The Journal of Web Science, 2018, 4: 34–52

Predicting Online Islamophobic Behavior after #ParisAttacks

Kareem Darwish[1], Walid Magdy[2], Afshin Rahimi[3], Timothy Baldwin[3] and Norah Abokhodair[4]

[1] *Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar*

[2] *School of Informatics, The University of Edinburgh, UK*

[3] *Dept. of Computing and Information Systems, The University of Melbourne, Australia*

[4] *Microsoft Corporation, Redmond, WA, USA*

ABSTRACT

The tragic Paris terrorist attacks of November 13, 2015 sparked a massive global discussion on Twitter and other social media, with millions of tweets in the first few hours after the attacks. Most of these tweets were condemning the attacks and showing support for Parisians. One of the trending debates related to the attacks concerned possible association between Muslims and terrorism, which resulted in a world-wide debate between those attacking and those defending Islam. In this paper, we use this incident as a case study to examine using online social network interactions prior to an event to predict what attitudes will be expressed in response to the event. Specifically, we focus on how a person's online content and network dynamics can be used to predict future attitudes and stance in the aftermath of a major event. In our study, we collected a set of 8.36 million tweets related to the Paris attacks within the 50 hours following the event, of which we identified over 900k tweets mentioning Islam and Muslims. We then quantitatively analyzed users' network interactions and historical tweets to predict their attitudes towards Islam and Muslims. We provide a description of the quantitative results based on the tweet content (hashtags) and network interactions (retweets, replies, and mentions). We analyze two types of data: (1) we use post-event tweets to learn users' stated stance towards Muslims based on sampling methods and crowd-sourced annotations; and (2) we employ pre-event interactions on Twitter to build a classifier to predict post-event stance. We found that pre-event network interactions can predict attitudes towards Muslims with 82% macro F-measure, even in the absence of prior mentions of Islam, Muslims, or related terms.

Keywords: Islamophobia, Paris attacks, Terrorist Attacks, Stance Prediction, Network Analysis, Twitter, Homophily, Social Networks

ISSN 2332-4031; DOI 10.1561/106.00000013

1 Introduction

In recent years, it has become increasingly common for a broad range of political actors and citizens to engage with one another on social media platforms like Twitter. This is all part of a movement towards a more networked society through sociopolitical technical mediums that are making such connections easier. Through these platforms, stakeholders are now able to engage in public discourse (e.g., political engagement) in a way that was not previously achievable, making it a rich target for research.

There is a rich tradition of research on social influence and homophily in the physical world Cialdini and Trost, 1998; Turner, 1991. More recently, there has been research examining social influence, homophily, and polarity in the context of social media, focusing on a variety of aspects including: utilizing social media as a tool for social influence to incite behavioral change Korda and Itani, 2013; Laranjo *et al.*, 2015, identifying influential users Dubois and Gaffney, 2014, determining the homogeneity of user sub-groups Himelboim *et al.*, 2013, ascertaining political leanings of users Cohen and Ruths, 2013, and utilizing co-follow relations in predicting biases and preferences Garimella and Weber, 2014. This paper extends on this work by examining the effect of online social network interactions — in terms of content and network dynamics — on future attitudes and stance in the aftermath of a major event. Specifically, we examine three primary research questions:

1. Can a user's social posts and interactions on Twitter be used to predict their stance on a given topic, even if they have never mentioned that topic?
2. What are the most predictive features/approaches for stance prediction?
3. Who are the primary influencers in the data, for different stances?

To answer these questions, we use people's expressed attitudes towards Muslims and Islam after the Paris terrorist attacks as a case study. The Paris attacks were carried out by the so-called Islamic State of Iraq and Syria (ISIS), also known as Daesh, over multiple locations in Paris on November 13, 2015. The attacks triggered a massive response on social media platforms such as Twitter, where posts covered a range of related subtopics, including posts showing attitudes towards Muslims: either blaming them for the attacks and linking terrorism to Islam, or defending them and disassociating them from the attacks. We focus on predicting the attitudes of Twitter users towards Muslims subsequent to the Paris terrorist attacks, based

on their interactions on Twitter prior to the attack. Specifically, we collected the Twitter profile information and timeline tweets of users who indicated a personal stance towards Muslims right after the Paris attacks, and we studied the possibility of using these users' interactions and tweets prior to the attacks to predict their expected stance after the attacks. We explored the effectiveness of three types of features for the prediction, namely: (1) content features (i.e., the body of the tweets from a user); (2) profile features (i.e., user-declared information such as name, location, and description); and (3) network features (i.e., user interactions within the Twitter community, through mentions, retweets, and replies).

Our dataset contains more than 145,000 users who posted at least one tweet about the Paris attacks within the 50 hours following the attacks, conveying either a positive or a negative stance towards Muslims. The dataset contains users' profile information and network interactions, in addition to a set of more than 12 million tweets collected from their timelines before the attacks. We manually annotated the polarity of user stance towards Muslims, and found that a majority (77%) of users showed a positive stance towards Muslims. On the other hand, a considerable number of tweets (23%) used language that blamed Muslims and Islam for these attacks.

Our results show that a user's pre-event network interactions are more effective in predicting a positive or a negative stance than content or profile features. Additionally, our results reveal that it is not necessary for the user to have mentioned the topic of interest in order to predict their stance. However, if they have mentioned the topic explicitly, this significantly boosts the accuracy of prediction (from a macro-averaged F-score of 0.77 to 0.85). Finally, our study provides analysis of how different features can affect the prediction performance, and discusses the implications of our findings.

This paper is an extension of earlier work by the authors Magdy *et al.*, 2016a, in the following ways: (1) we provide global-scale analysis of attitudes towards Muslims across a wide range of languages and countries; (2) we perform analysis of the most popular negative, positive and neutral tweets relating to Muslims after the Paris attacks; and (3) we extend our experiments on prediction of stance from just the US to include the UK and France, and complement the Twitter text features with user profile features and network modeling.

2 Background

2.1 The 2015 Terrorist Attacks on Paris

On the evening of 13 November 2015, several coordinated terrorist attacks occurred simultaneously in Paris, France. At 20:20 GMT, three suicide bombers struck near the stadium where a football match between France and Germany was being played. Other suicide bombings and mass shootings occurred a few minutes later at cafés, restaurants and a music venue in Paris Chung *et al.*, 2016; Hamaide, 2015; BBC, 2015.

The tragic events resulted in more than 130 deaths and 368 injured people, with 80–99 seriously injured. These attacks are considered the deadliest in France since World War II Syeed, 2015. The Islamic State of Iraq and Syria (ISIS)[1] claimed responsibility for the attacks Castillo *et al.*, 2015, as a response to French airstrikes on ISIS targets in Syria and Iraq.

2.2 Anti-Muslim rhetoric

Some studies in the literature refer to anti-Muslim speech or actions as "Islamophobia", although there is still debate as to the exact meaning and characteristics of this phenomenon. Some regard it as a type of hate speech and others as a type of racism Awan, 2014. In most cases, it refers to the phenomenon of negatively representing Muslims and Islam, generally based on limited or biased understanding of Islamic culture or historical events Runnymede Trust, 1997.

In this study we are interested in Islamophobia in the context of our case study regarding positive or negative views of Twitter users towards Muslims in the aftermath of the Paris attacks. In earlier work Magdy *et al.*, 2015, it was shown that the majority (72%) of tweets from around the world defended Muslims and Islam after the Paris attacks. The collection of tweets represented 58 countries, with the tweets defending Muslims outnumbering the ones attacking them for all but two countries. It was also shown that the US had the largest number of generated tweets, with 71% of the polarized tweets defending Muslims Magdy *et al.*, 2015. We extend on this work by examining the effects of social network interactions on future attitudes.

2.3 Political Polarization and Homophily

Much research has been done on predicting and estimating a person's political orientation Conover *et al.*, 2011; Cohen and Ruths, 2013; Himelboim *et al.*, 2013; Barberá, 2015. Barberá 2015 developed a Bayesian spatial following model that takes into account the Twitter follower network to estimate the political ideology of political leaders and average citizens in several countries, including the US, the UK, Spain, Italy, and the Netherlands. Barberá's model was successful in estimating a user's political orientation based on information gained from his/her Twitter network, together with their location. Subsequent work by Barberá expands and validates the results of his model Barberá *et al.*, 2015. His investigation builds on 12 political and non-political events to better understand whether social media platforms resemble "echo chambers", or provide spaces for pluralist debate. The results show that during certain political events (e.g., elections), individuals with similar political orientation were more likely to engage in a discussion together, creating an echo chamber. The opposite is true in the case of sudden events (e.g., terrorist attacks or sports events) where signs of a more pluralist debate were visible during the first hours of such events before deteriorating into an echo chamber later on Barberá *et al.*, 2015.

Similar behavior has been observed by others Himelboim *et al.*, 2013; Colleoni *et al.*, 2014. Golbeck and Hansen 2014 provide a direct estimate of audience political preferences by focusing on Twitter following relationships. Their results compares favorably to the results of others such as Groseclose and

[1] Also known as Islamic State of Iraq and the Levant (ISIL).

Milyo 2005, who do not factor in the information gained from someone's Twitter network (i.e., the general social media dynamics). The results of this study are aligned with our decision to account for network characteristics in our prediction model. Colleoni et al. 2014 utilized a combination of machine learning and social network analysis to categorize users as either Democrats or Republicans based on the political content they shared, and then investigated the level of homophily among these groups. Homophily is the propensity for individuals to interact with similarly-minded individuals. Their results show varying levels of homophily between the opposing groups. Political and ideological orientation has also been explored in non-Western countries such as Egypt Weber *et al.*, 2013; Borge-Holthoefer *et al.*, 2015. Our approach builds on previous work and examines the effect of both network and content features on prediction.

2.4 Consistency of Orientation

For opinion shifts during polarizing events, Borge-Holthoefer et al. 2015 provide insights and empirical evidence from the 2013 military coup in Egypt through the examination of tweets from two opposite perspectives, namely: secular vs. Islamist, and pro-military vs. anti-military intervention. The results of their study show little evidence of ideological or opinion shifts even after violent events. However, they observe changes in tweet volume between different camps in response to events. This is consistent with offline research conducted by Chenoweth and Stephan 2011 where they examined dozens of civil conflicts around the world. Also, the tracking of political polarization in the US between conservatives, liberals, and moderates has shown that the relative percentage of the different groups has changed by less than 2% since the 1970's to the 2000's Dalton, 2013 (ch. 6). Such consistency enables us to assume that Twitter users would have stable sociopolitical opinions over a span of a few months.

2.5 Stance Prediction

Our work can also be framed as an instance of stance detection, whereby the opinions of an individual on a specific topic are identified (as opposed to general political orientation), including congressional debates Thomas *et al.*, 2006; Burfoot *et al.*, 2011, online forums Anand *et al.*, 2011; Walker *et al.*, 2012; Sridhar *et al.*, 2014; Qiu *et al.*, 2015 and student essays Faulkner, 2014. Twitter is a very attractive source of data for the study of stance-taking, due to the large volume of users and the tendency for users to express opinions on a broad range of topics in real-time. This attractiveness, though, comes with its own challenges, as tweets are short and in some cases contain misspellings, informal and slang language Baldwin *et al.*, 2013. These challenges make the stance detection task over Twitter data much more difficult than is the case for conventional documents and speeches. Several features have been studied for determining stance detection on Twitter. Rao et al. 2010 used socio-linguistic features that include types of utterances (e.g., emoticons and abbreviations) and word n-gram features. They showed that they can distinguish between republicans and democrats with more than 80% accuracy. Pennacchiotti and Popescu 2011 extended the work of Rao et al. 2010 by introducing features based on profile information (screen name, profile description, followers, etc.), tweeting behavior, sociolinguistic features, network interactions, and sentiment.

The simplest approach to stance detection is to use polarity lexicons such as SentiWordNet Esuli and Sebastiani, 2006 to identify the ratio of positive and negative terms in a document. Lexicon-based approaches fail to adopt to the dynamic and noisy nature of Twitter, and are generally outperformed by supervised stance detection models Pang and Lee, 2008. Supervised models, on the other hand, require manually-annotated documents, making them costly and time-consuming to develop. Most work on Twitter stance detection has made use of a small number of labeled samples and tried to use different sources of information such as follower graphs Speriosu *et al.*, 2011 and retweets Wong *et al.*, 2013; Rajadesingan and Liu, 2014. Recent work on entity-centric sentiment analysis suggests that a sentiment analyzer can be used to bootstrap the learning process Zhang *et al.*, 2011. Perhaps this can be extended to stance detection. For our work, given our manually-annotated data, we use a supervised model and utilize both content (e.g., text and hashtags) and network features (e.g., retweets and mentions) as candidate predictors of user stance toward Islam.

In work closely related to this paper, Qiu et al. 2015 proposed a graphical model approach to predict unexpressed stances on debate forums, taking inspiration from work on collaborative filtering (similar users will have similar opinions), topic modelling (users with similar stances tend to have similar topic distributions), and network analysis (a positive interaction with a given user is strongly suggestive of shared values). Different to this research, however, they assume access to partial knowledge of the stance of a given user across a range of issues, that all content from a given user will be related to a closed set of issues, and that there will be direct interactions between users specifically related to the topics of interest. As such, while their model is certainly able to predict unexpressed opinions, it does so in a much more constrained setting than this paper. The scalability of the proposed model to the scale of data targeted in this research is also questionable.

2.6 Lifestyle Politics and Recommendations

An emerging area of research is targeted at predicting and explaining correlations between political views and personal preferences in such things as food, sports, and music. The paper "Why Do Liberals Drink Lattes?" by DellaPosta et al. 2015 is one example of such research. Such correlations seem to arise as a result of homophily and social influence within echo-chambers DellaPosta *et al.*, 2015. One method for discovering these correlations employs co-following relationships on Twitter Garimella and Weber, 2014, and can be used to recommend music to users Weber and Garimella, 2014. Using this method, Garimella and Weber 2014 show that conservatives are more likely to listen to the country singer Kenny Chesney, while liberals are more likely to listen to Lady Gaga. In this work we observe such correlations, but they are discovered using content analysis and mention/retweet relations.

3 Post-Attack Data Collection

3.1 Streaming Tweets on the Attacks

In the hours immediately after the Paris attacks, the trending topics on Twitter mostly referred to the attacks, expressing sympathy for the victims. We used these trending topics to formulate a set of terms for streaming tweets using the Twitter REST API. We also used general terms referring to terrorism and Islam, which were hot topics at that time. We continuously collected tweets between 5:26 AM (GMT) (roughly 7 hours after the attacks) on November 14 and 7:13 AM (GMT) on November 16 (approximately 50 hours in total). The terms we used for collecting our tweets were: *Paris*, *France*, *PorteOuverte*, *ParisAttacks*, *AttaquesParis*, *pray4Paris*, *prayers4Paris*, *terrorist*, *terrorism*, *terrorists*, *Muslims*, *Islam*, *Muslim*, *Islamic*. In total we collected 8.36 million tweets. Since we were using the public API, the results were down-sampled and subject to preset limits. However, since we were searching using focused keywords, we are confident of having captured a substantial proportion (if not the majority) of on-topic tweets. On average, we collected 140k to 175k tweets per hour. Subsequent to collection, we checked the counts of the terms we used for the search in Topsy,[2] based on which we estimate that the number of tweets that matched our search terms was slightly higher than 12 million. Also, since we were using mostly English words/hashtags and a few French ones, we expected to be collecting mostly English tweets, with some French tweets. However, as the primary term, *Paris*, is language independent for most languages that use the Latin alphabet, in practice, we were able to retrieve data from a large number of languages.

We used an open-source language identification system to classify each tweet to understand the distribution of languages in our collection.[3] Figure 1 shows the language distribution of our tweet collection. As shown, the majority of the tweets (64%) were in English, which is expected since English is the predominant language on Twitter and people tend to comment on high-impact global events in high-density languages. The second language was French, the language used at the location of the attacks. Surprisingly, the third language was Arabic, though all of the keywords used for crawling were based on the Latin alphabet (and Arabic is generally reported to account for no more than 2% of the total Twitter traffic Baldwin *et al.*, 2013). The cause for this was that Arabs were commenting on the topic in their own language and adding English hashtags to make their tweets discoverable.

3.2 Identifying Tweets on Islam

To identify tweets about Islam and Muslims, we filtered the tweets using terms that refer to Islam, such as *Islam*, *Muslim*, *Muslims*, *Islamic*, and *Islamist*. Like the word *Paris*, the word *Islam* is used as-is in many languages that use the Latin alphabet.[4] Out of the 8.36 million tweets, we extracted 912,694 tweets mentioning something about Islam. This constitutes 11% of the collected tweets, which shows that reactions to Muslims after the attacks were common.

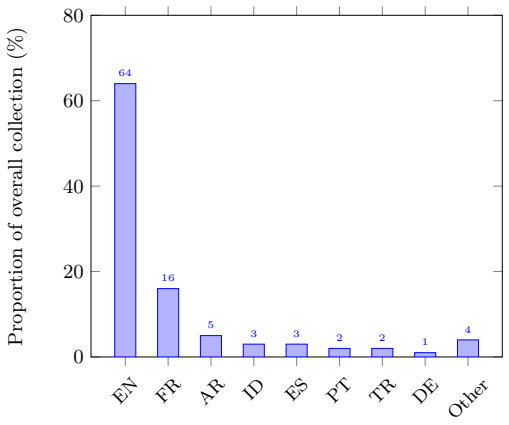

Figure 1: Language distribution of the tweet collection (based on ISO-639-2 language codes)

3.3 Sampling and Annotation of Tweets

The number of tweets pertaining to Muslims was too large to be fully manually annotated. In order to determine the attitudes expressed in the tweets, we sampled the data collection by getting a representative sample of tweets. We used a sample size calculator[5] to calculate the sample size that would lead to an estimation of the attitude distribution with error less than $\pm 2.5\%$ (confidence interval = 2.5%) and a confidence level of 95%. Table 1 shows per language counts and the size of the samples that we manually annotated. The extracted samples contained some duplicate tweets and retweets. Only unique tweets were annotated and the label is then propagated to duplicate tweets. The number of unique tweets in each sample is shown in Table 1.

For the manual annotation, we submitted the sampled tweets to CrowdFlower.[6] We asked annotators to label each of the tweets with one of three labels:

- **Defending**: the tweet is defending Islam and/or Muslims against any association to the attacks.
- **Attacking**: the tweet is attacking Islam and/or Muslims as being responsible for the terrorist attacks.
- **Neutral**: the tweet is reporting news, not related to the event, or talking about ISIS in specific and not Muslims in general.

In CrowdFlower, each tweet was annotated by at least 3 annotators, and majority voting was used to select the final label. A control set of 25 tweets was used to assess the quality of the annotators, whereby the data from low-quality annotators was discarded. The annotated tweet sample had an average inter-annotator agreement of 77.7%, which is considered high for a three-way annotation task annotated by at least three

[2] http://topsy.com/ *(currently unavailable)*

[3] https://github.com/shuyo/language-detection

[4] Although it did mean a big drop in the relative proportion of tweets in non-Latin script languages such as Arabic, and also, interestingly, languages which use the Latin script but are associated with countries with a large Muslim population, namely Indonesian (ID) and Turkish (TR).

[5] http://www.surveysystem.com/sscalc.htm

[6] http://www.crowdflower.com/

Table 1: Per language tweet count, sample size, and annotations for top 7 languages

Language	Size	Sample	Unique	Defend	Attack	Neutral
EN	753,476	1,534	1,167	880	324	328
FR	63,410	1,500	740	607	286	603
ES	15,726	1,400	705	681	351	368
DE	6,388	1,239	613	510	363	365
NL	4,406	1,139	586	208	773	158
IT	3,825	1,096	558	376	588	129
PT	2,194	904	235	661	139	104

different annotators. The percentage of disagreement among annotators shows that some tweets were not straightforward to label. This usually occurred between neutral and one of the other attitudes. Table 1 and Figure 2 provide the count and breakdown of tweets across the three classes.

Given that many of the tweets in our collection were actually retweets or duplicates of other tweets, we applied label propagation to label the tweets in our collection that have identical text to the labeled tweets. To detect duplicates and retweets, we normalized the text of the tweets by applying case folding and filtering out URLs, punctuation, and user mentions. Tweets in the collection that matched the annotated sample tweets after text normalization were then automatically assigned the same label. This label propagation process led to the labeling of 336,294 of the tweets referring to Islam in the collection.

3.4 Location Identification

To filter tweets by location, we used two different methods. The first uses the user-declared location, and the second uses the text of the tweets.

3.4.1 User-declared location

We extracted the user-declared locations to map them to their respective countries. The location field in Twitter is optional, so users can leave it blank. In addition, it is free text, which means that there is no standard way for declaring locations. This renders a large portion of the declared locations unusable, e.g., *in the heart of my mom*, *the 3rd rock from the son*, and *at my house*. This is a common problem in social media in general and in Twitter in particular, as demonstrated in Hecht et al. 2011.

In our work, we used a semi-supervised method to map out the user-declared locations to countries, as follows:

1. A list of the countries of the world and their most popular cities were collected from Wikipedia and saved in a database.
2. A list of the 50 states of the United States and their abbreviations, along with the top cities in each state, were then added to the database.
3. Location strings were normalized by case folding and removing diacritics and accents. For example, *México* is normalized to *mexico*.
4. If the location string contains a country name, it is mapped to the country. Otherwise, the string is searched for in our database, and mapped to its corresponding country in the case of a match. In the case of multiple countries/cities existing in the location string, we use the first-matching location.
5. All unmapped locations appearing at least 10 times are then manually mapped to countries where possible (noting that there are high-frequency junk locations, such as *earth*). All newly mapped locations are then added to the database, and an additional iteration of matching as in the previous step is applied.

With the initial application of our approach to the users who tweeted the 336,294 tweets, we found that 125,583 contained blank user-declared locations. In addition, 41,905 were locations of tweets labeled as "neutral", which were not of much interest in our analysis. The reason for this is that a neutral tweet does not necessarily mean that its author is neutral, but may mean that the authors did not express a position. The remaining tweets with non-blank user-declared locations numbered 168,807 (with 76,894 unique locations). Using the above algorithm, we managed to map 107,377 locations (42,140 unique) to countries.

3.4.2 Text-based geolocation

To expand the coverage of geolocated tweets for the users with blank or undefined location, we further exploit the linguistic content of the tweets. Previous research has shown that the geographical bias in the use of language can be utilized for the geolocation of documents and social media users Cheng *et al.*, 2010. Geographical bias is evident in countries with different languages, but also exists in the use of toponyms (e.g., city names, landmarks, popular figures) and regional dialects (e.g., *centre* vs. *center*). These linguistic features can be used in supervised classification models for geolocation Han *et al.*, 2014.

We used the supervised text-based geolocation model of Rahimi et al. 2015, trained on the Twitter-World dataset Han *et al.*, 2012, to geolocate the users. The dataset contains geotagged tweets from around 1.3M Twitter users from all over the world. Although the dataset is limited to English tweets, it contains some foreign language text. The model uses the aggregated tweets of a user, represented by a bag of unigrams and weighted by a variant of *TF-IDF* weighting in a l_1 regular-

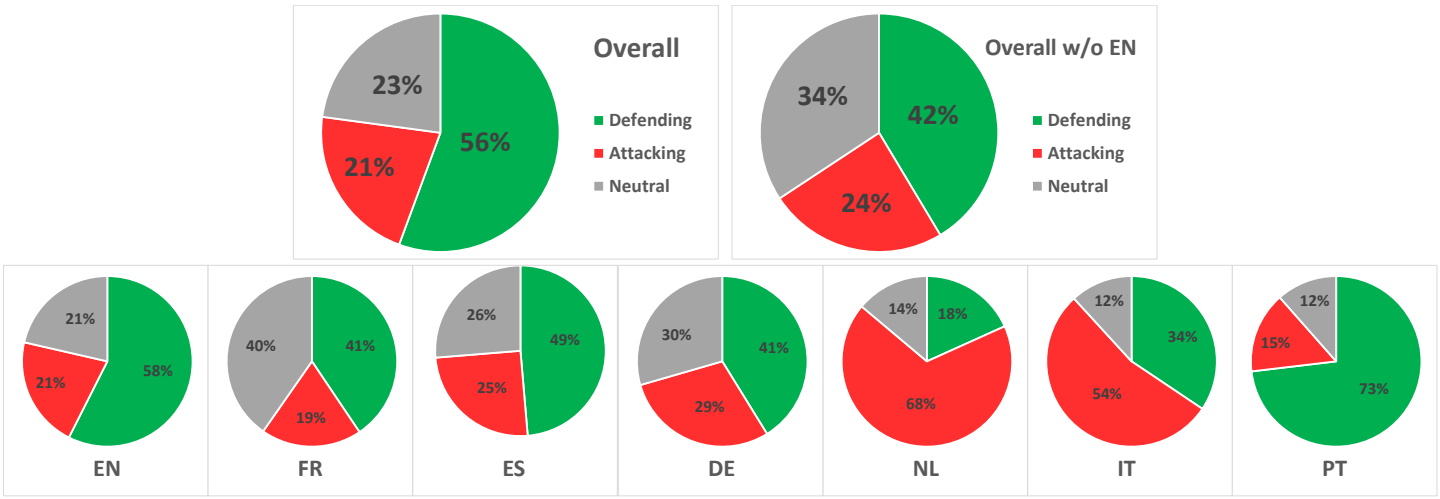

Figure 2: Stance distribution by language

ized logistic regression, to classify users into one of 171 home countries. The trained model is then applied to the users of the current dataset. The accuracy of the model in predicting the home country of a user is 90% for the test set of Twitter-World dataset.

To apply this algorithm to our data, we obtain the aggregated user tweets from their timelines using the Twitter API, as will be explained in the following section. We evaluate the geolocation model over the current dataset by comparing the predicted labels with the labels extracted from the location field. The model correctly identifies the home country of users with around 77% accuracy, substantially lower than the accuracy of the model over the test set of Twitter-World. The drop in accuracy can be a result of temporal differences in topics, different geographical coverage (e.g., inclusion of new countries in the current dataset), and linguistic bias in Twitter-World, due to the fact that all users of Twitter-World tend to geotag their tweets. Pavalanathan and Eisenstein 2015 report that Twitter users who geotag their tweets have demographic differences with those who just fill their location field, which reflects itself in their language.

We evaluated the performance of the text-based geolocation by comparing the prediction with the location of users who had a recoverable country in their location field. The accuracy over top 10 countries in terms of the number of users is shown in Table 2. The performance is lower for countries which are less represented in the training set of Twitter-World or have a shared language with another larger country (e.g., Canada vs. US).

We keep the top 50% most confident predictions for each country, in order to increase the accuracy at the expense of coverage. We assume that all tweets from the same user originate from the same country that is predicted by the geolocation model. Using this method, we increase the number of geolocated tweets from 107k to 177k. These 177k geolocated tweets account for around 147k unique users, of which 44k are predicted to originate from the US, which is the largest number among all countries.

Figure 3 provides a breakdown of the tweet collection, and

Table 2: Text-based geolocation accuracy of top 10 countries with the most number of users with recoverable self-declared location field.

Country	Accuracy
US	86
UK	78
France	94
Malaysia	95
India	91
Spain	92
Canada	79
Australia	69
Italy	81
Singapore	82

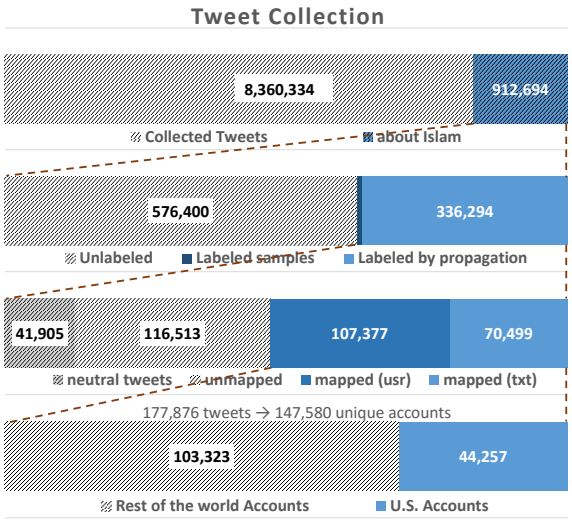

Figure 3: Summary of the tweet collection used in this study. The first three rows show the numbers of tweets; the final row shows the number of Twitter accounts.

all the steps applied to get the annotated data. The blue portion in each row of the figure represents the tweets used in the next stage of processing. Account information and timeline tweets were collected for each of these accounts for the prediction process described later.

4 Statistics on the Data

4.1 Distribution of Attitudes by Language

Figure 2 shows the distribution of attitudes towards Muslims for each language, and the overall distribution of all languages, which is estimated based on the size of each language in the collection. As shown, most of the tweets are positive towards Muslim and Islam, and disassociate them from the attacks. Portuguese (PT) had the highest proportion of positive tweets, and for only two languages — Dutch (NL) and Italian (IT) — negative tweets were more prevalent than positive tweets.

The language which has the largest percentage of neutral tweets was French (FR), which might be expected, since France was the scene of the attacks and people there were most likely more concerned with following the news and its updates compared to others. Many of these updates referred to Islamic State, which matched our query term *Islam*.

The overall finding of this analysis is that 21.5% of the tweets on the topic appeared to try to link the ISIS attacks on Paris to Islam. However, most tweets (55.6%) were defending Muslims and disassociating Islam from terrorism.

4.2 Attitudes by Country

As mentioned earlier, we automatically mapped out the location of 106K tweets that have non-neutral attitudes to 144 different countries, which shows the global impact of the terrorist attacks. Some of the countries had only a handful of tweets assigned to them, making it difficult to draw any real conclusions about general attitudes for these countries. Thus, in our analysis, we focus on countries which had at least 100 tweets assigned to them, resulting in 58 countries.

The United States (US) had the highest number of tweets, namely 36.5% of the mapped tweets, followed by the UK (12.5%), France (7.5%), Malaysia (6.7%),[7] India (6.6%), and Spain (3.4%). Each of the remaining countries had less than a 3% share.

Figure 4 lists the 58 countries that have more than 100 tweets mapped to them. For clarity, Figure 4 splits the graph into 4 parts according to the order of magnitude of the number of tweets. For each country, the green and red components of the bar represent positive and negative tweets towards Muslims, respectively. A rank for each country is displayed to the right of each bar according to the percentage of positive tweets.[8]

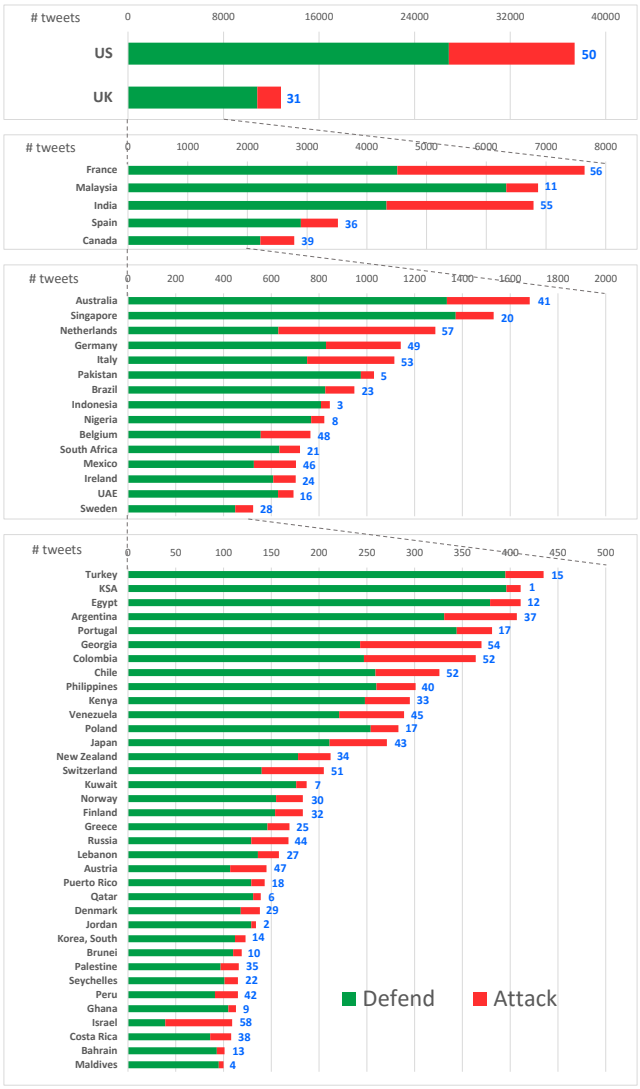

Figure 4: Attitude towards Muslims by country. The label beside each bar represents the rank of the country in terms of defending Muslims.

We calculated the confidence interval for each of the countries when setting the confidence level to 95%, because a sample of 100 tweets only is considered low to represent a country of populations in millions. It was found that most of the countries had a confidence interval of less than 5%, leading to estimation errors of less than ±5%. In Figure 4, the countries listed below New Zealand got a confidence interval ranging between 5% and 8.9%, indicating more expected errors in percentage estimation. Nevertheless, the numbers are at least indicative of an overall trend.

As shown in Figure 4, the countries with the highest percentages of positive tweets are mostly Muslim and/or Arab countries, such as Saudi Arabia (KSA), Jordan, Indonesia, Maldives, Pakistan, and Qatar. Only two countries had more

[7] Note that based on the automatic language identification, Indonesian was identified as being the language with the fourth-greatest number of tweets, and Malay was much further down the list. In practice, the strong similarities between Malay and Indonesian make them a common confusion pair for language identifiers Zampieri *et al.*, 2015, and the breakdown for this language pair may be somewhat noisy.

[8] We ranked according to the percentage of positive tweets, since it was the prevailing attitude.

negative than positive tweets, namely Israel and the Netherlands, at ranks 58 and 57 respectively. They were followed by France, India, Georgia, and Italy at ranks 56, 55, 54, and 53 respectively. US, which is the country with the largest number of tweets, comes in at the rank 50, while the UK, the country with the second highest number of tweets, comes in at rank 31, with 85% of positive tweets.

Our analysis shows large variations in attitudes between countries. As expected, predominantly Muslim countries had the highest percentages of positive tweets. However, neighboring countries such as Spain (rank 36) and Italy (rank 53) had dramatically different percentages of positive/negative tweets. This is also reflected in the percentage of Spanish and Italian language tweets, where roughly a quarter of Spanish tweets are negative, compared to more than half of Italian tweets. Similarly, the percentage of negative tweets is much higher in the Netherlands compared to Germany. The large variation between neighboring countries is worthy of further study. Further, the rank of the US is considerably low (rank 50). We analyze US tweets later in greater detail. Figure 4 also shows some non-Muslim countries with very small Muslim populations that are ranked quite high, such as South Korea (rank 10) and Portugal (rank 17). This also warrants further investigation.

4.3 Most Popular Tweets

The label propagation step that we applied showed that a large portion of the tweets in our collection are retweets. This refers to the presence of highly popular tweets that got retweeted thousands of times. Our last research question was who are the most influential accounts in the discussion with positive or negative stance. In other words, who was promoting anti-Islam sentiment on Twitter in the time after the Paris attacks, and who was opposing that sentiment. Here, we consider the 5 most retweeted tweets in each of the categories we identified earlier: neutral, positive, and negative. Figure 5 illustrates the 5 most retweeted tweets with the account handle in each of the three categories (attacking, defending, and neutral). For the purpose of this paper, we consider and discuss tweets in the list from celebrity-type accounts, i.e. people who have both high content influence and high account influence.[9] Using both qualitative and quantitative analyses, we found that most of the interesting results appear in the Negative category. However, we describe our observations across the three categories.

4.3.1 Top Neutral and Positive Tweets

The top 5 neutral tweets were mostly about news, as expected, with the exception of the top tweet, which received a large number of retweets (43,000+). This tweet comes from a seemingly Muslim female who has a moderate number of followers.[10] Her tweet was her reflection on the effect of the attacks on the Muslim community in the US, where she mentions that her young niece is afraid of telling her friends in school that she is Muslim. Although the tweet was most probably retweeted by those disassociating Muslims from the attacks, it is not overtly positive. The third tweet is concerned with a hate-crime that was perpetrated against a Muslim woman in London.

Regarding the most popular positive tweets, two of them were tweeted by accounts apparently owned by Muslims. The top 2 tweets mainly emphasize the importance of discriminating between ISIS and Islam. The third tweet is from a Muslim user who condemns the attacks. The fourth tweet wonders why people think ISIS represent Islam, given that ISIS also conducts similar attacks on Muslims. The last tweet mocks media outlets that generalize attacks perpetrated by a Muslim to all Muslims or an African American to all African Americans, while taking careful measures when the attacker is white.

4.3.2 Top Negative Tweets

As the 2016 US presidential candidate, Donald Trump topped the list of most retweeted negative tweets: *Why won't President Obama use the term Islamic Terrorism? Isn't it now, after all of this time and so much death, about time!*. Trump had another tweet in the top 5 revolving around anti-Muslim rhetoric in reference to the Paris Attacks. Here, Trump continues to slam the Democratic Party and President Obama for not referring to the ISIS attacks as "Islamic Terrorism". When looking at Trump's timeline, it becomes clear that this is one of many tweets along the same lines, where he blames Islam and Muslims worldwide for the Paris attacks.

Ted Cruz, another US presidential candidate, claimed a top 5 tweet linking Islam and terrorism. The appearance of another tweet from one of the conservative US politicians may indicate the political nature of the comments, and their ties to conservative right-wing mood in the US.

Following Trump's tweet is a tweet from Ayaan Hirsi Ali, a female activist based in the US with Somali origins, who is known for her critical view on Islam.[11] In her tweet, Ayaan writes, *As long as Muslims say IS has nothing to [do] with Islam or talk of Islamophobia they are not ready to reform their faith.* Ayaan calls on all Muslims around the world to recognize Islam as a source of terrorist ideology. Ayaan is affiliated with the American Enterprise Institute, a right-wing conservative think tank based in the US, which may indicate yet another link to US politics.

5 Pre-Attack Prediction

5.1 Prediction

Next, we experiment with using pre-event tweets, interactions and profile information of users to predict their post-event stance. We use the content, profile information and network features from the tweets posted by users before the Paris attacks to predict their stance toward Muslims after the attacks. For supervision, we use the annotated tweet labels and extend them to the user, based on the assumption that a user has a single stance which is invariant over the period of time of

[9] Tweets shown in Figure 5 were found to exist after more than a year of the Paris attacks. Thus we did not anonymize their authors.

[10] 1,826 followers at the time of writing the paper

[11] https://en.wikipedia.org/wiki/Ayaan_Hirsi_Ali

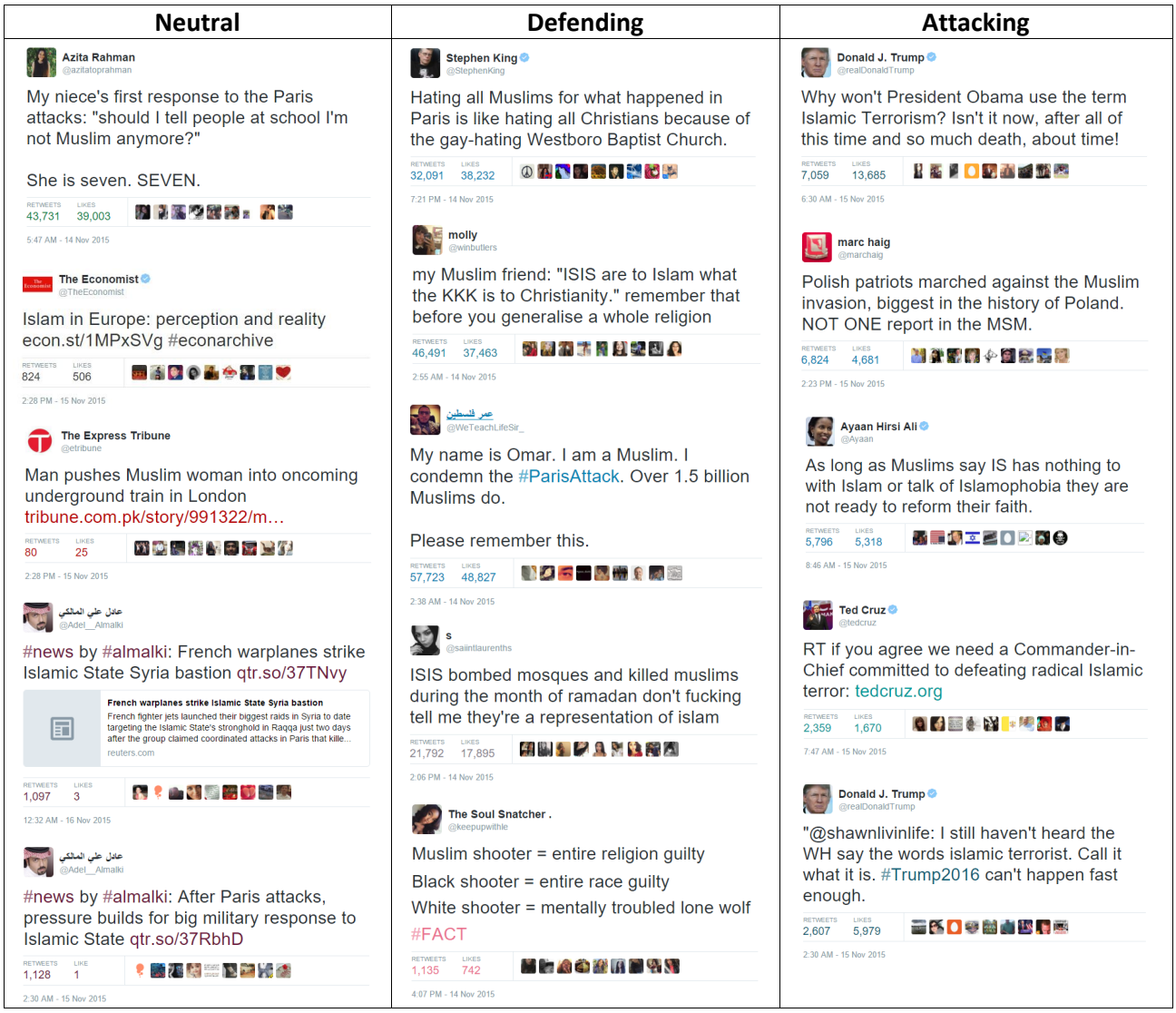

Figure 5: Most popular tweets for each attitude

our Twitter crawl (pre- and post-attack). Prior research has shown that the opinions of the vast majority of people persist over time Chenoweth and Stephan, 2011; Dalton, 2013; Borge-Holthoefer *et al.*, 2015. Besides the actual stance prediction, we are also interested in finding out what features strongly correlate with positive and negative stance toward Muslims. Subsequent qualitative analysis of these features can shed light on personal, social and political attributes that are predictive of a user's stance.

5.2 Pre-Attack Data Collection

We restricted our consideration to the top 3 countries, and performed expanded analysis on the US. The numbers of users with either positive or negative stance who were geolocated in the top 3 countries are as follows:

Country	User Count
US	44,257
UK	14,749
France	10,498

We used the Twitter API to crawl (up to) 200 tweets for each of these users that were posted before the attacks.[12] Some of these user accounts had so many tweets posted after the attacks that the Twitter API did not allow us to crawl any tweets for them before the specified attack date, since it does not allow retrieval of tweets outside the most recent 3,200 for a given user.

[12]The API supports user-level crawling by specifying a tweet ID, and returns the history of tweets of that user prior to the post.

Table 3: Results for US users

(a) US users who are positive (6,599 users)/negative (4,082 users) towards Muslims before the Paris attacks

		Content Features			Profile Features				Network Features				All
	BL	Hashtags	Text	All	Desc.	Name	Loc.	All	Mention	Reply	Retweet	All	Features
$\mathcal{AUC}$	0.5	0.87	0.88	0.89	0.73	0.58	0.60	0.75	0.89	0.77	**0.90**	0.89	0.89
Accuracy	0.61	0.83	0.79	0.83	0.71	0.64	0.62	0.73	**0.86**	0.75	**0.86**	**0.86**	0.85
$\mathcal{F}$	0.54	0.82	0.79	0.82	0.67	0.58	0.58	0.70	**0.85**	0.74	**0.85**	**0.85**	0.84
pos $\mathcal{P}$	0.61	0.88	0.89	0.84	0.73	0.67	0.67	0.75	**0.90**	0.80	**0.90**	0.89	0.89
pos $\mathcal{R}$	1.00	0.84	0.77	0.84	0.87	0.82	0.75	0.85	0.88	0.82	0.89	**0.90**	0.87
pos $\mathcal{F}$	0.76	0.86	0.83	0.84	0.79	0.74	0.71	0.80	**0.89**	0.81	**0.89**	**0.89**	0.88
neg $\mathcal{P}$	0.00	0.76	0.69	0.75	0.70	0.56	0.51	0.69	0.81	0.70	0.82	**0.83**	0.79
neg $\mathcal{R}$	0.00	0.82	**0.85**	**0.85**	0.47	0.36	0.42	0.54	0.83	0.66	0.83	0.82	0.83
neg $\mathcal{F}$	0.00	0.79	0.76	0.80	0.56	0.44	0.46	0.61	**0.82**	0.68	**0.82**	**0.82**	0.81

(b) US users who are positive (27,457)/negative (6,119) towards Muslims from only after the Paris attacks

		Content Features			Profile Features				Network Features				All
	BL	Hashtags	Text	All	Desc.	Name	Loc.	All	Mention	Reply	Retweet	All	Features
$\mathcal{AUC}$	0.5	0.76	0.77	0.80	0.65	0.60	0.57	0.67	0.80	0.68	**0.81**	0.80	0.82
Accuracy	0.81	0.83	0.83	0.84	0.79	0.74	0.7	0.79	**0.88**	0.81	**0.88**	**0.88**	0.87
$\mathcal{F}$	0.61	0.71	0.72	0.73	0.59	0.58	0.54	0.62	**0.77**	0.65	**0.77**	**0.77**	0.76
pos $\mathcal{P}$	0.81	0.89	**0.90**	**0.90**	0.84	0.85	0.84	0.86	**0.90**	0.87	**0.90**	**0.90**	**0.90**
pos $\mathcal{R}$	1.00	0.90	0.90	0.91	0.93	0.83	0.79	0.9	0.96	0.91	0.96	**0.97**	0.95
pos $\mathcal{F}$	0.90	0.89	0.90	0.90	0.88	0.84	0.81	0.88	**0.93**	0.89	**0.93**	**0.93**	0.92
neg $\mathcal{P}$	0.00	0.54	0.55	0.58	0.41	0.31	0.25	0.42	0.74	0.49	0.76	**0.79**	0.69
neg $\mathcal{R}$	0.00	0.51	**0.55**	0.51	0.24	0.34	0.31	0.33	0.54	0.39	0.52	0.51	0.53
neg $\mathcal{F}$	0.00	0.52	0.55	0.54	0.30	0.32	0.28	0.37	0.62	0.43	**0.62**	**0.62**	0.60

5.3 Prediction of Future Stance

For each country, we aggregated all pre-attack tweets for a user into a single (meta-)document, and labeled the document with the stance label of that user after the attacks. We used three different groups of features:

- *tweet content features*: word unigrams and hashtags. Content features help identify topics users are interested and their lexical choices when they discuss these topics.
- *profile features*: user-declared profile information, namely the name, profile description, and location. Profile features may provide hints on the stance of users. For example, users with a particular stance may cluster in specific geographic locals. Similarly, users often use words in their profile description that may indicate political leaning.
- *network features*: user interaction activities, namely other accounts that a user mentioned, retweeted, and replied to. Network features help capture information about a user's social network such as who they interact with and which other users and media sources they read. Users tend to prefer to interact with similarly minded users (homophily).

The content has the largest number of features, followed by network and profile. For example, for the results shown in Table 3 (a), the number of content, network and profile features are 50k, 15k and 1.7k respectively. The same pattern was seen in other experiments. We weighted the features by a variant of *TF-IDF* with sub-linear term frequency and l_2 normalization of samples. We excluded terms that occur in less than 10 tweets. For classification, we use a binary linear-kernel support vector machine (SVM) with l_2 regularization for stance prediction, and 10-fold cross-validation to tune the weighting scheme and regularization coefficient. We trained the model using each feature individually, as well as in combination. We evaluate the prediction performance using precision ("$\mathcal{P}$"), recall ("$\mathcal{R}$"), macro-averaged F-score ("$\mathcal{F}$"), and overall accuracy. Because we evaluate the method over three countries each with two sets of users (users who spoke on topic or not before the event), we evaluated the stance prediction method over each of the 6 datasets using the area under the curve of a ROC curve ("$\mathcal{AUC}$") so that the results can be compared over all the datasets.

Because it is easier to predict the stance of users who mentioned Muslims before the attacks compared to those who did not, we partition the users into two groups depending on whether they had used one of *Islam* or *Muslim* (case-insensitive; can match in the middle of another word) before the attacks (11k users) or not (33k users). For each of the two groups, we perform the training, evaluation and analysis of the most salient features separately. We compare the performance of each feature set with a majority-class baseline ("BL"), by classifying all accounts to positive stance.

Table 4: Results for UK users

(a) UK users who are positive (3,758 users)/negative (1,170 users) towards Muslims before the Paris attacks

	BL	Hashtags	Text	All	Desc.	Name	Loc.	All	Mention	Reply	Retweet	All	Features
$\mathcal{AUC}$	0.5	0.78	0.81	0.81	0.64	0.54	0.51	0.64	0.81	0.69	**0.82**	0.81	0.81
Acc	0.58	0.88	0.68	0.88	0.82	0.77	0.78	0.82	**0.89**	0.84	0.88	**0.89**	0.88
$\mathcal{F}$	0.43	0.72	0.75	0.77	0.59	0.56	0.52	0.59	0.77	0.68	**0.78**	**0.78**	0.77
pos $\mathcal{P}$	0.76	0.88	0.89	0.88	0.82	0.77	0.78	0.82	**0.89**	0.84	0.88	**0.89**	0.88
pos $\mathcal{R}$	1.00	0.83	0.86	0.92	0.74	**0.94**	0.31	0.70	0.89	0.87	0.92	0.91	0.92
pos $\mathcal{F}$	0.87	0.86	0.87	**0.90**	0.77	0.85	0.44	0.75	0.89	0.85	**0.90**	**0.90**	**0.90**
neg $\mathcal{P}$	0.00	0.54	0.59	0.70	0.36	0.40	0.25	0.34	0.64	0.53	**0.71**	0.69	**0.71**
neg $\mathcal{R}$	0.00	0.65	0.65	0.58	0.46	0.12	0.72	0.51	**0.66**	0.47	0.61	0.63	0.59
neg $\mathcal{F}$	0.00	0.59	0.62	0.64	0.40	0.18	0.37	0.41	0.65	0.49	**0.66**	**0.66**	0.65

(b) UK users who are positive (8,681)/negative (1,140) towards Muslims from only after the Paris attacks

		Content Features			Profile Features				Network Features				All
	BL	Hashtags	Text	All	Desc.	Name	Loc.	All	Mention	Reply	Retweet	All	Features
$\mathcal{AUC}$	0.5	0.62	0.66	0.67	0.53	0.54	0.52	0.54	0.70	0.59	0.68	0.70	**0.71**
Acc	0.78	0.83	0.80	0.83	0.80	0.80	0.79	0.80	0.83	0.82	0.83	0.83	**0.84**
$\mathcal{F}$	0.47	0.58	0.60	0.59	0.51	0.53	0.49	0.52	0.60	0.55	0.59	0.60	**0.61**
pos $\mathcal{P}$	0.88	**0.91**	0.90	0.90	0.89	0.89	0.88	0.89	**0.91**	0.90	**0.91**	**0.91**	0.90
pos $\mathcal{R}$	1.00	0.77	0.93	0.93	0.74	0.84	0.58	0.72	0.85	0.83	0.84	0.87	**0.96**
pos $\mathcal{F}$	0.94	0.83	0.92	0.92	0.81	0.86	0.70	0.80	0.88	0.86	0.87	0.89	**0.93**
neg $\mathcal{P}$	0.00	0.20	0.31	0.31	0.13	0.15	0.11	0.14	0.25	0.19	0.24	0.26	**0.39**
neg $\mathcal{R}$	0.00	**0.43**	0.25	0.24	0.29	0.23	0.40	0.35	0.38	0.30	0.38	0.35	0.20
neg $\mathcal{F}$	0.00	0.27	0.28	0.27	0.18	0.18	0.17	0.20	**0.30**	0.23	0.29	**0.30**	0.26

5.4 Results

Tables 3, 4, and 5 provide the classification results for users who expressed positive/negative stance towards Muslims prior to or only after the Paris attacks for the three countries under consideration. Not surprisingly, since the positive class was the majority class, the classification results for those who expressed a positive stance are on the whole higher than for those who expressed negative views, for all three countries. Further, the results for positive users without prior tweets about Muslims were consistently higher across countries compared to users with prior tweets. However, this is antithetical to the results for users with negative views. For those who expressed views towards Muslims before the attacks, content- and network-based features both yielded relatively high precision and recall in predicting stance after the attacks, with network-based features performing slightly better for the US and the UK. However, for those who did not express views towards Muslims prior to the event, network features consistently outperformed content features, except for the positive class in the UK with no prior tweets, where content features had a slight edge over network features (Table 4 (b)). Combining network and content features often did not yield better results than either one alone (Table 3 (a) and (b)). The performance is better for US and UK compared to France. Besides the training size which is larger for US and UK, the discussions in France are certainly more specific, detailed and contain more issues compared to the other countries as the events happened in France. The variation of discourse in the French dataset along with smaller number of training samples results in less generalization of the model over the French test set. We also repeated the experiment for US users who expressed their opinion before the event but this time removed the tweets which directly mentioned the topic and observed a 1% performance reduction both for accuracy and $\mathcal{F}$ over all features, and about 40% performance reduction over network features, which is indicative of the importance of network features within on-topic tweets. The performance over content features didn't change substantially.

We also evaluated the statistical significance of the results for each feature type using random permutations Ojala and Garriga, 2010 and found all the models to be significant at $p < 0.01$ level, except for the models that only use profile field features. This is not a surprise given that there is not enough signal about the sentiment of the users in profile fields (name, location and description).

The results above highlight the fact that network features that model user interactions on Twitter are either the most effective or slightly lower than the most effective features for predicting a user's stance on a given topic, particularly in the absence of prior discussion of this topic and for the minority class. This finding answers our first two research questions about the possibility of predicting unexpressed views, and the most effective features to achieve that.

Table 5: Results for French users

(a) French users who are positive (1,437 users)/negative (579 users) towards Muslims before the Paris attacks

		Content Features			Profile Features				Network Features				All
	BL	Hashtags	Text	All	Desc.	Name	Loc.	All	Mention	Reply	Retweet	All	Features
$\mathcal{AUC}$	0.5	0.68	0.70	0.69	0.59	0.51	0.54	0.60	0.69	0.56	**0.70**	0.69	**0.70**
Acc	0.51	0.81	0.58	0.79	0.76	0.72	0.73	0.78	0.80	0.75	**0.82**	**0.82**	0.78
$\mathcal{F}$	0.42	0.75	0.78	0.80	0.74	**0.81**	0.74	0.72	0.78	0.77	0.79	0.78	0.80
pos $\mathcal{P}$	0.71	0.81	0.81	0.79	0.76	0.72	0.73	0.78	0.80	0.75	**0.82**	**0.82**	0.78
pos $\mathcal{R}$	1.00	0.67	0.73	0.82	0.71	**0.98**	0.74	0.63	0.76	0.81	0.75	0.72	0.84
pos $\mathcal{F}$	0.83	0.73	0.77	0.80	0.73	**0.83**	0.74	0.69	0.78	0.78	0.78	0.77	0.81
neg $\mathcal{P}$	0.00	0.43	0.47	0.50	0.38	0.44	0.34	0.38	0.47	0.41	0.49	0.47	**0.52**
neg $\mathcal{R}$	0.00	**0.62**	0.59	0.45	0.45	0.04	0.33	0.55	0.54	0.32	0.59	0.61	0.42
neg $\mathcal{F}$	0.00	0.50	0.52	0.48	0.41	0.07	0.34	0.45	0.51	0.36	**0.53**	**0.53**	0.46

(b) French users who are positive (7,236)/negative (1,246) towards Muslims from only after the Paris attacks

		Content Features			Profile Features				Network Features				All
	BL	Hashtags	Text	All	Desc.	Name	Loc.	All	Mention	Reply	Retweet	All	Features
$\mathcal{AUC}$	0.5	0.60	0.65	0.63	0.55	0.51	0.54	0.56	0.63	0.53	**0.64**	0.63	**0.64**
Acc	0.73	0.87	0.75	**0.88**	0.85	0.83	0.85	0.85	**0.88**	**0.88**	0.86	**0.88**	0.87
$\mathcal{F}$	0.46	0.56	0.58	0.58	0.53	0.50	0.52	0.53	0.58	0.58	0.52	**0.59**	0.57
pos $\mathcal{P}$	0.85	**0.89**	0.88	0.88	0.86	0.85	0.87	0.87	0.88	0.88	0.86	**0.89**	0.87
pos $\mathcal{R}$	1.00	0.62	0.80	0.82	0.73	0.59	0.65	0.68	0.83	0.82	0.89	0.79	**0.92**
pos $\mathcal{F}$	0.92	0.73	0.84	0.85	0.79	0.70	0.74	0.76	0.85	0.85	0.88	0.84	**0.89**
neg $\mathcal{P}$	0.00	0.20	0.25	0.27	0.18	0.15	0.17	0.18	0.26	0.26	0.19	0.25	**0.30**
neg $\mathcal{R}$	0.00	**0.54**	0.39	0.37	0.34	0.41	0.41	0.41	0.35	0.37	0.15	0.42	0.20
neg $\mathcal{F}$	0.00	0.29	0.30	0.31	0.23	0.21	0.24	0.25	0.30	0.31	0.17	**0.32**	0.24

5.5 Analysis

Next, we were interested in understanding the underlying features that make the two groups separable. We focus here exclusively on US users. To this end, we interrogated the SVM classification model to identify the most discriminating features that the classifier used to determine if a person would have positive or negative views of Islam and Muslims post-Paris attacks. The results show that network level features — especially mentions and retweets — are better predictors of stance, particularly for the negative class and for the case where users did not mention Islam-related terms prior to the attacks.

Tables 6 and 7 show the top-mentioned/retweeted Twitter accounts and hashtags from users who expressed negative attitudes towards Muslims either before the attacks or only after the attacks, along with those that are shared between both groups. The common categories for both groups are as follows:

- conservative media outlets such as `@FoxNews`, `#theFive`, `@Drudge_Report`, `@theBlaze` and conservative accounts such as `@CloyDrivers`, `@RealJamesWood`, and `#TCOT` (top conservatives on Twitter). Fox News dominated the category with: official accounts (e.g., `@FoxBusiness` and `@FoxNews`) and Fox News presenters and shows (e.g., `@MegynKelly`, `@SeanHannity`, and `@Greta` (Greta Van Susteren); `#KellyFile`, `#Greta`, and `#Hannity`).
- Presidential primaries either on the Republican side (e.g., `#CNBCGopDebate`, `@TedCruz`, `@MarcoRubio`, `#Trump2016`, `@RealDonaldTrump`, and `#BC2DC16` (Ben Carson to DC)) or Democratic side (e.g., `#WhyImNotVotingForHillary`).
- evangelical Christian preachers (e.g., `@Franklin_Graham` and `@JoelOsteen`).
- foreign issues (e.g., `#ISIS`, `#Benghazi`)

Categories that distinguish the group who talked about Muslims before the attacks are:

- pro-Israel media and accounts (e.g., `@Jerusalem_Post` and `@Yair_Rosenberg`).
- atheists with strong anti-religious views (e.g., `#Atheism` and `@Sam-HarrisOrg`).
- secular Muslim activists with strong anti-Islamist views such as `@TarekFatah` and `@MaajidNawaz`.
- strictly anti-Islam/Muslim content such as `@AmyMek` and `@Ayaan`.
- issues relating primarily to abortion (e.g., `#ProLife`, `#PlannedParenthood`, and `#DefundPP`) and race relations (`#ISaluteWhitePeople` and `#BlueLivesMatter` (referring to policemen)).

What sets apart users with strictly post-attack views are sports-related mentions and hashtags (e.g., `@ESPN`, `@NFL`, `@NHL`, `#Patriots`, and `#Nascar`) and those promoting men's rights,

Table 6: Top 40 mentioned/retweeted accounts by users who expressed negative views towards Muslims before or only after after the attack or by both groups ("shared")

Pre-attack Negative

conservative - media/tweep:
@Greta, @Drudge_Report, @SeanHannity, @BreitbartNews, @PrisonPlanet, @DailyCaller, @theBlaze, @Ayaan, @LindaSuhler, @Christiec733, @CharlieDaniels
conservative - election:
@DanScavino (Trump adivsor), @WriteinTrump
atheist/anti-religion: @SamHarrisOrg, @AliAmjadRizvi
Muslim - secular:
@MaajidNawaz, @TarekFatah, @TaslimaNasreen
Israel - media/news:
@Yair_Rosenberg, @Jerusalem_Post, @coinabs
Other:
@AmyMek (Anti-Muslim tweep), @LemondeFR (French media), @TRobinsonNewEra (UK nationalist)

Shared

conservative - media/tweep:
@FoxNews, @MegynKelly, @FoxAndFriends, @AnnCoulter, @FoxBusiness, @NRO, @CloyDrivers, @RealJamesWoods, @ClayTravisBGID
conservative - election:
@RealDonaldTrump, @TedCruz, @JebBush, @MarcoRubio, @RandPaul
atheist/anti-religion: @RichardDawkins
Christian: @Franklin_Graham (Evangelist)

Post-attack Negative

conservative - election: @RealBenCarson
conservative - media/tweep:
@BenShapiro, @SCrowder, @NYPost, @GregGutfeld, @Nero
issues:
@USMC (US Marine Corp - military), @MeninistTweet (men's rights), @CauseWereGuys (men's rights)
Christian: @JoelOsteen (Evangelist)
media/satire: @cnbc, @IowaHawkBlog
sports:
@SportsCenter, @Yankees, @ESPNcfb, @TotalGolfMove, @MLB (baseball), @NFL (football), @DarrenRovell, @ESPN, @NHL (Hokey), @TimTebow (conservative commentator), @OldRowOfficial (conservative tweep)
music: @country_words

Table 7: Top 40 hashtags used by users who expressed negative views towards Muslims before or only after after the attack or by both groups ("shared")

Pre-attack Negative

conservative - elections:
#RNC, #AllInForJeb, #WhyImNotVotingForHillary
conservative - media: #theFive
issues:
{#ProLife, #PlannedParenthood, #DefundPP, #PPSellsBabyParts, #ShoutYourAbortion} (abortion), #ObamaCare (health care), #ISaluteWhitePeople (race relations), #BlueLivesMatter (race relations), #Military, #NeverForget (general), #Hamas (foreign)
music & pop culture:
#PreOrderPurpose, #Legend, #Cats, #Fallout4
sports: #MLB (Major League Baseball)
ideology: #Atheism

Shared

conservative - elections:
#Trump2016, #MakeAmericaGreatAgain, #BC2DC16 (Ben Carson to DC), #Trump, #StandWithRand, #CNBCGopDebate
conservative - media/tweep:
#KellyFile, #Greta, #Hannity, #TCOT (Top Conservatives On Twitter)
issues:
#MillionStudentMarch (education), #ISIS (foreign), #Benghazi (political), #Obama (political)
others:
#GamerGate (online harassment), #pray, #NationalOffendACollegeStudentDay, #CSLewis (author)

Post-attack Negative

conservative - media/tweep:
#PJNet (Patriot Journalist Network), #WakeUpAmerica, #CCOT (Conservative Christian on Twitter), #Merica
conservative - elections: #GOPDebate, #CruzCrew
Christian: #IamAChristian, #Jesus
sports:
#WorldSeries, #Mets, #SEC, #NFL, #Yankees, #OneFinalTeam, #Patriots, #Nascar, #Vols, #RollTide
issues: #ThankAVet (veterans)
other: #safespace, #TFM, #faith

such as @MeninistTweet (vs. feminist) and @CauseWereMen.

We also looked at the most distinguishing profile and content features. Unfortunately, the top profile features (account description, location, and screen name) and top words were not as readily explainable as network features or hashtags. This could be due to their observed relative weakness in distinguishing between the positive and negative classes. Hence, we placed our analysis of the top profile features and most distinguishing words in the Appendix (section A).

Tables 8 and 9 show the top-mentioned/retweeted Twitter accounts and top-used hashtags by users who expressed positive attitudes towards Muslims either before the attacks or only after the attacks, along with those that are shared between both groups. Common categories between the both groups of users are:

- liberal media outlets (e.g., @theDailyShow, @theNation, @NewYorker, @HuffPost, #LibCrib, and #UniteBlue)
- presidential primaries either on the Democratic side (e.g., @HillaryClinton, @BernieSanders, #ImWithHer (referring to Hillary Clinton), and #Bernie2016) or on the Republican side (#BenCarsonWikipedia and #TedCruz)
- indicative of the US president (e.g., @BarackObama or @POTUS (President of the US))
- social issues such as abortion (e.g., #P2), race relations (e.g., #AssaultAtSpringValleyHigh (police beating of

black student) and #BlackLivesMatter), same sex marriage (e.g., #LoveWins), and gun control (e.g., #NRA (National Rifle Assoc.))

- foreign media (e.g., @AJEnglish and @theDailyEdge).

Features that set apart the group who mentioned Muslims before the attacks are:

- Muslim academics (e.g., @RezaAslan), activists (e.g., @FreeLaddin), artists (e.g., @ShujaRabbani), and comedians (e.g., @DeanOfComedy).

- support for Muslims around the world (e.g., #Kunduz (an Afghan city, where a hospital was bombed by the US) and #Rohingya (a persecuted Muslim minority in Myanmar)) and attacks against Muslims in the US (e.g., #IStandWithAhmed (the student who was arrested for making a clock) and #ChapelHillShooting (a hate crime resulting in the death of Muslim students)).

- African American media and persons (e.g., @theRoot)

What discerns users with only post-attacks views are those pertaining to music (e.g., @ComplexMusic, @Acapella-Vids, #EDM (electronic dance music), and #AMAS (American Music Awards)). The prevalence of music and absence of sports for this group (the opposite of what we observed in the equivalent group with negative views) requires further investigation. Though it may seem surprising at first, there is evidence in the literature that food, sports, and music preferences are often correlated with political polarization DellaPosta *et al.*, 2015; Garimella and Weber, 2014.

6 Discussion

6.1 Methodology

Our approach for predicting the stance of individuals in this paper is based on past behavior on social media, focusing in part on users who have expressed no explicit opinion on a particular topic in the past. The methodology involves analyzing two types of data, namely: (1) post interactions (tweets and network activity), in which we are able to learn a user's stated stance towards an event, an issue, or a group based on sampling methods and crowd-sourced annotations; and (2) pre-interactions, which are used to build a classifier to predict stance which is expressed only later. For the specific case study in this paper, our results show that using a user's pre-attack network interactions can predict a user's positive or negative attitudes towards Muslims with 90% and 79% precision, respectively, even when they had not previously mentioned *Islam*, *Muslims*, or related terms. This work extends previous research in which content-based and network-based analysis was used to predict future support or opposition to an entity Magdy *et al.*, 2016b; Pennacchiotti and Popescu, 2011. Our work here suggests that network-based analysis may often be more reliable than content-based analysis.

Table 8: Top 40 mentioned/retweeted accounts by users who expressed positive views towards Muslims before or only after after the attack or by both groups ("shared")

Pre-attack Positive

liberal - media/tweep: @JohnFugelsang, @TheEconomist, @TheNation, @HuffPostRelig, @NewYorker, @MyDaughtersArmy, @Salon, @Libertea2012, @WilW
liberal - election/political: @HillaryClinton, @MoveOn
Muslim - academic/activist: @RezaAslan, @TariqRamadan, @FreeLaddin
Muslim - comedian/artist: @DeanOfComedy, @AzizAnsari, @ShujaRabbani
pop culture/science: @UncleRush, @TedTalks
sports: @KingJames (basketball)
actors: @MattMcgorry (US), @AnupAmpkher (India)
Other:
@AJEnglish (Aljazeera), @TheRoot (African American-media), @OhNoSheTwitnt (comedian), @BabyAnimalPics

Shared

liberal - media/tweep: @Bipartisanism, @TheDailyShow, @BuzzFeed, @NYTimes, @LOLGop
liberal - election: @BernieSanders, @SenSanders
liberal - US president: @POTUS
pop culture: @RollingStone
US-civil rights activist: @DeRay
Other:
@TheDailyEdge (foreign media), @Mark_Beech (UK actor), @JK_Rowling (UK liberal author), @DavidKWilliams (US business person)

Post-attack Positive

liberal - media/tweep: @HuffingtonPost, @Maddow, @ThinkProgress, @NeilTyson, @SarahKSilverman, @StephenKing
liberal - US president: @WhiteHouse, @BarackObama
music/media/TV/pop culture: @NPR, @VoxDotCom, @ComplexMusic, @FuckTyler, @JoeBudden, @AcapellaVids, @WSHHFans, @JonBuckhouse, @ColiegeStudent, @MattBellassai, @MrCocoyam, @AnnaKendrick47
US-civil rights activist: @_JonathanButler
sports: @Arsenal, @TSBible
foreign person: @DalaiLama (Bhuddist), @LoaiDeeb (tweep)
Other: @CuteEmergency

6.2 Homophily or Social Influence

As we can see from the results, network features — as primarily manifested in retweets and mentions — are strong predictors of a user's stance on a given topic, even when they have not mentioned that topic in their posts. For the presented case study, network features have a precision of 0.79 for the minority class (negative views towards Muslims) even for users who had not mentioned Muslims previously. The power of network features can be a result of either homophily — the propensity of individuals to interact with similarly minded individuals —

Table 9: Top 40 hashtags by users who expressed positive views towards Muslims before or only after after the attack or by both groups ("shared")

Pre-attack Positive

liberal - election:
#ImWithHer, #BenCarsonWikipedia, #Bernie2016
liberal - tweeps/media:
#GOPClownCar, #Maddow, #LibCrib, #UniteBlue, #inners, #DemForum
issues:
#P2 (abortion), #NRA (guns), #HumanRights (human rights), #ConcernedStudent1950 (race relations), #LBGT (gay rights)
pop culture & music:
#Emmys, #empire, #GreysAnatomy, #DoctorWho, #BackToTheFuture
support for Muslims worldwide:
#Kunduz, #Rohingya, #Palestine, #Gaza
conservative:
#TedCruz (election), #BB4SP (tweep)
anti-Muslim attack: #ChapelHillShooting
general:
#peace, #news, #TacoEmojiEngine
media & humor:
#MorningJoe, #IBDEditorials, #StuffHappens
Muslim specific: #EidMubarak

Shared

anti-Muslim act: #IStandWithAhmed
issues:
#StandWithPP (abortion), #AssaultAtSpringValleyHigh (race relations), #LoveWins (gay rights), #ActOnClimate (climate change)
liberal - election:
#FeelTheBern, #DebateWithBernie, #IAmWithHer

Post-attack Positive

issues:
#BookBoost (education), #nanowrimo (education), #AmWriting (education), #Afghanistan (foreign), #BlackLivesMatter (race relations), #SandraBland (race relations)
music:
#EDMA, #EDM, #EDMLifestyle, #EDMFamily, #EDMLife, #MadeInTheAM, #AMAS, #WomenInMusic, #DJSet
pop culture:
#arrow, #theFlash, #htgawm, #supernatural, #AllMyMovies, #StarGate, #MasterOfNone, #SuperGirl, #MockingJayPart2, #tvd
Muslim activist: #DrLoaiDeeb, #WeSupportGNRD
general:
#business, #lrt, #leadership, #gratitude, #halloween

or social influence — where individual attitudes are affected by the attitudes of others. For example, in our study we observe that individuals who follow conservative media outlets are more likely to harbor negative attitudes towards Muslims. Whether these individuals follow such media sources because they agree with their stance towards Muslims, or whether they started having anti-Muslim views because they tune in to such media, is unclear. Prior research has shown a strong tendency for homophily in social networks based, for example, on politics or ideology. It could be that individuals coalesce, for example, around broad political positions, but rely on others who share the same broad position to shape their position towards narrow topics. This warrants further investigation.

6.3 Prediction

The ability to predict a person's unstated stance (or probable stance) has many implications and applications, as outlined below.

6.3.1 Recommendation

As can be seen from the results, users who are closer together from a network standpoint may also share similar preferences. In this study, we were able to observe this not just in terms of positions towards an ethnic or religious group, but also in terms of preference of religion, media outlets, and potentially music and sports. Though choice of music and political stance may seem unrelated, recent work on so-called "lifestyle politics" suggest that such correlations are real DellaPosta *et al.*, 2015 and could be used by recommender systems Weber and Garimella, 2014. Thus, network information may aid in providing more accurate recommendations to users and better targeted advertising.

6.3.2 Ascertaining unspoken views

Users may avoid expressing positions explicitly for many reasons, such as fear of social judgment or political repression, especially under repressive regimes. As seen in our study, predicting unexpressed positions may be possible based not just on an individual's network interactions but also, as suggested by lifestyle politics research, preferences for specific music, sports, or food items. On the positive side, such predictions may be utilized to guess how a population may vote in elections or referenda. On the negative side, it can be used by oppressive regimes to identify potential dissidents, though they may not express their opposition publicly.

6.3.3 Population segmentation

As can be seen from the case study, those who expressed positive (or negative) views towards Muslims were not a homogeneous whole. For example, those with positive views included, inter alia, Muslims, liberals, and civil rights activists. The methodology that we employed provides the ability to ascertain underlying groups who may share a common position towards an issue. The ability to discover such groups (i.e., segment the population) can be helpful for a variety of applications. For example, marketers may be able to perform market segmentation. Similarly, political candidates, activists, or politicians can craft targeted messages to different constituent sub-groups.

7 Conclusion

In this paper, we presented a methodology for predicting a person's stance towards an issue, topic, or group in response

Figure 6: Top 20 terms in profile description indicating negative views

Figure 7: Top 20 terms in profile description indicating negative views

to an event and given previous activity on social media sites. As a case study, we used the views of Twitter users towards Muslims in the wake of the Paris terrorist attacks of Nov. 13, 2015. We show that previous Twitter interactions — particularly network-based interactions — serve as strong predictors of stance. Prediction is possible because users tend to congregate with like-minded users online (homophily) and are influenced by the views of others in their social network (social influence). Social media messages and networks therefore have profound influence on political attitudes and shape national and international policy. Therefore, the relative effects of homophily and social influence warrant further research for more accurate predictions of community response to crises and the drivers of policy change Colleoni *et al.*, 2014.

Successful prediction can facilitate much interesting research. One such area is so-called lifestyle politics, where the objective is to discover correlations between preferences (e.g., in music or sports) and political views. What correlations exist and why they exist are interesting lines of future work. Another area is the identification of the traits (e.g., political, ideological, economic, or religious) of people holding particular views. Such identification can help in areas such as population segmentation, which would have impact on other areas like automatic recommendation and targeted marketing. There has been some recent work on employing such user traits for recommendation Weber and Garimella, 2014, but this area is rather nascent and requires much further work.

Figure 8: Top 20 terms in location field indicating negative views

Figure 9: Top 20 terms in location field indicating positive views

A Top Features

As for the profile and content features, Figures 6 through 13 show tag clouds of the most distinguishing features per feature source. Figures 6 and 7 show the most discriminating words in profile descriptions for negative and positive classes respectively. For the negative class, the words indicating political leaning (e.g., *conservative* and *Trump*), religious persuasion (e.g., *Jesus*), and nationalism (e.g., *patriot*) stand out. For the positive class, the most notable terms were those indicating activism such as *feminist*, *community*, and *service*. Another interesting contrast is the presence of the words *retired* and *student* for the negative and positive classes respectively, which may indicate an age gap.

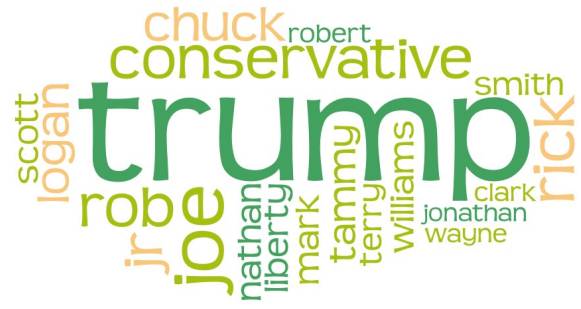

Figure 10: Top 20 terms in screen name field indicating negative views

Figure 11: Top 20 terms in screen name field indicating positive views

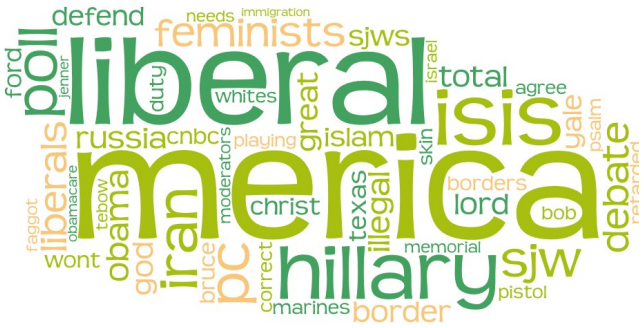

Figure 12: Top 20 terms in the text indicating negative views

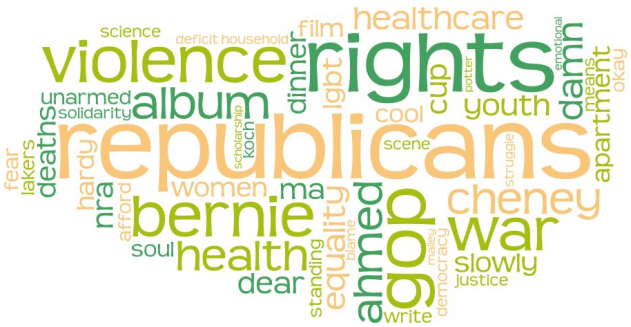

Figure 13: Top 50 terms in the text indicating positive views

For the terms in the location field, which yielded lower classification effectiveness, the most distinguishing terms for the negative class (Figure 8) prominently featured the words *southern* and *south* (noting that Southern states are typically more conservative), and names of states (or cities therein) that voted for Trump in the 2016 presidential election such as *Texas*, *Arizona*, and *Kentucky*. The positive class was dominated by traditionally democratic states (e.g., *New York*) and territories (e.g., *Puerto Rico*) and foreign locations (e.g., *Khobar (Saudi Arabia)* and *Korea*), but more conservative locales such as *Dakota* and *Denton (Texas)* were also present. For the terms in the user screen names, the most discernible terms were *Trump* and *conservative* for the negative class. We could not ascertain the relationship of other terms to classification. The top 50 most discriminating terms in the text of tweets for the negative class (Figure 12) were *merica* (slang for the US that used by prominent conservative Twitter users), traditional foes of conservatives (e.g., *Obama*, *liberal* and *feminist*), external enemies (e.g., *ISIS*, *Iran*, and *Russia*), conservative issues (e.g., *taxes* and *illegal (immigration)*), and religiously related terms (e.g., *God*). The positive class (Figure 13) was almost the polar opposite with prominent terms indicating traditional foes of liberals (e.g., *Republicans* and *(Dick) Cheney*) and liberal issues (e.g., *rights*, *healthcare*, and *equality*).

References

Anand, P., M. Walker, R. Abbott, J. E. F. Tree, R. Bowmani, and M. Minor. 2011. "Cats rule and dogs drool!: Classifying stance in online debate". In: *The 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. 1–9.

Awan, I. 2014. "Islamophobia and Twitter: A Typology of Online Hate Against Muslims on Social Media". *Policy & Internet*. 6(2): 133–150.

Baldwin, T., P. Cook, M. Lui, A. MacKinlay, and L. Wang. 2013. "How Noisy Social Media Text, How Diffrnt Social Media Sources?" In: *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)*. 356–364.

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

Barberá, P., J. T. Jost, J. Nagler, J. A. Tucker, and R. Bonneau. 2015. "Tweeting from Left to Right: Is Online Political Communication more than an Echo Chamber?" *Psychological Science*.

BBC. 2015. "Paris attacks: What happened on the night". *BBC*. Nov. URL: http://www.bbc.com/news/world-europe-34818994.

Borge-Holthoefer, J., W. Magdy, K. Darwish, and I. Weber. 2015. "Content and network dynamics behind Egyptian political polarization on Twitter". In: *CSCW 2015*. 700–711.

Burfoot, C., S. Bird, and T. Baldwin. 2011. "Collective Classification of Congressional Floor-Debate Transcripts". In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011)*. 1506–1515.

Castillo, M., M. Haddad, M. Martinez, and S. Almasy. 2015. "Paris suicide bomber identified; ISIS claims responsibility for 129 dead". *CNN*. Nov. URL: http://edition.cnn.com/2015/11/14/world/paris-attacks/.

Cheng, Z., J. Caverlee, and K. Lee. 2010. "You are where you tweet: a content-based approach to geo-locating Twitter users". In: *CIKM 2010*. 759–768.

Chenoweth, E. and M. J. Stephan. 2011. *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.

Chung, W.-T., K. Wei, Y.-R. Lin, and X. Wen. 2016. "The Dynamics of Group Risk Perception in the US After Paris Attacks". In: *International Conference on Social Informatics*. Springer. 168–184.

Cialdini, R. B. and M. R. Trost. 1998. "Social influence: Social norms, conformity and compliance". In: *The Handbook of Social Psychology*. Ed. by D. T. Gilbert, S. T. Fiske, and G. Lindzey. McGraw-Hill.

Cohen, R. and D. Ruths. 2013. "Classifying Political Orientation on Twitter: It's Not Easy!" In: *ICWSM 2013*.

Colleoni, E., A. Rozza, and A. Arvidsson. 2014. "Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data". *Journal of Communication*. 64(2): 317–332.

Conover, M. D., B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer. 2011. "Predicting the political alignment of twitter users". In: *The 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom)*. 192–199.

Dalton, R. J. 2013. *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies: Public Opinion and Political Parties in Advanced Industrial Democracies*. CQ Press.

DellaPosta, D., Y. Shi, and M. Macy. 2015. "Why Do Liberals Drink Lattes?" *American Journal of Sociology*. 120(5): 1473–1511. ISSN: 00029602, 15375390.

Dubois, E. and D. Gaffney. 2014. "The Multiple Facets of Influence Identifying Political Influentials and Opinion Leaders on Twitter". *American Behavioral Scientist*. 58(10): 1260–1277.

Esuli, A. and F. Sebastiani. 2006. "SentiWordNet: A publicly available lexical resource for opinion mining". In: *LREC 2006*. Vol. 6. 417–422.

Faulkner, A. 2014. "Automated classification of stance in student essays: An approach using stance target information and the Wikipedia link-based measure". In: *The 27th International Florida Artificial Intelligence Research Society Conference*. 174–179.

Garimella, V. R. K. and I. Weber. 2014. "Co-following on Twitter". In: *The 25th ACM Conference on Hypertext and Social Media*. 249–254. ISBN: 978-1-4503-2954-5.

Golbeck, J. and D. Hansen. 2014. "A method for computing political preference among Twitter followers". *Social Networks*. 36: 177–184.

Groseclose, T. and J. Milyo. 2005. "A measure of media bias". *The Quarterly Journal of Economics*: 1191–1237.

Hamaide, S. de la. 2015. "Timeline of Paris attacks according to public prosecutor". *Reuters*. Nov. URL: https://www.reuters.com/article/us-france-shooting-timeline/timeline-of-paris-attacks-according-to-public-prosecutor-idUSKCN0T31BS20151114.

Han, B., P. Cook, and T. Baldwin. 2012. "Geolocation Prediction in Social Media Data by Finding Location Indicative Words". In: *COLING 2012*. 1045–1062.

Han, B., P. Cook, and T. Baldwin. 2014. "Text-based Twitter User Geolocation Prediction". *Journal of Artificial Intelligence Research*. 49: 451–500.

Hecht, B., L. Hong, B. Suh, and E. H. Chi. 2011. "Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles". In: *SIGCHI Conference on Human Factors in Computing Systems*. 237–246.

Himelboim, I., S. McCreery, and M. Smith. 2013. "Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter". *Journal of Computer-Mediated Communication*. 18(2): 40–60.

Korda, H. and Z. Itani. 2013. "Harnessing social media for health promotion and behavior change". *Health Promotion Practice*. 14(1): 15–23.

Laranjo, L., A. Arguel, A. L. Neves, A. M. Gallagher, R. Kaplan, N. Mortimer, G. A. Mendes, and A. Y. Lau. 2015. "The influence of social networking sites on health behavior change: a systematic review and meta-analysis". *Journal of the American Medical Informatics Association*. 22(1): 243–256.

Magdy, W., K. Darwish, and N. Abokhodair. 2015. "Quantifying Public Response towards Islam on Twitter after Paris Attacks". *arXiv preprint arXiv:1512.04570*.

Magdy, W., K. Darwish, A. Rahimi, N. Abokhodair, and T. Baldwin. 2016a. "#ISISisNotIslam or #DeportAllMuslims? Predicting Unspoken Views". In: *Proceedings of the 8th International ACM Web Science Conference 2016 (WebSci 2016)*. 95–106.

Magdy, W., K. Darwish, and I. Weber. 2016b. "#FailedRevolutions: Using Twitter to study the antecedents of ISIS support". *First Monday*. 21(2).

Ojala, M. and G. C. Garriga. 2010. "Permutation tests for studying classifier performance". *Journal of Machine Learning Research*. 11: 1833–1863.

Pang, B. and L. Lee. 2008. "Opinion mining and sentiment analysis". *Foundations and Trends in Information Retrieval*. 2(1-2): 1–135.

Pavalanathan, U. and J. Eisenstein. 2015. "Confounds and Consequences in Geotagged Twitter Data". In: *EMNLP 2015*. 2138–2148.

Pennacchiotti, M. and A.-M. Popescu. 2011. "Democrats, Republicans and Starbucks afficionados: user classification in Twitter". In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 430–438.

Qiu, M., Y. Sim, N. A. Smith, and J. Jiang. 2015. "Modeling User Arguments, Interactions, and Attributes for Stance Prediction in Online Debate Forums". In: *Proceedings of the 2015 SIAM International Conference on Data Mining*. 855–863.

Rahimi, A., D. Vu, T. Cohn, and T. Baldwin. 2015. "Exploiting Text and Network Context for Geolocation of Social Media Users". In: *NAACL-HLT 2015*. 1362–1367.

Rajadesingan, A. and H. Liu. 2014. "Identifying Users with Opposing Opinions in Twitter Debates". In: *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. 153–160.

Rao, D., D. Yarowsky, A. Shreevats, and M. Gupta. 2010. "Classifying latent user attributes in twitter". In: *Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents*. 37–44.

Runnymede Trust, L. (K. 1997. *Islamophobia A challenge for us all*.

Speriosu, M., N. Sudan, S. Upadhyay, and J. Baldridge. 2011. "Twitter polarity classification with label propagation over lexical links and the follower graph". In: *The 1st Workshop on Unsupervised Learning in NLP*. 53–63.

Sridhar, D., L. Getoor, and M. Walker. 2014. "Collective stance classification of posts in online debate forums". *ACL 2014*: 109–117.

Syeed, N. 2015. "Paris Terror Attacks: Yes, Parisians are traumatised, but the spirit of resistance still lingers". *Independent.ie*. Nov. URL: http://goo.gl/toaabz.

Thomas, M., B. Pang, and L. Lee. 2006. "Get out the vote: Determining support or opposition from Congressional floor-debate transcripts". In: *EMNLP 2006*. 327–335.

Turner, J. C. 1991. *Social Influence*. Thomson Brooks/Cole Publishing Co.

Walker, M. A., J. E. F. Tree, P. Anand, R. Abbott, and J. King. 2012. "A Corpus for Research on Deliberation and Debate." In: *LREC 2012*. 812–817.

Weber, I., V. R. K. Garimella, and A. Batayneh. 2013. "Secular vs. islamist polarization in Egypt on Twitter". In: *The 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 290–297.

Weber, I. and V. R. K. Garimella. 2014. "Using Co-Following for Personalized Out-of-Context Twitter Friend Recommendation." In: *ICWSM 2014*.

Wong, F. M. F., C. W. Tan, S. Sen, and M. Chiang. 2013. "Quantifying Political Leaning from Tweets and Retweets." In: *ICWSM 2013*. 640–649.

Zampieri, M., L. Tan, N. Ljubešić, J. Tiedemann, and P. Nakov. 2015. "The Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects". In: *Overview of the DSL Shared Task 2015*. Hissar, Bulgaria. 1–9.

Zhang, L., R. Ghosh, M. Dekhil, M. Hsu, and B. Liu. 2011. "Combining lexicon-based and learning-based methods for Twitter sentiment analysis". In: *HP Laboratories Technical Report*.

Kareem Darwish is a senior scientist at the Arabic Language Technologies group (ALT) at the Qatar Computing Research Institute (QCRI) with interest in information retrieval, social computing, and natural language processing. Kareem Darwish worked as a researcher at the Cairo Microsoft Innovation Lab and the IBM Human Language Technologies group in Cairo. He also taught at the German University in Cairo and Cairo University.

Walid Magdy is a Lecturer (assistant professor) at the school of Informatics, the University of Edinburgh (UoE). His main research interests include computational social science, information retrieval, and data mining. Before joining UoE, Walid worked for five years as a scientist at Qatar Computing Research Institute (QCRI). He also worked in his early career for IBM and Microsoft as a research engineer.

Afshin Rahimi is a PhD candidate at the School of Computing and Information Systems, The University of Melbourne. His research focuses on Natural Language Processing, Social Network Analysis, and Machine Learning. He is specifically interested in exploiting both structured and unstructured data to gain insights about people and their interests. He also worked on combining text and network information in a semi-supervised setting to learn user demographics and specifically Twitter user geolocation.

Timothy Baldwin is a Professor in the School of Computing and Information Systems, The University of Melbourne. His primary research focus is on natural language processing (NLP), including social media analytics, computational lexical semantics, deep learning, and topic modeling.

Norah Abokhodair is an applied social scientist working in the area of human computer interaction, computer mediated communication, mobile/ubiquitous computing and social media. Her research targets two areas. First, she examines the information practices and behaviors of Arab social media users to inform the design of global and inclusive technology. Second, she researches the use and influence of social and political bots on social computing systems. Norah received her PhD from the University of Washington and is currently a researcher at Microsoft's Cloud and Enterprise division.