni, and G. Riotta. 2014. A multi-level geographical study of Italian political elections from Twitter data. *PLOS ONE*, 9(5):e95809, May.

Chung, J. and E. Mustafaraj. 2011. Can collective sentiment expressed on Twitter predict political elections? In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI-11)*, pages 1170–1171, San Francisco, CA, August. AAAI.

Congosto, M.L., M. Fernández, and E. Moro Egido. 2011. Twitter y política: Información, opinión y ¿predicción? *Cuadernos de Comunicación Evoca*, 4:11–15.

Conover, M.D, B. Gonçalves, A. Flammini, and F. Menczer. 2012. Partisan asymmetries in online political activity. *EPJ Data Science*, 1(1):Article: 6, December.

Conover, M.D., J. Ratkiewicz, M. Francisco, B. Gonçalves, A. Flammini, and F. Menczer. 2011. Political polarization on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, pages 89–96, Barcelona, Spain, July. AAAI.

Contractor, D. and T.A. Faruquie. 2013. Understanding election candidate approval ratings using social media data. In *WWW 2013 Companion*, pages 189–190, Rio de Janeiro, May. ACM Press.

Cotelo, J.M., F. Cruz, F.J. Ortega, and J.A. Troyano. 2015. Explorando Twitter mediante la integración de información estructurada y no estructurada. *Procesamiento del Lenguaje Natural*, 55:75–82, September.

Criado, J.I., G. Martínez-Fuentes, and A. Silván. 2013. Twitter en campaña: las elecciones municipales españolas de 2011. *Revista de investigaciones Políticas y Sociológicas*, 12(1):93–113.

Deltell, L. 2012. Estrategias de comunicación política en las redes sociales durante la campaña electoral del 2011 en España: el caso de eQuo. In *II Jornadas de la Asociación Madrileña de Sociología*, Madrid, March.

Deltell, L., F. Claes, and J.M. Osteso. 2013. Predicción de tendencia política por Twitter: Elecciones andaluzas 2012. *Ambitos: Revista internacional de comunicación*, 22:91–100.

DiGrazia, J., K. McKelvey, J. Bollen, and F. Rojas. 2013. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLOS ONE*, 8(11):e79449, November.

Effing, R., J. van Hillegersberg, and T. Huibers. 2011. Social media and political participation: Are Facebook, Twitter and YouTube democratizing our political systems? In E. Tambouris, A. Macintosh, and H. de Bruijn, editors, *Electronic Participation*, volume 6847 of *Lecture Notes in Computer Science*. Springer, Berlin and Heidelberg, pages 25–35.

Fink, C., N. Bos, A. Perrone, E. Liu, and J. Kopecky. 2013. Twitter, public opinion, and the 2011 Nigerian presidential election. In *Proceedings of SocialCom / PASSAT / BigData / EconCom / BioMedCom 2013 (SocialCom 2013)*, pages 311–320, Washington, D.C., USA, September. IEEE Computer Society.

Gamallo, P., M. García, and S. Fernández Lanza. 2013. TASS: A naive-Bayes strategy for sentiment analysis on Spanish tweets. In A. Díaz Esteban, I. Alegría Loinaz, and J. Villena Román, editors, *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, pages 126–132, Madrid, Spain, September.

Gaurav, M., A. Srivastava, A. Kumar, and S. Miller. 2013. Leveraging candidate popularity on Twitter to predict election outcome. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis (SNA-KDD 2013)*, page Article No. 7, Chicago, IL. ACM.

Gayo-Avello, D. 2011. Don't turn social median into another 'literay digest' poll. *Communications of the ACM*, 54(10):121–128, October.

Gayo-Avello, D. 2012. No, you cannot predict elections with Twitter. *IEEE Internet Computing*, 16(6):91–94, November/December.

Gayo-Avello, D., P.T. Metaxas, and E. Mustafaraj. 2011. Limits of electoral predictions using Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, pages 490–493, Barcelona, Spain, July. AAAI.

Golbeck, J. and D.L. Hansen. 2011. Computing political preference among Twitter followers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2011)*, pages 1105–1108, Vancouver, BC, Canada, May. ACM.

Huberty, M. 2013. Multi-cycle forecasting of congressional elections with social media. In *Proceedings of the 2nd workshop on Politics, elections and data (PLEAD 2013)*, pages 23–29, San Francisco, CA, October. ACM.

Jensen, M.J. and N. Anstead. 2013. Psephological investigations: Tweets, votes and unknown unknowns in the republican nomination process. *Policy & Internet*, 5(2):161–182, June.

Jungherr, A., P. Jürgens, and H. Schoen. 2012. Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment". *Social Science Computer Review,*, 30(2):229–234.

Lampos, V., D. Preoţiuc-Pietro, and T. Cohn. 2013. A user-centric model of voting intention from social media. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 993–1003, Sofia, Bulgaria, August. ACL.

Lassen, D.S. and A.R. Brown. 2011. Twitter: The electoral connection? *Social Sience Computer Review*, 29(4):419–436, November.

Livne, A., M.P. Simmons, E. Adar, and L.A. Adamic. 2011. The party is over here: Structure and content in the 2010 election. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, pages 201–208, Barcelona, Spain, July. AAAI.

Makazhanov, A. and D. Rafiei. 2013. Predicting political preference of Twitter users. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pages 298–305, Niagara, ON, Canada, August. ACM.

Metaxas, P.T., E. Mustafaraj, and D. Gayo-Avello. 2011. How (not) to predict elections. In *Proceedings of 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing — PASSAT/SocialCom 2011—*, pages 165–171, Boston, Massachusetts, USA, October. IEEE Computer Society.

Mohammad, S.M., X. Zhu, S. Kiritchenko, and J. Martin. 2015. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing and Management*, 51(4):480–499, July.

Monti, C., A. Rozza, G. Zapella, M. Zignani, A. Arvidsson, and E. Colleoni. 2013. Modelling political disaffection from Twitter data. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM 2013)*, page Article No. 3, Chicago, IL, USA. ACM.

Nooralahzadeh, F., V. Arunachalam, and C. Chiru. 2013. 2012 presidential elections on Twitter — an analysis of how the US and French election were reflected in tweets. In I. Dumitrache, A.M. Florea, and F. Pop, editors, *Proceedings of the 19th International Conference on Control Systems and Computer Science (CSCS 2013)*, pages 240–246, Bucharest, Romania, May. IEEE Computer Society.

O'Connor, B., R. Balasubramanyan, B. Routledge, and N.A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In W.W. Cohen and S. Gosling, editors, *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*, pages 122–129, Washington, DC, May. AAAI.

Park, S. 2015. Sentiment classification using sociolinguistic clusters. In J- Villena-Román, J. García-Morera, M.Á. García-Cumbreras, E. Martínez-Cámara, M.T. Martín-Valdivia, and L.A. Ureña-López, editors, *TASS 2015 Workshop on Sentiment Analysis at SEPLN*, volume 1397 of *CEUR Workshop Proceedings*, pages 99–104, Alicante, Spain, September. CEUR-WS.org.

Peña-López, I., M. Congosto, and P. Aragón. 2011. Spanish indignados and the evolution of the 15M: Towards networked para-institutions. In J. Balcells, A. Cerrillo-i-Martínez, M. Peguera, I. Peña-López, M.J. Pifarré, and M. Vilasau, editors, *Big Data: Challenges and Opportunities. Proceedings od the 9th International Conference on Internet, Law & Politics*, pages 359–386, Barcelona, June. UOC-Huygens Editorial.

Pennacchiotti, M. and A.-M. Popescu. 2011. Democrats, Republicans and Starbucks afficionados: User classification in Twitter. In *KDD'11. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 430–436, San Diego, CA, USA, August. ACM.

Pennebaker, J.W., M.E. Francis, and R.J. Booth. 2001. *Linguistic inquiry and word count: LIWC 2001*. Lawrence Erlbaum Associates, Mahway.

Pla, F. and L.-F. Hurtado. 2013. ELiRF-UPV en TASS-2013: Análisis de sentimientos en Twitter. In A. Díaz Esteban, I. Alegría Loinaz, and J. Villena Román, editors, *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, pages 220–227, Madrid, Spain, September.

Prasetyo, N.D. and C. Hauff. 2015. Twitter-based election prediction in the developing world. In *Proceedings of the 26th ACM Conference on Hypertext and Social Media (HT 2015)*, Middle East Technical University Northern Cyprus Campus, Cyprus, September. ACM.

Razzaq, M.A., A.M. Qamar, and H.S.M. Bilal. 2014. Prediction and analysis of pakistan election 2013 based on sentiment analysis. In X. Wu, M. Ester, and G. Xu, editors, *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 700–703, Beijing, China, August. IEEE.

Romero-Frías, E. and L. Vaughan. 2010. European political trends viewed through patterns of web linking. *Journal of the American Society for Information Science and Technology*, 61(10):2109–2121, October.

Romero-Frías, E. and L. Vaughan. 2012. Exploring the relationships between media and political parties through web hyperlink analysis: The case of Spain. *Journal of the American Society for Information Science and Technology*, 63(5):967–976, May.

Thelwall, M., K. Buckley, and G. Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, January.

Tumasjan, A., T.O. Sprenger, P.G. Sandner, and I.M. Welpe. 2010. Predicting elections in Twitter: What 140 characters reveal about political sentiment. In W.W. Cohen and S. Gosling, editors, *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*, pages 178–185, Washington, DC, May. AAAI.

Vilares, D., Y. Doval, M.A. Alonso, and C. Gómez-Rodríguez. 2015. LyS at TASS 2015: Deep learning experiments for sentiment analysis on Spanish tweets. In J. Villena-Román, J. García-Morera, M.Á. García-Cumbreras, E. Martínez-Cámara, M.T. Martín-Valdivia, and L.A. Ureña-López, editors, *TASS 2015 Workshop on Sentiment Analysis at SEPLN*, volume

1397 of *CEUR Workshop Proceedings*, pages 47–52, Alicante, Spain, September. CEUR-WS.org.

Vilares, D., M. Thelwall, and M.A. Alonso. 2015. The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *Journal of Information Science*, 41(6):799–813, December.

Villena-Román, J. and J. García-Morera. 2013. TASS 2013 — workshop on sentiment analysis at SEPLN 2013: An overview. In A. Díaz Esteban, I. Alegría Loinaz, and J. Villena Román, editors, *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013). TASS 2013 - Workshop on Sentiment Analysis at SEPLN 2013*, pages 112–125, Madrid, Spain, September.

Villena-Román, J., J. García-Morera, M.Á. García-Cumbreras, E. Martínez-Cámara, M.T. Martín-Valdivia, and L.A. Ureña-López. 2015. Overview of TASS 2015. In J. Villena-Román, J. García-Morera, M.Á. García-Cumbreras, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña-López, editors, *TASS 2015 Workshop on Sentiment Analysis at SEPLN*, volume 1397 of *CEUR Workshop Proceedings*, pages 13–21, Alicante, Spain, September. CEUR-WS.org.

Wang, H., D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. 2012. A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 115–120, Jeju, Republic of Korea, July. ACL.