Political alignment and emotional expression in Spanish Tweets

Posicionamiento político y expresión emocional en tweets en castellano

David García
ETH Zurich
Zurich, Switzerland
dgarcia@ethz.ch

Mike Thelwall
University of Wolverhampton
Wolverhampton, UK
m.thelwall@wlv.ac.uk

Resumen: Presentamos un estudio sobre el discurso político y la expresión emocional en tweets en castellano. Analizamos la posición políotica de los cuatro partidos mayoritarios a través de su actividad en Twitter, descubriendo que el discurso político en Twitter depende de percepciones subjetivas, y que representa el espacio político de España. Proponemos un sencillo método basado en léxicos para identificar los temas de un tweet, el cual funciona especialmente bien para detectar contenido político en tweets. Adaptamos SentiStrength al castellano, traduciendo y convirtiendo un léxico establecido de la valencia de palabras. Bajo ciertas condiciones, esta herramienta tiene mejores resultados que un clasificador al azar, con amplias posibilidades para su mejora. Para terminar, combinamos tres conjuntos de datos para analizar los sentimientos expresados en los tweets políticos de los cuatro partidos mayoritarios de España, encontrando diferencias relacionadas con el status quo y el clíma político Español.
Palabras clave: Análisis de sentimientos, detección de temas, política en Internet

Abstract: We present a study political discourse and emotional expression through a dataset of Spanish tweets. We analyze the political position of four major parties through their Twitter activity, revealing that Twitter political discourse depends on subjective perception, and resembles the political space of Spain. We propose a simplified lexicon-based method to identify the topics of a tweet, which works especially well to detect the political content of tweets. Furthermore, we adapted SentiStrength to Spanish, by translating and converting an established lexicon of word valence. Under certain design decisions, this tool performs better than random, with ample room for improvement. Finally, we combined three datasets to analyze the sentiment expressed in the political tweets of four major Spanish parties, finding differences related to the status quo, and the Spanish political climate.
Keywords: Sentiment analysis, topic detection, Internet politics

1 Introduction

1.1 Politics and social media

The increasing adoption of Twitter as a communication tool offers both great opportunities and risks for individuals and media outlets. On one hand, individuals can to interact with large amounts of users without the access to traditional mass media channels, mediating in the creation of social and political movements (González-Bailón et al., 2011). On the other hand, the fast spreading of information in Twitter can lead to large-scale reactions, in which discussions around hashtags and topics are unpredictable (El Huffington Post, 2013). The singer David Bisbal is a paradigmatic example of how a single tweet can change a celebrity's reputation (Galaz, 2011), leading to concepts like the Bisbal index (Arias and Jimenez, 2012), which measures the size of mistakes in Twitter.

In Spain, Twitter serves as a commonly known platform for political discussions, providing an interaction medium between politicians and citizens. The influence of individual tweets and campaigns can receive lots of attention in discussions among Spanish politicians. Single tweets can create official political reactions, like the conflict between PP and IU for a tweet by *@GLlamazares* (Europa Press, 2013), or the dismissal of a diplomat due to the content expressed in a single tweet (El Pais, 2013).

Recent studies in communication theory have investigated the impact of computer-mediated communication on the mechanism leading to a dominant *public opinion.* (Liu and Fahmy, 2011) found for instance that in the online world there is a weaker feeling of isolation for holders of a minority opinion compared to face-to-face interactions. (Schulz and Roessler, 2012) have recently studied the question of the influence of online communication on the expression of opinions, and how *subjective* factors change the individual's perception of the prevailing climate of opinions.

Online communication might change how groups exchange ideas and opinions, and empirical evidence shows that the mechanisms leading to group polarization are different in computer-mediated communication than in face-to-face discussions (Taylor and Macdonald, 2002). Consistently, works on product reviews (Wu and Huberman, 2010) and Twitter (Yardi and Boyd, 2010) found that the phenomenon of group polarization is present also in online settings. Furthermore, a recent finding by (Mieghem, 2011) suggests nontrivial relations between positive and negative opinions at the collective level in Reddit, calling for psychological explanations for this phenomenon.

A currently emergent field of research is the study of political science from digital traces. Initial works showed the relevance of blogs (Adamic and Glance, 2005; Dodds and Danforth, 2009) in political discussions. While the usage of these online data sources to predict the outcome of elections is still under debate (Gayo Avello, Metaxas, and Mustafaraj, 2011), recent works show that user behavior in Twitter can predict political alignment (Conover et al., 2011), and the party asymmetries in social interaction (Conover et al., 2012). Furthermore according to (Sobkowicz and Sobkowicz, 2010) emotional interaction in political fora also reveals patterns of behavior regarding hate and political topics, with users taking the lead in controversial discussions. Digital traces can be used to understand the effectiveness of political campaigns, and the patterns of collective interaction of different political audiences (Garcia et al., 2012). Additionally, Eurovision voting patterns have being used to create a macroscope of the relation between European countries, measuring the impact of EU economic policies on cultural polarization (Garcia and Tanase, 2013).

1.2 Sentiment analysis and the social web

The study of emotional expression on the Internet requires the usage of tools from sentiment analysis (Pang and Lee, 2008). While most of these tools are developed for opinion mining, part of the sentiment analysis focuses on the extraction of emotional content from text. Some supervised approaches can be trained on large datasets (De Choudhury, Counts, and Gamon, 2012). The current state-of-the-art tools apply unsupervised lexicon-based analysis techniques (Taboada, Brooke, and Voll, 2011; Thelwall et al., 2010; Thelwall, Buckley, and Paltoglou, 2012). This tools use lexica of annotated emotional-bearing terms and syntax rules, building on previous survey studies on emotional words (Bradley and Lang, 1999; Pennebaker, Francis, and Booth, 2001). These kind of lexica are subject of extension and analysis, covering additional languages (Garcia, Garas, and Schweitzer, 2011), and words (Warriner, Kuperman, and Brysbaert, 2013).

These kind of techniques have been successfully applied to Twitter messages to study daily mood changes (Golder and Macy, 2011), emotional well-being (O'Connor et al., 2010; Bollen et al., 2011), happiness (Dodds and Danforth, 2009), and collective emotions (Thelwall, Buckley, and Paltoglou, 2011). The flexibility of this approach allowed the application to emotional expression in chatrooms (Garas et al., 2012) and Yahoo answers (Kucuktunc et al., 2012). Recent studies show how this kind of analysis might depend on the context of a discussion (Thelwall et al., 2013), and how it can be used to predict the decision of users to leave an online community (Garcia, Zanetti, and Schweitzer, 2013).

2 Party alignment in Twitter

2.1 Manual annotation of political alignment

To locate the position of political parties in policy space, we compare two different sources of subjective ratings on the position of users in Twitter:

1. TASS task 4 dataset: for which a human coder assigned to each one of 158 a tag corresponding to the political alignment of the user. This alignment was expressed as a value from the set *Right*, *Left*, *Centre*, or *Undefined*.

2. Party discourse dataset: an independent rater with political experience was consulted to review the accounts of the 158 users from the previous dataset. The rater tagged each account in a space corresponding to the major political parties of the Spanish system: *PP* (Partido Popular), *PSOE* (Partido Socialista Obrero Español), *UPyD* (Unión Progreso y Democracia), *IU* (Izquierda Unida), and the additional *none* value for users that do not discuss about politics.

The combination of these two sets of human annotations gives us two values per Twitter account: a value of political position from the first dataset, and a value of party alignment from the second.

2.2 Perceived party positions

We aggregate the position in political space of each of the studied Spanish parties through their ratios of Left, Right, and Centre users. This way, for each party we have three values corresponding to the amounts of users of each class, divided over the total amount of users of the party that were not tagged as *Undefined*. This implies that the sum of all three ratios equals to 1, and it can be visualized as a 2 dimensional projection. Figure 1 shows a point for each party, located with a distance to the three edges of the triangle proportional to each ratio of aligned users.

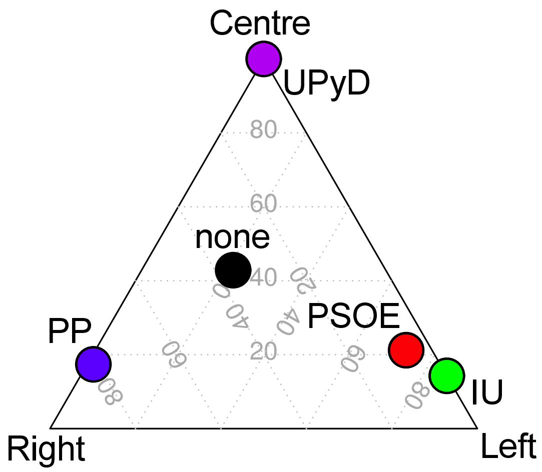

Figure 1: 2D-simplex of the ratios of right, left, and center users assigned to each party.

The parties *PP* and *IU* are located in symmetrical positions along the vertical axis, with 20% *Centre* users and 80% *Right* and *Left* users respectively. The party *UPyD* is located at the extreme *Centre* position, as the two accounts it had assigned were identified as *Centre*. The party *PSOE* is located at a close distance to the triangle vertex corresponding to *Left* users, but its softer position compared to *IU* reveals that some of its users were classified as either *Centre* or *Right*. Finally users without a clear party alignment are homogeneously distributed among the three alignments, having a position close to the barycenter of the triangle.

2.3 Coder agreement

The above analysis of user political discourse and party alignment suggests the grouping of parties in three sets: A right set composed of *PP*, a centre one with *UPyD*, and a left one composed of *PSOE* and *IU*. Under this simplification, we can measure the level of agreement of the original coding into party positions, and the converted values given by the second coder. As both coders rated each one of the 158 accounts, we use Cohen's kappa measure (Carletta, 1996):

$$\kappa = \frac{P(agreement) - P(random)}{1 - P(random)} \tag{1}$$

where $P(agreement)$ is the ratio of agreeing scores over the total amount of scores, and $P(random)$ is the chance of agreement under a randomized version of the user ratings.

For the coded political positions of the users in the dataset, we calculated $\kappa = 0.623$. This reveals a substantial agreement between coders, yet far from perfect agreement. Similarly to the subjective content of opinions and emotions in text, the perceived political position of a Twitter user is a subjective matter, for which ground truths can only be approximations to a more complex phenomenon.

3 Detecting political tweets

3.1 Lexicon model for topics

For topic detection, we used a lexicon-based approach that detects stemmed keywords assigned to each topic. This method is a simplified version of more general approaches based on clustering on the term frequencies of documents. Since tweets are very short texts, we considered a presence-based approach a good starting point, as term frequency within the text could be approximated by its binary transformation, assigning 0 to absent terms, and 1 to present ones. For the construction of these lexica, we proceeded as follows: first we stemmed the text of all tweets in the training dataset using an implementation[1] of Porter's method (Van Rijsbergen, Robertson, and Porter, 1980); then we created lexica for each topic, in which a lemma appears in the lexicon if its frequency is at least 2; and finally we optimized the lexica to increase precision and recall.

Lexicon optimization was performed through a greedy iterative method. Terms were sorted at random and assigned the value 0. The optimization loop would range over all the terms, measuring if its inclusion in the classification lexicon would increase accuracy. The loop was repeated until no change was done, creating a lexicon of relevant lemmas for each topic.

3.2 Evaluation metrics

To evaluate the performance of our topic detection method, we calculated the following metrics:

- Base rate: $B_c = \frac{\sum_x \delta[t(x)=c]}{N}$
- Classification rate: $C_c = \frac{\sum_x \delta[c(x)=c]}{N}$
- Precision: $P_c = \frac{\sum_x \delta[t(x)=c(x)=c]}{\sum_x \delta[c(x)=c]}$
- Recall: $R_c = \frac{\sum_x \delta[t(x)=c(x)=c]}{\sum_x \delta[t(x)=c]}$
- F_1 measure: $F_c = 2 * P_c * R_c / (P_c + R_c)$

where $t(x)$ is the test value of tweet t, $c(t)$ is the classified value of t, and N is the size of the test dataset. In general, we aim at a precision higher than the base rate of each class, taking recall as the measure of relevance of our results. The classification rate will be used to detect systematic errors when compared to the base rate, and the F_1 measure will be taken to compare the performance of the tool for different classes.

3.3 Results

Table 1 shows the results of the lexicon-based topic detection method explained above. While the precision value for all the classes is above their base rate, these are still far from the ideal case of 1. One notable exception is the performance of the classifier for the topic of politics, for which precision is close to 80% and recall to 60%. Since the lexica for each class are independent, this result suggests that tweets of political content are easier to detect than those of other topics. This might be thanks to mentions to parties and politicians, whose presence in the tweet would unequivocally imply a political context.

Table 1: **Topic detection results**

class	B_c	P_c	R_c	F_c	C_c
pol.	0.494	0.797	0.599	0.684	0.371
cine	0.009	0.155	0.167	0.161	0.011
dep.	0.002	0.084	0.355	0.136	0.009
eco.	0.041	0.246	0.216	0.230	0.036
ent.	0.089	0.343	0.453	0.390	0.117
fút.	0.013	0.257	0.410	0.316	0.021
lit.	0.001	0.085	0.225	0.123	0.004
mús.	0.024	0.390	0.353	0.371	0.022
otr.	0.463	0.614	0.217	0.320	0.163
tec.	0.004	0.147	0.299	0.197	0.009

The classification rate for the politics topic is lower than its base rate, suggesting that the lexicon found here is not complete, i.e. a significant amount of tweets of political content do not contain any of the words contained in the lexicon.

4 Lexicon-based sentiment analysis

4.1 Converting valence into sentiment scores

In our exercise of sentiment analysis, we adapted the latest version of SentiStrength (Thelwall, Buckley, and Paltoglou, 2012) to Spanish. SentiStrength takes a set of lexica as an input: a negative term list, a booster word list, and an emotion lexicon. We created the first two lexica by translating the negation and boosting terms in the English version of SentiStrength. We created a Spanish emotion lexicon from a valence lexicon (Warriner, Kuperman, and Brysbaert, 2013), which includes valence annotations for 13,915 English lemmas. We proceeded as follows:

1. We used Google translate to create a Spanish adaptation of the terms in the lexicon. We manually inspected the translation and deleted mistranslations and anomalous terms. After that, we extracted the stem of each word, averaging the valence scores for words with the same stem. The final set of words consists of 8201 different lemmas with their corresponding valence values.
2. We mapped the valence values of the lexicon to the scale of SentiStrength, in which terms have a value assigned between -5 and +5. The conversion table used is shown in Table 2.

[1] `sourceforge.net/projects/stemmer-es`

Table 2: **Mapping of valence to SentiStrength scores**

Valence	Score	Valence	Score
$[1, 2)$	-5	$[5, 6.5)$	1
$[2, 2.5)$	-4	$[6.5, 7)$	2
$[2.5, 3)$	-3	$[7, 7.5)$	3
$[3, 3.5)$	-2	$[7.5, 8)$	4
$[3.5, 5)$	-1	$[8, 9)$	5

4.2 SentiStrength at the global level

For each tweet, SentiStrength provided two values: a score of negative sentiment between -5 and -1, and a score of positive sentiment from +1 to +5. To provide empirical validation to this tool in Spanish, we converted its output to the five levels of the TASS test dataset.

Table 3: **SentiStrength output conversion**

N/P	1	2	3	4	5
-1	NONE	NONE	P	P	P+
-2	NONE	NONE	P	P	P+
-3	N	N	NEU	NEU	NEU
-4	N	N	NEU	NEU	NEU
-5	N+	N+	NEU	NEU	NEU

Table 4 reports the results of our translation of SentiStrength in 3 levels (*P*, *N*, *NEU*) and *NONE* (noted as -), using the metrics explained in Section 3.2. Our tool achieves a precision above the base rate for each class, performing better than a random classifier even for the minority class *NEU*. Recall values are around 50% for each class, and close to 35% for *NONE*, suggesting that the lexicon used here is not sufficient to cover the wealth of emotional words of Spanish language.

Table 4: **Valence lexicon results** - 3 levels

Class	B	P	R	F_1	C
P	0.366	0.676	0.555	0.609	0.299
NEU	0.021	0.034	0.457	0.064	0.283
N	0.260	0.563	0.494	0.526	0.228
NONE	0.352	0.662	0.352	0.460	0.187

Table 5 shows the results for five levels, calculated over each level separately. Again, precision values are above base rates, but recall values are too low to provide a satisfactory output. The comparison between base rates and classification rates in both evaluations provide an insight for this problem: the *NEU* class has a classification rate of 0.283, while its base rate is just 0.024. This suggest that the transformation matrix of Table 3 gives too much weight to the *NEU* class, which can be improved in further works. In addition, the *NONE* class has a classification rate much lower than its base rate (0.187 vs 0.352), which could be improved by pushing the lower limits to map SentiStrength values to positive and negative classes. Furthermore, the balance between the classes *P* and *P+* shows that our tool misclassifies some strong positive tweets as slightly positive, which could be related to the negation lexicon of SentiStrength.

Table 5: **Valence lexicon results** - 5 levels

Class	B	P	R	F_1	C
P+	0.341	0.824	0.282	0.352	0.116
P	0.024	0.053	0.397	0.093	0.183
NEU	0.024	0.034	0.456	0.064	0.283
N	0.185	0.395	0.359	0.376	0.168
N+	0.075	0.353	0.282	0.314	0.059
NONE	0.352	0.662	0.352	0.460	0.187

5 *Party sentiment*

Finally, we combine the datasets on user party alignment, topics, and sentiment to study the sentiment expressed by the supporters of each party. To measure sentiment, we ignored all tweets that did not have the politics topic, and calculated the ratios of tweets classified as *P*, *N*, *NEU*, and *NONE*.

Table 6 shows these ratios and the politics tweet counts for each party, as well as for the political tweets from users that were reported as not aligned with any party. These statistics reveal two patterns. First, the sentiment expressed by supporters of *IU* show a significantly larger ratio of negative tweets, and lower ratios of positive and objective tweets. Second, users not aligned with any party show a similar but softer pattern, with a higher ratio of negative tweets but a similar ratio of positive tweets as the rest of the parties.

This shows a larger difference between *IU* and *PSOE* that did not appear in the mapping to political space of Section 2.2. This difference in emotional expression for two left parties can be interpreted as their relation to the *status quo*: *PSOE*, the previous ruling party, has an interest in showing a positive state of the society, while a smaller party like *IU* shows a more critical and negative view of current events.

Table 6: **Sentiment ratios for political tweets from each party**

party	P	N	NEU	NONE	count
PP	0.27	0.35	0.03	0.35	15495
UPyD	0.24	0.36	0.02	0.38	1183
PSOE	0.28	0.35	0.03	0.34	6962
IU	0.17	0.51	0.04	0.28	4407
none	0.24	0.41	0.03	0.32	2020

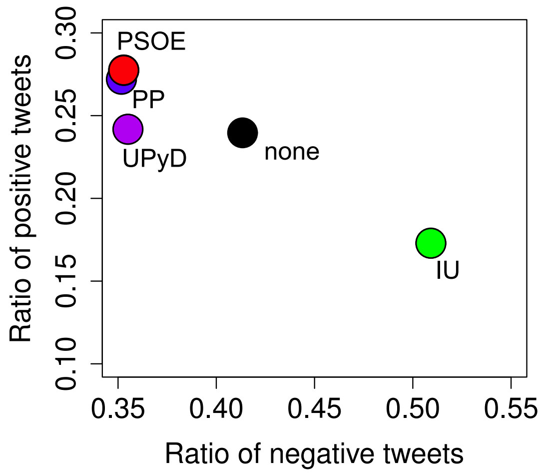

Figure 2: Ratios of positive and negative tweets for each party.

6 Conclusions

In this article, we presented a study of political discourse and emotional expression in Twitter.

Using two independent codings of Twitter user accounts, we related political parties with their position in political space. These two annotation sources allow us to measure the similarity in subjective perception of the political alignment of a Twitter user, finding some disagreement between coders. This suggests that the perception of the alignment of a user is a subjective matter, but when averaged, the policy position of major parties represent the political space of a country.

We described a lexicon-based method for topic detection, making use of the short nature of Tweets. While this method performed slightly better than random for most of the topics, it showed a very good performance to detect tweets with political content. We followed by adapting SentiStrength to Spanish, creating a new Spanish lexicon based on a previous English one (Warriner, Kuperman, and Brysbaert, 2013). We designed a way to validate the output of SentiStrength with the TASS datasets, reaching precision values above the base rate of each sentiment class. Nevertheless, the precision and recall values can be improved, as suggested by the comparison of base rates and classification rates. Our design detects too many tweets as simultaneously positive and negative, an event that is rarely observed in Twitter.

We combined three annotated datasets to analyze the sentiment expressed by the four major Spanish parties. We found that the sentiment expressed by left-wing parties differs, suggesting this measure as a way to understand their relation to election results and government policies. These results call for a deeper political science analysis, matching our observations with previous theories and works on party discourse and emotionality.

7 Acknowledgments

This research received funding from the Swiss National Science Foundation under the grant CR21I1_146499/1 and from the European Community's Seventh Framework Programme FP7-ICT-2008-3 under grant agreement no 231323 (CYBEREMOTIONS).

References

Adamic, Lada A. and Natalie Glance. 2005. The political blogosphere and the 2004 u.s. election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, LinkKDD '05, pages 36–43, New York, NY, USA. ACM.

Arias, Alfredo and Pepo Jimenez. 2012. Como perder tu credibilidad en redes sociales. *FIMP'12*.

Bollen, Johan, Bruno Gonçalves, Guangchen Ruan, and Huina Mao. 2011. Happiness is assortative in online social networks. *Artificial life*, 17(3):237–51, January.

Bradley, Margaret M and Peter J Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida.

Carletta, Jean. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2):249–254, June.

Conover, Michael D, Bruno Gonc, Alessandro Flammini, and Filippo Menczer. 2012. Partisan Asymmetries in Online Political Activity. *EPJ DataScience*, 1(6):1–17.

Conover, Michael D., Bruno Goncalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the Political Alignment of Twitter Users. In *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing*, pages 192–199. IEEE, October.

De Choudhury, Munmun, Scott Counts, and Michael Gamon. 2012. Not All Moods Are Created Equal! Exploring Human Emotional States in Social Media. In *Sixth International AAAI Conference on Weblogs and Social Media*, pages 66–73.

Dodds, Peter Sheridan and Christopher M. Danforth. 2009. Measuring the Happiness of Large-Scale Written Expression: Songs, Blogs, and Presidents. *Journal of Happiness Studies*, 11(4):441–456, July.

El Huffington Post. 2013. Orgullosopp: El hashtag de gonzalez pons que se ha vuelto en contra del pp.

El Pais. 2013. El 'numero dos' de la marca españa cesa tras insultar a los catalanes en twitter.

Europa Press. 2013. Victoria condena los "miserables" comentarios de llazamares en twitter tras el accidente de cifuentes.

Galaz, Mabel. 2011. El resbalon de bisbal en twitter. `http://elpais.com/elpais/2011/02/01/actualidad/1296546530_850215.html`.

Garas, Antonios, David Garcia, Marcin Skowron, and Frank Schweitzer. 2012. Emotional persistence in online chatting communities. *Scientific Reports*, 2:402, May.

Garcia, David, Antonios Garas, and Frank Schweitzer. 2011. Positive words carry less information than negative words. page 13, October.

Garcia, David, Fernando Mendez, Uwe Serdült, and Frank Schweitzer. 2012. Political polarization and popularity in online participatory media. In *Proceedings of the first edition workshop on Politics, elections and data - PLEAD '12*, pages 3–10, New York, New York, USA, November. ACM Press.

Garcia, David and Dorian Tanase. 2013. Measuring Cultural Dynamics Through the Eurovision Song Contest. *Advances in Complex Systems (Forthcoming)*, pages 1–27.

Garcia, David, Marcelo Serrano Zanetti, and Frank Schweitzer. 2013. The Role of Emotions in Contributors Activity: A Case Study on the GENTOO Community. June.

Gayo Avello, Daniel, Panagiotis T. Metaxas, and Eni Mustafaraj. 2011. Limits of Electoral Predictions using Twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*, pages 490–493, Barcelona, Spain.

Golder, Scott A. and Michael W. Macy. 2011. Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures. *Science*, 333(6051):1878–1881, September.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. The dynamics of protest recruitment through an online network. *Scientific reports*, 1:197, January.

Kucuktunc, Onur, B. Barla Cambazoglu, Ingmar Weber, and Hakan Ferhatosmanoglu. 2012. A large-scale sentiment analysis for Yahoo! answers. In *Proceedings of the fifth ACM international conference on Web search and data mining - WSDM '12*, page 633, New York, New York, USA, February. ACM Press.

Liu, Xudong and Shahira Fahmy. 2011. Exploring the spiral of silence in the virtual world: lndividuals' willingness to express personal opinions in online versus offline settings. *Journal of Media and Communication Studies*, 3:45–57, February.

Mieghem, Piet Van. 2011. Human psychology of common appraisal: The reddit score. *IEEE Transactions on Multimedia*, 13(6):1404–1406.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R Routledge, and Noah A Smith. 2010. From Tweets to Polls : Linking Text Sentiment to Public Opinion Time Series. In *Fourth International AAAI Conference on Weblogs and Social Media*, pages 122–129.

Pang, Bo and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(12):1–135, January.

Pennebaker, James W., Martha E. Francis, and Roger J. Booth. 2001. *Linguistic Inquiry and Word Count: LIWC 2001*. Erlbaum Publishers.

Schulz, Anne and Patrick Roessler. 2012. The spiral of silence and the internet: Selection of online content and the perception of the public opinion climate in computer-mediated communication environments. *International Journal of Public Opinion Research*.

Sobkowicz, P. and A. Sobkowicz. 2010. Dynamics of hate based Internet user networks. *The European Physical Journal B*, 73(4):633–643, February.

Taboada, Maite, Julian Brooke, and Kimberly Voll. 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2):267–307.

Taylor, Jacqui and John Macdonald. 2002. The effects of asynchronous computer-mediated group interaction on group processes. *Social Science Computer Review*, 20(3):260–274.

Thelwall, Mike, Kevan Buckley, George Paltoglou, Marcin Skowron, David Garcia, Stephane Gobron, Junghyun Ahn, Arvid Kappas, Dennis Küster, and A Janusz. 2013. Damping Sentiment Analysis in Online Communication : Discussions , Monologs and Dialogs. In *to appear in: Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 1–12.

Thelwall, Mike, Kevan Buckley, and Georgios Paltoglou. 2011. Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418.

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

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, December.

Van Rijsbergen, C. J., S. E. Robertson, and M. F. Porter. 1980. *New models in probabilistic information retrieval.* Computer Laboratory, University of Cambridge, Cambridge, England.

Warriner, Amy Beth, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, February.

Wu, Fang and Bernardo a. Huberman. 2010. Opinion formation under costly expression. *ACM Transactions on Intelligent Systems and Technology*, 1(1):1–13, October.

Yardi, Sarita and Danah Boyd. 2010. Dynamic debates: An analysis of group polarization over time on twitter. *Bulletin of Science, Technology and Society*, 30(5):316–327.

A Sentiment lexicon optimization

We applied the lexicon optimization technique explained in (Thelwall, Buckley, and Paltoglou, 2012), to adapt the lexicon to the training dataset. The results are summarized in Table 7 and Table 8, giving values comparable to those achieved without optimization, and not strictly better.

Table 7: **Optimized lexicon results** - 3 levels

Class	B	P	R	F_1	C
P	0.365	0.628	0.620	0.624	0.361
NEU	0.021	0.034	0.431	0.064	0.266
N	0.260	0.564	0.448	0.499	0.206
NONE	0.352	0.651	0.305	0.416	0.165

Table 8: **Optimized lexicon results** - 5 levels

Class	B	P	R	F_1	C
P+	0.341	0.780	0.185	0.299	0.081
P	0.024	0.042	0.484	0.077	0.280
NEU	0.024	0.034	0.431	0.064	0.266
N	0.185	0.359	0.097	0.153	0.050
N+	0.075	0.217	0.453	0.293	0.156
NONE	0.352	0.651	0.305	0.416	0.165

B *Relation between global and entity level*

It is argued that tweets are too short to express complex sentiments towards different entities. To test this concept, we reported the output of SentiStrength at the global level to compare it with the sentiment of tweets at the entity level. The results are shown in Table 9, revealing significant drops in precision under this approximation. This calls for the need of sentiment analysis techniques at the entity level, as 140 characters are enough to express different opinions and emotions towards different events, things, and people.

Table 9: **Relation between entity and global levels**

Class	B	P	R	F_1	C
P	0.374	0.471	0.306	0.371	0.243
NEU	0.228	0.223	0.318	0.262	0.325
N	0.308	0.364	0.335	0.349	0.284
NONE	0.091	0.114	0.187	0.142	0.148