Negativity spreads faster: A large-scale multilingual twitter analysis on the role of sentiment in political communication

Dimosthenis Antypas [*], Alun Preece, Jose Camacho-Collados

Cardiff NLP, School of Computer Science and Informatics, & Crime and Security Research Institute, Cardiff, CF24 3AA, Wales, United Kingdom

ARTICLE INFO

ABSTRACT

Social media has become extremely influential when it comes to policy making in modern societies, especially in the western world, where 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 agendas aiming to influence voter behaviour. In this paper, we attempt to analyse tweets of politicians from three European countries and explore the virality of their tweets. Previous studies have shown that tweets conveying negative sentiment are likely to be retweeted more frequently. By utilising state-of-the-art pre-trained language models, we performed sentiment analysis on hundreds of thousands of tweets collected from members of parliament in Greece, Spain and the 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 differences between political parties as well as between politicians and the general population.

1. Introduction

In recent years social media have come to resemble a 'battleground' between politicians who constantly aim to reach out to 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 [1]. 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 [2], to such an extent that political accounts are often more active than non-political ones [3]. Due to their inherent interest and thanks to the openness of Twitter's API, political tweets have been extensively studied. Using tweets extracted from politicians and their followers, researchers have attempted to answer questions such as: is it possible to predict elections results [4]? or what makes a politician popular on social media [5]?

In general, discovering what makes a message viral has been a popular research topic. Sentiment analysis is one of the tools that have been extensively used to answer this question. Findings in related literature suggest that negatively charged tweets have a bigger network penetration than average [6–8]. However, even though there have been studies of sentiment in tweets revolving around politics/elections [9–11], to our knowledge there has not been a large-scale multilingual

analysis 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 fine-grained analysis, a distinction is made between politicians from different political parties as we attempt to identify differences in their tweeting activities. At the same time, we investigate whether politicians' behaviour regarding their tweet sentiment is independent of their home country and language, and also whether the behaviour is consistent or evolves over time.

Overall, we bring together and assess the validity of these and other research questions (Section 2) 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 (Section 4); (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 (Section 5); and (3) performing a multi-faceted analysis to investigate relations between sentiment and virality of tweets (Section 6).

* Corresponding author.
  E-mail addresses: antypasd@cardiff.ac.uk (D. Antypas), preecead@cardiff.ac.uk (A. Preece), camachocolladosj@cardiff.ac.uk (J. Camacho-Collados).

In addition to the specific sections where our main research questions are addressed, we included relevant control experiments for robustness across different aspects in Section 7. These control experiments help put in context the prior research questions and provide extra insights about those. Finally, all our findings linked to these research questions are summarised in Section 8.

2. Research questions

The aims of this paper are summarised by the following three research questions.

RQ1. What is the best sentiment analysis classifier when dealing with political content in social media? This question is mainly answered in Section 5. To answer this question, we first constructed an ad hoc sentiment analysis dataset with Twitter messages of members of parliament in Greece, Spain and United Kingdom. Then, we compared a wide range of settings and sentiment analysis models, including language models.

RQ2. Is there a correlation between virality and sentiment in political communication in social media? This question is mainly answered in Section 6.1. To answer this question, we relied on state-of-the-art sentiment analysis models (as tested in RQ1) and analysed a large corpus of almost one million tweets comprising all Twitter posts by members of parliament from Greece, Spain and United Kingdom. Then, we ran various statistical tests to measure and understand the possible correlation between virality and sentiment.

RQ3. Is there a difference in the sentiment of social media posts between (a) government and opposition, and (b) devolved parliaments? These questions are mainly answered in Sections 6.2(a) and 6.3(b). To answer these questions, in addition to analyse the data by all MPs in the main governments, we extracted and analysed data split by political party and extended to members of devolved parliaments in Spain (Catalonia and Basque Country) and United Kingdom (Northern Ireland, Scotland and Wales).

3. 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 [12]. 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 [13]. Given the rapid growth of politicians' engagement through Twitter, there is plenty of research on how the platform is used for political communication.

Many studies 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 [14–16]. 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 re-posted more rapidly and frequently compared to neutral ones [17]. Negative messages in particular are likely to be more retweeted than positive and neutral ones [18]. However, other studies show that the relationship between sentiment and virality in Twitter is more complex and is related to subject area [19]. Literature suggests that sentiment occurring in politically relevant tweets has an effect on their propagation [20,21].

When considering techniques used to extract sentiment from political text in social media it is common to utilise dictionary based approaches [22–24] or in cases where the platform offers the functionality to react in one's post (e.g., Facebook's like/angry/happy reactions)

then an aggregation of such reactions is used to determine the sentiment of the post [25,26]. These approaches have the benefit of not requiring to train machine learning models for the sentiment analysis task which itself can be a time consuming process while also requiring previously annotated data. Other researchers chose to train their own ML models instead and often utilise neural network architectures such as LSTM and GRU networks [27,28]. Despite the variety of methods utilised, there seems to be a lack of usage of state-of-the-art NLP methods like language models such as BERT [29] and RoBERTa [30]. In Section 5, we show how this makes an important difference in practice given the important improvement attained by language models in NLP in recent years.

Finally, politics and social media is a popular research area in multiple countries across the globe, particularly where platforms such as Twitter and Instagram have a strong presence. For example, studies focusing in the US political scene have tried to identify popularity measures [31], trending topics [22], what drives public engagement [24] and relations between sentiment and politicians popularity [32]. Similar research can be found in other countries such as UK [33], but also in non-English speaking countries such as Russia [27], Mexico [26], Italy [34], Germany [21], Austria [35] as well as in cross-European settings [23].

4. 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[1]: Northern Ireland, Scotland and Wales in the UK and the Basque Country and Catalonia in Spain for 2021.

Countries selection. The above countries were chosen because even though they share similar characteristics (e.g., they are all European) they have distinct attributes that makes a comparison between them more interesting. These distinctions are both linguistic, where English, Spanish and Greek are not closely related languages, and also socioeconomic.

For example, the UK has a population of 67.33 million and a GDP per capita of 47,334$ and is the largest and richest country of the group, followed by Spain with a population of 47.33 million and GDP per capita of 30,115$, and finally Greece which is the least populated (10.67 million) and with a smaller economy (20,276$ GDP per capita).[2] At the same time, the three countries seem to hold different views to social subjects such as religion [36] and LGBTI inclusion [37]. Socioeconomic differences as the above, have been shown to affect public perception on national issues (e.g., sense of solidarity [38]) but also the way politicians engage with the public [39].

As a final note, due to the nature of the analysis, it is beneficial that our research team includes researchers that have a good understanding of the languages as well as awareness of the political environments of the countries studied.

Time period selection. Aiming to minimise any potential biases that could occur during election periods, January to December 2021 was chosen as the time period for our main analysis. The last national elections for the three main parliaments took place in 2019 and thus we can assume that any campaigning rhetoric will be absent from the tweets. Finally, this time period is particularly interesting as all three countries had to deal with the common issue of the COVID-19 pandemic.

---

[1] We decided to include these devolved parliaments due to their size, idiosyncrasy and nationalist identity.

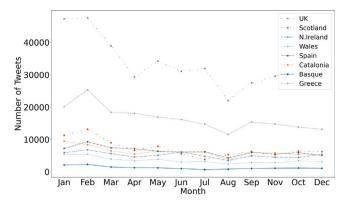[2] Demographic data extracted from: https://data.worldbank.org/.

Fig. 1. Number of MP tweets collected through 2021.

Country	#	@	emoji	url	up/low
UK	0.05	0.23	0.23	0.66	0.10
Spain	0.05	0.33	0.25	0.70	0.11
Greece	0.09	0.27	0.14	0.72	0.12
Overall	0.05	0.27	0.23	0.68	0.11

Control data. In order to establish relevant comparison points with respect to the specific time period examined 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 4.2 for more details).

Collection method. In total 2,933,143 tweets were scraped from 157,333 users of which 2213 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 [40] and Twarc [41] libraries. Retweets were ignored; only original tweets were considered, as the main metric to measure 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.

4.1. 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. Fig. 1 displays the total number of tweets collected for each of the months under study for all considered parliaments.[3]

4.1.1. Parliaments of Greece, Spain and UK

Our main analysis focuses on tweets from members of the UK's, Spain's, and Greece's parliaments from January to December 2021, will be referred to as: *2021 Main Dataset*. For this time period, we scraped a continuous collection of tweets from 1040 members of parliaments (UK: 577, Spain: 279, Greece: 184). The Twitter accounts of the MPs were manually retrieved and verified using either their respective parliament official website or Google search. There are several cases where MPs do not own a Twitter account or they have a protected account which makes the retrieval of their tweets impossible. Consequently, our dataset is not necessarily proportional to the actual parliaments distributions, e.g., in the UK the governing Conservative party has a 56% of the total seats whereas in our UK dataset the Conservative party represents 54% of the total MPs. Finally, it is worth noting that Greek MPs tend to be the least involved with Twitter with only 61% of them actually having an active account in contrast to the UK's and Spanish MPs with 96% and 85% active accounts respectively.

We further analyse the data gathered and explore text features that according to previous research [42,43] are related to the virality of tweets. Table 1 displays text statistics extracted from tweets of the three main parliaments. Specifically, the percentage of tweets that include emojis, hashtags, mentions, and URLs as well as the percentage of upper case characters are shown. The overall majority of the tweets (68%) contain at least one URL which is not surprising as sharing news and commenting on them is one of the main use case of Twitter. It is interesting to note that Greek MPs tend to use fewer emojis than their UK and Spain counterparts (44% and 40% less), while they include hashtags more often (44%). The above statistics are going to be utilised in our regression analysis in Section 6.1.

4.1.2. Devolved parliaments

Tweets from members of the devolved parliaments of the UK (N.Ireland, Scotland, Wales) and Spain (Basque Country, Catalonia) were also collected for the same time period. These subsets, will be referred to as: *2021 Devolved Dataset*, are used to identify potential differences in tweets between the main and the devolved parliaments regarding the prevailing sentiments. Again, a manual search and verification was applied for every MP in order to retrieve his/her Twitter handle. Specifically for Catalonia, care was taken when aggregating the MPs handles due to the elections that took place on February 14th, 2021. Text statistics for tweets gathered from the devolved parliaments are presented in Table A.10 of Appendix.

4.2. Control datasets

4.2.1. 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, will be referred as the *UK 2014 & 2016* dataset. In this case, the MPs and their respective handles were collected by utilising the SPARQL [44] endpoint of Wikidata [45]. In total 202,954 and 429,538 tweets were extracted for 2014 and 2016 respectively. It should be mentioned that the number of Twitter handles scraped for each year was smaller than that of 2021; 156 and 474 handles for 2014 and 2016 respectively in comparison to 577 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.

4.2.2. Random & verified users 2021

To compare politicians' tweets with those of the general population for 2021, two distinct sets of tweets of random and verified[4] users were also collected (each for any of the country's parliament studied — Spain, Greece, the UK), will be referred to as *Random & Verified 2021*. 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

---

[3] Table A.9 of the Appendix provides more detailed results.

[4] According to Twitter, an account is deemed verified if it is authentic, notable and active (https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts).

their respective country shown in Fig. 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 known verified users and a combination of keywords[5] alongside with information on the location from user profiles metadata was applied to ensure the accounts resided in the countries studied.

5. Sentiment analysis: Evaluation and model selection

Over the years there have been multiple approaches of dealing with sentiment analysis in text data. Varying from the use of sentiment lexicons [46–48] to utilising linear machine learning models [49] and, most recently, by applying transformer models like BERT [29]. One of the most challenging problems appears when we have to deal with multilingual data. Acquiring a model that is able to perform well in a multilingual setting is a difficult task that often requires large labelled corpora. This is especially true if low resource languages are taken into consideration [50]. 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 [51]. It is also important to note that despite the architecture being used, an important factor to achieve accurate sentiment classification is the domain of the training and the target corpus [52].

For our purposes, we select a number of pre-trained language models, both monolingual and multilingual, which we attempt to further finetune and evaluate them using a manually labelled tweets dataset (Section 5.1) aiming to find the best suitable classifier for each language (English, Spanish, Greek) studied.

5.1. Sentiment annotation

Tweets from the *2021 Main Dataset* (Section 4.1.1) for each language included in the study were sampled from their respective parliaments. This way three datasets were collected and annotated based on their sentiment for the English, Spanish and Greek languages (will be referred to as: *Annotated Set*).

In sentiment annotation tasks, annotators are asked to either evaluate the overall polarity of the text on a scale, e.g., 1 to 5 [53] or to distinct positive/neutral/negative classes [54]. For simplicity and to follow current state-of-the-art sentiment analysis models [55,56], 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 do not 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

---

[5] https://github.com/cardiffnlp/politics-and-virality-twitter/blob/main/data/keywords.csv

Table 2
Inter-annotator agreement based on Cohen's Kappa.

	UK	Spain	Greece
Including Indeterminate entries	0.72	0.69	0.72
Excluding Indeterminate entries	0.74	0.75	0.76

Table 3
Distribution of labels in the annotated dataset for each country.

Label	Country		
	UK	Spain	Greece
Positive	471	366	275
Neutral	240	155	326
Negative	253	415	364
Indeterminate	36	64	35

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 [57]. 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 'Indeterminate' 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.

Having establish an acceptable agreement between the annotators each one was given 300 extra tweets to label. In total, 964, 936 and 965 tweets were collected and labelled for English, Spanish and Greek respectively (the final numbers slightly vary given the different number of discarded tweets with the indeterminate label). The label distribution of the annotated data appears to be different for each language (Table 3) with negative tweets being the most dominant class in Spanish and Greek, in contrast to English where most tweets were labelled as positive.

5.2. Experimental setting

5.2.1. 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 5.1), is being utilised along with another language specific dataset for each language (*out-domain*): the sentiment analysis dataset from 'SemEval-2017 Task 4' [58] is acquired for English (*out-domain* for English); the InterTASS corpus [59] 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) [60]. 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 anonymised by replacing them by '@username'.

5.2.2. 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. The stratification process was based on existing functionality provided by scikit-learn[6] and aims to preserve the percentage of tweets of each country and not necessary of each sentiment in every fold. This cross-validation experiment is set to complement the evaluation done in the single train/test split, which may be limiting [61].

5.2.3. Comparison systems

Multiple transformer-based language models, including domain-specific models which are state of the art in Twitter-related classification tasks [62] are selected for our experiment.

1. XLM-R [63], is pre-trained on a large multilingual corpus containing 100 languages (CommonCrawl [64]).
2. XLM-T [56] further trained on 198 million multilingual tweets.
3. XLM-T-Sent [56], similar training to XLM-T but further finetuned for multilingual sentiment analysis.
4. Three different RoBERTa-Base implementations for each language studied: roberta-base [30], roberta-base-bne [65] and palobert-base-greek-uncased-v1 [66] for English, Spanish and Greek respectively.
5. TweetEval-Sent [62] trained on 58 million English tweets and finetuned for sentiment analysis with the English SemEval-2017 dataset [58].
6. Bertweet-Sent, based on the Bertweet [55] implementation and further finetuned for sentiment analysis.

All the models are based on the implementations of the uncased versions provided by Hugging Face [67], 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, three baseline models were tested: (1) an SVM model using a combination of frequency features, TF–IDF, and semantic, based on the average of word embeddings (2) a neural network with two LSTM layers while also utilising pre-trained word embeddings[7] and (3) a lexicon based approach by utilising VADER [69].[8]

5.2.4. Optimisation

All models were trained with the same set of hyper parameters. Specifically, Adam optimiser [70] and a linear scheduler with warm-up 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.

As our aim was not to necessarily find the best sentiment classifier but instead to acquire a robust classifier with good overall performance across all the languages studied no hyper-parameter tuning was performed.

---

[6] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html

[7] As pre-trained word embeddings, we used 300-dimensional fasttext embeddings [68] trained on Wikipedia data.

[8] Spanish and Greek texts were first translated to English with Google's Translate API prior to using VADER.

Table 4

Average F1 scores of positive and negative classes ($F1^{PN}$) when trained/evaluated with 5-fold cross-validation. The results displayed are the averages of three runs. Models were trained on both 'in' and 'out' of domain. Both multilingual and monolingual train-setting results are reported.

Train setting	Model	UK	Spain	Greece
Monolingual	XLM-T	80	80	73
	XLM-T-Sent	80	**81**	**75**
	XLM-R	74	80	73
	RoBERTa-Base	80	80	49
	TweetEval-Sent	81	–	–
	Bertweet-Sent	**82**	–	–
	SVM	61	58	44
	LSTM	67	60	57
	VADER	64	63	58
Multilingual	XLM-T	81	80	75
	XLM-T-Sent	80	**81**	**77**
	XLM-R	78	76	76
Random - Baseline		33	30	27

5.2.5. Evaluation metrics

We report results both in the usual macro-average F1 and the F1 average between positive and negative classes ($F1^{PN}$ henceforth). For sentiment analysis tasks the average of the F1 scores of the Positive and Negative classes is often used as an evaluation measure [58] 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 subsequent analysis.

5.3. Results

Table 4 displays the average F1 score for only the Positive and Negative classes, $F1^{PN}$, of the models trained in 5-fold cross-validation experimental setting (More detailed results can be found in Table B.11 in the Appendix). 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 $F1^{PN}$ of 83%. For Spanish tweets the multilingual version of 'XLM-T-Sent' performs the best with $F1^{PN}$ cross-validation score of 81%. Finally, when considering the Greek dataset the results are not as clear as in a cross-validation setting the multilingual implementation of 'XLM-R' seems to perform better ($F1^{PN} = 78\%$) while in the train/test split setting the implementation of 'XLM-R' trained only on both Greek datasets (*out-domain* and *Annotated Set*) performs the best while achieving a similar score.

5.4. Model selection

Considering the focus of our study, classifying tweets from members of parliaments in different countries, we decided against the use of mono-lingual models such as 'Bertweet-Sent'[9] 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; 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.[10] Our choice is further justified as

---

[9] In Section 7.4 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.

[10] This XLM-T-Sent model is publicly available at https://huggingface.co/cardiffnlp/xlm-twitter-politics-sentiment.

Table 5
Percentiles of retweet counts for UK, Spanish and Greek parliaments from January to December 2021. Highlighted are the percentiles used for determining if a tweet belongs in the 'Head' or 'Tail'.

| Parliament | Pct | | | | | |
	0%	25%	**50%**	75%	**95%**	100%
UK	0	0	2	9	52	36689
Spanish	0	1	5	36	185	25132
Greek	0	0	3	9	28	3357

the selected option produces consistently strong results in all countries (73%, 87%, 76% $F1^{PN}$ score for UK, Spain and Greece, respectively, when considering the train/test split setting), including state-of-the-art results for Spanish and Greek.

Robustness across languages:. As a final check we aimed to evaluate 'XLM-T-Sent' performance on languages other than English, Spanish and Greek. To this end, we utilised our Spanish test set which contains 51 entries in Catalan (5.4% of the whole set) and considered the model's performance on these entries. Our results (5-fold CV) indicate that the performance remains satisfactory with an overall Accuracy of 86.6% and $F1^{PN}$ score of 88.9% when considering only the Catalan tweets. For reference, when looking only at the Spanish tweets in the test set, 'XLM-T-Sent' achieves 87.4% and 90% Accuracy and $F1^{PN}$ scores respectively.

The above results, along with the fact that 'XLM-T-Sent' has originally been trained in a total of 100 languages including Catalan, Basque and Welsh [63], and previous work indicating XLM-R's cross-lingual transfer learning capabilities [56,71,72], provide further reassurances that our classifier is capable on handling the majority of tweets collected to satisfying degree.

Discussion. Overall, our results show that language models perform consistently better than traditional machine learning approaches in our setting. Even though our experiments do not necessarily provide a clear answer to RQ1 they do help us acquire a stable and robust sentiment classifier that can be used across multiple languages. The selected multilingual model (i.e., 'XLM-T-Sent' which has been further fine-tuned on our labelled in-domain data) is then utilised for the rest of the experiments, and in particular for attempting to answer research questions RQ2 and RQ3.

6. Analysis

Having acquired a suitable sentiment analysis classifier capable to successfully distinguish sentiment polarity in MP tweets (see Section 5.4) we applied it to our collected Twitter corpora (see Section 4) 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 5 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 5th 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 50th percentile.

6.1. Sentiment and virality

To directly answer the research question RQ2, we investigate whether there exists a correlation between retweets count and sentiment based on our collected data (see Section 4) and sentiment analysis classifiers (see Section 5).

6.1.1. Correlation analysis

As an initial step, we ran several correlation experiments based on statistical testing and regression analysis. For the purposes of these experiments, only tweets from the *2021 Main Dataset* (see Section 4.1) were utilised.

Statistical testing. 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 sentiment labels ($p\text{-value} < 10^{-16}$, $a = 0.05$).[11] 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 retweet count distributions differ between sentiment labels.

Regression analysis. To complement the statistical tests multiple regression models are fitted and the significance of sentiment is examined.[12] In our experimental setting, the retweet count is set as the dependent variable while the existence (or not) of positive and negative sentiment constitutes the independent variables. Neutral sentiment is not taken into account in order to avoid potential collinearity problems. Furthermore, we utilise four lexical statistics (Table 1) as control variables. Specifically, the presence of emojis, URLs, hashtags (#), and mentions (@) in a tweet are used as binary variables.[13] We did not consider features such as number of followers, favourites count or time posted as our main focus was to identify the importance of features extracted from text.

Due to the nature of the problem at hand, i.e., modelling a count variable, and to the highly skewed distribution of the target (see Table 5) we test a negative binomial regression model (*NBR*) [74] and a zero inflated variation of it (*zero-NBR*). Previous research suggests that *NBR* models and the zero inflated variations can be successfully used to estimate tweets popularity (i.e., retweet count) [75–77]. *NBR* is a generalised linear model used for modelling count data where the dependent variable is assumed to follow the negative binomial distribution. In contrast to Poisson regression, which is also used to model count variables, *NBR* does not make the assumption that the mean is equal to variance making it more robust when dealing with over-dispersed data.[14]

For our setting we train six different *NBR* models using both the combined dataset of the UKs, Spanish and Greek parliaments and data from each individual parliament. The results of both *NBR* and *zero-NBR* models (Table 6) indicate that both negative and positive sentiment are statistically significant in predicting the retweet count of a tweet. At the same time, when considering the coefficients of the two variables we can observe that negative sentiment is assigned a greater weight than the positive sentiment with both models assigning coefficient over 5 times larger to negative sentiment when trained on all three parliaments. In fact, for the UK parliament positive sentiment has even an overall negative impact on the retweet count. Finally, it is interesting to note that even though every feature tested appears to be statistically significant, the negative sentiment appears to be the most effective on spreading a tweet.

---

[11] *p-value* is lower than the minimum accepted value in Python.
[12] All regression models were based on the implementations of Python's 'statsmodels' library [73].
[13] The percentage of uppercase characters was not used in order to avoid influencing the rest of the binary predictors.
[14] The zero inflated variations are based on the assumption that the dependent variable's distribution displays an excess of zero-valued observations and combines two processes: one process (logistic regression) that generates zeros and a second one that generates counts according to the negative binomial model or Poisson model.

Table 6
Coefficients for the Negative Binomial Regression (NBM) and zero inflated NBM.

Variable	Overall		UK		Spain		Greece	
	NBR	zero NBR	NBR	zero NBR	NBR	zero NBR	NBR	zero NBR
const	3.0756*	3.4951*	3.2976*	3.7691*	3.373*	3.6191*	2.2644*	2.3446*
emojis	−0.1868*	−0.1886*	−0.2776*	−0.3187*	−0.2431*	−0.1643*	−0.2524*	−0.2535*
has_url	0.4031*	−0.0417*	0.0864*	−0.4065*	0.8029*	0.5433*	−0.1167*	−0.2282*
hashtags	−0.3745*	−0.5002*	−0.2173*	−0.3105*	−0.402*	−0.442*	−0.2642*	−0.2833*
mentions	−0.9496*	−0.6461*	−1.1553*	−0.6113*	−1.0098*	−0.8653*	−0.7287*	−0.6088*
neg	1.4476*	1.3516*	1.3666*	1.231*	1.1611*	1.1297*	0.8549*	0.8589*
pos	0.2766*	0.2438*	−0.0796*	−0.1364*	0.4683*	0.3971*	0.3472*	0.3525*

*Indicates p-value < 0.05.

To verify our regression results we also test Poisson regression [78] and zero inflated Poisson regression[14] models. Poisson regression is also a generalised linear model that assumes that dependent variable follows the Poisson distribution. Similar to *NBR* it is used to model count data but makes stricter assumptions (*mean = variance*). Both implementations confirm our original results indicating that while both negative and positive sentiments are significant, negative sentiment appears to be larger factor on the tweet popularity. Finally, a simple ordinary least square model (*OLS*) is used while using the softmax outputs of *XLM-T-sent* as dependent variable. For each tweet we select the highest softmax score (only negative or positive scores are considered) and assign negative and positive signs to each sentiment respectively. By combining the two variables we acquire an overview on the overall impact of sentiment and model confidence (not only the final labels) on the tweets popularity. Similar to the previous experiments the *OLS* model identifies the sentiment score as significant and assigns a negative weight implying that the more negative a tweet is the more likely is to be shared. Detailed results of all the models tested can be found in the Appendix (Tables C.13 & C.12).

To sum up, the results of the statistical tests and regression analysis show a strong correlation between negative sentiment and virality, which help us to answer RQ2. Specifically, model's coefficients show that the existence of negative sentiment in a tweet has a major positive impact on its popularity independently of the country from where it was posted.

6.1.2. Head and tail distribution analysis

For this extended analysis, 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 following the 'Head' and 'Tail' partitions presented at the beginning of the section. Fig. 2 displays the normalised 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.

6.2. Governing party vs opposition

To answer the first part of the researcher question RQ3 we consider the sentiment distribution on the party level for each parliament. Fig. 3 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
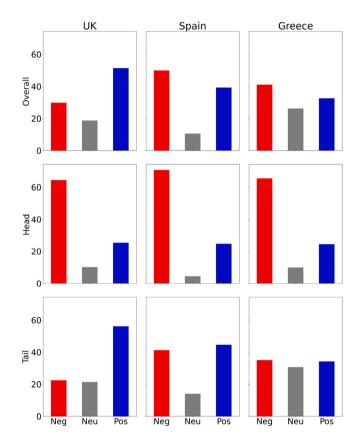

Fig. 2. Sentiment distribution (Overall, Head, Tail) for tweets from the UK, Spanish, Greek parliaments in 2021. From left to right: Negative sentiment (red), Neutral sentiment (grey), Positive sentiment (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(UK: Conservatives, Spain: Spanish Socialist Workers' Party (PSOE) & 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% more positive tweets 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. 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
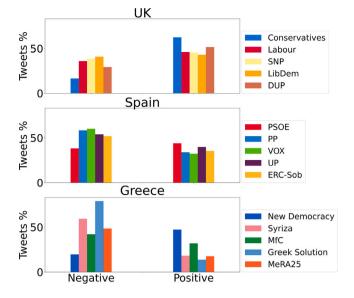
Fig. 3. Sentiment per party for tweets from the UK, Spanish, Greek parliaments 2021. Top five parties displayed, ordered from left to right by their seat number.

Table 7
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
	Labour	Keir Starmer	35	7	58
Spain	PSOE	Pedro Sánchez	13	3	84
	PP	Pablo Casado Blanco	63	6	30
Greece	ND	Kyriakos Mitsotakis	17	6	77
	Syriza	Alexis Tsipras	59	12	29

opposition parties (Table 7). 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%).

Therefore, these observations support the hypothesis that politicians use Twitter to promote their agenda. 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.

6.3. Devolved parliaments

Aiming to acquire a more detailed representation for each country, as well as to answer the second part of the research question RQ3, we applied our sentiment analysis pipeline to the devolved parliaments, *2021 Devolved Dataset*, of the UK (N.Ireland, Scotland, Wales) and Spain (Catalonia, Basque country). Table 8 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 to its main parliament, with more positive than negative tweets overall. In contrast, for the Spanish parliaments there is no consistent pattern with the Basque and main parliaments being dominated by negatively charged tweets whereas in the Catalan parliament the tweets tend to be more positive inclined.

Irrespective of these differences, all the devolved parliaments seem to follow the same general trend where tweets conveying negative sentiment travel further. Similar to their respective main counterparts, we observe that in the 'Head' of each devolved parliament negative tweets tend to be more numerous than positive ones. The only exception being the Welsh parliament (Senedd) where positive tweets are the majority (49% positive to 37% negative). These findings provide more evidence to the hypothesis of a higher network penetration of negatively charged tweets. It is worth noting that both for the Catalan and Basque parliaments our sentiment model classifies as Neutral only 7% and 9% of the total entries which may indicate a higher number of polarised tweets. Moreover, in these regions we can find a more frequent use of less-resourced languages which the sentiment analysis model may find it harder to deal with: Catalan (62% of all entries) and Basque (27% of all entries).

7. Control experiments

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

7.1. 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 normalised metrics. To achieve this, three 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. Finally, we also consider a third popularity metric [79] where a tweet is considered to be viral if its retweet count is at least two standard deviations above the mean for its creator. These three metrics offer an alternative and more normalised view of popularity.

Fig. 4 displays the results of the sentiment distribution for 'Head' and 'Tail' for the UK's, Spanish and Greek parliaments using the different popularity metrics. The trends for the four metrics are largely similar, with negative-charged tweets being more popular in all cases. Having established that all four popularity metrics verify the underlying phenomenon, the total retweets count is used as a metric for the rest of our analysis.

Discussion. As there is no clear definition of what a viral tweet is, it is challenging to investigate the effect of sentiment on the popularity of a tweet. By testing multiple popularity metrics we attempt to strengthen our analysis. The results provide further evidence on the effect of negative sentiment on tweet propagation, irrespective of the chosen popularity metric.

7.2. 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 (*Random & Verified 2021* dataset, see Section 4.2.2 for more details). Fig. 5 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

Table 8
Overall, Head & Tail distributions of sentiment for the datasets utilised. Numbers in the table correspond to the percentage of tweets labelled as negative and positive for each subset.

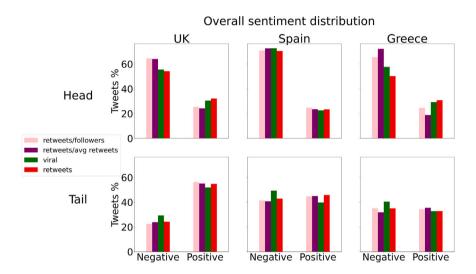| Source	Year	Country	Overall		Head		Tail	
			Neg	Pos	Neg	Pos	Neg	Pos
Parliament	2014	UK	29	41	43	24	27	46
	2016		29	46	48	31	27	47
	2021		30	51	65	25	22	56
		-N.Ireland	36	44	48	34	39	41
		-Scotland	31	53	60	27	26	57
		-Wales	22	60	37	49	23	60
		Spain	50	39	71	25	41	45
		-Catalonia	41	52	68	28	34	58
		-Basque	56	34	75	22	54	36
		Greece	41	33	66	24	35	34
Random		UK	30	51	27	54	31	51
		Spain	46	39	43	43	46	38
		Greece	25	26	28	40	26	25
Verified		UK	23	43	34	38	21	45
		Spain	31	39	45	34	31	38
		Greece	38	42	38	45	39	41

Fig. 4. Sentiment distribution for tweets from the UK, Spanish, Greek parliaments 2021 based on different definitions of *tweet popularity*. (1) 'number of retweets': pink — left bar, (2) 'number of retweets/ number of followers': purple — middle bar, (3) 'viral >= 2 ∗ std retweet count', (4) 'number of retweets/ number of average retweets': green — right bar. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 13% 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 seems 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 no 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.

Discussion. This control experiment was aimed at calibrating the predictions made by the fine-tuned sentiment analysis model. By comparing the overall predictions made on our MP datasets with that of the general politicians and verified users, we certified that the trends for sharing positive and negative tweets are indeed different and not widespread Twitter artefacts. In particular, this analysis showed that the proportion of popular negatively charged tweets is higher when coming from an MP in comparison to the rest of the population.

7.3. Temporal analysis

Continuing our analysis, we explore whether the tendency where politicians' negatively charged tweets are more influential is constant
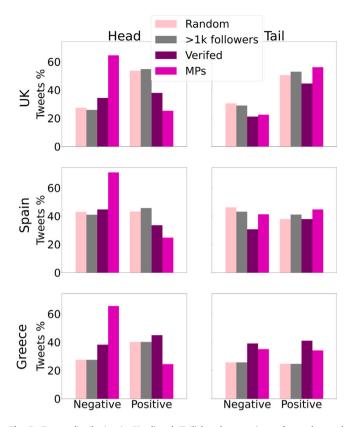
Fig. 5. Tweets distribution in 'Head' and 'Tail' based on sentiment for random and verified Twitter users and Members of Parliament. Bars from left to right: (1) Random users: pink, (2) Random users with more than 1000 followers: grey, (3) Verified users: purple, (4) Members of parliaments: fuchsia. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

through time. To this end, our UK (*2021 Main Dataset*) (Section 4.1.1) along with the *UK 2014 & 2016* dataset (Section 4.2.1) are utilised.[15] Fig. 6 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 6.1.

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.

7.4. 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 B.11 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' 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.

8. General discussion & summary of findings

In this section we summarise the main findings of the paper with respect to the research questions (Section 8.1), contextualise them in the light of the control experiments (Section 8.2), and add a general discussion of the implications of the research (Section 8.3).

8.1. Research questions

Having tested several machine learning models and conducted multiple sentiment analysis experiments, we attempt to answer the research questions we set in Section 2.

Sentiment analysis classifier (RQ1). When considering which is the best sentiment analysis classifier for our particular use case, language models such as 'Bertweet-Sent' and 'XLM-T-Sent' prove to be the most suitable. As shown in our evaluation (Section 5), language models outperform lexicon-based approaches and traditional machine learning models. However, there is no clear winner when selecting between language models. In our use case, we decided to use 'XLM-T-Sent' for most of our experiments as its performance is consistent for all languages studied while also achieving the best scores for Greek and Spanish. Finally, as our dataset contains tweets in other languages too, i.e., Turkish, Catalan, Welsh, using a multilingual model like 'XLM-T-Sent' is more convenient to ensure that minority languages are also represented. On the other hand, if our dataset was made only by English tweets then our results indicate that 'Bertweet-Sent' would have been a better choice.

Virality and sentiment correlation (RQ2). Similarly to what previous research has indicated, our experiments confirm that the sentiment of tweets from MPs is an important factor for predicting their popularity. By utilising the sentiment output of 'XLM-T-Sent' we performed statistical tests (Chi-square and Kruskal–Wallis H-test) which show that there is a statistically significant difference in the distribution of retweets between positive and negative sentiment (being negative tweets more directly correlated with popularity). These results are supported by the Negative Binomial and Poisson regression models that were tested while using as a dependent variable the retweet count and the sentiment as a predictor. The regression results also reveal that negative sentiment has a larger impact on the tweet's network penetration ability. More details about the regression analysis can be found in Section 6.1.
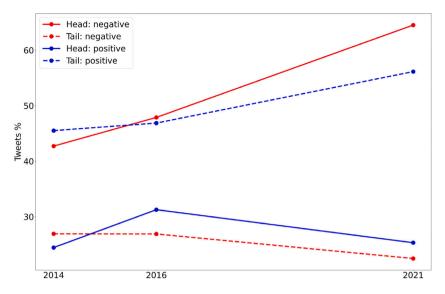
Fig. 6. Sentiment in 'Head' & 'Tail' for tweets from the UK 2014, 2016 and 2021 parliaments.

Sentiment differences in parliaments (RQ3). We continued our analysis with the aim to explore differences in sentiment of tweets made by different parties. Our results reveal that tweets made by the governing parties tend to be more positive in contrast to tweets made by the opposition parties which are in general more negatively charged (Section 6.2). This behaviour stands across all of the three main parliaments studied (UK, Spain and Greece).

Moreover, after investigating potential differences in sentiment between the main and devolved parliaments (Section 6.3) we conclude that even though the overall distribution may differ (i.e., Basque parliament tweets being mostly negative) there is a consistent trend across all parliaments where negatively charged tweets are more popular than tweets conveying positive sentiment.

8.2. Control experiments

To test the robustness of our results, several control experiments were performed. Firstly, we considered different popularity metrics (Section 7.1) including the retweets to followers ratio and the retweets to average user's retweets. Our results indicate that the same general trends, where negatively charged tweets spread further, are consistent independently of which popularity metric is taken into account.

Potential temporal changes were also explored (Section 7.3) by comparing tweets popularity and sentiment from the 2014 and 2016 UK's parliaments. Our experiment confirms that tweets displaying negative sentiment have a bigger network penetration in every year studied while also revealing a trend where the percentage of negatively charged tweets increases through time.

A comparison between politicians and public twitter accounts was also performed (Section 7.2). Our analysis indicate that tweets from public accounts tend to be more positive. However, when considering only tweets from verified accounts differences in the retweet behaviour between countries were identified where in Spain negatively charged tweets were more popular in contrast to UK and Greece where tweets conveying positive sentiment were more retweeted.

Finally, we compare the performance of our model ('XLM-T-Sent') with the best performing English model on the 2021 UK dataset ('Bertweet-Sent') (Section 7.4). Our results show that the models reach a high percentage of agreement (80% agreement on all tweets) and more importantly 'Bertweet-Sent' results confirm that negatively charged tweets are in general more popular than tweets conveying positive sentiment.

8.3. General discussion

Even though in recent years pre-trained language models have been successfully used on various tasks, when dealing with political communication in Twitter, relevant literature suggests a lack of their utilisation. For our experiments we do take advantage of these state-of-the-art language models along with the data manually annotated to improve our sentiment analysis results. This enables us to reliably answer complex research questions at a large scale. Moreover, we establish a clear and robust experimental setting that is easy to replicate and apply on different countries and languages. Our analysis confirms the initial hypothesis that politicians tweets charged with negative sentiment are more popular independently of the country and language studied.

Our results also shed light on the trend where politicians messages tend to be more negative (Section 7.3) in recent times, while also highlighting the polarised political environments present where compromises and agreement between parties appear to be rare (Section 6.2).

It is beyond the scope of this study to consider the wider implications of this trend towards increased negativity in online political communication, or what interventions might mitigate it towards more positive discourse. However, the tools and approach that we have taken can provide a 'dashboard' to monitor and detect any changes in this trend across countries, political spheres, and languages over time, including changes in response to any attempts to mitigate the negative discourse.

We offer both our methodology and released multilingual models in the hope that these can be leveraged in future work for subsequent large-scale sociopolitical studies, including other topics such as studying public reaction to climate change and natural disasters [80]. Finally, our methodology can be easily extended to languages not studied in this work by either applying the sentiment model provided directly or by gathering a small annotated corpus and further finetuning the classifier for more precise results.

9. 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 utilising state-of-the-art transformer-based language models, which we also finetune to the

Table A.9
Number of MP tweets collected through 2021.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
UK	47,296	47,606	38,965	29,336	34,215	31,081	31,970	22,044	27,509	29,557	31,330	29,026	399,935
-Scotland	11,282	13,175	8,988	6,622	7,890	5,600	3,856	5,295	6,285	5,301	6,427	6,241	86,962
-Wales	5,381	5,462	3,903	3,386	3,874	3,013	3,191	2,393	2,864	2,826	3,410	3,084	42,787
-N.Ireland	5,945	6,788	5,640	4,579	5,127	6,138	4,798	3,484	4,947	4,505	4,452	5,391	61,794
Spain	20,119	25,343	18,407	18,073	16,938	16,167	14,732	11,595	15,376	14,753	13,823	13,175	198,501
-Catalonia	9,512	8,386	6,463	5,537	6,389	6,030	6,047	3,729	5,940	5,856	5,376	5,255	74,520
-Basque	2,103	2,299	1,478	1,287	1,230	989	685	843	1,040	1,118	1,161	1,091	15,324
Greece	7,234	9,298	7,421	7,136	6,345	6,140	6,239	4,300	6,060	5,458	5,941	5,083	76,655
ALL	108,872	118,357	91,265	75,956	82,008	75,158	71,518	53,683	70,021	69,374	71,920	68,346	956,478

Table A.10
Percentage of tweets that include hashtags (#), mentions (@), emojis, and URLs in *2021 Devolved Dataset*. The upper to low case ratio (upp/low) average of tweets texts is also reported.

Country	#	@	emoji	url	up/low
N.Ireland	0.04	0.39	0.25	0.45	0.11
Scotland	0.04	0.30	0.29	0.54	0.11
Wales	0.04	0.26	0.31	0.59	0.12
Catalonia	0.07	0.40	0.33	0.69	0.11
Basque	0.05	0.33	0.27	0.63	0.09
Overall	0.05	0.34	0.29	0.58	0.11

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.[16]

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 (i.e., Greece, Spain and United Kingdom). Our findings are further verified by the control experiments performed. Among these control experiments, a temporal analysis suggests that the trend of negative tweets being 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 follow our methodology to 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 utilise 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 (with the exception of the public party leaders). Our experiments therefore 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 (Fig. 3) 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.

CRediT authorship contribution statement

Dimosthenis Antypas: Conceptualization, Investigation, Methodology, Software, Formal analysis, Visualization, Data curation, Writing – original draft. **Alun Preece:** Supervision, Conceptualization, Methodology, Writing – review & editing. **Jose Camacho-Collados:** Supervision, Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jose Camacho Collados reports financial support was provided by UK Research and Innovation.

Data availability

A link to a github repo with the data/code used is included in the paper.

Appendix A. Data collection

Table A.9 displays the number of tweets collected from each parliament for the main time period studied, January to December 2021.

Table A.10 displays text statistics extracted from tweets of the devolved parliaments of N.Ireland, Scotland, Wales, Catalonia and Basque country. The percentage of tweets that include emojis, hashtags, mentions, and URLs as well as the percentage of upper case characters are displayed. In general, the devolved parliaments seem to follow a similar trend to their respective main bodies (Section 4.1.1, Table 1).

---

[16] https://github.com/cardiffnlp/politics-and-virality-twitter

Table B.11

Average F1 scores along with the average of F1 scores of positive and negative classes (F1^{PN}) when trained/evaluated with cross-validation (CV) and on the ~900/100 train/test split. The results displayed are the averages of three runs. Training domain indicates the data used for training the models. 'In' domain indicates models trained on the labelled data collected (Section 5.1), 'Out' indicates the model has been finetuned on other sentiment analysis datasets (see Section 5.2). Finally, 'Mono' indicates that the model was trained with data from only one language, while 'Multi' shows models that were trained with data from all available languages.

Type	Model	Train Lang	Train domain		United Kingdom				Spain				Greece			
					F1		F1^{PN}		F1		F1^{PN}		F1		F1^{PN}	
			Out	In	CV	split	CV	split	CV	split	CV	split	CV	split	CV	split
Multilingual	XLM-T	Mono		✓	70	72	78	75	67	66	80	80	74	76	73	**74**
		Mono	✓	✓	74	65	80	76	70	76	80	81	75	66	73	61
		Multi		✓	75	67	81	77	**72**	78	80	84	76	**77**	75	75
	XLM-T-Sent	Mono	✓		72	72	78	79	68	71	77	82	65	69	64	68
		Mono	✓	✓	74	67	80	74	70	70	**81**	78	77	76	75	**74**
		Multi	✓	✓	74	68	80	75	70	**79**	**81**	**87**	76	70	77	70
	XLM-R	Mono		✓	69	72	77	75	60	65	75	80	74	75	73	73
		Mono	✓	✓	70	67	74	74	66	65	80	78	74	74	73	70
		Multi		✓	74	66	80	73	69	70	80	80	**78**	72	**78**	70
		Multi	✓	✓	70	57	78	67	67	70	76	80	76	72	76	72
Monolingual	RoBERTa-Base	Mono		✓	68	73	78	77	64	70	78	86	56	57	50	52
		Mono	✓	✓	75	70	80	75	70	78	80	**87**	56	62	49	58
	TweetEval-Sent	Mono	✓		73	72	80	79								
		Mono	✓	✓	76	67	81	74								
	Bertweet-Sent	Mono	✓		77	**76**	83	**83**								
		Mono	✓	✓	**78**	75	82	78								
–	SVM	Mono		✓	33	36	50	47	40	40	60	61	60	63	56	60
–		Mono	✓	✓	47	47	61	57	42	38	58	57	52	55	44	51
–	LSTM	Mono		✓	52	48	66	55	51	48	60	61	59	62	56	61
–		Mono	✓	✓	58	55	67	63	49	42	60	56	59	60	57	59
–	VADER				55	47	64	61	54	57	63	66	57	51	58	53
–	Baseline				22	22	33	33	20	21	30	32	17	20	27	30

Table C.12

Coefficients for the OLS model when using tweets from the combined 2021 UK, Spain, Greece parliaments and for each individual parliament. Dependent variable is the retweet count while XLM-T's softmax scores (negative and positive sentiment) are utilised (sent score) as independent variables.

Variable	Overall	UK	Spain	Greece
const	72.8914*	75.9854*	96.6305*	96.6305*
emojis	−5.1085*	−7.9482*	−9.7004*	−9.7004*
has_url	−1.3613	−17.524*	26.56*	26.56*
hashtags	−20.0335*	−7.811*	−41.0476*	−41.0476*
mentions	−41.0511*	−40.4977*	−63.0461*	−63.0461*
max_score	−43.16*	−42.5478*	−39.615*	−39.615*

*Indicates p-value < 0.05.

the rest of the data was used for training (approximately 900 tweet for each language) (Section 5.1). The results indicate the *XLM-T-Sent-multi* performs consistently in both the CV and train/test settings and achieves satisfactory performance particularly for Spanish.

Finally, our results demonstrate the importance that a specialised corpus can have when training the models. In particular, when looking at the $F1$ scores for the 'CV' setting the models display an increased performance when trained further with the 'In' domain, labelled data collected (Section 5.1). The performance of 'XLM-T-Sent', 'TweetEval-Sent' and 'Bertweet-Sent' models is increased (or at worst stays the same), across all languages tested when the 'In' domain data are added to their training corpus.

Appendix B. Extended classification results

Table B.11 displays the macro average F1 scores along with the average F1 scores of positive and negative classes (F1^{PN}) for all the experiments performed. The experiments were performed by using both a cross-validation setting (CV) and a train/test split approach. For the train/test approach the set of tweets that was annotated by all coders was used (approximately 100 tweets for each language) while

Appendix C. Extended regression results

Table C.13 displays the results (coefficients) of the Poisson and zero inflated Poisson regression models while using the retweet count as a dependent variable. Similarly to the *NBR* models used (Section 6.1: *Regression Analysis*) both Poisson models indicate that the existence of negative sentiment influences the retweet count greatly than that of positive sentiment.

Table C.13

Coefficients for the Poisson Regression and zero inflated Poisson Regression when using tweets from the combined 2021 UK, Spain, Greece parliaments and for each individual parliament. Dependent variable is the retweet count while the independent variables *neg* and *pos* indicate the sentiment according to XLM-T results.

Variable	Overall		UK		Spain		Greece	
	Poisson	zero Poisson	Poisson	zero Poisson	Poisson	zero Poisson	Poisson	zero Poisson
const	3.4035*	4.0809*	3.7006*	4.4486*	3.61*	4.1586*	2.3372*	2.8278*
emojis	−0.193*	−0.1769*	−0.3277*	−0.2978*	−0.2114*	−0.1946*	−0.3298*	−0.3639*
has_url	0.013*	−0.3695*	−0.3798*	−0.8422*	0.41*	0.0827*	−0.1871*	−0.305*
hashtags	−0.5219*	−0.6281*	−0.4001*	−0.5168*	−0.5878*	−0.6266*	−0.3372*	−0.41*
mentions	−1.0543*	−0.7944*	−1.4615*	−0.8382*	−0.929*	−0.8291*	−0.9082*	−0.7696*
neg	1.4129*	1.1709*	1.3014*	1.0566*	1.1818*	1.0028*	0.9232*	0.6859*
pos	0.2295*	0.1629*	−0.1486*	−0.1855*	0.5513*	0.4503*	0.3229*	0.2315*

*Indicates p-value < 0.05.

Table C.12 shows the coefficients for each predictor when using the softmax scores of the positive and negative sentiments of the *XLM-T* model instead of the final label. For each entry the highest absolute softmax score is considered and is assigned positive or negative sign based on the sentiment. The results indicate, i.e., negative coefficient, that the "more" negative a tweet the more is retweeted.

References

[1] European Commission, Social media influences our political behaviour and puts pressure on our democracies, new report finds, 2020, EU Science Hub - European Commission, https://joint-research-centre.ec.europa.eu/jrc-news/social-media-influences-our-political-behaviour-and-puts-pressure-our-democracies-new-report-finds-2020-10-27_en. (Accessed 10 July 2021).

[2] S. Stier, A. Bleier, H. Lietz, M. Strohmaier, Election campaigning on social media: Politicians, audiences, and the mediation of political communication on Facebook and Twitter, Political Commun. 35 (1) (2018) 50–74.

[3] W.J. Grant, B. Moon, J. Busby Grant, Digital dialogue? Australian politicians' use of the social network tool Twitter, Aust. J. Political Sci. 45 (4) (2010) 579–604.

[4] A. Tumasjan, T. Sprenger, P. Sandner, I. Welpe, Predicting elections with Twitter: What 140 characters reveal about political sentiment, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 4, 2010, pp. 178–185.

[5] C. Vaccari, R.K. Nielsen, What drives politicians' online popularity? An analysis of the 2010 US midterm elections, J. Inf. Technol. Politics 10 (2) (2013) 208–222.

[6] S. Tsugawa, H. Ohsaki, On the relation between message sentiment and its virality on social media, Soc. Netw. Anal. Min. 7 (1) (2017) 1–14.

[7] N. Naveed, T. Gottron, J. Kunegis, A.C. Alhadi, Bad news travel fast: A content-based analysis of interestingness on Twitter, in: Proceedings of the 3rd International Web Science Conference, 2011, pp. 1–7.

[8] S.M. Jiménez-Zafra, A.J. Sáez-Castillo, A. Conde-Sánchez, M.T. Martín-Valdivia, How do sentiments affect virality on Twitter? R. Soc. Open Sci. 8 (4) (2021) 201756.

[9] J. Chung, E. Mustafaraj, Can collective sentiment expressed on Twitter predict political elections? in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 25, 2011, pp. 1770–1771.

[10] J.H. Parmelee, S.L. Bichard, Politics and the Twitter Revolution: How Tweets Influence the Relationship Between Political Leaders and the Public, Lexington books, 2011, pp. 22–26.

[11] C. Llewellyn, L. Cram, Brexit? Analyzing opinion on the UK-EU referendum within Twitter, in: Tenth International AAAI Conference on Web and Social Media, 2016, pp. 760–761.

[12] M.J. Kushin, M. Yamamoto, Did social media really matter? College students' use of online media and political decision making in the 2008 election, in: New Media, Campaigning and the 2008 Facebook Election, Routledge, 2013, pp. 63–86.

[13] S. Hong, D. Nadler, 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, 2011, pp. 182–186.

[14] M. Taddy, Measuring political sentiment on Twitter: Factor optimal design for multinomial inverse regression, Technometrics (2013).

[15] D. Vilares, M. Thelwall, M.A. Alonso, The megaphone of the people? Spanish SentiStrength for real-time analysis of political tweets, J. Inf. Sci. 41 (6) (2015) 799–813.

[16] K.L. Kermanidis, M. Maragoudakis, Political sentiment analysis of tweets before and after the Greek elections of May 2012, Int. J. Soc. Netw. Min. 1 (3–4) (2013) 298–317.

[17] S. Stieglitz, L. Dang-Xuan, Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior, J. Manage. Inf. Syst. 29 (4) (2013) 217–248.

[18] S. Tsugawa, H. Ohsaki, Negative messages spread rapidly and widely on social media, in: Proceedings of the 2015 ACM on Conference on Online Social Networks, 2015, pp. 151–160.

[19] L.K. Hansen, A. Arvidsson, F.Å. Nielsen, E. Colleoni, M. Etter, Good friends, bad news-affect and virality in Twitter, in: Future Information Technology, Springer, 2011, pp. 34–43.

[20] S. Stieglitz, L. Dang-Xuan, The role of sentiment in information propagation on Twitter – An empirical analysis of affective dimensions in political tweets, in: ACIS 2011 Proceedings - 22nd Australasian Conference on Information Systems, 2011.

[21] S. Stieglitz, L. Dang-Xuan, 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, IEEE, 2012, pp. 3500–3509.

[22] A. Karami, A. Elkouri, Political popularity analysis in social media, in: International Conference on Information, Springer, 2019, pp. 456–465.

[23] T. Heidenreich, J.-M. Eberl, F. Lind, H. Boomgaarden, Political migration discourses on social media: A comparative perspective on visibility and sentiment across political Facebook accounts in Europe, J. Ethn. Migration Stud. 46 (7) (2020) 1261–1280.

[24] S. Rathje, J.J. Van Bavel, S. van der Linden, Out-group animosity drives engagement on social media, Proc. Natl. Acad. Sci. 118 (26) (2021).

[25] R. Sandoval-Almazan, D. Valle-Cruz, Facebook impact and sentiment analysis on political campaigns, in: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 2018, pp. 1–7.

[26] R. Sandoval-Almazan, D. Valle-Cruz, Sentiment analysis of Facebook users reacting to political campaign posts, Digital Gov.: Res. Pract. 1 (2) (2020) 1–13.

[27] K. Svetlov, K. Platonov, Sentiment analysis of posts and comments in the accounts of Russian politicians on the social network, in: 2019 25th Conference of Open Innovations Association, FRUCT, IEEE, 2019, pp. 299–305.

[28] S. Dorle, N. Pise, Political sentiment analysis through social media, in: 2018 Second International Conference on Computing Methodologies and Communication, ICCMC, IEEE, 2018, pp. 869–873.

[29] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, 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, in: Long and Short Papers, vol. 1, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186, http://dx.doi.org/10.18653/v1/N19-1423, URL https://aclanthology.org/N19-1423.

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

[31] Y. Peng, What makes politicians' Instagram posts popular? Analyzing social media strategies of candidates and office holders with computer vision, Int. J. Press/Politics 26 (1) (2021) 143–166.

[32] B. Joyce, J. Deng, Sentiment analysis of tweets for the 2016 US presidential election, in: 2017 IEEE Mit Undergraduate Research Technology Conference, Urtc, IEEE, 2017, pp. 1–4.

[33] A. Mee, E. Homapour, F. Chiclana, O. Engel, Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit, Knowl.-Based Syst. 228 (2021) 107238.

[34] A. Ceron, G. d'Adda, E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election, New Media Soc. 18 (9) (2016) 1935–1955.

[35] E. Kušen, M. Strembeck, Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian presidential elections, Online Soc. Netw. Media 5 (2018) 37–50.

[36] Pew Research Center, How religious commitment varies by country among people of all ages, 2018.

[37] A.R. Flores, Social Acceptance of LGBTI People in 175 Countries and Locations: 1981 to 2020, UCLA School of Law, Williams Institute, 2021, p. 2022, Retrieved February 4.

[38] S. Vasilopoulou, L. Talving, Poor versus rich countries: A gap in public attitudes towards fiscal solidarity in the EU, West Eur. Politics 43 (4) (2020) 919–943.

[39] J.L. Brown-Iannuzzi, K.B. Lundberg, S. McKee, The politics of socioeconomic status: How socioeconomic status may influence political attitudes and engagement, Current Opin. Psychol. 18 (2017) 11–14.

[40] J. Roesslein, Tweepy documentation, 2009, [Online] http://Tweepy.Readthedocs.Io/En/V3.

[41] E. Summers, Archive tweets from the command line, 2013, [Online] https://Pypi.Org/Project/Citepy/.

[42] E. Pancer, M. Poole, The popularity and virality of political social media: Hashtags, mentions, and links predict likes and retweets of 2016 US presidential nominees' tweets, Soc. Influ. 11 (4) (2016) 259–270.

[43] K. Nanath, G. Joy, Leveraging Twitter data to analyze the virality of Covid-19 tweets: A text mining approach, Behav. Inf. Technol. (2021) 1–19.

[44] W3C, SPARQL 1.1 query language, 2013, https://www.w3.org/TR/sparql11-query/. (Accessed 10 August 2021).

[45] D. Vrandečić, M. Krötzsch, Wikidata: A free collaborative knowledgebase, Commun. ACM (2014).

[46] T. Inui, M. Yamamoto, Applying sentiment-oriented sentence filtering to multilingual review classification, in: Proceedings of the Workshop on Sentiment Analysis where AI Meets Psychology, SAAIP 2011, 2011, pp. 51–58.

[47] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Comput. Linguist. (2011) 267–307.

[48] C. Banea, R. Mihalcea, J. Wiebe, A bootstrapping method for building subjectivity lexicons for languages with scarce resources, in: LREC, vol. 8, 2008, pp. 2–764.

[49] M. Ahmad, S. Aftab, I. Ali, Sentiment analysis of tweets using SVM, Int. J. Comput. Appl. 177 (5) (2017) 25–29.

[50] J. Barnes, R. Klinger, S. Schulte im Walde, Bilingual sentiment embeddings: Joint projection of sentiment across languages, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, in: Long Papers, vol. 1, Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 2483–2493, http://dx.doi.org/10.18653/v1/P18-1231, URL https://aclanthology.org/P18-1231.

[51] E.F. Can, A. Ezen-Can, F. Can, Multilingual sentiment analysis: An RNN-based framework for limited data, 2018, arXiv preprint arXiv:1806.04511.

[52] M. Peng, Q. Zhang, Y.-g. Jiang, X.-J. Huang, Cross-domain sentiment classification with target domain specific information, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, in: Long Papers, vol. 1, 2018, pp. 2505–2513.

[53] N. Al-Twairesh, H. Al-Khalifa, A. Al-Salman, Y. Al-Ohali, AraSenTi-Tweet: A corpus for Arabic sentiment analysis of Saudi tweets, Procedia Comput. Sci. 117 (2017) 63–72.

[54] P. Patwa, G. Aguilar, S. Kar, S. Pandey, S. Pykl, B. Gambäck, T. Chakraborty, T. Solorio, A. Das, Semeval-2020 task 9: Overview of sentiment analysis of code-mixed tweets, in: Proceedings of the Fourteenth Workshop on Semantic Evaluation, 2020, pp. 774–790.

[55] D.Q. Nguyen, T. Vu, A. Tuan Nguyen, BERTweet: A pre-trained language model for English Tweets, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 9–14, http://dx.doi.org/10.18653/v1/2020.emnlp-demos.2, URL https://aclanthology.org/2020.emnlp-demos.2.

[56] F. Barbieri, L.E. Anke, J. Camacho-Collados, XLM-T: A multilingual language model toolkit for Twitter, 2021, arXiv preprint arXiv:2104.12250.

[57] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1) (1960) 37–46.

[58] S. Rosenthal, N. Farra, P. Nakov, SemEval-2017 task 4: Sentiment analysis in Twitter, in: Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval-2017, Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 502–518, http://dx.doi.org/10.18653/v1/S17-2088, URL https://aclanthology.org/S17-2088.

[59] M.C. Díaz-Galiano, E. Martínez-Cámara, M.Á.G. Cumbreras, M.G. Vega, J.V. Román, The democratization of deep learning in TASS 2017, Proces. Del Leng. Natural 60 (2018) 37–44.

[60] A. Tsakalidis, S. Papadopoulos, R. Voskaki, K. Ioannidou, C. Boididou, A.I. Cristea, M. Liakata, Y. Kompatsiaris, Building and evaluating resources for sentiment analysis in the Greek language, Lang. Resour. Eval. 52 (4) (2018) 1021–1044.

[61] K. Gorman, S. Bedrick, We need to talk about standard splits, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, 2019, pp. 2786–2791, http://dx.doi.org/10.18653/v1/P19-1267, URL https://aclanthology.org/P19-1267.

[62] F. Barbieri, J. Camacho-Collados, L. Espinosa Anke, L. Neves, TweetEval: Unified benchmark and comparative evaluation for tweet classification, in: Findings of the Association for Computational Linguistics: EMNLP 2020, Association for Computational Linguistics, Online, 2020, pp. 1644–1650, http://dx.doi.org/10.18653/v1/2020.findings-emnlp.148, URL https://aclanthology.org/2020.findings-emnlp.148.

[63] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8440–8451, http://dx.doi.org/10.18653/v1/2020.acl-main.747, URL https://aclanthology.org/2020.acl-main.747.

[64] G. Wenzek, M.-A. Lachaux, A. Conneau, V. Chaudhary, F. Guzmán, A. Joulin, É. Grave, CCNet: Extracting high quality monolingual datasets from web crawl data, in: Proceedings of the 12th Language Resources and Evaluation Conference, 2020, pp. 4003–4012.

[65] A. Gutiérrez-Fandiño, J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C.P. Carrino, A. Gonzalez-Agirre, C. Armentano-Oller, C. Rodriguez-Penagos, M. Villegas, Spanish language models, 2021.

[66] G. Alexandridis, I. Varlamis, K. Korovesis, G. Caridakis, P. Tsantilas, A survey on sentiment analysis and opinion mining in Greek social media, Information 12 (8) (2021) 331.

[67] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2020, pp. 38–45.

[68] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, European Language Resources Association (ELRA), Miyazaki, Japan, 2018, URL https://aclanthology.org/L18-1550.

[69] C. Hutto, E. Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, 2014, pp. 216–225.

[70] I. Loshchilov, F. Hutter, Decoupled weight decay regularization, 2017, arXiv preprint arXiv:1711.05101.

[71] I. Li, P. Sen, H. Zhu, Y. Li, D. Radev, Improving cross-lingual text classification with zero-shot instance-weighting, in: Proceedings of the 6th Workshop on Representation Learning for NLP, RepL4NLP-2021, 2021, pp. 1–7.

[72] A. Rusli, M. Shishido, On the applicability of zero-shot cross-lingual transfer learning for sentiment classification in distant language pairs.

[73] S. Seabold, J. Perktold, Statsmodels: Econometric and statistical modeling with Python, in: Proceedings of the 9th Python in Science Conference, vol. 57, Austin, TX, 2010, pp. 10–25080.

[74] J.M. Hilbe, Negative Binomial Regression, Cambridge University Press, 2011.

[75] S. Bhattacharya, P. Srinivasan, P. Polgreen, Engagement with health agencies on Twitter, PLoS One 9 (11) (2014) e112235.

[76] J. Sutton, C.B. Gibson, N.E. Phillips, E.S. Spiro, C. League, B. Johnson, S.M. Fitzhugh, C.T. Butts, A cross-hazard analysis of terse message retransmission on Twitter, Proc. Natl. Acad. Sci. 112 (48) (2015) 14793–14798.

[77] J. Samuel, M. Garvey, R. Kashyap, That message went viral?! exploratory analytics and sentiment analysis into the propagation of tweets, 2020, arXiv preprint arXiv:2004.09718.

[78] S. Coxe, S.G. West, L.S. Aiken, The analysis of count data: A gentle introduction to Poisson regression and its alternatives, J. Personal. Assess. 91 (2) (2009) 121–136.

[79] Y. Han, T. Lappas, G. Sabnis, The importance of interactions between content characteristics and creator characteristics for studying virality in social media, Inf. Syst. Res. 31 (2) (2020) 576–588.

[80] D.C. Mojica, D. Antypas, J. Camacho-Collados, S. Wilkinson, Sentiment analysis (SA) (supervised and unsupervised classification) of original Twitter data posted in English about the 10th anniversary of the 2010 Haiti Earthquake, 2022, http://dx.doi.org/10.25405/data.ncl.19688040.v1, URL https://data.ncl.ac.uk/articles/dataset/Sentiment_analysis_SA_supervised_and_unsupervised_classification_of_original_Twitter_data_posted_in_English_about_the_10th_anniversary_of_the_2010_Haiti_Earthquake/19688040.