

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

5-2016

Are you Charlie or Ahmed? Cultural pluralism in Charlie Hebdo response on Twitter

Jisun AN

Haewoon KWAK

Singapore Management University, hkwak@smu.edu.sg

Yelena MEJOVA

Sonia Alonso Saenz DE OGER

Braulio Gomez FORTES

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Databases and Information Systems Commons](#), and the [Social Media Commons](#)

Citation

AN, Jisun; KWAK, Haewoon; MEJOVA, Yelena; DE OGER, Sonia Alonso Saenz; and FORTES, Braulio Gomez. Are you Charlie or Ahmed? Cultural pluralism in Charlie Hebdo response on Twitter. (2016). *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016, Cologne, Germany, May 17-20*. 2-11.

Available at: https://ink.library.smu.edu.sg/sis_research/5337

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Are you Charlie or Ahmed?

Cultural pluralism in Charlie Hebdo response on Twitter

**Jisun An, Haewoon Kwak,
Yelena Mejova***

Qatar Computing Research Institute, Qatar
jan,hkwak,ymejova@qf.org.qa

Sonia Alonso Saenz De Oger

Georgetown University School
of Foreign Service, Qatar
sal197@georgetown.edu

Braulio Gomez Fortes

Deusto University, Bilbao, Spain
braulio.gomez@deusto.es

Abstract

We study the response to the Charlie Hebdo shootings of January 7, 2015 on Twitter across the globe. We ask whether the stances on the issue of freedom of speech can be modeled using established sociological theories, including Huntington's culturalist Clash of Civilizations, and those taking into consideration social context, including Density and Interdependence theories. We find support for Huntington's culturalist explanation, in that the established traditions and norms of one's "civilization" predetermine some of one's opinion. However, at an individual level, we also find social context to play a significant role, with non-Arabs living in Arab countries using #JeSuisAhmed ("*I am Ahmed*") five times more often when they are embedded in a mixed Arab/non-Arab (mention) network. Among Arabs living in the West, we find a great variety of responses, not altogether associated with the size of their expatriate community, suggesting other variables to be at play.

Introduction

On the 7th of January 2015 the Paris offices of the French satirical weekly newspaper Charlie Hebdo were assaulted by two brothers, French citizens born to Algerian parents, who killed 11 persons and injured 11 more. The brothers claimed to belong to Al Qaeda's branch in Yemen. Charlie Hebdo is a controversial magazine, partly due to the paper's highly secularist, and even openly anti-religious, articles making fun of Catholicism, Judaism and Islam. The terrorist attack against Charlie Hebdo was therefore widely interpreted as an attack against freedom of expression and freedom of the press, core principles of liberal democratic societies.

The social media, and Twitter in particular, reacted immediately upon the attack. The hashtags #CharlieHebdo and #JeSuisCharlie ("*I am Charlie*") became an explicit endorsement of freedom of expression and freedom of the press and travelled fast and wide in Twitter. Qualifying or directly opposing this endorsement, other hashtags soon followed: #JeSuisAhmed ("*I am Ahmed*") and #JeNeSuis-PasCharlie ("*I am not Charlie*"). The latter was used not

only by people who were against the editorial line of Charlie Hebdo for being offensive to Islam but also, paradoxically, by representatives of radical right parties as a statement against the Islamization of Europe.

The objective of this paper is to understand the social factors that contribute to online individual behavior. In particular we use Charlie Hebdo as a case study of three prominent sociological theories modeling attention and opinion, ranging in the assumptions about the formation of individual's opinion:

- **Clash of civilizations** – "the great divisions among humankind (...) will be cultural" (Huntington and others 1993). Individuals' opinions and behavior are determined by the culture in which they are socialized. Cultures, on the other hand, organize around different civilizations, such as the Western Christian and the Islamic civilizations.
- **Density theory** – individuals' opinions are influenced by the socio-demographic and/or cultural density of their offline social context. The amount of interaction between Muslims and non-Muslims, both in the West and Middle East, as expat communities become integrated into the cultural fabric of its host nation, may affect the opinion of one about the other.
- **Interdependence theory** – individuals' opinions are influenced by the structure of online interactions within their social network. The personal connections the individuals have, including those online, may change the their worldview.

The aim of this work is both to re-examine the above theories in the context of culturally-charged online discussion, and to better understand the actors within the online phenomena of #JeSuisCharlie.

Modeling Opinion Formation

Scholars of political behavior have long demonstrated that individual political behavior changes as a function of social context (Allardt and Pesonen 1967; Huckfeldt 2009a; Huckfeldt 2009b; Przeworski 1974; Wright 1976). Studies of voting behavior, for example, have shown that vote choice is not the result of an individual decision taken in isolation from the characteristics of the social context in which the individual is embedded. As Przeworski put several decades

ago: “In order to understand political behavior, it is necessary to treat individuals within the context of their social interactions” (Przeworski 1974).

Theories we chose are well-established in social science community, and their use in big data analysis extends both computational and social science fields.

We begin with a macro-scale, deterministic cultural explanation offered by Huntington’s civilizational theory. In the **Clash of Civilizations** seminal paper (Huntington and others 1993), Samuel Huntington argues that “the fault lines between civilizations will be the battle lines of the future”. A *civilization* is defined as a cultural entity, the highest among humans, and the broadest level of cultural identity. Religion is a major civilizational component. Two of the major civilizations discussed by Huntington are the Christian Western civilization and the Islamic civilization. According to Huntington, the Islamic civilization is incompatible with democratic values such as freedom of speech and freedom of the press. An individual’s opinion on the Charlie Hebdo attack will therefore be determined by the civilization she belongs to, irrespective of the offline social context and online structure of interactions:

[H1] *Opinions expressed about the Charlie Hebdo shootings are divided along “civilizational” faultlines, with a higher proportion of pro-free speech tweets by users in Western Christian civilization countries, and a higher one of pro-Muslim tweets by users in the Islamic civilization countries.*

Clash of Civilizations theory has been previously tested in international communications networks in both social media and email by State et al. (State et al. 2015), who conclude its continuing endurance: “a bottom-up analysis confirms the persistence of the eight culturally differentiated civilizations posited by Huntington, with the divisions corresponding to differences in language, religion, economic development, and spatial distance”. Going beyond communication volume, we test the Clash of Civilization theory by examining individuals’ behavior as captured from the usage of different hashtags.

Next, we turn to meso-scale dynamics with **Density theory**, which postulates that individual behavior is “density dependent and hence varies as a function of aggregate population characteristics” (Huckfeldt 2009b). First applied in the context of urbanization in 1938, Wirth uses density to describe the behavioral pressures social heterogeneity puts on individuals (Wirth 1938). These pressures continue to be central to the study of opinion and behavior, including as expressed online (see for example “Bowling alone but tweeting together” (Antoci, Sabatini, and Sodini 2014) or “Online social networks and trust” (Sabatini and Sarracino 2015)).

Accordingly, the reaction to the Charlie Hebdo attack does not depend exclusively on individual beliefs or geographic distance, but also on the offline *social context* in which the individual is embedded. Concretely, the proportion of people from a different culture surrounding an individual may prompt a shift in one’s beliefs and attitudes. The prominence of Muslim diasporas in the Western countries may prompt two possible reactions: (1) the heightened interaction with Muslim population provides a common ground

in Westerners for understanding and empathy, or (2) the majority feels threatened by increasing minority. Thus we pose two hypotheses for *Western* users, reflecting the two alternatives (with mirror theories possible for Muslim users):

[H2a] *The higher the proportion of Muslims in the population, the higher the proportion of pro-Islam tweets.*

[H2b] *The higher the proportion of Muslims in the population, the lower the proportion of pro-Islam tweets.*

Finally, the personal connections we have may contribute the most to our view of the world. By interacting with others on a daily basis we negotiate relationships in order to derive some benefit, and in this process we change ourselves. **Interdependence theory** is a social exchange theory that postulates that people weigh costs to achieve the greatest benefits out of their relationships (Thibaut and Kelley 1959). Rewards may come from both similarities and differences in the dyad, as long as both parties are equally able and willing to provide rewards for others. Thus, we formulate the last hypothesis:

[H3] *Within mixed Arab/non-Arab networks, users are likely to tweet similar content to that of their neighborhood.*

Related Work

Recently, Twitter and other online media have been utilized to re-examine longstanding sociological theories. Providing unprecedented scale, and capturing behaviors heretofore unattainable by standard sociological methods, big social data initiated a new field of computational social science (Lazer et al. 2009). Below we describe works on communication and opinion formation most relevant to this paper, and direct the reader to (Mejova, Weber, and Macy 2015) for a comprehensive view of the field.

Analyses of responses to salient political events on Twitter have ranged from Occupy Wall Street protests (Conover et al. 2013) and same-sex marriage debates (Zhang and Counts 2015) in US, to, more internationally, Mexican drug wars (De Choudhury, Monroy-Hernandez, and Mark 2014), Ferguson unrest (Jackson and Foucault Welles 2015), and the Arab Spring protests (Bruns, Highfield, and Burgess 2013; Lotan et al. 2011; Wolfsfeld, Segev, and Sheaffer 2013). Although Twitter is often associated with social movements, as Wolfsfeld et al. point out, “politics comes first” (Wolfsfeld, Segev, and Sheaffer 2013), and is followed by discussion on social media. Due to the international nature of social media, this discussion, Burns et al. state, is often by the “outsiders looking in” (Bruns, Highfield, and Burgess 2013). It is these “outsiders” – both in the West and Middle East – who are the focus of our present work.

Among the theories we consider, Clash of Civilizations has been revisited by State et al. (State et al. 2015) using Twitter, who found the clusters of countries in the international communication network to resemble the “civilizations” defined by Huntington. Other works on interpersonal interaction, including hashtag usage propagation (Romero, Meeder, and Kleinberg 2011), health behavior (Abbar, Mejova, and Weber 2014), vote turnout (Bond et al. 2012), and news (Kwak et al. 2010), use immediate user

	#JeSuisAhmed	#JeSuisCharlie
muslim	<i>years, year, old, remembering, outside, attackers, cartoonists, while, guy, shot</i>	<i>jew, christian, frankdeleeuw, merry, jewish, russia, christmas, customers, jews, zionists</i>
islam	<i>love muhammad, war, isis, wrong, islamicstate, truth, anti, obama, nd</i>	<i>christianity, judaism, islamism, bible, kkk, religionkills, atheism, reform, teaches, teachings</i>
freedom	<i>democracy, double, comes, support, liberty, religious, offensive, insulting, women, without</i>	<i>free, democracy, includes, principle, cornerstone, principles, trumps, limits, essential, speech</i>
press	<i>insulting, values, law, called, liberty, line, double, democracy, offensive, women</i>	<i>defenders, claiming, speech, slams, censor, while, principle, defence, countries, advocate</i>
terror	<i>protest, since, tomorrow, mosques, pictures, new, tag, war, wake, days</i>	<i>terrorist, fatah, deadly, chechnya, terrorism, attacks, savage, senegal, gatestoneist, warns</i>

Table 1: Word associations produced by word2vec for #JeSuisAhmed and #JeSuisCharlie hashtag collections.

neighborhood to predict behavior, inadvertently challenging Interdependence theory, wherein social relationships are negotiated for some mutual benefit. For example, (Abbar, Mejova, and Weber 2014) model the health value of users Twitter feed by considering the number of network connections they have who post unhealthy content. Fewer studies have been done on an intermediate community-level scale. Community socio-economic well-being has been studied by (Quercia et al. 2012), who apply sentiment analysis to tweets from London, and show a significant correlation between the Index of Multiple Deprivation and the “Gross Community Happiness” score they define. We go a step further, characterizing the mixing of communities, and the effect this mixing has on their opinions, as expressed on Twitter. For both personal as well as larger scales, our work is a contribution to the ongoing effort to re-examine existing sociological theories in the sphere of social media.

Recently, hashtags concerning Charlie Hebdo, and specifically “Je Ne Suis Pas Charlie” (“*I am not Charlie*”), have been examined by Giglietto & Lee (Giglietto and Lee 2015), who found a high proportion of retweets and image sharing, with a unique practice of retweeting nothing but the hashtag itself (in 2% of the cases). This hashtag, the authors conclude, is a “discursive device that facilitated users to form, enhance, and strategically declare their self-identity”. In this work, we attempt to uncover the mechanisms underlying such self-identification.

Reactions to the Charlie Hebdo attack

Reactions to the Charlie Hebdo attack have clustered around the following hashtags which we use in our study. We describe each and outline other prominent hashtags associated with them:

#CharlieHebdo – Sympathy towards the victims of the attack, general condemnation of the attack. It is associated with informational tags mentioning *Paris*, the *cartoonists killed*, and the *suspects*. (In shorthand, #CH.)

#JeSuisCharlie (*I am Charlie*) – Endorsement of freedom of speech and freedom of the press under any circumstances. It is more focused on *freedom*, and the responses to the event in form of the *tributes*, many of them draw-

ings of *pens* as symbols of writers, and especially popular tweets by the urban artist *banksy*. (#JSC)

#JeNeSuisPasCharlie (*I am not Charlie*) – May convey two meanings: Rejection of freedom of speech and freedom of the press when the message is offensive towards Islam. Alternatively, rejection of freedom of speech for Muslims in Christian countries. It is associated with prominent reporters as *Max Blumenthal* and *Benjamin Norton*. (#JNSPC)

#JeSuisAhmed (*I am Ahmed*) – Reactions that differentiate between Islam and terror; emphasis on the fact that among those defending freedom of speech there are also Muslims, such as Ahmed, one of the policemen killed by the terrorists. It is associated with the *murdered* policeman *Ahmed*, who was tweeted to be “*protecting* free speech” or other *french* people. Note that this stance is not necessarily in opposition to #JeSuisCharlie, in fact 76.5% of those tweeting #JeSuisAhmed also mention #JeSuisCharlie (though only 6.17% do the opposite). (#JSA)

Throughout this project, we focus in particular on #JeSuisCharlie and #JeSuisAhmed – hashtags representing two distinct positions. The former is a radical defense of freedom of speech; the latter is a defense of the compatibility between Islam and freedom of speech (though in some cases limited freedom). These positions are not altogether mutually exclusive, but they do emphasize two different, sometimes opposing, aspects of the same phenomenon.

To take a closer look at the stances associated with these hashtags, we use *word2vec* (Mikolov et al. 2013), a computational framework that learns a vector representation of words by taking a text corpus as input. Table 1 lists the words associated with a selection of topics. Both #JeSuisCharlie and #JeSuisAhmed hashtags connect freedom with democracy. Clearly, *freedom* is understood by all as a *democratic* value. However, for #JeSuisAhmed users, freedom is also attached to more negative meanings, such as *offense* against Islam, whereas for #JeSuisCharlie users freedom is treated as an *essential* principle that should not be *trumped* by any other.

Data & Methodology

Western and Islamic “civilizations”

To compare the Western and Islamic cultures, we focus on the 39 countries, 20 countries including Western Europe and the USA, which represent the Western civilizational culture, and 19 countries from the Middle-East, which represent (not exhaustively) the Islamic civilizational culture. These countries are listed in Table 2, along with each country’s proportion of Muslim population in parentheses (CIA 2010). The two groups have a wide difference in the proportion of Muslims, with most Middle Eastern countries having > 70% and Western < 8%, with notable exceptions such as Cyprus (at 25.3%), which has a distinct population composition.

Region	Country (Muslim population (%))
Middle East (19)	Morocco (99.9), Iran (99.5), Tunisia (99.5), Yemen (99.1), Iraq (99), Turkey (98), Algeria (97.9), Palestine (97.6), Jordan (97.2), Libya (96.6), Egypt (94.4), Saudi Arabia (93), Syria (92.8), Oman (85.9), United Arab Emirates (76.9), Kuwait (74.1), Bahrain (70.3), Qatar (67.7), Lebanon (61.3)
Western (20)	Cyprus (25.3), France (7.5), Netherlands (6), Belgium (5.9), Germany (5.8), Switzerland (5.5), Austria (5.4), Greece (5.3), Sweden (4.6), United Kingdom (4.4), Denmark (4.1), Italy (3.7), Norway (3.7), Luxembourg (2.3), Spain (2.1), Ireland (1.1), Finland (0.8), Portugal (0.6), Iceland (0.2), USA (0.9)

Table 2: Selected Middle Eastern and Western countries (with % Muslim population).

Twitter data

Before focusing on the individuals within countries, however, we collect tweets concerning the Charlie Hebdo incident using two sources: (1) Nick Ruest’s collection of tweets which track #JeSuisCharlie, #JeSuisAhmed, and #CharlieHebdo, and (2) a Topsy.com collection tracking #JeNeSuisPasCharlie and #JeSuisPasCharlie.

Nick Ruest collection. We use a collection created by Nick Ruest¹, who has collected tweets that include one of the following three hashtags – #JeSuisCharlie, #JeSuisAhmed, and #CharlieHebdo – from 2015-01-07 11:59:12 UTC to 2015-01-28 18:15:35 UTC using Twitter’s search API.

We “hydrated” (i.e. collected metadata for) the released tweet IDs² using Twitter public API, collecting 11,367,987 tweets (7.1M tweets with #CharlieHebdo, 6.5M with #JeSuisCharlie, and 264,097 with #JeSuisAhmed) posted by 3,081,039 unique users (2M users for #CharlieHebdo, 2M

users for #JeSuisCharlie, and 169,598 users for #JeSuisAhmed). Although the volume of tweets is less than the original, since some of the tweets no longer exist, aggregate statistics are similar to what Nick Ruest has reported – for example, in this dataset, 76.74% of tweets are retweets and 1.77% of them are replies, with the most retweeted tweets having images.

The activity level of the users varies widely across our dataset – up to the maximum of 35,418 tweets by one user, with the median of 1 and the mean of 3.69. To remove abnormally active users, who are likely to be spammers, we discard users who tweeted more than 148 times, which is the 99th percentile of the distribution. This filters out 0.1% users (2,787) with 9% of total tweets (881,100) from the dataset. Then to further focus on those users who show their stance regarding the CharlieHebdo incident strongly and somewhat unambiguously, we only consider users with two or more tweets in this dataset, resulting in a dataset of 1,389,673 users with 8,796,872 tweets.

To map the tweets to their respective countries of origin, we geo-locate the data in two ways. First, we look at whether the tweet is geo-tagged and, if it is, we use it as user’s location. In a case where a user’s tweets are in different countries, we discard these users to avoid ambiguity. If geo-tagging is not available, then we apply Yahoo! PlaceMaker³ to the location field in their bio on Twitter. Yahoo! PlaceMaker is a web service which, given a text, returns best matched location. For example, with the sentence “I live in New York”, it returns “New York, New York, USA”. The service is especially suitable for our data, as it supports languages beyond English.

Among 1,389,673 users, we successfully located 688,651 (45,717 users from geo-tagged tweets and 642,934 Yahoo! PlaceMaker)⁴. These users are mostly from North America and Europe – the top five countries are US, France, UK, Spain and Canada. Note that we discard users with two or more locations (e.g., India/Paris, Dubai/Paris) – 221 users when using geo-tagged tweets and 17,352 (2.6%) users when using Yahoo! PlaceMaker.

Finally, among those located users, 464,176 are in the 39 countries of our interest. In the forthcoming analysis, we focus on these 464,176 users, who are engaged with the Charlie Hebdo shootings, have expressed an opinion on it, and could be located geographically, along with their 3,030,558 tweets (1.37M of #CH, 1.62M of #JSC, and 42,029 of #JSA). These users are mostly located in five countries, in order of magnitude: France, United States, United Kingdom, Spain, and Italy.

We expand this collection by crawling the most recent tweets (maximum of 3,200) of each of these users. By detecting mentions in these tweets (handles of other Twitter users), we then build an ego mention network for the users in our dataset. We collect 932,003,251 tweets in total and we extract 23,406,770 mentioned users in those tweets.

¹<http://goo.gl/fi0QPU>

²<http://dataverse.scholarsportal.info/dvn/dv/rnuest/faces/study/StudyPage.xhtml?globalId=hdl:10864/10830>

³<https://developer.yahoo.com/boss/geo/>

⁴We have 13,823 users who got located by both Geo-tagged tweets and Yahoo PlaceMaker. For the 92.3% users (12,756), two methods are resulted in the same location.

We attempt to locate these mentioned users using their geo-tagged tweets and self-described location (as above) and successfully find 4,326,045 users' location (18.4%). Among 274,152,345 links between our seeding users to mentioned users, 0.6% (1,779,086) of them are reciprocal links between users who tweeted CH.

Topsy collection. We collect the tweets containing one of the two hashtags not available in the Nick Ruest collection – #JeNeSuisPasCharlie (JNSPC) and #JeSuisPasCharlie (JSPC), both versions of “*I am not Charlie*” – using Topsy⁵ from 7th to 28th January 2015. Topsy is a certified partner of Twitter for offering social search and social analytics, such as Twitter Oscar Index⁶ and Twitter political Index⁷. Topsy indexes every public tweet and allows users to search them by certain keywords since 2013⁸. This means that our analysis is based on the entire set of public tweets instead of small-sized samples. While Topsy offers the public interface to access to tweets even after it was acquired by Apple in 2013, Apple finally shutdowns the service as of December 2015.

We initially gather 35,966 tweets (tweet id, screen name of users, and text) from Topsy. Then using Twitter API, we collect 32,315 tweets (30,638 (JNSPC) and 5,379 (JSPC)) with 21,276 users. We then filter out users who have high activity level (512) and users who have only one tweet (16,919). Among the 4,356 remaining users, 395 users are live in one of 39 countries of our interest. We focus on these 395 users and their 1,404 tweets for the analysis. We then crawl 945,762 recent tweets posted by these 395 users. Out of 159,028 users mentioned in those tweets, 12.29% of users (19,529) are located.

These tweets are coming from locations that are somewhat different from our previous dataset. The top 5 countries where these users are located are France, Algeria, United States, Morocco, and Belgium.

The normalized temporal volume (showing percentage of total hashtag volume) of the final collection (after user geolocation and selection) can be found in Figure 1, and a raw volume can be found as an inset plot. The vast majority of activity happens within 3 days of the event, with #CharlieHebdo dominating the volume. The use of #JeSuisAhmed peaks on the day after the attack.

Arabic identification. As our hypotheses deal with users' religious identities, we need to differentiate the Muslims from the non-Muslims among the users in our dataset. Since Twitter users usually do not declare their religious identities in their profiles, we proceed with the – admittedly rough – assumption that Arabic speakers, or users with Arabic names, are Muslim. All other names or languages are non-Muslim. Considering that approximately 94% of Arabs are Muslims (CIA 2010), the assumption can be reasonably accepted. Also, it is worth to mention that Iran and Turkey are Muslim countries (99.5% and 98% of populations are

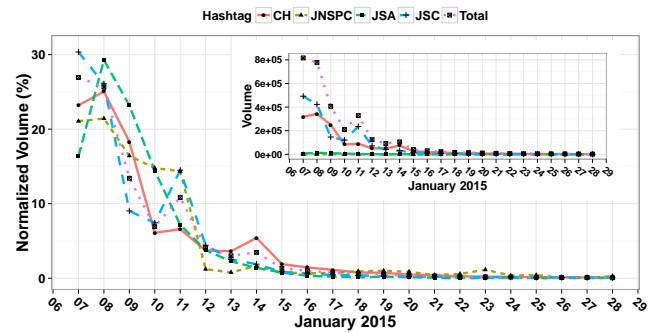


Figure 1: Daily tweet volume mentioning each of four hashtags (#CH, #JSC, #JSA, and #JNSPC). Insert: original data without user selection.

Muslims, respectively, as in Table 2), but they are non-Arab countries. We thus exclude users from Turkey and Iran to avoid bias in the experiments using Arab and non-Arab distinction.

We firstly detect any user who tweets in Arabic, has a name in Arabic, or set their language on Twitter as Arabic. To detect the language of each tweet, we use three widely-used libraries for language detection, which are CLD2 (embedded in Google Chrome)⁹, langid.py (Lui and Baldwin 2012), and LangDetect¹⁰, and mark language by simple majority voting. It is known that this ensemble approach consistently outperforms any individual system, including Twitter's language metadata (VRL 2014). If the name is not in Arabic, then we check it against a dictionary of 4,401 Arabic names in English (2,160 male names, 2,151 female names, and 100 neutral names), which we build using baby name lexicons¹¹. The list of names used in the analysis is available at ¹².

In our seeding dataset, we find that 5.3% of users (23,924) pass the above filters. Among them, 69.8% of users (16,705) are detected by the name-based approach, while 27.0% of users (6,469) are detected by their language use. Only 750 users are detected by both methods.

For the rest of the paper, we will use Arab/non-Arab distinction for the users identified via the above method, not to confuse it with other sources of religious identity (such as that identified by CIA Fact Book and listed in Table 2).

Language. The languages used in our collections are shown in Table 3. For Non-Arab users, French is the most used language at 47.93% of all tweets with English at 35.57%. For Arab users, English is the most used language at 49.57% of all tweets, with French at 25.85%, and Arabic at only 9.89%. The latter statistic is understandable, since

⁹<http://blog.mikemccandless.com/2011/10/accuracy-and-performance-of-googles.html>

¹⁰<https://github.com/shuyo/language-detection/blob/wiki/ProjectHome.md>

¹¹http://www.searchtruth.com/baby_nameshttp://www.urduseek.com/names

¹²<https://goo.gl/Nam1ts>

⁵<http://topsy.com/>

⁶<http://oscars.topsy.com/>

⁷<https://election.twitter.com/>

⁸<http://about.topsy.com/2013/09/04/>

every-tweet-ever-published-now-at-your-fingertips

Hashtag	Language	Arab	Non-Arab
CharlieHebdo	Arabic	11,846 (13.74%)	0 (0.00%)
	English	44,316 (51.42%)	513,918 (40.45%)
	French	17,479 (20.28%)	493,908 (38.87%)
	Others	12,551 (14.56%)	262,750 (20.68%)
JeSuisCharlie	Arabic	2,008 (3.94%)	0 (0.00%)
	English	22,531 (44.16%)	461,233 (30.98%)
	French	18,428 (36.11%)	832,313 (55.90%)
	Others	8,060 (15.80%)	195,314 (13.12%)
JeSuisAhmed	Arabic	207 (4.11%)	0 (0.00%)
	English	3,702 (73.48%)	18,929 (53.66%)
	French	824 (16.36%)	13,299 (37.70%)
	Others	305 (6.05%)	3,050 (8.65%)
JNSPC	Arabic	41 (15.77%)	0 (0.00%)
	English	98 (37.69%)	272 (29.73%)
	French	111 (42.69%)	547 (59.78%)
	Others	10 (3.85%)	96 (10.49%)
Total	Arabic	14,102 (9.89%)	0 (0.00%)
	English	70,647 (49.57%)	994,352 (35.57%)
	French	36,842 (25.85%)	1,340,067 (47.93%)
	Others	20,926 (14.68%)	461,210 (16.50%)

Table 3: The fraction of tweets in different languages by Arabs and non-Arabs. (Classified using pooled language detection.)

the queries were made using French hashtags, and in latin alphabet, which surely excluded those tweets written purely in Arabic (for more on this limitation, see the Discussion section).

The large amount of users classified as Arabs that use languages other than Arabic is also probably due to the fact that the largest number of tweets are concentrated in countries like France, United Kingdom, and the USA. This means that a lot of users with an Arab background living in these countries are tweeting in English and French, not Arabic.

Results

In this section we present our findings regarding the three posed theories modeling the formation of opinions expressed wherein.

Clash of civilizations theory

Under Huntington’s thesis, the major fault lines in post-Cold War geo-politics lie along cultural and religious identities. In this study, the users we consider can be roughly divided as belonging to two “civilizations” – the Western Christian civilization and the Islamic civilization. Huntington poses that Muslims, by virtue of belonging to the Islamic civilization, will be more wary of defending freedom of speech than Westerners. Here we test this hypothesis.

[H1] *Opinions expressed about the Charlie Hebdo shootings are divided along “civilizational” faultlines, with a higher proportion of pro-free speech tweets by users in Western Christian civilization countries, and a higher one of pro-Muslim tweets by users in the Islamic civilization countries.*

Table 4 shows the proportion in the use of hashtags by users identified as Arab and all others (Non-Arab). Both

Hashtag	Non-arab	Arab	Western	Middle east
JSC	97.67	88.09	97.65	90.51
JSA	2.27	10.88	2.29	8.94
JNSPC	0.06	1.03	0.07	0.07

Table 4: Percentage of tweets mentioning each hashtag.

groups use the largely topical #CharlieHebdo hashtag, and very little #JeNeSuisPasCharlie. However, the relative proportion of #JeSuisCharlie to #JeSuisAhmed is strikingly different, with one #JeSuisAhmed to every 10 #JeSuisCharlie for the Arab users, and one to 43 for non-Arab ones. Similar distinction is evident when we segment users by geographical locations (Western vs. Middle East).

Thus, we find some support for H1, although both populations use #JeSuisCharlie more than #JeSuisAhmed, and this cannot be explained by the Clash of Civilizations theory. The wordclouds in Figure 2 show how Non-Arab and Arab users use #JeSuisCharlie, with Arabs mentioning *Ahmed*, *God*, and *solidarity* while both focusing on *freedom*.

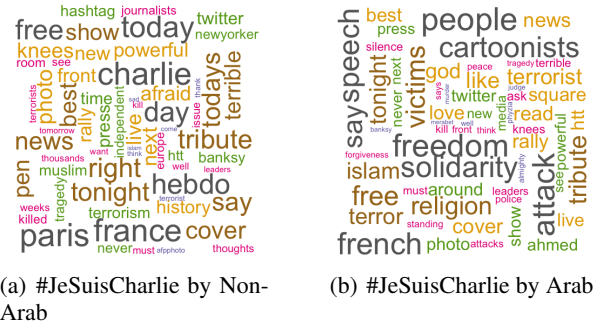


Figure 2: Wordclouds for #JeSuisCharlie collection by Non-Arab vs. Arab users.

Density theory

Density theory claims that the population densities of culturally diverse groups in the individual’s offline social context are important factors in the formation of opinion. In this case, population densities are characterized by the size of groups sharing the same “civilizational” culture within one country. Is it possible that a diaspora of Arabs in the West, and Westerners in the Arab world, affects the understanding of and attitudes toward Charlie Hebdo event? Such effects may be simultaneous and contradictory: on one side, they could be promoting empathy and understanding by co-habitation, on another, they could encourage hostility to an increasingly visible minority (from the point of view of Muslims in the Middle East or Westerners in the West) or towards an unfriendly majority (from the point of view of Muslims in Western countries or Westerners in the Middle East). Two alternatives arise in the face of minority/majority interactions (here, for Western users):

[H2a] *The higher the proportion of Muslims in the population, the higher the proportion of pro-Islam tweets.*

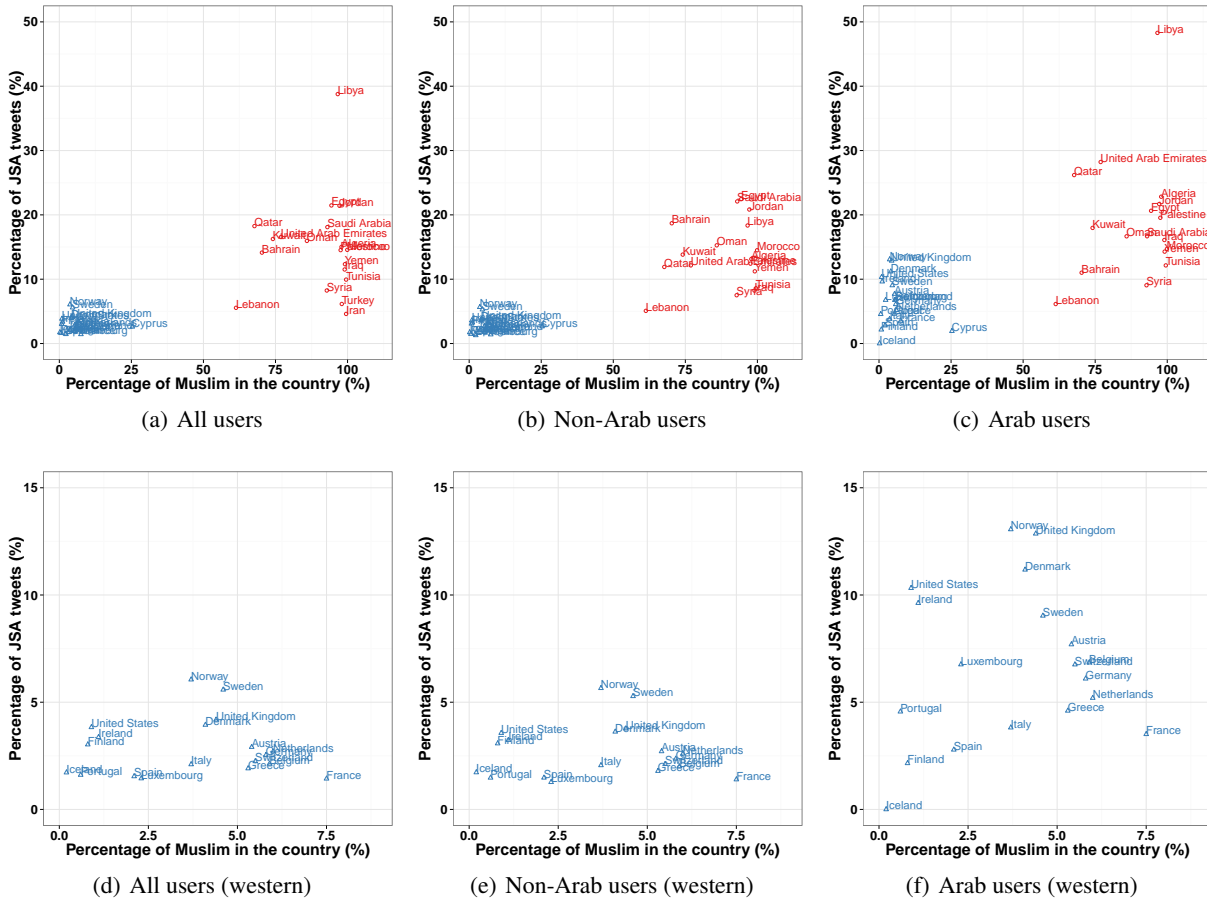


Figure 3: The percentage of JSA tweets over JSA+JSC tweets by Muslim population of the country, comparing 3 different groups: all users, non-Arab, and Arab. Arab countries are colored in red and non-Arab in blue. Due to Arab filter design, Turkey and Iran are removed from figures b, c, e, and f.

[H2b] *The higher the proportion of Muslims in the population, the lower the proportion of pro-Islam tweets.*

Mirror hypotheses can be posed for Arab users. However, our conclusions are more sound for Western population due to the languages of our dataset, so we focus on this group of users.

Here, we take advantage of The World Factbook’s proportion of Muslim residents, as described in Data Section. Figure 3(a) plots the percent of #JeSuisAhmed (#JSA) tweets over the combined total of #JeSuisAhmed and #JeSuisCharlie (y-axis) against the proportion of Muslims in the country (x-axis). Figure 3(d) shows a zoom of the bottom left corner of Figure 3(a), where Western countries are clustered (except Cyprus, which has 25.3% Muslim population).

To compare the behavior of Arab and non-Arab users (as defined in Data Section), we present the two user populations in Figures 3(b,e) for non-Arab users and Figures 3(c,f) for Arab ones. In these graphs, we *exclude* Turkey and Iran to eliminate bias, as users from these countries are not Arabs but are Muslims nevertheless.

Table 5 shows Pearson product-moment correlation r and

Spearman rank correlation coefficient ρ between the percentage of #JSA tweets and the percentage of Muslims in the country’s population in various slices of data. As Figure 3(b) shows, there is a clear positive correlation (Pearson $r=0.845$, $p < 0.001$), suggesting that Westerners who live in Middle Eastern countries tend to tweet more with #JSA than those who live in the West. There is, therefore, a clustered division along the two “civilizations” described by Huntington. However, the story is more complicated when we go deeper and pay attention to the social context.

According to the Clash of Civilizations theory, non-Arabs (i.e. Westerners) living in the Middle-East should behave in a similar way to non-Arabs living in the West; after all, they are all non-Muslims and they belong to the Western “civilizational” culture. Figure 3(b), according to Huntington, should show all countries clustered on the left bottom corner. The graph shows, on the contrary, that non-Arabs living in the Middle East, where they are surrounded by large majorities of Muslims, are much more likely to use #JSA than non-Arabs living in the West. A similar observation can be made for Arabs in the West (which all should cluster at the

	All countries		Western		Arab	
	Person (r)	Spear. (ρ)	r	ρ	r	ρ
All users	0.745***	0.698***	-0.004	0.136	0.064	-0.300
Non-Arab	0.845***	0.740***	0.021	0.130	0.193	-0.010
Arab	0.675***	0.675***	-0.186	0.097	0.157	-0.022

Significance: $p < 0.0001$ ***, $p < 0.001$ **, $p < 0.01$ *

Table 5: Pearson and Spearman correlations of % of JSA tweets to the % of Muslims in the country.

top right of Figure 3(c), but do not).

If we now turn to users living in the West, we also see that the density of the social context matters. For both non-Arabs and Arabs the correlation is extremely weak (see “Western” column of Table 5). However, Figure 3(d) seems to suggest that the relationship between the number of #JSA hashtags and the percentage of Muslims in the country might not be linear, but concave downwards. At between 0 and 3.5% of Muslims in the country, non-Arabs are more likely to use #JSA the larger the number of Muslims that live in the country; after a tipping point of 3.5% of Muslims in the country, however, non-Arabs are less likely to hashtag JSA the larger the number of Muslims surrounding them. Therefore, the Muslim minority helps non-Muslims to be more emphatic as far as this minority is not too large. The tipping point at which non-Arabs become less emphatic and more fearful of the Arab point of view is approximately at 3.5% of Muslim population. Italy would seem to be the only clear outlier of this concave relationship.

To verify the robustness of these figures, we model this behavior using a measure of religiosity (indication of how important religion is to a country’s residents). Indeed, religiosity, as measured by Gallup in 2009¹³, is highly correlated with the proportion of #JSA tweets at $r = 0.7085$. However, when a linear regression is fitted using both religiosity and rate of Muslim population, the effect of religiosity is lost.

Interdependence theory

Whereas density theory concerns the aggregate level of countries, we now turn to the individual level of analysis, in which individuals build interpersonal relationships which affect both parties. Interdependence theory concerns the effect of online interactions on the individual’s online behavior:

[H3] *Within mixed Arab/non-Arab networks, users are likely to tweet similar content to that of their neighborhood.*

As mentioned in the Data & Methodology Section, we build a mention network for each user in our dataset. This network contains all users whose Twitter handles have been mentioned in the tweets of our users. Those users were then also labeled as Arab or not. These mentions signify a user’s connection to, or at least awareness of, other Twitter users, and in the case they are Arab, increased awareness of the Arab point of view.

¹³<http://www.gallup.com/poll/142727/religiosity-highest-world-poorest-nations.aspx>

We divide users into two groups: (1) users who have not mentioned any Arabs in their tweets at all (28,939, denoted as “No Mentions”) and (2) users who have mentioned an Arab user at least once (338,430, denoted as “Some mentions”). We then compare the use of #JeSuisAhmed between the groups, and find that the mixed group uses #JSA more than twice as much as the homogeneously non-Arab group, with 3.61% compared to 1.31% likelihood, respectively. A Welch’s t test confirms that the difference in two groups is statistically significant ($t_{44,164} = 38.80$, $p < 0.001$).

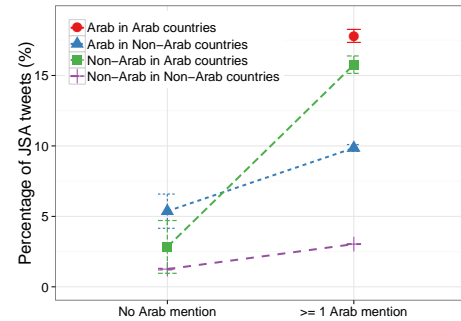


Figure 4: Mean percentage of JSA tweets for four user groups in conditions with and without Arab mentions.

To understand better whether the mention network effect is confounded by any offline effect, such as country of residence, we now look at four different user groups: a) Arabs living in Arab countries, b) Arabs living in Non-Arab countries, c) Non-Arabs living in Arab countries, and d) Non-Arabs living in Non-Arab countries and examine to what extent the online factor plays a role. In Figure 4 we show the behavior of each group. Mention network factor plays a role for all user groups except Arabs in Arab countries, which due to sparsity we do not consider (there are only 24 users, all of whom mention some Arab users).

Since the majority of users is in the “non-Arabs in non-Arab countries” group, the result is similar to what we observe when we consider all users (see earlier paragraph). The means of No Mention and Some Mentions are 1.26 and 3.03, respectively ($t_{42,018} = 30.19$, $p < 0.001$). The next strongest relationship is for “Arabs in non-Arab countries” user group, with the likelihood of tweeting #JSA almost doubling from 5.37 to 9.86 ($t_{320} = 3.62$, $p < 0.001$). The last group, “non-Arabs in Arab countries” also shows a strong pattern, with the difference between 2.83 and 15.76 having $p < 0.001$ ($t_{72} = 6.56$).

Overall, we observe that the personal mentions in users’ interactions do affect the likelihood of expressing an opinion favorable to #JeSuisAhmed. Note, however, that the “of-line” distinction – that is, where the user lives – is a stronger predictor of online behavior.

Among the 27.27M users, including the mentioned users, our filters detect 4.8% (1,312,008) Arab users. We find that the links in mention network are mostly to non-Arab Twitter users – 90.07% links from non-Arabs and 5.54% links from Arabs. Only 3.73% links mention Arab users (2.67% from

	Person (p)	Spearman (r)
All users	0.215***	0.153***
Non-Arab users	0.208***	0.130***
Non-Arab in Non-Arab countries	0.171***	0.118***
Non-Arab in Arab countries	0.134***	0.121***
Arab users	0.156***	0.191***
Arab in Arab countries	0.134***	0.122***
Arab in Non-Arab countries	0.171***	0.118***
Significance: $p < 0.0001$ ***, $p < 0.001$ **, $p < 0.01$ *		

Table 6: Pearson and Spearman correlations of % of JSA tweets to % of Arab mentions in the mention network by different user groups.

non-Arabs and 1.06% from Arabs). Thus the discussion in our dataset is focused on the Western world.

Table 6 shows how the percentage of Arab mentions in one’s mention network is associated with the percentage of JSA tweets. We find a positive relationship across all different user groups, weak but statistically significant.

Discussion

The results of this study must be seen in the light of two technical limitations, both of which would serve as important future directions of research.

The data we have considered here has been collected using French hashtags, and in Latin alphabet. Although many other languages, including English, were captured, this method has surely missed relevant Arabic content. Capturing the multilingual response to international news is an important technical challenge for the worldwide opinion tracking community. Another challenge is the identification of religious affiliation purely from online data. Automatic classification, such as the one proposed by Nguyen & Lim (Nguyen and Lim 2014), may provide access to users whose religion does not statistically follow from their name or language, as we have assumed in this research.

Above limitations aside, the insights in this study have several implications for human-centric application design. While it has been studied extensively in the political context, our study is the first which empirically shows that exposure to other views affects user behavior in the cultural context. Diversity is one of the key elements for a healthy society, yet there is much polarization in both online and offline worlds – with echo chambers limiting the views of both sides (Gilbert, Bergstrom, and Karahalios 2009). Our findings support the design of more pluralistic discourse efforts.

As noted by (Giglietto and Lee 2015), “Je Suis...” hashtags aid users in self-identification as a part of a group. This kind of behavior has been reported in various contexts (Chen, Sun, and Hsieh 2008). For instance, in online games, guild (small group) members explicitly show their guild names in their handle names (Nardi and Harris 2006). In a virtual world, expressing oneself and having a group membership is vital to sustain online communities and offer better user experience. The affordances for self-

identification, thus, are important for successful online communities.

As we mentioned earlier, scholars have long worked on understanding social responses to political events, especially on social media (Bruns, Highfield, and Burgess 2013; Conover et al. 2013; Zhang and Counts 2015). Our work is aligned with the study by Burns et al. in that it takes into account the global characteristics of social media around the Arab Spring (Bruns, Highfield, and Burgess 2013). As their study shows information flow between Arabic and non-Arabic user groups by looking into reply and retweets, our work illustrates, instead, how such data could be used for international-scale verification of existing hypotheses developed in social and political science.

Analyses described in this work, as in most social behavior studies, must be interpreted within **correlation is not causation** warning. The captured phenomena is likely in part due to homophily, wherein more tolerant people would connect to a more diverse sphere of friends. The next step, then, is to simulate, or indeed perform, experimental evaluation in order to verify the causal links between interaction with diverse communities and opinion change. Social media giants such as Facebook and Twitter are in a unique opportunity to monitor the readership behaviors of their users, however a strict adherence to privacy and non-manipulation considerations must be implemented (for such studies as (Bakshy, Messing, and Adamic 2015), for example).

Finally, the role of mass media may play a central stage in the opinion formation and propagation in social media – an important dimension for future study. Moreover, as (Lin and Margolin 2014) show, in social media attention tends to converge on few hashtags which signal the topic. Thus, more fine-grained topic analysis may find stances in line with #JeSuisAhmed in the #JeSuisCharlie stream.

Conclusion

Our work presents a systematic application of sociological opinion formation theories to the analysis of the Twitter response to the Charlie Hebdo shootings of January 2015. The theory of the Clash of Civilizations first seemed to be confirmed at face value by the data, but when we look deeper, paying attention to the social context (i.e. the country and its socio-demographic composition) and the structure of online interactions between users (culturally mixed or culturally homogeneous), we see that Clash of Civilizations needs to be rejected, or at least qualified, in favor of Density theory and Interdependence theories. Culture – and religion as a fundamental part of it – matters a great deal, as Huntington argues, but it matters in much more subtle ways than those advanced by the Clash of Civilizations theory.

Social media data makes it possible to model an individual’s interaction with both mainstream and minority cultures, allowing us to model individual behavior change. As geo-political developments unfold, and greater number of cultures will come in contact, this data will increasingly present opportunities for verifying old and forming new theories on opinion formation in pluralistic societies.

References

- [Abbar, Mejova, and Weber 2014] Abbar, S.; Mejova, Y.; and Weber, I. 2014. You tweet what you eat: Studying food consumption through twitter. *SIGCHI*.
- [Allardt and Pesonen 1967] Allardt, E., and Pesonen, P. 1967. Cleavages in finnish politics. In Lipset, S., and Rokkan, S., eds., *Party Systems and Voter Alignments*. New York: Free Press.
- [Antoci, Sabatini, and Sodini 2014] Antoci, A.; Sabatini, F.; and Sodini, M. 2014. Bowling alone but tweeting together: the evolution of human interaction in the social networking era. *Quality & Quantity* 48(4):1911–1927.
- [Bakshy, Messing, and Adamic 2015] Bakshy, E.; Messing, S.; and Adamic, L. 2015. Exposure to diverse information on facebook. *Facebook Research Blog*.
- [Bond et al. 2012] Bond, R. M.; Fariss, C. J.; Jones, J. J.; Kramer, A. D.; Marlow, C.; Settle, J. E.; and Fowler, J. H. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298.
- [Bruns, Highfield, and Burgess 2013] Bruns, A.; Highfield, T.; and Burgess, J. 2013. The arab spring and social media audiences english and arabic twitter users and their networks. *American Behavioral Scientist* 57(7):871–898.
- [Chen, Sun, and Hsieh 2008] Chen, C.-H.; Sun, C.-T.; and Hsieh, J. 2008. Player guild dynamics and evolution in massively multiplayer online games. *CyberPsychology & Behavior* 11(3):293–301.
- [CIA 2010] CIA, E. 2010. The world factbook 2010. *Central Intelligence Agency, Washington, DC*.
- [Conover et al. 2013] Conover, M. D.; Ferrara, E.; Menczer, F.; and Flammini, A. 2013. The digital evolution of occupy wall street.
- [De Choudhury, Monroy-Hernandez, and Mark 2014] De Choudhury, M.; Monroy-Hernandez, A.; and Mark, G. 2014. Narco emotions: affect and desensitization in social media during the mexican drug war. In *SIGCHI Conference on Human Factors in Computing Systems*. ACM.
- [Giglietto and Lee 2015] Giglietto, F., and Lee, Y. 2015. To be or not to be charlie: Twitter hashtags as a discourse and counter-discourse in the aftermath of the 2015 charlie hebdo shooting in france. *Workshop on Making Sense of Microposts at the 24th International World Wide Web Conference*.
- [Gilbert, Bergstrom, and Karahalios 2009] Gilbert, E.; Bergstrom, T.; and Karahalios, K. 2009. Blogs are echo chambers: Blogs are echo chambers. In *System Sciences*, 1–10. IEEE.
- [Huckfeldt 2009a] Huckfeldt, R. 2009a. Citizenship in democratic politics: Density dependence and the micro-macro divide. In *Comparative Politics: Rationality, Culture, and Structure*. New York: Cambridge University Press. 291–313.
- [Huckfeldt 2009b] Huckfeldt, R. 2009b. Interdependence, density dependence, and networks in politics. In *American Politics Research*, volume 37. 921–950.
- [Huntington and others 1993] Huntington, S. P., et al. 1993. The clash of civilizations?
- [Jackson and Foucault Welles 2015] Jackson, S. J., and Foucault Welles, B. 2015. #ferguson is everywhere: initiators in emerging counterpublic networks. *Information, Communication & Society* 1–22.
- [Kwak et al. 2010] Kwak, H.; Lee, C.; Park, H.; and Moon, S. 2010. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, 591–600. ACM.
- [Lazer et al. 2009] Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabási, A.-L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; Jebara, T.; King, G.; Macy, M.; Roy, D.; and Van Alstyne, M. 2009. Computational social science. *Science* 323(5915):721–723.
- [Lin and Margolin 2014] Lin, Y.-R., and Margolin, D. 2014. The ripple of fear, sympathy and solidarity during the boston bombings. *EPJ Data Science* 3(1):1–28.
- [Lotan et al. 2011] Lotan, G.; Graeff, E.; Ananny, M.; Gaffney, D.; Pearce, I.; et al. 2011. The arab spring, the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International journal of communication* 5:31.
- [Lui and Baldwin 2012] Lui, M., and Baldwin, T. 2012. langid.py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*, 25–30. Association for Computational Linguistics.
- [Mejova, Weber, and Macy 2015] Mejova, Y.; Weber, I.; and Macy, M. W. 2015. *Twitter: A Digital Socioscope*. Cambridge University Press.
- [Mikolov et al. 2013] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.
- [Nardi and Harris 2006] Nardi, B., and Harris, J. 2006. Strangers and friends: Collaborative play in world of warcraft. In *CSCW*, 149–158. ACM.
- [Nguyen and Lim 2014] Nguyen, M.-T., and Lim, E.-P. 2014. On predicting religion labels in microblogging networks. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 1211–1214. ACM.
- [Przeworski 1974] Przeworski, A. 1974. Contextual models of political behaviour. In *Political Methodology*, volume 1. 27–61.
- [Quercia et al. 2012] Quercia, D.; Ellis, J.; Capra, L.; and Crowcroft, J. 2012. Tracking gross community happiness from tweets. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, 965–968. ACM.
- [Romero, Meeder, and Kleinberg 2011] Romero, D. M.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *WWW*, 695–704. ACM.
- [Sabatini and Sarracino 2015] Sabatini, F., and Sarracino, F. 2015. Online social networks and trust. *Munich Personal RePEc Archive*.
- [State et al. 2015] State, B.; Park, P.; Weber, I.; and Macy, M. 2015. The mesh of civilizations in the global network of digital communication. *PLoS ONE*.
- [Thibaut and Kelley 1959] Thibaut, J. W., and Kelley, H. H. 1959. The social psychology of groups.
- [VRL 2014] VRL, N. 2014. Accurate language identification of twitter messages. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)@ EACL*, 17–25.
- [Wirth 1938] Wirth, L. 1938. Urbanism as a way of life. *American journal of sociology* 1–24.
- [Wolfsfeld, Segev, and Sheaffer 2013] Wolfsfeld, G.; Segev, E.; and Sheaffer, T. 2013. Social media and the arab spring politics comes first. *The International Journal of Press/Politics* 18(2):115–137.
- [Wright 1976] Wright, G. C. 1976. Community structure and voter decision making in the south. In *Public Opinion Quarterly*, number 40. 201–215.
- [Zhang and Counts 2015] Zhang, A. X., and Counts, S. 2015. Modeling ideology and predicting policy change with social media: Case of same-sex marriage. In *SIGCHI*.