Politics and Virality in the Time of Twitter: A Large-Scale Cross-Party Sentiment Analysis in Greece, Spain and United Kingdom

Dimosthenis Antypas, Alun Preece, Jose Camacho-Collados

Cardiff NLP, School of Computer Science and Informatics & Crime and Security Research Institute
Cardiff University, United Kingdom
{antypasd,preecea,camachocolladosj}@cardiff.ac.uk

Abstract

Social media has become extremely influential when it comes to policy making in modern societies especially in the western world (e.g., 48% of Europeans use social media every day or almost every day). Platforms such as Twitter allow users to follow politicians, thus making citizens more involved in political discussion. In the same vein, politicians use Twitter to express their opinions, debate among others on current topics and promote their political agenda aiming to influence voter behaviour. Previous studies have shown that tweets conveying negative sentiment are likely to be retweeted more frequently. In this paper, we attempt to analyse tweets from politicians from different countries and explore if their tweets follow the same trend. Utilising state-of-the-art pre-trained language models we performed sentiment analysis on multilingual tweets collected from members of parliament of Greece, Spain and United Kingdom, including devolved administrations. We achieved this by systematically exploring and analysing the differences between influential and less popular tweets. Our analysis indicates that politicians’ negatively charged tweets spread more widely, especially in more recent times, and highlights interesting trends in the intersection of sentiment and popularity.

1 Introduction

In recent years social media have come to resemble a ‘battleground’ between politicians who constantly aim to reach out to more people and win their votes. This behaviour is not surprising as the influx of people onto social media aiming to get updated on the latest news keeps growing, especially in the west with 48% of Europeans using social media on a regular basis (Cattaneo 2020). More specifically, Twitter seems to have become established as the main online platform where politicians attempt to engage with the public in regards to social and political commenting (Stuer et al. 2018), to such an extent that political accounts are often more active than non-political ones (Grant, Moon, and Busby Grant 2010).

At the same time, thanks to the openness of Twitter’s API, Twitter data have been a major source for academic research in numerous fields. One of those fields deals with the sentiment analysis of tweets. Sentiment analysis, which is itself an important Natural Language Processing (NLP) topic, has been utilized in Twitter with various degrees of success. One of the findings in related literature suggests that negatively charged tweets have a bigger network penetration than average (Tsugawa and Ohsaki 2017; Naveed et al. 2011; Jiménez-Zafra et al. 2021). As to be expected, political tweets, due to their inherent interest, have been also extensively studied. However, even though there have been studies of sentiment in tweets revolving around politics/elections (Tumasjan et al. 2010; Chung and Mustafaraj 2011; Parmelee and Bichard 2011; Llewellyn and Cram 2016), to our knowledge there has not been a large-scale study of the relation between sentiment and the propagation of politicians’ tweets.

In this paper, we focus on politicians’ tweets to understand the relation between their sentiment and virality. By performing a more finegrained analysis, a distinction is made between politicians from different political parties and we attempt to identify differences in their tweeting activities. At the same time, an investigation is performed on whether politicians’ behaviour, regarding their tweet sentiment, is independent of the country where they are politically active, and also whether it is consistent or evolves over time.

For this purpose, we bring together and assess the validity of these and other research questions with regards to politicians and Twitter by (1) collecting a large-scale dataset of recent tweets by members of parliament (MPs) in three different countries (Greece, Spain and United Kingdom) and multiple languages; (2) establishing a robust evaluation making use of state-of-the-art multilingual sentiment analysis models powered by the recent successes of transformer-based language models; and (3) performing a multi-faceted analysis including control experiments for robustness across different aspects including parties, time and location, among others.

2 Related Work

The popularity of social media platforms such as Facebook and Twitter is transforming the way citizens and politicians communicate with one another. Political candidates and voters use Twitter to discuss social and political issues, sharing information and encouraging political participation (Kushin and Yamamoto 2010). Politicians in particular, especially in recent years, have eagerly embraced social media tools to
self-promote and communicate with their electorate, seeing in these tools the potential for changing public opinion especially during election campaigns (Hong and Nadler 2011). Given the rapid growth of politician’s engagement through Twitter, there is plenty of research on how the platform is used for political communication.

Many studies across the globe focus on the classification of tweets referring to politicians by sentiment — positive, negative, or neutral — to investigate popularity and voting intention, and whether there is a correlation between post sentiments and political results (Taddy 2013; Vilares, Theilwall, and Alonso 2015; Kermanidis and Maragoudakis 2013). Moreover, sentiment is considered to affect message diffusion in social media. Research suggests that the virality of a message — the probability of it being retweeted — is affected by its polarity, as emotionally charged messages tend to be reposted more rapidly and frequently compared to neutral ones (Stieglitz and Dang-Xuan 2011). Negative messages in particular are likely to be more retweeted than positive and neutral ones (Tsugawa and Ohsaki 2015). However, other studies show that the relationship between sentiment and virality in Twitter is more complex and is related to subject area (Hansen et al. 2011). Literature suggests that sentiment occurring in politically relevant tweets has an effect on their propagation (Stieglitz and Dang-Xuan 2011).

In our paper, we attempt to validate these claims and others in the recent political landscape. To this end, we collect a large-scale dataset (see Section 3) and investigate the usage of state-of-the-art multilingual sentiment analysis classifiers based on transformers (see Section 4 for more details and a more detailed background on sentiment analysis), effectively running the first study of this scale.

3 Data Collection

For the purposes of this study, tweets were collected from the MPs (members of parliament) of three sovereign European countries, namely, the United Kingdom (UK), Greece and Spain, including several of their devolved parliaments in Northern Ireland, Scotland and Wales in the UK and the Basque Country and Catalonia in Spain for 2021.

Furthermore, in order to establish relevant comparison points with respect to the specific time period examined, January to December 2021, and the general population, we collected additional Twitter corpora. These included (1) tweets from MPs from the UK’s national (London) parliament for 2014 & 2016, (2) tweets from random users from Greece, UK and Spain and (3) tweets from verified users residing in the aforementioned countries see Section 5.2 for more details.

In total, 2,933,143 tweets were scraped from 157,333 users of which 2,213 represented members of the parliaments for the aforementioned countries. The total number of tweets acquired from politicians were 1,588,970. The collection of the tweets was achieved using Twitter’s API while utilising Python’s Tweepy (Roesslein 2009) and Twarc (Summers 2013) libraries. Retweets were ignored and only original tweets were considered as the main metric to consider the popularity of a tweet (the retweet count) is not available on retweeted messages, which is crucial for our sentiment exploration. Finally, tweets that do not contain meaningful text, e.g., text contains only URLs, were also discarded.

3.1 Main 2021 MP Twitter Corpus

To compile our 2021 dataset, we extracted tweets from the members of parliament of the three sovereign countries analysed, i.e., Greece, Spain and UK and their devolved parliaments. Table 1 displays the total number of tweets collected for each of the months under study for all considered parliaments.

According to Table 1, MPs from the main parliaments (UK, Greece, and Spain) were the most active, with the total number of tweets being 577,279, and 184, respectively. Devolved parliaments had fewer tweets, with Northern Ireland (52), Scotland (59), and Wales (56) being the most active. The total number of tweets for each month is also shown in Table 1. The number of tweets is highest in July and August for most parliaments.

3.2 Control Datasets

UK parliament 2014 & 2016 Aiming to explore the sentiment trends we added a temporal element in our analysis by collecting tweets from UK’s MPs for the years of 2014 and 2016. In this case, the MPs and their respective handles were collected by utilising the SPARQL (W3C 2013) endpoint of Wikidata (Vrandečić and Krötzsch 2014). In total, 202,954 and 429,538 tweets were extracted for 2014 and 2016 respectively. It should be noted that the number of Twitter handles scrapped for each year was smaller than that of 2021; 156 and 474 handles for 2014 and 2016 respectively in comparison to 579 for the 2021 UK dataset. The increased popularity of Twitter over the years can potentially
explain the larger number of handles for 2016 and 2021 over 2014. However, this might also be due to limitations of the Wikidata archive.

Random & Verified Users 2021 To compare politicians’ tweets with those of the general population for 2014, two distinct sets of tweets of random and verified users were also collected (each for any of the country’s parliament studied — Spain, Greece, the UK). Verified users accounts usually belong to recognisable figures such as brands, organisations, and influential persons (e.g. athletes and artists) whose Twitter activity can be deemed to be closer to that of MPs. Each set of random users follows the same distribution of tweets of their respective country shown in Table 1. The geolocalization of tweets was achieved using Twitter’s API country filter (‘place.country’). An assumption was made that tweets belong to users that reside to the country they are posting from. The verified users set was constructed by extracting tweets only from a list of verified users and combination of keywords, alongside with information on the location from user profiles metadata was applied to ensure the accounts resided in the countries studied.

3.3 Sentiment Annotation

Tweets from the 2021 Main Dataset (Section 3.1) for each language included in the study were sampled for their respective parliaments. This way three datasets were collected and annotated based on their sentiment for the English, Spanish and Greek languages (Annotated Set).

In sentiment annotation tasks, annotators are asked to either evaluate the overall polarity of the text on scale, e.g., 1 to 5 (Al-Twairesh, Al-Salman, and Al-Ohali 2017) or to distinguish positive/neural/negative classes (Patwa et al. 2020). For simplicity and to follow current state-of-the-art sentiment analysis models (Nguyen, Vu, and Tuan Nguyen 2020; Barbieri, Anke, and Camacho-Collados 2021), in our setting annotators were asked to indicate the sentiment of each tweet and classify it in one of the following classes:

- **Positive**: Tweets which express happiness, praise a person, group, country or a product, or applaud something.
- **Negative**: Tweets which attack a person, group, product or country, express disgust or unhappiness towards something, or criticise something.
- **Neutral**: Tweets which state facts, give news or are advertisements. In general those which don’t fall into the above 2 categories.
- **Indeterminate**: Tweets where it is not easy to assess sentiment or sentiments of both polarities of approximately the same strength exist. Tweets annotated with the indeterminate label were discarded from our analysis.

For each set of tweets three native speakers were assigned as annotators. Initially 100 tweets were sampled for each language and were given to each group of annotators. The annotators were advised to consider only information available in the text, e.g., to not follow links present, and in cases where a tweet includes only news titles to assess the sentiment of the news being shared.

Table 2 displays the inter-annotation agreement based on Cohen’s Kappa (Cohen 1960). It is observable that for all three language sets the agreement between annotators is satisfactory with the lowest score, 0.69, being in the Spanish set when considering entries labelled as ‘Unidentified’ too. It is also worth noting that the divergence between positive and negative labels (which could be the most problematic in our subsequent analysis) was extremely low. Only 9% (Greece), 3% (Spain) and 7% (UK) of all annotated tweets had a contrasting positive/negative or negative/positive labels between any annotator pair. Finally, in order to consolidate the annotations, the final label of each tweet was decided by using the two annotators agreement in each group and in cases of differences the third annotator was used as a tiebreaker.

| Country | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Total |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
| UK      | 47296 | 47606 | 38965 | 29336 | 34215 | 31081 | 31970 | 22044 | 27509 | 25957 | 31330 | 29026 | 399935 |
| Scotland| 11282 | 13175 | 8988 | 6622 | 7890 | 5600 | 3856 | 5295 | 6285 | 5301 | 6427 | 6241 | 86996 |
| Wales   | 5381 | 5462 | 3903 | 3386 | 3874 | 3013 | 3191 | 2393 | 2864 | 2826 | 3410 | 3084 | 42787 |
| N.Ireland| 5945 | 6788 | 5640 | 4579 | 5127 | 6138 | 4798 | 3484 | 4947 | 4505 | 5255 | 6463 | 74520 |
| Spain   | 20119 | 25343 | 18407 | 18073 | 16938 | 16167 | 14732 | 11595 | 1118 | 5462 | 4452 | 6239 | 5856 |
| Catalonia| 9512 | 8386 | 6463 | 5537 | 6389 | 6030 | 6047 | 3729 | 5940 | 5856 | 5376 | 5255 | 74520 |
| Basque  | 2103 | 2299 | 1478 | 1287 | 1230 | 989 | 685 | 843 | 1040 | 1118 | 1161 | 1091 | 15324 |
| Greece  | 7234 | 9298 | 7421 | 7136 | 6345 | 6140 | 6239 | 4300 | 6060 | 5458 | 5941 | 5083 | 76655 |
| ALL     | 108872 | 118357 | 91265 | 75956 | 82008 | 75158 | 71518 | 53683 | 70021 | 69374 | 71920 | 68346 | 956478 |

Table 1: Number of MP tweets collected through 2021.

Table 2: Inter-annotator agreement based on Cohen’s Kappa

| Country | Including Unidentified entries | Excluding Unidentified entries |
|---------|-------------------------------|--------------------------------|
| UK      | 0.72                          | 0.74                           |
| Spain   | 0.69                          | 0.75                           |
| Greece  | 0.72                          | 0.76                           |
4 Sentiment Analysis: Evaluation and Model Selection

Sentiment analysis is an important NLP task often used to create detailed profiles of users’ behaviours and their world views. It has been extensively used by companies to infer user’s preferences and their attitudes towards products and services (Kauffmann et al. 2020) in order to implement an optimal marketing strategy. With the explosion of popularity of social media, and the inevitable turn of marketing funds from traditional media to them, sentiment analysis remains an invaluable tool for companies and organisations to acquire an understanding of their customers’ loyalty and how they’re brand is perceived to the public (Neri et al. 2012). However, the value of sentiment analysis is not only appreciated in commercial settings. It has been successfully used on numerous social topics varying from tracking the public’s perspective towards the Coronavirus pandemic (Barkur and Vibhaj 2020) and to investigate gender stereotypes (Bhaskaran and Bhallamudi 2019) to provide insights on how voters view their political representatives (Oliveira, Bermejo, and dos Santos 2017).

Over the years there have been multiple approaches of dealing with sentiment analysis in text data. Varying from the use of sentiment lexicons (Inui and Yamamoto 2011) to utilizing linear machine learning models (Ahmad, Aftab, and Ali 2017) and, most recently, by applying transformer models like BERT (Devlin et al. 2019). One of the most challenging problems appears when we have to deal with multilingual data. Acquiring a model that is able to perform well on a multilingual setting is a difficult task that often requires large labelled corpora. This stands especially if low sourced languages are taken into consideration (Barnes, Klinger, and Schulte im Walde 2018). Some cross-lingual approaches, such as language models, deal with this issue by making use of the large amount of training data available in major languages, e.g. English, to essentially transfer sentiment information to low resource languages, e.g. Greek (Can, Ezen-Can, and Can 2018). It is also important to note that despite which architecture is being used an important factor to achieve accurate sentiment classification is the domain of the train and the target corpus (Peng et al. 2018).

In our use case, we select a number of pre-trained language models, both monolingual and multilingual, which we attempt to further finetune and evaluate them using the manually labelled tweets dataset retrieved (Section 3.3) aiming to find the best suitable classifier for each language (English, Spanish, Greek) studied.

4.1 Experimental Setting

Data For the training purposes six different datasets are sourced. For each language their respective set of annotated tweets, Annotated Set (in-domain, see Section 3.3), is being utilised along with another language specific datasets for each language (out-domain): the sentiment analysis dataset from ‘SemEval-2017 Task 4’ (Rosenthal, Farra, and Nakov 2017a) is acquired for English (out-domain for English); the InterTASS corpus (Díaz Galiano et al. 2018) for Spanish (out-domain for Spanish); and a sentiment dataset constituted of tweets related to the 2015 Greek elections for Greek (out-domain for Greek) (Tsakalidis et al. 2018). These additional sources are used only for training purposes. All the datasets have been constructed for the specific task of Twitter sentiment analysis where each tweet is classified as either Positive, Negative, or Neutral. All Twitter handles are anonymized by replacing them by ‘@username’.

Training/Evaluation We consider two different approaches to evaluate our models. Firstly, a train/validation/test split method is applied. The testing data are the subset of tweets from Annotated Set that are cross-annotated by all annotators (approximately 100 tweets for each language), the rest of the entries (approximately 900 tweets per language) are used for training and validation with a 85/15 train/validation ratio. The cross-annotated subsets are used for testing as it is assumed to be more precise than the larger subset annotated by a single person.

Secondly, due to the relatively small size of the Annotated Set a 5-fold cross-validation method is also applied where the whole dataset is used. In cases where a multilingual model is trained, the combined Annotated Set is used along with a stratification method that ensures that all languages are equally represented in each fold. This cross-validation experiment is set to complement the evaluation done in the single train/test split, which may be limiting (Gorman and Bedrick 2019).

Comparison systems Multiple transformer-based language models, including domain-specific models which are state of the art in Twitter-related classification tasks (Barbieri et al. 2020) are selected for our experiment.

1. XLM-R (Conneau et al. 2020), is pre-trained on a large multilingual corpus containing 100 languages (CommonCrawl (Wenzek et al. 2020)).
2. XLM-T (Barbieri, Anke, and Camacho-Collados 2021) further trained on 198 million multilingual tweets.
3. XLM-T-Sent (Barbieri, Anke, and Camacho-Collados 2021), similar training to XLM-T but further finetuned for multilingual sentiment analysis.
4. Three different RoBERTa-Base implementations for each language studied: roberta-base (Liu et al. 2019), roberta-base-bne (Gutierrez-Fandino et al. 2021) and palobert-base-greek-uncased-v1 (Alexandridis et al. 2021) for English, Spanish and Greek respectively.
5. TweetEval-Sent (Barbieri et al. 2020) trained on 58 million English tweets and finetuned for sentiment analysis with the English SemEval-2017 dataset (Rosenthal, Farra, and Nakov 2017a).
6. Bertweet-Sent, based on the Bertweet (Nguyen, Vu, and Tuan Nguyen 2020) implementation and further finetuned for sentiment analysis.

All the models are based on the implementations of the uncased versions provided by Hugging Face (Wolf et al. 2020) and, further finetuned and tested for each language individually, as well as in a multilingual setting using the data collected.
In order to assess the difference between these recent transformer models with more traditional approaches, an SVM model was trained using a combination of frequency features, TF-IDF, and semantic, based on the average of word embedding.\footnote{As pre-trained word embeddings, we used 100-dimensional fasttext embeddings (Bojanowski et al. 2017) trained on Twitter data (Camacho-Collados et al. 2020).}

### Optimization

For all the models’ training the same set of hyper parameters were used. Specifically, Adam optimizer (Loshchilov and Hutter 2017) and a linear scheduler with warmup are applied. We warm up linearly for 50 steps with a learning rate of 5e-5, while a batch size n=16 is used. The models are trained up to 20 epochs, with a checkpoint in every epoch, while an early-stop callback stops the training process after 3 epochs without a performance increase of at least 0.01. We select the best model out of all the checkpoints based on their performance on the validation set.

### Evaluation metrics

We report results both in the usual macro-average F1 and the F1 average between positive and negative classes (F1PN henceforth). For sentiment analysis tasks the average of the F1 scores of the Positive and Negative classes is often used as an evaluation measure (Roththal, Farra, and Nakov 2017b) instead of other metrics such as Accuracy. This is mainly justified as firstly F1 scores are more robust to class imbalance, and secondly due to the fact that classifying correctly Positive and Negative classes is more crucial than the Neutral class, especially in our subsequence analysis.

### 4.2 Results

Table 3 displays both the macro average F1 scores and the average F1 scores for only the Positive and Negative classes F1PN of the models trained for both the train/test split and cross-validation (CV) experimental settings. The performance of the classifiers varies depending not only in their architecture but also on the data that they are trained on. In the UK dataset, the default implementation of ‘Bertweet-Sent’ outperforms all the other models achieving a F1PN of 83%. For Spanish tweets the multilingual version of ‘XLM-T-Sent’ performs the best with F1PN=78% while in the train/test split setting the implementation of ‘XLM-R’ seems to perform better (F1PN=78%) than the ‘Bertweet-Sent’ on the same data set.

### 4.3 Model selection

Considering our use case, classifying tweets from members of parliaments in different countries, we decided against the use of mono-lingual models such as ‘Bertweet-Sent’\footnote{In Section 6.4 we further present a control analysis in which we compare the trends between our selected multilingual model and the best performing ‘Bertweet-Sent’ in English.} for two reasons: (1) there is no certainty that a tweet will follow the same language as the main language of the parliament, e.g., Welsh tweets in the UK parliament, Catalan tweets in the Spanish parliament, Turkish tweets in the Greek parliament, etc.
ment; and (2) using a multilingual model will make the comparison across countries easier. As such, for the purposes of our experiment, the multilingual implementation of ‘XLM-T-Sent’ fine-tuned on our in-domain data is selected as a classifier and applied across all of the data collected. Our choice is further justified as the selected option produces consistently strong results in all countries (73%, 87%, 76% F1 score for UK, Spain, and Greece, respectively, when considering the train/test split setting), including state-of-the-art results for Spanish and Greek.

5 Analysis

Having acquired a suitable sentiment analysis classifier capable to successfully distinguish sentiment polarity in MP tweets (see Section 4.3), we applied it to our collected Twitter corpora (see Section 3) and perform an in-depth analysis aiming to explore whether politicians tweets containing negative sentiment have a bigger network penetration than tweets that are positive or neutral.

Initially, we attempt to establish what is considered a ‘popular’ tweet in the context of our analysis. Table 4 displays the percentiles of how many times a tweet has been retweeted for the UK, Spanish, and Greek parliaments (2021). It is noticeable that the vast majority of tweets are retweeted only a few times; 75% of tweets having a retweet count below 40 across all parliaments, indicating a long tail in the distribution of the retweets count. In our analysis we consider a tweet to be ‘popular’ and to belong in the ‘Head’ of the distribution when it is included in the top 5% percentile of the retweets count. On the other hand, a tweet is labelled as ‘not popular’ and it belongs in the ‘Tail’ of the distribution when it falls under the 50% percentile.

| Parliament | Pct 0% | 25% | 50% | 75% | 100% |
|------------|-------|-----|-----|-----|------|
| UK         | 0     | 0   | 2   | 9   | 36689|
| Spanish    | 0     | 1   | 5   | 36  | 25132|
| Greek      | 0     | 0   | 3   | 9   | 3357 |

Table 4: Percentiles of retweet counts for UK, Spanish and Greek parliaments from January to December 2021

5.1 Sentiment and Virality

To answer one of the main research questions in this paper, we investigate whether there exists a correlation between retweets count and sentiment. As the comparison includes numerical (retweets count) and categorical values (sentiment) approaches such as the Spearman’s correlation are not suited. Instead we perform a Chi-square test which indicates the existence of a dependency between popularity and sentiment. Then the Kruskal-Wallis H-test is performed on the retweets count populations for Positive, Negative and Neutral charged tweets to test whether their median values differ. Our test clearly confirms the existence of a difference in the distributions of the populations among sentiments ($p$-value < $10^{-10}$, $a=0.05$). This test was performed in our 2021

$p$-value is lower than the minimum accepted value in Python

6

Main Dataset (see Section 5.1). Even though there is no evidence for direct correlation, we manage to establish that there is a relation between sentiment and popularity, and also that retweets count distributions differ between sentiments.

Following this, we consider the tweets made from politicians of the main parliaments of the UK, Spain, and Greece separately, and compare them based on their popularity; ‘Head’ and ‘Tail’ parts. Figure 1 displays the normalized counts of positive, negative and neutral tweets in the 2021 Main Dataset. Looking at the overall tweets distribution, negative account for a higher percentage for the Spanish and Greek parliaments whilst the reverse is true for the UK. However, when comparing the most ‘popular’ tweets (‘Head’) to those having only a small number of retweets count (‘Tail’), there is clear pattern with negative charged messages being more numerous for all parliaments. For the UK parliament when comparing the ‘Head’ and ‘Tail’ sets, the proportion of negative tweets is higher for the most ‘popular’; ‘Tail’ tweets with negative emotion are 65% less than those of the ‘Head’, whilst positive tweets display a 121% increase between ‘Head’ and ‘Tail’. Similarly, negative tweets are 71% and 88% more numerous in the Spanish and Greek parliaments, respectively, when comparing most to least popular tweets.

Figure 1: Sentiment distribution (Overall, Head, Tail) for tweets from the UK, Spanish, Greek parliaments in 2021. Negative sentiment: red, Positive sentiment: green, Neutral sentiment: grey.

5.2 Governing Party vs Opposition

Figure 2 presents the difference in sentiment for the top five major parties (based on the number of MPs) in the UK’s, Spain’s and Greece’s parliaments for 2021. Taking into account the governing parties of each country (UK: Conservatives, Spain: Spanish Socialist Workers’ Party (PSOE) &
Table 5: Overall, Head & Tail distributions of sentiment for the datasets utilized. Numbers in the table correspond to the percentage of tweets labeled as negative, neutral and positive for each subset.

Table 5: Overall, Head & Tail distributions of sentiment for the datasets utilized. Numbers in the table correspond to the percentage of tweets labeled as negative, neutral and positive for each subset.

Unidas Podemos (UP), Greece: New Democracy) a distinctive trend appears. In each country, the main governing parties tweets tend to be significantly more positive than those of the opposition. In the UK, there are 46% positive tweets more than negative posted by the ruling party (Conservatives) whereas only a 10% difference for the main opposition (Labour party). The same pattern appears for Greece with positive charged tweets posted by New Democracy being 28% more than the negative. At the same time, the opposition party, Syriza, being on the ‘attack’ has 59% of its total tweets posted classified as negative. Similarly, for Spain, even though to a smaller degree, PSOE display the biggest contrast between positive and negative tweets with only a 6% difference in favour of positive tweets. On a final note, it is interesting to observe that the VOX and Greek Solution, two right-wing parties, display the largest percentages of negative tweets for their respective parliaments (60% and 79%).

This behaviour becomes even more apparent when we take into consideration only the tweets from the leaders of the governing and opposition parties (Table 5). The UK’s Prime Minister (Boris Johnson) shares a considerably larger percentage of positive tweets compared to the opposition leader Keir Starmer (79% vs. 58%) and only 5% of his tweets being negative, in contrast to the 35% for Keir Starmer. An even bigger contrast is seen between the political leaders in Spain where 84% of all tweets from President Sánchez are negative, compared to only 30% from the opposition leader Pablo Casado. In Greece the trend is similar, with a high contrast of positive/negative tweets between the Prime Minister Mitsotakis and the opposition leader Tsipras (77%/17% vs. 29%/59%).

These observations support the hypothesis that politicians use Twitter to promote their agenda and influence the public. Not surprisingly, the governing parties try to depict a positive image of the state of their country. In contrast, the opposition parties challenge the same positions by using negatively-charged tweets.

5.3 Devolved Parliaments

Aiming to acquire a more detailed representation for each country, we applied our sentiment analysis pipeline to the devolved parliaments of the UK (N.Ireland, Scotland, Wales) and Spain (Catalonia, Basque country). Table 5 displays the sentiment distribution (Overall, ‘Head’ and ‘Tail’) in the main and devolved parliaments for each country. The devolved parliaments of the UK follow a similar pattern as...
Table 6: Sentiment distribution for government leaders (Johnson, Sanchez, Mitsotakis) and opposition (Starmer, Blanco, Tsipras) leaders.

| Parliament | Party    | Name              | Neg | Neu | Pos |
|------------|----------|-------------------|-----|-----|-----|
| UK         | Conservatives | Boris Johnson       | 5   | 16  | 79  |
| UK         | Labour   | Keir Starmer        | 35  | 7   | 58  |
| Spain      | PSOE     | Pedro Sánchez       | 13  | 3   | 84  |
| Spain      | PP       | Pablo Casado Blanco | 65  | 6   | 30  |
| Greece     | ND       | Kyriakos Mitsotakis | 17  | 6   | 77  |
| Greece     | Syriza   | Alexis Tsipras      | 59  | 12  | 29  |

6 Control Experiments

In this section, we present four additional control experiments to test the robustness of our evaluation.

6.1 Popularity metrics

In addition to the raw retweet counts, we also tested additional metrics of popularity to divide the tweets in ‘Head’ and ‘Tail’. As the retweet count is an absolute measure that does not take into account the existing popularity of the user posting a tweet, it may be skewed to favour users with a big number of followers. We attempt to incorporate the popularity of each user and explore whether there are differences in the sentiment trend when using normalized metrics. To achieve this, two new metrics are introduced: (1) the ratio between the retweets count and the follower count of the user; and (2) the ratio of the retweets count to the average number of retweets of the user. This way, a heavily shared tweet from a user that tends to get only few retweets will be considered more ‘popular’ than a similar tweet originating from a user that is retweeted often. These two metrics offer an alternative and more normalized view of popularity.

Figure 3 displays the results of the sentiment distribution for ‘Head’ and ‘Tail’ for the UK’s, Spanish and Greek parlia-
ments using the different popularity metrics. The trends for the three metrics are largely similar, with negative-charged tweets being more popular in all cases. Having established that all three popularity metrics verify the underlying phenomenon, the total retweets count is used as a metric for the rest of our analysis.

6.2 Politicians Vs General Population

We continue our exploration by investigating whether politicians’ tweets spread in the same manner to the general public by comparing the 2021 Main Dataset with a collection of tweets from random and verified users (see Section 3.2 for more details). Figure 4 displays the difference in the sentiment of the tweets amongst these groups for the ‘Head’ and ‘Tail’ of popularity distribution. In contrast to politicians’ tweets, the general population seems to post more positive tweets overall. Moreover, positive tweets are significantly more retweeted in comparison to negative by a big margin. When only considering the most popular tweets (‘Head’), in the UK positive tweets for random users are 26% more numerous than negative, and the same stands for Greece with positive tweets being 15% more populous while in Spain we observe a small difference in favour of positive (0.3%). A similar trend occurs when looking at more influential users (random users with more than 1000 followers) where their most shared tweets are mostly positive; 55%, 46%, 40% portion in ‘Head’ for UK, Spain and Greece respectively.

The above results seem to be contradicting the trends observed for MPs. Even though politicians’ negative tweets are being shared more often it is not the same case for an average/random user. This suggests that users tend to retweet more easily a negatively-charged tweet posted by a politician than from another random user.

On the other hand, there is not a clear distinction based on sentiment when considering only verified users among all countries. UK and Greek verified users tweet positive
messages more than negative. However, the opposite phenomenon is observed for Spain where the proportion of negative charged tweets are significantly higher in the ‘Head’ when the opposite is true when looking at less popular tweets (‘Tail’). This could be evidence that Twitter users are more likely to share negatively charged content when it is originating from widely recognisable and influential accounts (artists, athletes, organisations, etc.) or from figures of authority such as politicians, whose negative messages seem to spread faster overall in all countries analysed.

6.3 Temporal Analysis

Continuing our analysis, we explore whether the tendency where politicians’ negatively charged tweets are more influential is constant through time. To this end, our UK (2021 Main Dataset) (Section 3.1) along with 2014 and 2016 UK datasets (Section 3.2) are utilized. Figure 5 displays the fluctuation of sentiment (Negative and Positive) in tweets from MPs of the UK’s parliament through time.

Again, the tweets are separated in ‘Head’ and ‘Tail’ based on their ‘popularity’. When considering the ‘Head’ of the distribution, it is clear that tweets with negative sentiment polarity are more numerous than those with positive sentiment throughout the three years studied. As a possibly worrying trend, we can observe how the negativity of tweets in the ‘Head’ grows over time, with a 65% of all 2021 MP tweets being negative. On the other hand, in the distribution of tweets for the ‘Tail’ the opposite stands, where positively charged tweets outnumber negative tweets by a large margin in all three years, further confirming the main trends discussed in Section 5.1.

Discussion Even if the trend is clearly negative, the idiosyncrasy of each year could potentially explain this trend. For instance, the large discrepancy between Positive and Negative sentiment both in the ‘Head’ and ‘Tail’ of the distribution that is observed on 2021 could be justified due to the Coronavirus pandemic which affected the UK during that time. It is also interesting to note the general increase in negatively charged tweets from 2014 to 2016, 5% which could be justified due to the talks that took place that year in relation to Brexit and the eventual referendum that took place on June 2016. Further investigation should be required to explain these sociological aspects, not studied in our quantitative research.

6.4 Sentiment Analysis Models Consistency

In order to ensure the validity of our results, a comparison is made between our selected model (multilingual ‘XLM-T-Sent’) and the best performing model for English, ‘Bertweet-Sent’ (see Table 3 for our main sentiment analysis results). Using the 2021 UK dataset as a test ground, the two models agree on the classification of 80% of all tweets while reaching a 0.68 agreement score (Cohen’s kappa). Besides their overall similar performance, it is important to note that the general trend where ‘popular’ tweets tend to have a bigger network penetration still stands when using ‘Bertweet-Sent’. More specifically, when considering the ‘Head’ of the dataset tested, both models indicate that the majority of the tweets convey negative sentiment, with 65% and 62% of tweets, for ‘multilingual XLM-T-Sent’ and ‘Bertweet-Sent’ respectively, being classified as negative. Similarly, when inspecting the ‘Tail’ of the data, again the models seem to be in agreement with the ‘multilingual XLM-T-Sent’ and ‘Bertweet-Sent’ respectively, being classified as negative. Similarly, when inspecting the ‘Tail’ of the data, again the models seem to be in agreement with the ‘multilingual XLM-T-Sent’ classifying 22% of tweets as Negative and 56% as Positive, while ‘Bertweet-Sent’ classifies 18% as Negative and 52% as Positive. The above results, provide further evidence to the robustness of the sentiment analysis classifier.

7 Conclusion

We have presented an analysis of the relation between sentiment and virality when considering politicians’ tweets. By performing an exhaustive search for a successful sentiment classifier we obtained a robust multilingual model capable of accurately identifying sentiment in politicians’ tweets. This
is achieved by utilizing state-of-the-art transformer-based language models, which we also finetune to the domain-specific task at hand. Both the model used in our analysis and the collected dataset of manually annotated tweets used for training and evaluation are made publicly available[7].

Our analysis indicates that there is a strong relationship between the sentiment and the popularity of politicians’ tweets, with negatively charged tweets displaying a larger network penetration than tweets conveying positive sentiment. This phenomenon seems to be consistent across all three sovereign countries analysed, and is independent of location as it is observable in every parliament considered in this study. Our findings are further verified by the control experiments performed. Among these control experiments, a temporal analysis suggests that the trend, where negative tweets are more influential, is consistent and getting more pronounced over time. Finally, while tangential to the main research questions, we observe a clear distinction between government and opposition parties irrespective of their ideology, with government parties and leaders being more positive overall.

Future work could extend our analysis with additional parliaments and a wider set of time periods taken into consideration. Furthermore, a more fine-grained classification process, where sentiment is classified on a scale, e.g., 1–5 or considering various aspects, might be useful to discover subtler relations between sentiment and virality that are unseen when using a three class classification (negative, positive, neutral). Finally, we hope that both our methodology and released multilingual models can be leveraged in future work for subsequent large-scale sociological studies, including other topics.

Ethics Statement In this paper we explore the relation between sentiment and virality in politicians’ tweets with the aim to acquire a better understanding of how politicians utilize twitter and their relations with the public. As the aim of this work was to identify general trends, only aggregate statistics are displayed and no attempt is made to identify or focus upon individual MPs. In this way our experiments respect the privacy of individuals and also comply with Twitter’s policies (https://developer.twitter.com/en/developer-terms). At the same time, as the main focus of our analysis deals with public figures (MPs) the content of our analysis is by definition addressed to the general population. All of the data used in the experiments are public and accessible through Twitter and are also made available in our repository where we share the tweet IDs that are used.

We note that some of our experiments identify specific groups and aim to differentiate between them. Mainly, a comparison is made between political parties (Figure 2) and their usage of sentiment in tweets; however, our analysis remains neutral, considering only the objective differences between governing and opposition parties, irrespective of their political stances. Nonetheless, we are aware that our analysis, and potential use of our dataset by others, can be potentially used in a politically-biased way.

Acknowledgements

Jose Camacho-Collados is supported by a UKRI Future Leaders Fellowship.

We also acknowledge the help of Ángela Collados Aís, Carla Perez Almendros, Dimitra Mavridou, Mairi Antypa, David Owen, Matthew Redman and David Humphreys in the annotation task.

References

Ahmad, M.; Aftab, S.; and Ali, I. 2017. Sentiment analysis of tweets using svm. Int. J. Comput. Appl.

Al-Twairesh, Al-Khalifa, H.; Al-Salman, A.; and Al-Ohali, Y. 2017. Arasenti-tweet: A corpus for arabic sentiment analysis of saudi tweets. Procedia Computer Science.

Alexandridis, G.; Varlamis, I.; Korovesis, K.; Caridakis, G.; and Tsantilas, P. 2021. A Survey on Sentiment Analysis and Opinion Mining in Greek Social Media. Information.

Banea, C.; Mihalcea, R.; and Wiebe, J. 2008. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In LREC.

Barbieri, F.; Anke, L. E.; and Camacho-Collados, J. 2021. XLM-T: A Multilingual Language Model Toolkit for Twitter. arXiv preprint arXiv:2104.12250.

Barbieri, F.; Camacho-Collados, J.; Espinosa Anke, L.; and Neves, L. 2020. TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification. In Findings of the Association for Computational Linguistics: EMNLP 2020.

Barkur, G.; and Vibha, G. B. K. 2020. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian journal of psychiatry.

Barnes, J.; Klinger, R.; and Schulte im Walde, S. 2018. Bilingual Sentiment Embeddings: Joint Projection of Sentiment Across Languages. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

Bhaskaran, J.; and Bhallamudi, I. 2019. Good Secretaries, Bad Truck Drivers? Occupational Gender Stereotypes in Sentiment Analysis. In Proceedings of the First Workshop on Gender Bias in Natural Language Processing.

Bojanowski, P.; Grave, E.; Joulin, A.; and Mikolov, T. 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics.

Camacho-Collados, J.; Doval, Y.; Martínez-Cámara, E.; Espinosa-Anke, L.; Barbieri, F.; and Schockaert, S. 2020. Learning Cross-lingual Embeddings from Twitter via Distant Supervision. In Proceedings of ICWSM.

Can, E. F.; Ezen-Can, A.; and Can, F. 2018. Multilingual sentiment analysis: An rmn-based framework for limited data. arXiv preprint arXiv:1806.04511.
Cattaneo, B. 2020. Social media influences our political behaviour and puts pressure on our democracies, new report finds.

Chung, J. E.; and Mustafaraj, E. 2011. Can collective sentiment expressed on twitter predict political elections? In Twenty-fifth AAAI conference on artificial intelligence.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*.

Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, E.; Ott, M.; Zettlemoyer, L.; and Stoyanov, V. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.

Díaz Galiano, M. C.; Martínez Cámara, E.; García Cumbreras, M. À.; García Vega, M.; and Villena Román, J. 2018. The democratization of deep learning in TASS 2017. *Procesamiento de Lenguaje Natural*.

Gorman, K.; and Bedrick, S. 2019. We Need to Talk about Standard Splits. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*.

Grant, W. J.; Moon, B.; and Busby Grant, J. 2010. Digital dialogue? Australian politicians’ use of the social network tool Twitter. *Australian journal of political science*.

Gutiérrez-Fandiño, A.; Armengol-Estapé, J.; Pàmies, M.; Llop-Palao, J.; Silveira-Ocampo, J.; Carrino, C. P.; González-Agirre, A.; Armentano-Oller, C.; Rodríguez-Penagos, C.; and Villegas, M. 2021. *Spanish Language Models*.

Hansen, L. K.; Arvidsson, A.; Nielsen, F. Á.; Colleoni, E.; and Etter, M. 2011. Good friends, bad news-affect and virality in twitter. In *Future information technology*.

Hong, S.; and Nadler, D. 2011. Does the Early Bird Move the Polls? The use of the social media tool Twitter by US politicians and its impact on public opinion. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*.

Inui, T.; and Yamamoto, M. 2011. Applying Sentiment-oriented Sentence Filtering to Multilingual Review Classification. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAIIP 2011)*.

Jiménez-Zafra, S. M.; Sáez-Castillo, A. J.; Conde-Sánchez, A.; and Martín-Valdivia, M. T. 2021. How do sentiments affect virality on Twitter? *Royal Society Open Science*.

Kauffmann, E.; Peral, J.; Gil, D.; Ferrández, A.; Sellers, R.; and Mora, H. 2020. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*.

Kermainidis, K. L.; and Maragoudakis, M. 2013. Political sentiment analysis of tweets before and after the Greek elections of May 2012. *International Journal of Social Network Mining*.

Kushin, M. J.; and Yamamoto, M. 2010. Did social media really matter? College students’ use of online media and political decision making in the 2008 election. *Mass communication and society*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Llewellyn, C.; and Cram, L. 2016. Brexit? analyzing opinion on the UK-EU referendum within Twitter. In *Tenth International AAAI Conference on Web and Social Media*.

LoSshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Naveed, N.; Gottron, T.; Kunegis, J.; and Alhadi, A. C. 2011. Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proceedings of the 3rd international web science conference*.

Neri, F.; Aliprandi, C.; Capeci, F.; Cuadros, M.; and By, T. 2012. Sentiment analysis on social media. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

Nguyen, D. Q.; Vu, T.; and Tuan Nguyen, A. 2020. BERTweet: A pre-trained language model for English Tweets. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*.

Oliveira, D. J. S.; Bermejo, P. H. d. S.; and dos Santos, P. A. 2017. Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. *Journal of Information Technology & Politics*.

Parmelee, J. H.; and Bichard, S. L. 2011. *Politics and the Twitter revolution: How tweets influence the relationship between political leaders and the public*. Lexington books.

Patwa, P.; Aguilar, G.; Kar, S.; Pandey, S.; Pykl, S.; Gambäck, B.; Chakraborty, T.; Solorio, T.; and Das, A. 2020. Semeval-2020 task 9: Overview of sentiment analysis of code-mixed tweets. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*.

Peng, M.; Zhang, Q.; Jiang, Y.-g.; and Huang, X.-J. 2018. Cross-domain sentiment classification with target domain specific information. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Roesslein, J. 2009. *Tweepy Documentation*. [Online] http://tweepy.readthedocs.io/en/v3/.

Rosenthal, S.; Farra, N.; and Nakov, P. 2017a. *SemEval-2017 task 4: Sentiment analysis in Twitter*. In *Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017)*.
Rosenthal, S.; Farra, N.; and Nakov, P. 2017b. SemEval-2017 Task 4: Sentiment Analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).

Stieglitz, S.; and Dang-Xuan, L. 2011. The role of sentiment in information propagation on Twitter—An empirical analysis of affective dimensions in political tweets. *ACIS 2011 Proceedings - 22nd Australasian Conference on Information Systems*.

Stieglitz, S.; and Dang-Xuan, L. 2012. Political communication and influence through microblogging—An empirical analysis of sentiment in Twitter messages and retweet behavior. In *2012 45th Hawaii international conference on system sciences*.

Stieglitz, S.; and Dang-Xuan, L. 2013. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of management information systems*.

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*.

Summers, E. 2013. Archive tweets from the command line. Online] https://pypi . org/project/citepy/ .

Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; and Stede, M. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics* 267–307.

Taddy, M. 2013. Measuring political sentiment on Twitter: Factor optimal design for multinomial inverse regression. *Technometrics*.

Tsakalidis, A.; Papadopoulos, S.; Voskaki, R.; Ioannidou, K.; Boididou, C.; Cristea, A. I.; Liakata, M.; and Komnatsiaris, Y. 2018. Building and evaluating resources for sentiment analysis in the Greek language. *Language resources and evaluation*.

Tsugawa, S.; and Ohsaki, H. 2015. Negative messages spread rapidly and widely on social media. In *Proceedings of the 2015 ACM on conference on online social networks*.

Tsugawa, S.; and Ohsaki, H. 2017. On the relation between message sentiment and its virality on social media. *Social Network Analysis and Mining*.

Tumasjan, A.; Sprenger, T.; Sandner, P.; and Welpe, I. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Vilares, D.; Thelwall, M.; and Alonso, M. A. 2015. The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets. *Journal of Information Science*.

Vrandečić, D.; and Krötzsch, M. 2014. Wikidata: A Free Collaborative Knowledgebase. *Commun. ACM*.

W3C. 2013. SPARQL 1.1 Query Language. https://www.w3.org/TR/sparql11-query/. (Accessed on 10/08/2021).

Wenzek, G.; Lachaux, M.-A.; Conneau, A.; Chaudhary, V.; Guzmán, F.; Joulin, A.; and Grave, E. 2020. CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data. In *Proceedings of the 12th Language Resources and Evaluation Conference*.

Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; Davison, J.; Shleifer, S.; von Platen, P.; Ma, C.; Jernite, Y.; Plu, J.; Xu, C.; Scao, T. L.; Gugger, S.; Drame, M.; Lhoest, Q.; and Rush, A. M. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.