ISISisNotIslam or DeportAllMuslims?: Predicting Unspoken Views

#ISISisNotIslam or #DeportAllMuslims? Predicting Unspoken Views

Walid Magdy, Kareem Darwish
Qatar Computing Research Institute
Hamad bin Khalifa University
Doha, Qatar
{wmagdy, kdarwish}@qf.org.qa

Norah Abokhodair
The Information School
University of Washington
Seattle, USA
noraha@uw.edu

Afshin Rahimi,Timothy Baldwin
Dept. of Computing and Information Systems
The University of Melbourne
Victoria 3010, Australia
{arahimi, tbaldwin}@unimelb.edu.au

ABSTRACT

This paper examines the effect of online social network interactions on future attitudes. Specifically, we focus on how a person's online content and network dynamics can be used to predict future attitudes and stances in the aftermath of a major event. In this study, we focus on the attitudes of US Twitter users towards Islam and Muslims subsequent to the tragic Paris terrorist attacks that occurred on November 13, 2015. We quantitatively analyze 44K users' network interactions and historical tweets to predict their attitudes. We provide a description of the quantitative results based on the content (hashtags) and network interaction (retweets, replies, and mentions). We analyze two types of data: (1) we use post-event tweets to learn users' stated stances 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 stances. We found that pre-event network interactions can predict someone's attitudes towards Muslims with 82% macro F-measure, even in the absence of prior mentions of Islam, Muslims, or related terms.

CCS Concepts

•**Social and professional topics** → **User characteristics;** •**Computing methodologies** → Model development and analysis; •**Applied computing** → *Law, social and behavioral sciences;*

Keywords

Network analysis, Twitter data analysis, Stance prediction, Paris attacks, Homophily, Social influence

WebSci '16, May 22-25, 2016, Hannover, Germany

DOI: http://dx.doi.org/10.1145/2908131.2908150

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 wasn't 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 subgroups [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 this work in examining the effect of online social network interactions — in terms of content and network dynamics — on future attitudes and stances in the aftermath of a major event. Specifically, we examine two primary research questions, namely:

1. Can users' social posts and interactions be used to predict their stance on a given topic, even if they have never mentioned that topic?
2. What are the most predictive feature/approaches for stance prediction?

To answer these two 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) 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 US

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 and timeline tweets of users who indicated a personal stance towards Muslims 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 with the Twitter community, through mentions, retweets, and replies).

Our dataset contains more than 44,000 US-based 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 77% of users have a positive stance towards Muslims (and 23% negative).

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. Also, our results show 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).

Additionally, We provide analysis of how different features can affect the prediction performance, and discuss the implications of our findings.

2. BACKGROUND

2.1 The Terrorist Attacks on Paris 2015

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 German was being played. Other suicide bombings and mass shootings occurred a few minutes later at cafés, restaurants and a music venue in Paris [de la Hamaide 2015, BBC 2015].

The tragic events resulted in more than 130 deaths and 368 injured people, 80–99 seriously. 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 its 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 in Western media, 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 US 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. Given tweets from 58 countries, tweets defending Muslims outnumbered tweets attacking them for all, but two countries. It was also shown that 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, specifically for US users.

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 follow 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 bear a resemblance to an "echo chamber", or provides a space for a 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 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 of 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.

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

2.4 Consistency of Orientation

In terms of 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 opinions over a span of at least 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] 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 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 contain misspellings, informal language, and slang [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.

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]. 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 stances toward Islam.

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] have shown 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 use Latin alphabet, in practice, we were able to retrieve data for a large number of languages.

An open-source language identification system was then applied to each of the tweets 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%) are 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. 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. To keep our analysis focused, we were interested exclusively in English tweets originating from the US, which is the country with the highest number of tweets in our collection. Thus, we excluded all non-English tweets. In Section 3.4 we discuss how we filtered the tweets by country.

[2]http://topsy.com/ *(currently unavailable)*
[3]https://github.com/shuyo/language-detection

Positive	Negative
#MuslimsAreNotTerrorist (34,925)	#IslamIsTheProblem (3,154)
#MuslimAreNotTerrorist (17,759)	#RadicalIslam (1,618)
#NotInMyName (4,728)	#StopIslam (1,598)
#MuslimsStandWithParis (1,228)	#BanIslam (460)
#MuslimsAreNotTerrorists (1,106)	#StopIslamicImmigration (333)
#ThisisNotIslam (781)	#IslamIsEvil (290)
#NothingToDoWithIslam (619)	#IslamAttacksParis (280)
#ISISareNotMuslim (316)	#ImpeachTheMuslim (215)
#ExtremistsAreNotMuslim (306)	#KillAllMuslims (206)
#ISISisNotIslam (243)	#DeportAllMuslims (186)

Table 1: Examples of the top hashtags that refer to positive and negative attitudes towards Muslims. The frequency of each hashtag in the data is provided in parentheses

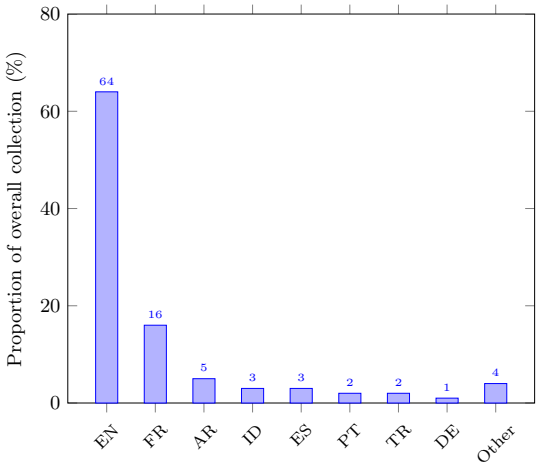

Figure 1: Language distribution of the tweet collection (based on ISO-639-2 language codes)

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*. Out of the 8.36 million tweets, we extracted 753,476 English tweets mentioning something about Islam. This constitutes 14% of the collected English tweets, which shows that reactions to Muslims after the attacks was one of the most popular topics.

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 attitude expressed in the tweets, we sampled the data collection by getting a representative sample of tweets. We used a sample size calculator[4] 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%. We then selected a set of tweets of this size — namely 1,534 tweets — at random from the full collection, for manual annotation.

For the manual annotation, we submitted the sampled tweets to CrowdFlower.[5] 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 different annotators. The percentage of disagreement among annotators shows that some tweets are not straightforward to label. Usually this occurs between neutral and one of the other attitudes.

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 by 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. After label propagation, 61% of the labeled tweets conveyed positive attitude towards Muslims, 17% negative, and the rest neutral.

Table 1 provides some examples of the most frequent hashtags in our collection that refer to positive and negative attitudes towards Muslims. As shown, the hashtags on the left side disassociate between ISIS/terrorism and Islam, while those on the right side mainly focus on a call to ban Muslims from entering western countries, and some of them go as far as to call for the extermination of all Muslims (#KillAllMuslims). The volume of positive hashtags is much larger than those with negative attitude.

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.

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

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 for declaring the location. 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 the user-declared locations to countries, as follows:

1. A list of 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, was 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 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 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, 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 regularized 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.

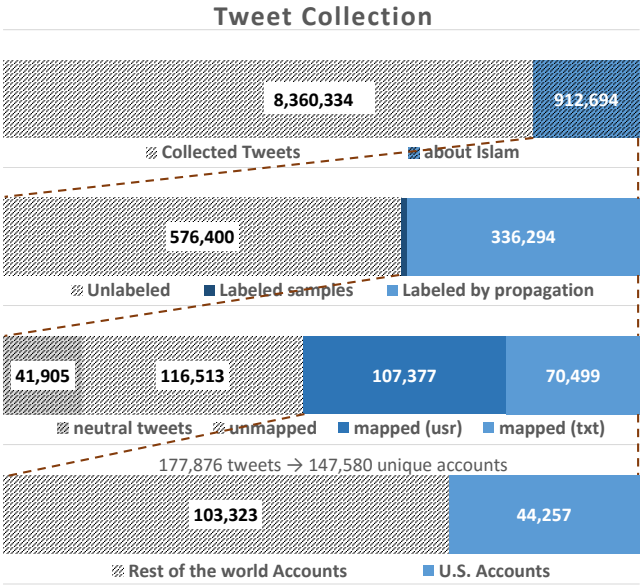

Figure 2: Summary of the tweet collection used in this study. The first three rows show the numbers of tweets; the final row shows the number of Twitter accounts.

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 keep the top 50% of the 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 U.S., which is the largest number among all countries.

Figure 2 provides a breakdown of the tweet collection, and 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. The final set of 44k U.S. accounts are the ones used in this study. Account information and tweet timelines were collected for each of these accounts for the prediction process described in the following section.

4. PRE-ATTACK PREDICTION

4.1 Prediction

Next, we try to use the pre-event Twitter interactions and profile information of users to predict their post-event stance. We use content, profile 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 it to the user, based on the assumption that the stance of a user will be consistent over time (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 stances toward Muslims. Subsequent qualitative analysis of these features can shed light on personal, social and political attributes that are predictive of the stance a user would adopt.

4.2 Pre-Attack Data Collection

The number of users with either positive or negative stance who were geolocated in the U.S. is 44k. We used the Twitter API to crawl (up to) 200 tweets for each of these users that were posted before the attacks.[6] 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. The total number of collected tweets for the 44k users was 12,574,882 *pre-attack* tweets.

4.3 Prediction of Future Stances

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.
- *profile features*: user-declared profile information, namely the name, profile description, and location.
- *network features*: user interaction activities, namely other accounts that a user mentioned, retweeted, and replied to.

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 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 occur 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.

[6]The API allows specifying tweet ID to get history tweets of a given user posted before it.

4.4 Results

Tables 2 and 3 provide the classification results for users who expressed positive/negative stance towards Muslims prior to or only after the Paris attacks, respectively. For those who expressed views towards Muslims before the attacks, content- and network-based features both yielded high precision ("$\mathcal{P}$") and high recall ("$\mathcal{R}$") in predicting their stance after the attacks, with network-based features performing slightly better. The results for those who expressed a positive stance are on the whole higher than for those who expressed negative views. This is due in large part to the class imbalance (users who expressed positive views outnumbered those who expressed negative views by approximately 3 to 2). For those who did not express views towards Muslims prior to the attacks, content features yielded comparable precision and lower recall compared to network features for those who expressed positive views. However, the effectiveness of the content based features may be attributed to the fact that positive users outnumbered negative users by more than 4 to 1. For those who expressed negative views only after the attacks, content-based features yielded much lower precision (0.58) than network features (0.79). Also, using both content and network features together led to lower results than using the network features alone.

The results above highlight the fact that network features that model user interactions on Twitter are the most effective for predicting a user's stance on a given topic, even in the absence of prior discussion of this topic. This finding answers both our research questions about the possibility of predicting unexpressed views, and the most effective features to achieve that.

4.5 Analysis

Next, we were interested in understanding the underlying features that make the two groups separable. To this end, we interrogated the SVM classification model to identify the most distinguishing features that the classifier used to determine if a person would have positive or negative views of Islam and Muslims post-Paris attacks. As the results show, 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 4 and 5 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`, `@Drudge_Report`, `@theBlaze`, `#theFive` and conservative accounts such as `@CloyDrivers`, `@RealJamesWood`, and `#TCOT` (top conservatives on Twitter). Fox News dominated the category with: official accounts (e.g., `@FoxNews` and `@FoxBusiness`) 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

		Content Features			Profile Features				Network Features				All
	BL	hashtags	text	All	Desc.	Name	Loc.	All	mention	reply	retweet	All	Features
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

Table 2: U.S. 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
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

Table 3: U.S. users who are positive (27,457)/negative (6,119) towards Muslims from only after the Paris attacks

(e.g., `@RealDonaldTrump`, `@TedCruz`, `@MarcoRubio`, `#Trump2016`, `#BC2DC16` [Ben Carson to DC], and `#CNBCGopDebate`) or on the Democratic side (e.g., `#WhyImNotVotingForHillary`).

- evangelical Christian preachers (e.g., `@Franklin_Graham` and `@JoelOsteen`).
- political and foreign issues (e.g., `#ISIS`, `#Benghazi`, `#Obama`)

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 who have strong anti-religion views (e.g., `@SamHarrisOrg` and `#Atheism`).
- 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`), 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, such as `@MeninistTweet` (counter to feminist) and `@CauseWereMen`.

Tables 6 and 7 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., `@theNation`, `@NewYorker`, `@theDailyShow`, `@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` [black student beaten by police] and `#BlackLivesMatter`), same sex marriage (e.g., `#LoveWins`), and gun control (e.g., `#NRA` [National Rifle Assoc.])
- foreign media outlets (e.g., `@AJEnglish` and `@theDailyEdge`).

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

- Muslim academics (e.g., `@Reza Aslan` and `@TariqRamadan`), activists (e.g., `@FreeLaddin`), comedians (e.g., `@DeanOfComedy`), and artists (e.g., `@ShujaRabbani`)
- 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 sets apart users with strictly post-attacks views are those pertaining to music (e.g., `@ComplexMusic`, `@AcapellaVids`, `#EDM` [electronic dance music], and `#AMAS` [American

Pre-attack Negative

conservative - media/tweep:
@Greta, @Drudge_Report, @SeanHannity, @BreitbartNews, @PrisonPlanet, @DailyCaller, @theBlaze, @Ayaan, @Linda-Suhler, @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, @Clay-TravisBGID
conservative - election:
@RealDonaldTrump, @TedCruz, @JebBush, @MarcoRubio, @Rand-Paul
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), @OldRowOf-ficial (conservative tweep)
music: @country_words

Table 4: 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 - elections:
#RNC, #AllInForJeb, #WhyImNotVotingForHillary
conservative - media: #theFive
issues:
{#ProLife, #PlannedParenthood, #DefundPP, #PPSellsBaby-Parts, #ShoutYourAbortion} (abortion), #ObamaCare (health care), #ISaluteWhitePeople (race relations), #BlueLives-Matter (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), #Beng-hazi (political), #Obama (political)
others:
#GamerGate (online harassment), #pray, #NationalOffendA-CollegeStudentDay, #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

Table 5: Top 40 hashtags used by users who expressed negative views towards Muslims before or only after after the attack or by both groups ("shared")

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 are indications in the literature that food, sports, and music preferences are often correlated with political polarization [DellaPosta et al. 2015, Garimella and Weber 2014].

5. DISCUSSION

5.1 Methodology

In this work we presented a method for predicting the stance of individuals based on their past behavior on social media, focusing in part on users who have explicitly expressed no 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 users' stated stances towards an event, an issue, or a group based on sampling methods and crowdsourced annotations; and (2) pre-interactions, which are used to build a classifier to predict stances which are 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 only content-based analysis was used to predict future support or opposition to an entity [Magdy et al. 2016]. Our work here suggests that network-based analysis may often be more reliable than content-based analysis.

5.2 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 person'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 view 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

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

Table 6: 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 - 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

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

similarly minded individuals — 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 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 requires further investigation.

5.3 Prediction

The ability to predict a person's position (or likely position) when the person has not stated an explicit stance has many implications and applications, such as:

5.3.1 Recommendation

As can be seen from the results, users who are closer together from a network standpoint may also share similar preferences. We are 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 stances 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.

5.3.2 Ascertaining unspoken views

Users may avoid expressing positions explicitly for many reasons, such as fear of social judgment or political repres-

sion, 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.

5.3.3 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 that 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.

6. CONCLUSION

In this paper, we presented a methodology for predicting a person's stance towards an issue, topic, or group in response to an event and given previous activity on social media sites. As a case study, we used the views of US Twitter users towards Muslims in the wake of the Paris terrorist attacks of Nov. 13, 2015. As shown in our work, 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 [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.

7. REFERENCES

[Anand et al. 2011] Pranav Anand, Marilyn Walker, Rob Abbott, Jean E Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. 1–9.

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

[Baldwin et al. 2013] Timothy Baldwin, Paul Cook, Marco Lui, Andrew MacKinlay, and Li 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 2015] Pablo Barberá. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* 23, 1 (2015), 76–91.

[Barberá et al. 2015] Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. Tweeting from Left to Right: Is Online Political Communication more than an Echo Chamber? *Psychological Science* (2015).

[BBC 2015] BBC. 2015. Paris attacks: What happened on the night. *BBC* (Nov. 2015). http://www.bbc.com/news/world-europe-34818994

[Borge-Holthoefer et al. 2015] Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. Content and network dynamics behind Egyptian political polarization on Twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 700–711.

[Burfoot et al. 2011] Clinton Burfoot, Steven Bird, and Timothy 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)*. Portland, USA, 1506–1515.

[Castillo et al. 2015] Mariano Castillo, Margot Haddad, Michael Martinez, and Steve Almasy. 2015. Paris suicide bomber identified; ISIS claims responsibility for 129 dead. *CNN* (Nov. 2015). http://edition.cnn.com/2015/11/14/world/paris-attacks/

[Cheng et al. 2010] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: a content-based approach to geo-locating Twitter users. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010)*. 759–768.

[Chenoweth and Stephan 2011] Erica Chenoweth and Maria J Stephan. 2011. *Why civil resistance works: The strategic logic of nonviolent conflict*. Columbia University Press.

[Cialdini and Trost 1998] Robert B Cialdini and Melanie R Trost. 1998. Social influence: Social norms, conformity and compliance. In *The Handbook of Social Psychology*, Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey (Eds.). McGraw-Hill.

[Cohen and Ruths 2013] Raviv Cohen and Derek Ruths. 2013. Classifying Political Orientation on Twitter: It's Not Easy!. In *Proceedings of ICWSM 2013*.

[Colleoni et al. 2014] Elanor Colleoni, Alessandro Rozza, and Adam Arvidsson. 2014. Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal*

of Communication 64, 2 (2014), 317–332.

[Conover et al. 2011] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *Proceedings of 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 2013] Russell J Dalton. 2013. *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies: Public Opinion and Political Parties in Advanced Industrial Democracies*. Cq Press.

[de la Hamaide 2015] Sybille de la Hamaide. 2015. Timeline of Paris attacks according to public prosecutor. *Reuters* (Nov. 2015). http://www.reuters.com/article/us-france-shooting-timeline-idUSKCN0T31BS20151114

[DellaPosta et al. 2015] Daniel DellaPosta, Yongren Shi, and Michael Macy. 2015. Why Do Liberals Drink Lattes? *Amer. J. Sociology* 120, 5 (2015), 1473–1511. http://www.jstor.org/stable/10.1086/681254

[Dubois and Gaffney 2014] Elizabeth Dubois and Devin Gaffney. 2014. The Multiple Facets of Influence Identifying Political Influentials and Opinion Leaders on Twitter. *American Behavioral Scientist* 58, 10 (2014), 1260–1277.

[Esuli and Sebastiani 2006] Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *LREC-2006*, Vol. 6. 417–422.

[Faulkner 2014] Adam Faulkner. 2014. Automated classification of stance in student essays: An approach using stance target information and the Wikipedia link-based measure. In *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2014)*. 174–179.

[Garimella and Weber 2014] Venkata Rama Kiran Garimella and Ingmar Weber. 2014. Co-following on Twitter. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*. ACM, 249–254. DOI: http://dx.doi.org/10.1145/2631775.2631820

[Golbeck and Hansen 2014] Jennifer Golbeck and Derek Hansen. 2014. A method for computing political preference among Twitter followers. *Social Networks* 36 (2014), 177–184.

[Groseclose and Milyo 2005] Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics* (2005), 1191–1237.

[Han et al. 2012] Bo Han, Paul Cook, and Timothy Baldwin. 2012. Geolocation Prediction in Social Media Data by Finding Location Indicative Words. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*. Mumbai, India, 1045–1062.

[Han et al. 2014] Bo Han, Paul Cook, and Timothy Baldwin. 2014. Text-based Twitter User Geolocation Prediction. *Journal of Artificial Intelligence Research* 49 (2014), 451–500.

[Hecht et al. 2011] Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 237–246.

[Himelboim et al. 2013] Itai Himelboim, Stephen McCreery, and Marc 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 (2013), 40–60.

[Korda and Itani 2013] Holly Korda and Zena Itani. 2013. Harnessing social media for health promotion and behavior change. *Health promotion practice* 14, 1 (2013), 15–23.

[Laranjo et al. 2015] Liliana Laranjo, Amaël Arguel, Ana L Neves, Aideen M Gallagher, Ruth Kaplan, Nathan Mortimer, Guilherme A Mendes, and Annie YS 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 (2015), 243–256.

[Magdy et al. 2015] Walid Magdy, Kareem Darwish, and Norah Abokhodair. 2015. Quantifying Public Response towards Islam on Twitter after Paris Attacks. *arXiv preprint arXiv:1512.04570* (2015).

[Magdy et al. 2016] Walid Magdy, Kareem Darwish, and Ingmar Weber. 2016. #FailedRevolutions: Using Twitter to study the antecedents of ISIS support. *First Monday* 21, 2 (2016).

[Pang and Lee 2008] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 1-2 (2008), 1–135.

[Pavalanathan and Eisenstein 2015] Umashanthi Pavalanathan and Jacob Eisenstein. 2015. Confounds and Consequences in Geotagged Twitter Data. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.

[Rahimi et al. 2015] Afshin Rahimi, Duy Vu, Trevor Cohn, and Timothy Baldwin. 2015. Exploiting Text and Network Context for Geolocation of Social Media Users. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL HLT 2015)*. 1362–1367.

[Rajadesingan and Liu 2014] Ashwin Rajadesingan and Huan Liu. 2014. Identifying Users with Opposing Opinions in Twitter Debates. In *Social Computing, Behavioral-Cultural Modeling and Prediction*. Springer, 153–160.

[Runnymede Trust 1997] London (United Kingdom); Runnymede Trust. 1997. *Islamophobia A challenge for us all*.

[Speriosu et al. 2011] Michael Speriosu, Nikita Sudan, Sid Upadhyay, and Jason Baldridge. 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the 1st Workshop on Unsupervised Learning in NLP*. 53–63.

[Sridhar et al. 2014] Dhanya Sridhar, Lise Getoor, and Marilyn Walker. 2014. Collective stance classification of posts in online debate forums. *Proceedings of ACL 2014* (2014), 109–117.

[Syeed 2015] Nafeesa Syeed. 2015. Paris Terror Attacks: Yes, Parisians are traumatised, but the spirit of resistance still lingers. *Independent.ie* (Nov. 2015). http://goo.gl/toaabz

[Thomas et al. 2006] Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. 327–335.

[Turner 1991] John C Turner. 1991. *Social influence.* Thomson Brooks/Cole Publishing Co.

[Walker et al. 2012] Marilyn A Walker, Jean E Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A Corpus for Research on Deliberation and Debate.. In *Proceedings of LREC 2012*. 812–817.

[Weber and Garimella 2014] Ingmar Weber and Venkata Rama Kiran Garimella. 2014. Using Co-Following for Personalized Out-of-Context Twitter Friend Recommendation.. In *Proceedings of ICWSM 2014*.

[Weber et al. 2013] Ingmar Weber, Venkata R Kiran Garimella, and Alaa Batayneh. 2013. Secular vs. islamist polarization in Egypt on Twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 290–297.

[Wong et al. 2013] Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, and Mung Chiang. 2013. Quantifying Political Leaning from Tweets and Retweets.. In *Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM 2013)*. 640–649.