Studying outbreak, successes and failures of the "Arab Spring" with news media and Twitter

Kenneth Joseph, Matthew Benigni, Wei Wei, Kathleen M. Carley, *Fellow, IEEE,* and Mia Bloom

***Abstract*—asdfsadf**

I. Introduction

On December 17th, 2010, Mohamed Bouazizi immolated himself in Sidi Bou Said, Tunisia in response to harassment from a local policewoman and local municipality officers. Though Bouazizi was not the first to engage in this form of protest, for some reason his case resonated with other Tunisians, who took to the streets in protest of constant persecution and victimization by a corrupt government. Although early protests were relatively small and were met with violence by government forces, social media sites like Twitter, Facebook and YouTube recorded the events and displayed them to the broader public. These events are widely considered to be the beginning of what has come to be known, for better or worse (Gelvin, 2015), as the Arab Spring.

There is little doubt that social media, and new media Baym (2010) more generally, played an important role in Arab Spring. However, conventional wisdom that emphasized social media as the cause of *the* revolutions has been proven to be overblown (Bruns et al., 2013; Comunello and Anzera, 2012; Goldstone, 2013). Recent research has focused instead on how social media may have aided certain aspects of the revolutions in important ways for different people (Gallé et al., 2013; Lotan et al., 2011; Starbird and Palen, 2012; Tufekci and Wilson, 2012), and whether social processes that were carried out via new media are reflective of those that occurred "offline" (Comunello and Anzera, 2012). Additionally, data from newspaper articles written during the time of the Arab Spring also may be of use in better understanding these processes better (REMOVED FOR BLIND REVIEW). Thus, social media and news media coverage should be appreciated as both pieces and reflections of a complex system of causal structures at play.

K. Joseph, M. Benigni, W. Wei, and Prof. K.M. Carley are with the Societal Computing Program, Carnegie Mellon University, Pittsburgh, PA, USA e-mail: kjoseph,mbenigni,weiwei,kathleen.carley@cs.cmu.edu.

Prof. M. Bloom is Professor of Communications at Georgia State University, GA, USA email: mbloom3@gsu.edu

Prior work on this subject has largely considered how the news media or Twitter are useful in understanding the social processes at play during particular events across a small set of countries (e.g. Borge-Holthoefer et al., 2014; Lotan et al., 2011) or a series of events in a particular country, most often Egypt (e.g. Tufekci and Wilson, 2012). Further, few studies have considered traditional news sources and Twitter data side-by-side, leaving questions as to the similarities and differences in how media responded to, or possibly influenced, long and short term social processes during the Arab Spring.

In the present work, we use a corpora of approximately 70M tweets and around 700K newspaper articles to provide an initial overview of the change in topical focus over time in sixteen countries relevant to the Arab Spring in both news and Twitter. We take a breadth-over-depth approach, attempting to reconcile patterns in Twitter usage and news media coverage over a wide range of countries and multiple time periods. Additionally, we employ the same methodology for both datasets, allowing us to compare results across social and print media. Specifically, we have focused on the following two research questions:

- **RQ1:** How did the topical focii of our news and Twitter data differ over time and across different countries and categories?
- **RQ2:** Did significant changes in focus on different topics reflect on-the-ground processes?

In addition to these research questions, we provide a case study that explores how changes in the topical foci of news and Twitter users may be useful in understanding how topical discussions clustered around particular countries in interesting ways, especially with respect to discussions of protests.

In order to perform our analysis, we begin by developing a set of human-curated topical themes of interest based on a review of the literature. For each theme, we determined a set of terms that, when mentioned, were relevant to these themes. We searched for these terms across all of our Twitter and newspaper data. Where we found a term used in a particular tweet or news article, we determined the time at which the content was produced and the particular country(ies) in the Arab World to which the content was relevant. We were left with a set of counts, over time, of the discussion of our different themes in different countries in the Arab world.

In theory, this count data, or rates directly calculated from it, could be used to address our research questions. As detailed in recent work by Eisenstein et al. (2014), however, the direct utilization of count data, or rate data based on these counts, is a methodologically unsound decision. As Eisenstein et al. (2014) discuss with regards to Twitter, term counts may be biased by unknown irregularities in the way Twitter provides tweets through its API (Morstatter et al., 2013) or via unique properties of the keyword or spatial queries researchers construct to obtain data from the API (REMOVED FOR BLIND REVIEW). Similarly, superfluous coverage by news media on specific countries may lead to artificial increases in counts or, if focused on themes not of interest for a study lead to superfluous decreases in rates. In the present work, we adapt a slightly modified version of the statistical model developed and employed by Eisenstein et al. (2014). This model controls for spatial and temporal patterns in the rate at which data is obtained, thereby removing many of the important biases associated with term count data. While a plethora of issues must be considered when analyzing, in particular, social media data (Tufekci, 2014), Eisenstein et al.'s (2014) model allows us to move beyond the statistical irregularities in the data and gives us more freedom to draw inferences about the relationship between rates of change in thematic content and actual events occurring during the Arab Spring.

After explaining this model in greater detail and how it was adapted for the present work, we consider analyses that address our two primary research questions, and then discuss results from the case study. With respect to RQ1, we find that Twitter and traditional news media were more highly correlated on certain topics and in certain countries than others. More precisely, topics relevant to social change and in countries in which massive social change occurred showed high levels of correlation between the social and print media, while others showed significantly less cohesion in topical focus across the two media. With respect to RQ2, we find that outlier data, as determined by our model, matched quite well with important time periods during the Arab Spring, providing further evidence that news media and Twitter data are important tools for the study of social change. Finally, with respect to our case study, we find evidence that countries clustered in thematic discussions along dimensions of social change, i.e. that countries with more similar levels of social change appear to have more similar levels of discussion across the themes we studied. We also see evidence that temporal patterns in discussions of protest provide initial support for qualitative claims made in the political science and communications literature on news media, Twitter and protests during the Arab Spring.

II. Related work

A. Overview of the Arab Spring

A host of historical factors led to the conditions in 2011 that made many Arab countries ripe for protest (Gelvin, 2015). One long-standing issue was the increasing extent of economic problems caused by ineffective, corrupt and state-run economies. These issues led to elevated levels of unemployment and high inflation (Dewey et al., 2012) as well as to food shortages and massive hikes in food prices – all of which contributed to high levels of civil unrest (Comunello and Anzera, 2012; Goldstone, 2011). The effect of high unemployment rates was particularly a problem because it heavily affected young, well-educated populations, i.e. individuals who had often been promised jobs as a result of their education (Dewey et al., 2012; Gelvin, 2015). This, combined with a "youth bulge" in which a disproportionate percentage of the population was between the ages of 15-29 in many of the MENA region countries, provided a fodder of civil unrest that required only a spark to ignite and a gust of wind to spread throughout the region. This spark came, as noted above, in the form of Mohamed Bouazizi. While Bouazizi may have provided the spark, it is generally agreed upon that social media, or more aptly, the existence of communication infrastructures that supported all forms of new media (Tufekci and Wilson, 2012; Wolfsfeld et al., 2013), served as the wind to fan the flames to Egypt, Libya and several other countries in the Arab world. In addition to work focusing on why the revolutions occurred, several scholars have developed rationals for why some revolutions succeeded (e.g., in Egypt, Libya and Tunisia) whereas others failed (e.g., in Bahrain and Saudi Arabia) and still others remain unresolved (e.g., in Syria). In interviews, Gerald Feierstein, Assistant Secretary for Near Eastern Affairs at the U.S. State Department, attributes the variation of success in the Arab Spring to the strength of each countries' civil society (REMOVED FOR BLIND REVIEW, 2015).

Other scholars have attributed the structure of governance to be the deciding factor in success or failure of the Arab Spring revolutions. Goldstone (2013) suggests that chief amongst the predictors of success is the structure of the ruling regime. Goldstone defines a *personalist* regime as one in which a single individual – who may have begun as an elected leader, or head of a military or even party regime – takes total or nearly total control of the national government. In his work he provides qualitative evidence that "the single best key to where regimes in MENA have been overturned or faced massive rebellions is where personalist regimes have arisen" (pg. 14). Goldstone's (2013) work argues that personalist regimes were the most susceptible because their power was tied to the leaders' ability to provide the necessary economic and political incentives to their constituency, particularly in rentier countries that depended on oil production.

The status of economic conditions played a role not only in fomenting revolution but also in different levels of success, as economic challenges in countries like Tunisia prevented personalist regimes from being able to "buy their way out" of the protests. It is likely that personalist regimes needed to retain their power over the military more than any other segment of societyBattera (2014); Comunello and Anzera (2012). Bellin (2012) argues that the relationship between revolutionary success can be narrowed down to the relationship between the ruling regime and the military - whether or not the regime was able to convince the military to shoot at the protestors and whether the soldiers carried out these orders. In countries where the military made the decision not to use force, protesters were able to flood the streets without fear of the full wrath of the state. This lack of impunity emboldened tougher protests that eventually led to the downfall of the Tunisian and Egyptian regimes. In contrast, in Saudi Arabia, the military opted to side with the regime and use force to quell protests, regimes managed to maintain power[1].

In making the decision of whether or not to fire on protestors, one important factor was the extent to which previously disparate social groups formed a cross-class coalition in their protests and revolutionary efforts Goldstone (2011). A unified coalition of protestors made it more difficult for the military to justify the use of force in their response to the protest for two reasons. First, the combination of various social groups lessened the military's ability to claim that violent actions were a response to a particular out-group in the interest of protecting the "nation". Second, the sheer size of such a coalition would relegate military action to being viewed as "illegitimate slaughter" (Bellin, 2012, pg. 132).

The development of these unified coalitions across various social groups also provided an opportunity for the news media to characterize protestors under a national identity, rather than as protests via one specific subgroup. As we discuss shortly, the portrayals of the revolutions by news media thus had an important impact on revolutionary outcomes in that their coverage ultimately influenced a final factor in revolutionary success, which was the extent and type of involvement in the revolutions in the various nations from international powers Comunello and Anzera (2012); Goldstone (2013). This applies both to the actions of the West during the revolutions as well as the actions taken by Arab nations themselves.

Social media had much the same effect, increasing both the level of information spread and the extent to which individuals felt compelled to participate in protests (Bellin, 2012; Tufekci and Wilson, 2012; Wolfsfeld et al., 2013). Similarly, new social media may have provided a forum for national identities to emerge (Cottle, 2011). More generally, however, there were important similarities between news media and social media not only in that they served similar purposes, but that they existed within a symbiotic relationship throughout the Arab Spring Cottle (2011). As we will see, the intertwinement of news media content and content from social media was significant, but existed more heavily within discussions related to particular themes and particular countries.

B. *Using new media to study the Arab Spring*

Having intimated our position on whether or not social media played a role in the revolutions, we now turn to how data from social media has been used to better understand the processes inherent to the revolutions as a whole. In particular, we focus on Twitter data in the present study. The earliest work that examines the relationship between new media and the Arab Spring was that of Lotan et al. (2011). The authors stressed the strong interplay between Twitter and news media, and showed that the relationships between mainstream media outlets, activists, journalists and bloggers differed in datasets collected from Tunisia and Egypt. Lotan et al. (2011) considered data that used the hashtags #egypt or #libya, finding that Egypt and Libya displayed differing amounts of tweets in the two major languages of interest, Arabic and English. Egypt had significantly more tweets in Arabic than in English, and in Libya the reverse was true. Subsequent work by Bruns et al. (2013) observed that the English-speaking world lost interest in the events of the Arab Spring much earlier than those tweeting in Arabic. (AUTHOR, REMOVED FOR BLIND REVIEW) in their examination of the Benghazi consulate and Cairo embassy attacks set against the context of the Arab Spring, observed that news attended less to Libya than to Egypt, that the volume of tweets was lowest in Libya than in Egypt, and that there was no connection between Arabic and the non Arabic tweets. As we focus largely on English-language keywords, such language effects must be kept in mind.

Perhaps most relevant to the present work, however, are the efforts of Borge-Holthoefer et al. (2014). These authors used Twitter data to test the extent to which individuals switched between Secularist and Islamist and pro and anti-military "camps" in Egypt during the Arab

[1]After the protests leading to the Fall of the Shah of Iran, inconsistent military response can also have deleterious effects. Thus the middle category of occasionally violent occasionally peaceful responses by the military also emboldens protesters and can destablize the regime Sick (1985)

Spring. While the authors focused on Arabic tweets, their efforts show that Twitter data provides a unique lens through which important and interesting social processes relevant to the revolutions can be studied. We extend their efforts to new questions, new data and new methods in the present work.

C. The role of the news media

As argued by (Hussain and Howard, 2013, pg. 60), international news media organizations were vital during Arab Spring as they focused world-wide attention to the events in the region, helping to "stave off overtly violent reactions from security forces." (Cottle, 2011, pg. 656) argues that this was particularly important in "alerting world opinion to repressive and potentially prosecutable acts of inhumanity". We would expect that the degree of coverage for particular topics, especially relating to violence and revolution, should be causal or correlated to the success and/or failure of uprisings across the Arab world. Other work by (REMOVED FOR BLIND REVIEW) has considered quantitative approaches to using news media data to better understand the Arab Spring. Specifically, the authors use news media to instantiate a dynamic network agent-based model and use model output to predict successful revolutions. While their use of newspaper data chiefly incorporates off-the-shelf statistical tools, it is an important indicator that news media data can be used in large quantities to provide accurate models of underlying social processes.

III. Data

The present work focuses on discussions of human-curated topical categories in the large corpora of news and Twitter data. Data collection and analysis was focused on thirteen Arab countries, plus Iran. The countries are listed in Table I, along with a brief description of how the Arab Spring affected that particular state and what was happening in the country in May 2015 . Below we describe in greater detail the Twitter and news datasets. For more information on the Twitter data, we refer the reader to (REMOVED FOR BLIND REVIEW) . For information on the news data, we refer the reader to (REMOVED FOR BLIND REVIEW). Note that we use a subset of the Twitter data described in previous work, and an expanded news dataset collected as the data used in previous work.

A. Twitter data

The present work utilizes a corpora of approximately 70M tweets collected between March 2011 and December 2012. The Twitter data was collected from two sources. The first utilized a variety of approaches, including geospatial bounding boxes, keyword searches and searches for specific Twitter users, to capture material that data collectors deemed relevant to the Arab Spring. As the data was obtained from an outside source, an exact listing of terms and spatial regions from which this set of data was collected is unavailable to us. However, most of the collection is known to have focused on events relevant to Egypt, Libya, Syria, Tunisia and Yemen. The second source of data is obtained from a 10% sample of all Twitter data made available to researchers at our university. From this sample, we extracted all tweets that were geo-tagged from the Arab World and Iran. Any geo-tagged tweet from a particular country is assumed to be relevant to that country. In addition, any tweet that mentioned the country's name, in Arabic or in English, or a tweet that used the name of any of the country's five major (most populous) cities in English or in Arabic, was assumed to be relevant to that country. Note that the same tweet may be considered relevant to multiple countries.

B. News data

The present work utilizes a collection of approximately 700K newspaper articles collected during the same time period as the Twitter data. All newspaper articles were extracted from the LexisNexis archive of major, English-speaking news agencies. We collected data that LexisNexis' proprietary indexing algorithms labelled as being relevant to one of the thirteen countries of interest. Text analysis was conducted only on the body of the text for each article.

IV. Methodology

There are two major components to the methodology utilized in this paper. First we explain the manner in which we narrowed down the themes, or categories, of interest and how we selected the terminology or proxy measures for each of the the categories. We then describe the methodology to determine the extent to which each category was assessed in different countries over time.

A. Categories

As the data we use in this article only covers periods after the first revolts in both Tunisia and Egypt, tthe categories we considered in our analysis were chiefly focused on identifying relevant themes regarding the spread of protests and the extent to which revolutions "succeeded". After discussion among the co-authors, three of whom are subject matter experts (SMEs), we initially settled on twenty categories of interest. After delineating the subset of categories of interest, our next task was to identify specific terms[2] that, if mentioned in

[2]Note that a term may comprise multiple words, e.g. "Michael Jordan" might be a single term for the category "sport"

Country name	Level of protests	Government overthrown?	Government response if not overthrown?	Current Situation
Bahrain	High	No	Violent	Ongoing protests
Egypt	High	Yes (2/11, 7/13)		Ongoing protests
Iran	Low	No	Police intervention	Idle
Iraq	Low	No	Police intervention	Civl war
Jordan	Moderate	No	Some reform	Idle
Kuwait	Moderate	No	Some reform	Idle
Lebanon	Moderate	No	Some reform	Idle
Libya	High	Yes (9/11)		Civil war (involving ISIS)
Saudi Arabia	Low	No	Violent	Idle
Syria	High	No	Violent	Civil war
Tunisia	High	Yes (1/11)		Idle
United Arab Emirates (UAE)	Low	No	Limited	Idle
Yemen	High	Yes (2/11)		Ongoing protests and violence

TABLE I: The thirteen countries studied in the present work, along with a (very) brief description of important events within the countries during the Arab Spring

a tweet or news article, suggested a discussion pertinent to that topic. This list of terms, was partly derived from dictionaries developed over several years of study on the region, and was further refined by the authors to finalize the set of relevant terms for each category. Finally, every tweet and news article in our dataset was searched for these terms. Due to the relatively data-hungry nature of the statistical model described in the following section, we chose to omit any category that was found in fewer than 5,000 news articles or 5,000 tweets. This left us with a final set of eleven categories that we considered in our analysis.

Table II presents the set of categories used in the present work, along with a description of why they were chosen, the number of terms that map to this category and a set of search terms representative of the category. Initially we sought a set of general indicators for social processes involved in revolution, particularly indicators that demonstrated varying levels of unrest (protest, violence, and war) and indicators for processes of change (adaptation). In addition, whether a terrorist organization existed in the country was included as theme as the presence of terrorist groups tends to be an indicator of instability (and is highly correlated with violence and revolution). Highly unstable countries also tend to have 'ungovernable spaces' absent of government control that are potential areas in which terrorist organizations can gain a foothold.

We also included several concepts inspired by the work of Goldstone and others (e.g. Cottle, 2011) suggesting that "virtually all successful revolutions were forged by cross-class coalitions...pitting society as whole against the regime" (Goldstone, 2011, pg. 457). Finally, we used a set of three "noise categories" to asses the existence of spurious correlations that the methodology we utilize here may induce. These noise categories were constructed by taking the top 25 most frequently used nouns and adjectives, as provided by Wikipedia[3] and randomly splitting them into three categories. Of these 50 terms, we removed five relevant nouns (man, woman, child, life, government) and one verb (young), leaving us with 44 terms split between the three noise categories.

With respect to the number of terms used, it is clear that there is significant variation across all of the categories. This imbalance is to be expected - while certain categories, e.g., Terrorist Organizations and Ethnic identities, can contain a limitless number of terms, many of the other categories (e.g. national identity) are restricted to a small number of possible terms. There are two reasons why one should not expect these differences to affect our analyses. First, it is important to realize that the vast majority of terms in these two categories were found in only a few, if any, tweets and news articles. Second, as we show in the next sections, categories were analyzed independently and only compared based on rates of change.

B. Statistical Approach

The statistical model we used is drawn from the work of Eisenstein et al. (2014). Eisenstein and his colleagues were interested in understanding the extent to which new words diffused on Twitter across metro areas in the United States. To capture the extent to which words in one metro area r_1 spread to a different area r_2, they formulated an autoregressive model that captured the extent to which the level of "activation" for the words in their dataset at time $t-1$ in metro area r_1 predicted the level in r_2 at time t. As noted above, a naïve use of count data, or proportions based on count data, would provide a biased analysis. Briefly, this bias can be broken into two general, exogenous factors that influence fluctuations in both counts and proportions beyond the desired influence relationship. First, as Eisenstein et al.

[3]http://en.wikipedia.org/wiki/Most_common_words_in_English

Category	Reason for Use	Num. Terms	Representative Terms
Terrorist Orgs	Indicator of instability	3619	Al-qaeda, terrorism
Adaptation	Indicator of change	130	adjustments, amendment
Protest	Indicator of unrest	19	protests, demonstrations
Violence	Indicator of unrest	39	violence, violent
War	Indicator of unrest	46	war, wartime
National identities (Nationality)	Goldstone (2011)	245	Algerian, Syrian
Youth	Goldstone (2011)	10	young person, youngster
Non-national, ethnic identities (Ethnic groups)	Goldstone (2011)	1489	Shi'a, Sunni
Stopwords 1	Check for spuriousness	14	good, person, problem
Stopwords 2	Check for spuriousness	16	good, person, problem
Stopwords 3	Check for spuriousness	14	good, person, problem

TABLE II: Categories Used

(2014) notes, sampling rates in the data may differ *over time* due to fluctuations in the rate at which Twitter's API provided data over time. These temporal patterns may also lead to random spikes in the data, spikes which may require smoothing. Second, there also exists an exogenous *spatial* factor on count data that one does not actually wish to model, particularly in that particular countries (metro areas) have much larger populations than others [4].

Our interest in the present work is not to understand diffusion, but rather to capture the extent to which different categories we are interested in were being discussed in different countries at different times. Thus, large portions of the work performed by Eisenstein et al. (2014) still apply, in particular the way in which they partial out temporal and spatial biases in their data. While the autoregressive portion of their model is not used here to estimate a diffusion network, we find that it still allows for a desirable smoothing of estimates over time. As we will discuss, it also makes it easy to remove temporal autocorrelation from the final estimates.

Table III presents an overview of the variables used in our description of the statistical model. Note that each model variable is differentiated by media via the subscript m. In all cases, results for the two media are estimated independently. Thus, to ease notation in this section, we will focus on Twitter, and consequently drop the m subscript in our discussion. Note that the analog to a Twitter user in our discussion is a single newspaper article. All data is first aggregated by month and by country. From this aggregated data, we extract two primary values. First, the variable $c_{w,r,t}$ gives the number of Twitter users (newspaper articles) who used a term in category w in at least one of their tweets related to the country r during month t. Second, $s_{r,t}$ represents the total number of users who sent one or more tweets about *anything* in country r during month t. Thus, it is straightforward to represent $c_{w,r,t}$ as being distributed binomially, $c_{w,r,t} \sim \text{Bin}(s_{r,t}, p)$.

[4]For more detail on these biases, we refer the reader to the original article

It is obvious to see that the MLE for p, $\hat{p} = \frac{c_{w,r,t}}{s_{r,t}}$ in the naive model. To address the biases above, Eisenstein et al. (2014) introduce a logistic model for the p parameter that allows an additive structure that incorporates parameters to estimate these biases, along with the "true" indicator of the popularity of category w in country r at time t. Equation 1 defines their basic model, where the logistic function $\text{Logistic}(x)$ is $\frac{1}{1+\exp(x)}$.

$$c_{w,r,t} \sim \text{Binomial}(s_{r,t}, \text{Logistic}(\eta_{w,r,t} + v_{w,t} + \mu_{r,t})) \tag{1}$$

In Equation 1, the three parameters in the logistic function affect the log-odds of an increase in $c_{w,r,t}$. An increase in a parameter represents an increase in these log-odds, and thus the higher the value of each parameter, the more likely a user is to use the term. Eisenstein et al. (2014) use the term "activation" to describe these increases, presumably in the context of the logistic model frequently used in cognitive activation theory (Anderson et al., 1997). We follow this terminology here. In Equation 1, the parameter $v_{w,t}$ is the overall activation of category w at month t, and $\mu_{r,t}$ is the activation of the country r at time t. These terms control for temporal and spatial biases, respectively. The parameter of interest to our analysis is thus $\eta_{w,r,t}$, which represented the "unbiased" activation for word w at time t in region r.

As noted above, we also would like to smooth our estimates of word activations over time to better understand longer term trends in the data. As $\eta_{w,r,t}$ is our parameter of interest, the smoothing is enacted as an autoregressive model on η, as described in Equation 2. Here, the parameter $A_{w,r}$ is the lagged influence variable, estimated for each category w and each country r.

$$\eta_{w,r,t} \sim \text{N}(A_{w,r} * \eta_{w,r',t-1}, \sigma^2_{w,r}) \tag{2}$$

The full autoregressive model can be specified as a Markov model with an *observation model* (on c) and a

Parameter	Explanation
$c_{w,r,t,m}$	The count, or number, of users (newspaper articles) that mentioned category w related to country r at time t in media m
$s_{r,t,m}$	The total number of users (newspaper articles) related to country r at time t in media m
$\eta_{w,r,t,m}$	The *activation rate* of category w related to country r at time t in media m
$v_{w,t,m}$	The activation rate of w across all countries at time t in media m
$\mu_{r,t,m}$	The base activation rate of all users interested in country r at time t in media m
$A_{w,m}$	The autoregressive component for each w across all countries and times in media m
$\sigma_{w,r,m}$	The standard deviation of η draws for a particular w and r in media m

TABLE III: table

dynamics model on η. Overall, the model we thus wish to estimate is:

$$P(\eta, c|s; A, \sigma^2, \mu, v) = P(c|\eta, s; \mu, v)P(\eta; A) \tag{3}$$

In order to do so, we adopted the same estimation process as was used by Eisenstein et al. (2014). We first considered the estimation of $v_{w,t}$ and $\mu_{r,t}$ at each time t assuming η is 0. To do so, we utilize a stepwise procedure. We first obtain a simplified $\bar{v_w}$ as the inverse logistic function ($\log(\frac{x}{1-x})$) of the total proportion of users that utilized word w across all time steps, t. Using this value, we now would like to compute the maximum likelihood estimate of each $\mu_{r,t}$ using $\bar{v_w}$ as the value for each $v_{w,t}$ (and again, setting all $\eta = 0$). Below we derive the MLE for a particular μ at an implicit region r at a single time point t (where the t and r subscripts are thus implicit). This derivation is absent (presumably due to space) from Eisenstein et al. (2014), and thus we find it useful to outline here. We first derive the form for the MLE:

$$\begin{aligned}\hat{\mu} &= \text{argmax}_{\mu} \prod_w P(c_w|s;\mu, v_w) \\ &= \text{argmax}_{\mu} \prod_w \tbinom{c_w}{s} \frac{1}{1+\exp(-(v_w+\mu_r))}^{c_w} \\ &\qquad (1 - \frac{1}{1+\exp(-(v_w+\mu_r))})^{s-c_w} \\ &= \text{argmax}_{\mu} \sum_w \log\tbinom{c_w}{s} + \log(\frac{1}{1+\exp(-(v_w+\mu))}^{c_w}) \\ &\qquad + \log((1 - \frac{1}{1+\exp(-(v_w+\mu))})^{s-c_w}) \\ &= \text{argmax}_{\mu} \sum_w \log\tbinom{c_w}{s} - c_w \log(1+\exp(-(v_w+\mu)) \\ &\qquad + (s-c_w)\log(1 - \frac{1}{1+\exp(-(v_w+\mu))})\end{aligned} \tag{4}$$

As there is no straightforward closed-form solution to this expression, we compute the derivative of the above expression and use it for a gradient descent approach to maximizing the function above. The gradient can be derived as follows:

$$\begin{aligned}&\frac{\partial}{\partial \mu} \sum_w \log\tbinom{c_w}{s} - c_w \log(1+\exp(-(v_w+\mu)))+ \\ &\qquad (s-c_w)\log(1 - \frac{1}{1+\exp(-(v_w+\mu))}) \\ &= \sum_w \frac{(c_w - s)\exp(v_w+\mu)}{1+\exp(v_w+\mu)} + \frac{c_w}{1+\exp(v_w+\mu)}\end{aligned} \tag{5}$$

An analogous derivation can be constructed for each v, and thus we excluded it from the present work. After solving for all $w_{r,t}$ and all $v_{w,t}$, we can estimate values for parameters in the autoregressive portion of the model, $\eta_{w,r,t}$, A_w and $\sigma^2_{w,r}$. Because the observation model here is non-Gaussian, the traditional Kalman Filter algorithm cannot be used to learn model parameters. Instead, we resort to Bayesian methods to perform approximate draws from the distribution of the ηs over time. We can then use maximum likelihood estimation to update values for A_w and $\sigma^2_{w,r}$. We then update expectations for all A and σ^2 parameters and iterate again to generate a draw from an updated version of the distribution for $\eta_{w,r,t}$.

As outlined, this process is a Monte-Carlo EM algorithm (Bishop and others, 2006, p. 439) that can be used to estimate A_w, $\sigma^2_{w,r}$ and $\eta_{w,r,t}$. In the **E** step, we get an expectation for η using the Forwards-Filtering Backwards Sampling (FFBS) algorithm (Godsill et al., 2004). To do so, we need an initial distribution, which we can use to begin the forward-filtering step in the algorithm. Eisenstein et al. (2014) do not specify an initial distribution for η_0, so we draw from $\mathrm{N}(\hat{eta}_0, 1)$, where $\hat{\eta_0}$ is the MLE for eta_0, given in Equation 6, where $x = \frac{c}{s}$:

$$\begin{aligned}\hat{\eta_0} &= \text{argmax}_{\eta_0} p(\eta_0|\sigma^2, c_{t=0}, s_{t=0}, \mu_{t=0}, v_{t=0}) \\ &= \log(\frac{x}{1-x}) - v_{t=0} - \mu_{t=0}\end{aligned} \tag{6}$$

After obtaining a value for $\eta_{t=0}$, we proceed with FFBS with the proposal distribution $Q(x)$ equal to the transition distribution. In contrast to Eisenstein et al. (2014), who use a simple particle filter in the forward-filtering step, we found our estimates were much more stable when we used an additional resampling step during the filtering process (Arulampalam et al., 2002).

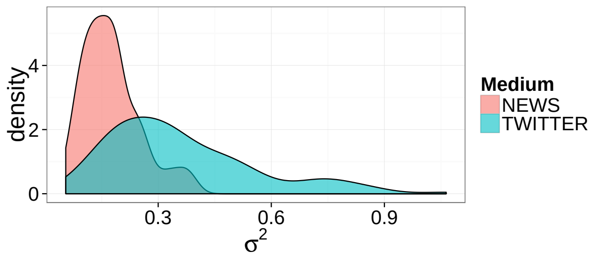

Fig. 1: Distribution of Sigmas

Thus, for each time point, we construct weights for each sample $\eta^{(m)}_{w,r,t}$ as:

$$\omega^{(m)}_{w,r,t} = \omega^{(m)}_{w,r,t-1} * P(c_{w,r,t}|\eta_{w,r,t}, s_{w,t}; \mu_{r,t}, v_{w,t}) \quad (7)$$

Next, we resampled from this distribution, producing a set of samples for $\eta_{w,r,t}$ from a discrete approximation of the true distribution. This process, when run on each $\eta_{w,r,t}$, completes the "filtering" step on the forward pass. On the backwards pass, we draw our samples for $eta_{w,r,t}$ using weights defined by the likelihood of each sample given the sample's likelihood as defined by the observation model. This completes the E step, or in other words, provides our samples for $\eta_{w,r,t}$. The M step updates the MLEs for A_w and $\sigma^2_{w,r}$. The update for A_w is simply determined via least-squares estimation. The MLE for $\sigma^2_{w,r}$ can be solved in closed form, $\frac{1}{T}\sum_t^T(\eta_{w,r,t} - a_{\tilde{w}}, r\eta_{w,r,t-1})$.

The code for this estimation process is available at (REMOVED FOR BLIND REVIEW). Note that the primary differences between our work and Eisenstein et al. (2014) are two-fold. First, Eisenstein et al. (2014) were interested in constructing a diffusion network, and thus continue with a further estimation step to approximate a full transition matrix A across all regions. Second, while Eisenstein et al. (2014) focus on specific terms, we focused on collections of terms. However, the generalization is trivial, as the sum of a set of independent Binomial random variables is still binomially distributed.

V. Results

In this section, we consider data for each research questions in more detail.

A. **RQ1**: *Comparing news and Twitter*

While a host of comparisons could be made between the news and Twitter data we study, we restricted ourselves here to a subset of questions of particular interest. First, we are interested in the level of variance over time in discussions of particular themes in particular countries on Twitter as compared to in news. Conveniently, our statistical model provides us with a parameter, $\sigma^2_{w,r}$ which captures the extent of this variability. Figure 1 shows a density plot detailing the distribution of values for $\sigma^2_{w,r}$ for the news and Twitter data. As is clear, Twitter data is much more variable in the news, both in the average variability across a particular theme/country time series and with respect to the expected value of $\sigma^2_{w,r}$ in general. This finding shows that news coverage of the Arab Spring was more stable than discussion on Twitter. Our results correlate with the idea that while the propagation of content is similar in form in both Twitter and news media, the social elements of Twitter introduce important variability in the diffusion and longevity of discussion (Yang and Leskovec, 2011). Perhaps more importantly, our findings show that the extent to which a change in thematic focus is "interesting" differs when analyzing news versus Twitter data. Researchers using social media data to interpret change in social processes, our data suggests, must be more cautious than those with access to news data.

In addition to levels of variability in the two media, we are also interested in the extent to which news and Twitter data focused on the same topics at the same time. That is, we are interested in the extent to which the time series of $\eta_{w,r,t}$ values for Twitter and news are correlated for a particular theme/country combination. Importantly, this correlation cannot be measured directly with our $\eta_{w,r,t}$ - baked into our model is the assumption of autocorrelation with a one-month lag (Equation 2), which will lead to spuriously high correlation values. To infer correlations between news and Twitter data, we therefore must remove the autocorrelation by subtracting the prior month's value from each activation score. More specifically, for a given country r and category w, we construct a vector of observations $\eta_{w,r,t} - \eta_{w,r,t-1} \quad \forall t > 1$ for both Twitter and news data. These vectors are, via the assumptions of our model, void of autocorrelation and can therefore be compared directly using, for example, a simple Pearson correlation.

The first question is whether or not there is a significant correlation between the news and Twitter across all countries and themes. Importantly, we would expect this correlation to be positive and significant for the non-stopword data, and to be close to zero for our stopword themes. Using the bootstrap procedure with 10000 iterations, we construct 99% confidence intervals using a pivot interval (Wasserman, 2003) on the Pearson correlation between news and Twitter data. That is, we determine the extent to which the changes in activation from the previous month in Twitter data are linearly correlated with changes from the previous month in news data, taking as input all observations across all countries. We perform this analysis twice, once for the non-noise (non-Stopword) themes and once for the noise-only

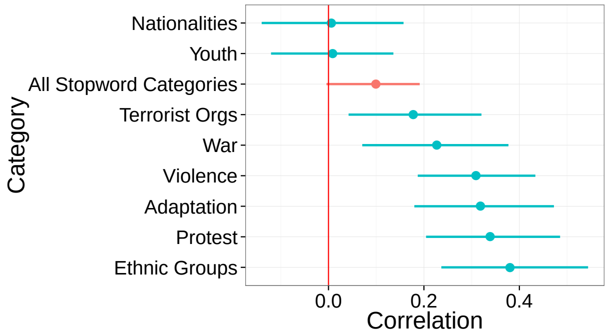

Fig. 2: News/Twitter correlations by category. Confidence intervals are 99% bootstrapped CIs calculated with a pivot interval and 10000 iterations

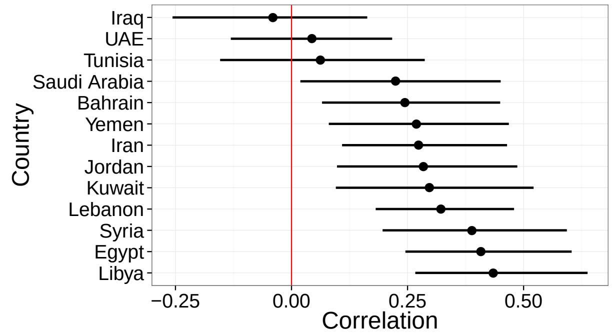

Fig. 3: News/Twitter correlations by country

themes.

The correlation for the non-noise data is .23 [.17,.28], indicating a moderate but significant correlation between news and Twitter. There are several reasons for the existence of this correlation. Perhaps most importantly, it is well known that news agencies send tweets about their key stories, and retweets of these stories comprise an important component of discussion on Twitter. Thus news and Twitter should be correlated, as they form a part of a symbiotic media system. In some sense, it is therefore interesting that the correlation between the two media is not higher. One obvious reason is that personal, socially-focused discussions do of course occur on Twitter. Additionally, a large volume of tweets are known to be sent by bots on Twitter, many of which are just tweeting nonsense phrases.

The correlation for the noise-only data was .1 [-.001, .19]. Results suggest that correlations between non-noise themes is (significantly) greater than noise themes, and that the noise data was not significantly different from 0 at $\alpha = .01$. However, results do indicate that there may still be some level of correlation in the residual noise in our data between news and Twitter. This may be due to a variety of factors, perhaps most obviously the fact that we treat each country independently. While future work may consider means of addressing these biases, we instead choose here only to account for the possible existence of this spurious correlation when considering further results. We now turn to explore which themes had the strongest and weakest correlations between news and Twitter (across all countries).

Figure 2 presents the correlations between news and Twitter data for each category, with data aggregated across countries. Additionally, we have aggregated all noise categories into a single category, entitled "All Stopword Categories". Confidence intervals are constructed in the same way as above, and are shown in the figure via the lines for each category. Mean estimates are shown with a dot. Figure 2 shows that the strongest correlations between news and Twitter occur in reference to general indicators of revolution, specifically, protest and violence. The general level of correlation across these categories supports the idea that news and Twitter discussions specific to the revolution fed off of each other (Comunello and Anzera, 2012; Cottle, 2011). Additionally, we can expect that this may be due to the fact that the Tweets by news agencies are focusing in this area. In addition, we observe strong correlations between media in discussions of adaptation and change and to ethnicity.

On the other hand, in comparison to the Stopword categories, there is no significant correlation[5] for discussions of national identity, youth movements, terrorist organizations and war. With reference to the “selecting on the dependent variable” problem in social media research (Tufekci, 2014), our findings indicate that scholars focusing too heavily on particular topics may over or under-estimate the correlations between topical focuses of news and Twitter. Perhaps most interesting is the fact that while discussion of Ethnic, or non-national, identities was highly correlated between the two media, discussion of national identities show little, if any, correlation between media This difference is important in considering the extent to which news or Twitter data can be used to support the idea of an evolving national identity in contrast to factionalized identities, and suggests that the evolution of a national identity may be a factionated process that is not well monitored by using a single media source. Thus, in future study of Goldstone’s (2011) hypotheses on how the proliferation of national identity was important to the success of revolutions, one must consider both news media and Twitter separately in attempting to understand how the existence of a national identity played out in the media landscape surrounding the Arab Spring.

Finally, Figure 3 shows correlations between news and

[5] at a level of $\alpha = .001$ using the parametric test described in (Zou, 2007)

Twitter across countries, aggregating across all non-noise themes. Here, as there is no noise baseline to compare to, we only consider differences from zero in our interpretation. In light of previous work, the most interesting point to be taken from Figure 3 is that countries where data has been compared across news and Twitter in the past differ in the ways in which news relates to Twitter. In particular, while there are high levels of correlation between news and Twitter in Libya, Syria and Egypt, no significant correlation exists between these two media in Tunisia. This implicates the fact that consideration of a single country may also lead to biased conclusions on the relationship between news and Twitter during the Arab Spring. More specifically, it seems that in countries with sustained levels of civil unrest (Syria, Egypt and Libya), news and Twitter are more likely to converge on thematic discussions. In contrast, in countries like Tunisia, where the majority of the civil unrest was short-lived, thematic discussions appear to deviate between the two media. Similarly, correlations in nations where no massive government change occurred show only weak or non-existent correlations between the media. This may be because in these countries the discussion in the news was on political, economic an global events, but in Twitter it focused on social, cultural and personal events.

B. **RQ2:** *Relating to events on the ground*

In this section, we explore the extent to which large, sudden changes in activation rates in particular theme/country pairs were indicative of important real-world changes occurring in the Arab world. Table IV lists the ten most "surprising" jumps in terms of change in activation from one month to the next. We ranked the extent to which events were surprising using absolute change in activation, controlling for the overall level of variance for a particular theme/country pair. That is, for month t in country r for theme w, surprise was calculated as $\frac{|\eta_{w,r,t}-\eta_{w,r,t-1}|}{\sigma^2_{w,r}}$. Table IV also shows whether the change was positive or negative (in the column "+/- change") and briefly details the major events we believe led to this change.

Model output shows that the beginning of the Syrian Civil War had the strongest impact on news media. Discussion of war decreased significantly relative to the previous month, as reporters sought to cover the increasing volume and intensity of protests throughout the country. These events, of course, marked the beginning of a civil war that is still well under way (as of the time of writing), showing the model's propensity to capture important changes in events via news media. In general, sizeable changes in news media's topical coverage were relevant to emergent protests. Interestingly, this focus appears to manifest in our model as a decrease in discussion of war as opposed to an increase in the discussion of protest. That is, the shift of news coverage away from war or the possibility of war is as indicative of critical protest events as is discussion of the protests themselves. We will see that this effect is significant in interpreting news discussion with respect to Libya, as we explore in our case study.

In contrast to news media, the most prominent events that caused observable large shifts on Twitter were elections. In Libya and Tunisia, elections led to differences in discussions of adaptation and change, a sign that these elections may have promoted an influx of hopeful discussion of a new era in these countries. In contrast, in Egypt, elections led to increased discussions on organizations with known connections to terrorism, for example, the Muslim Brotherhood or al-Qaeda. The difference in thematic focus in these different elections is an important implication of the differences regarding sentiments towards the elections in each country, and something of interest to future work In addition to elections, the gender equality protests in Tunisia, staged heavily on social media (Yuce et al., 2015), appear as a signal in the Twitter data but not in the news data.

Finally, with respect to the Twitter data, two entries relating to the UAE in Table IV, both of which are relevant to the publication of a European Union report condemning human rights conditions for certain ethnic groups in the country, show up as events which captured a surprising level of discussion relative to the general level of focus on the themes of ethnic identity and adaptation to change in the UAE. Similarly, in the news data, a surprising level of discussion focuses on the attack of a U.S. consulate in Benghazi, Libya. These issues were relatively minor with respect to their implications for the Arab Spring, but relatively major with respect to the interest drawn to them by U.S. and U.K. media. The change in discussion that resulted from these events is thus a reminder that the Western-media centric data utilized for this study presents a possible roadblock in

VI. Case Study

Our case study is divided into two sections. In the first, we focus on understanding how countries cluster in thematic space, particularly with respect to protest. In the second, we consider in greater detail how patterns in the protest theme differentiated by media provide support for claims explored by Cottle (2011).

A. *Connecting countries and themes*

Figure 4 shows two plots, one for news (left) and one for Twitter (right). Each plot depicts a heatmap of the mean activation values ($\overline{\eta_{w,r,t}}$) over all months for each country/theme combination. The higher the mean activation score above zero for a particular theme/country

N	Category	Country	Media	Date	+/- Change	Major Event
1	War	Syria	News	3/2011	-	Week of March 15-21 is considered to be the beginning of the Syrian uprising[6]
2	Protest	Syria	News	3/2011	+	Week of March 15-21 is considered to be the beginning of the Syrian uprising
3	Ethnic Groups	UAE	Twitter	10/2012	+	EU report condemning human rights climate in UAE
4	Protest	Tunisia	Twitter	8/2012	+	Tunisian women stage large protests[7]
5	War	Lebanon	News	3/2011	-	Major rallies throughout the nation on various days[8]
6	Adaptation	Libya	Twitter	8/2012	+	First free elections were held in Libya in July, 2012
7	Adaptation	Tunisia	Twitter	10/2011	-	First free election in Tunisia since 1956[9]
8	Terrorist Orgs	Libya	News	9/2012	+	Attack on U.S. consulate in Benghazi, Libya
9	Terrorist Orgs	Egypt	Twitter	11/2012	+	Major protests in Tahrir Square and the beginning of Parliamentary Elections
10	Adaptation	UAE	Twitter	10/2012	-	EU report condemning human rights climate in UAE

TABLE IV: The top 10 most surprising changes in activation from the previous month. The column "+/- change" indicates if the change was an increase or decrease in activation. The column "Major Events" gives an explanation of the major events that occurred during that month or the previous month (as change is relative to the previous month), or none if no such event occurred.

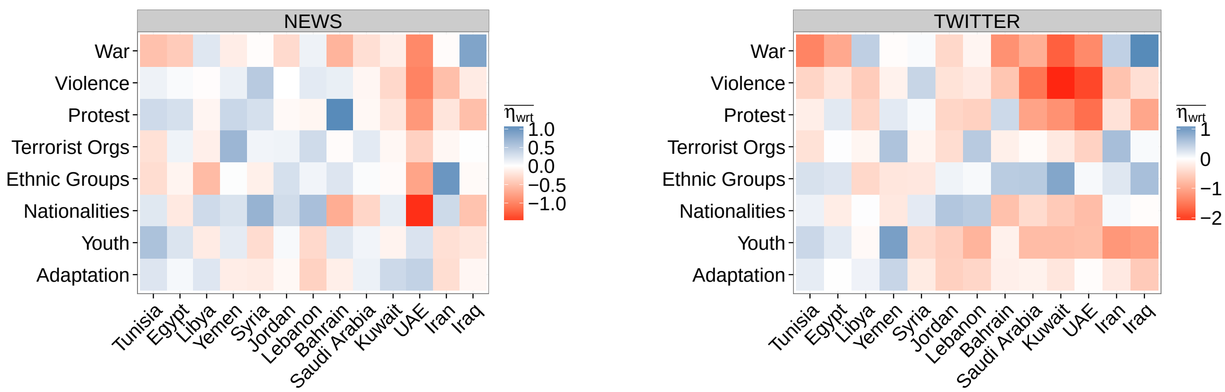

Fig. 4: Mean activation rate over all months for each country and category combination. The higher the value above 0, the darker the blue. The lower the value below 0, the darker the red.

combination, the darker blue the square. The lower the value below zero, the darker the red. Additionally, note that values are interpretable relative to global activity of each topic/country combination. That is, the dark blue square in the top right of the Twitter plot shows that, on average, relative to other countries, discussions of war were more frequent in Iraq than the discussion of other topics.

Figure 4 suggests two fascinating patterns in how countries clustered along themes explored here. First, we considered how countries clustered along perhaps the most interesting theme with respect to the Arab Spring, that of protest. Figure 4 shows that for countries in which protests occurred at relatively low levels (Iran, Iraq, Saudi Arabia and the UAE), discussion of protests were low in both the news and Twitter data. Interestingly, however, even in countries where relatively high levels of protest occurred, a distinction can be made in the level of discussion of protest between countries where revolutions are still ongoing or succeeded in overthrowing the government (Egypt, Yemen, Tunisia, Libya, Bahrain and Syria) versus those where little social change occurred (Jordan, Kuwait, Lebanon). In all but Libya, countries where social change followed protests showed positive (above zero, on average) levels of discussion about protest in the news media data. In contrast, countries where protests failed to produce significant results showed negative activation rates for this theme, on average, across time. In the Twitter data the same patterns hold, although discussion of protest in Tunisia, who's government fell before the beginning of our data collection, is also negative.

In sum, results show that in both Twitter and news media, levels of discussions about protest clustered into countries in which important social change occurred during the period, where positive activation scores for protest were observed, and countries where little or no social change occurred, where negative activation scores

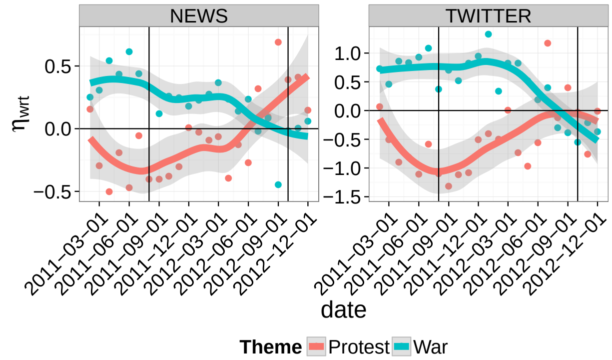

Fig. 5: Relationship between the protest and war themes over time in Libya. Points represent estimates from the model, non-parametric smoothed estimates with 95% confidence intervals are provided as lines to ease visual observation of trends.

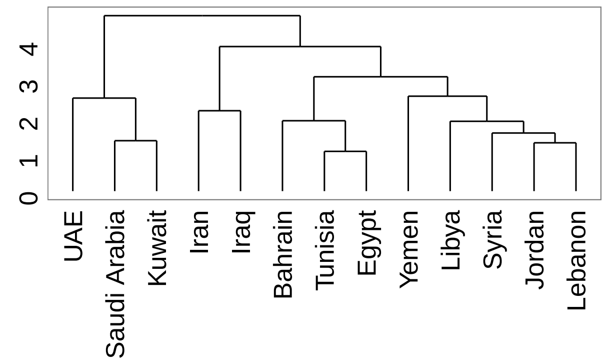

Fig. 6: Hierarchical clustering of our 13 countries based on mean activation levels across all categories for both news and Twitter

were seen. The lone exception to this categorization was Libya, where a six month civil war eventually led to the downfall of the ruling regime. We can understand why low levels of discussion on the subject of protest occurred by looking at the way in which protest was discussed over time in Libya. Figure 5 shows the activation rates of the categories protest and war over time in Libya. The first vertical black line represents, approximately, the point at which the ruling regime was overthrown, the second black line represents the attack on the U.S. consulate in Benghazi. As we can see, the focus in Libya on protest may largely have been mediated by a focus on the civil war and its aftermath. However, as unrest grew again in the face of complaints about the new regime, we do see a renewed trend towards protest, which could be seen as an initial indicator of a return to civil unrest. This intricate relationship between discussions of war and protest is of particular interest in understanding how unrest leads to organized violence.

Figure 4 also suggests there exists clusters of countries that exist at a larger level than simply distinctions across protests. To better evaluate the extent to which this clustering exists, we perform a standard, complete-linkage, agglomorative clustering, where each country is represented by sixteen features (the eight categories shown in Figure 4 for both news and Twitter). Figure 6 presents a dendrogram of the resulting clustering. Figure 6 shows that the oil-rich nations of Saudi Arabia, Kuwait and UAE, where little social change occurred, showed similar thematic structures. Similarly, nations with which the United States has the strongest tensions, Iran and Iraq, cluster together. Figure 6 shows that these five countries are heavily separated in thematic content from nations where social change occurred, although they are also separated from perhaps the more similar nations of Lebanon and Jordan. Within the cluster holding nations

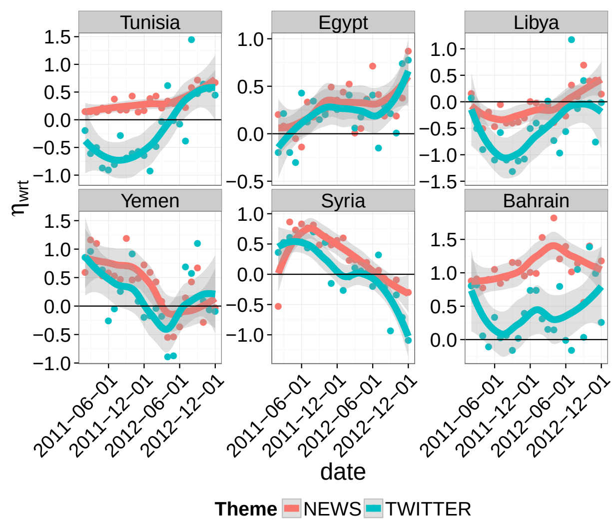

Fig. 7: Activation Rates of protest over time for six countries of interest. Points represent estimates from the model, nonparametric smoothed estimates with 95% confidence intervals are provided as lines to ease visual observation of trends.

where large-scale social change occurred, we observe that Tunisia and Egypt are the most tightly connected, an indication of the extent to which these two revolutions were tied together.

B. Temporal patterns in protests

Having noted that important clustering appears in average levels of activation with respect to protests, we turn now to how temporal patterns of protest played out in both news and Twitter for nations where high levels of social change occurred. Figure 7 illustrates activation rates for the theme of protest in news media and Twitter for Bahrain, Egypt, Libya, Syria, Tunisia and Yemen. Our findings, even over time, are consistent with the idea presented by Cottle (2011) that new media played an important role in discussions of protests during the Arab Spring. Perhaps more interestingly, Cottle (2011) also argues that the relationship between social media and news media facilitated international recognition and protest legitimization, and provided human rights surveillance. Our analysis is consistent with this assertion. In contrasting results for Egypt, with its partially free press, to more restricted countries like Syria, Tunisia, and Bahrain, we observe that states with more oppressive regimes tend to show less of a relationship between social and news media. However, even in the more oppressive regimes, uptakes in Twitter activation with respect to protest always lead to corresponding changes in news activation rates which implies a strong interaction between the two media types. This relationship between news media and social media activation rates reinforce Cottle's (2011) premises.

Since these clusters of countries we extract in this section are based on both news and Twitter, they cannot be attributed to just a western view of similarity among the countries. Rather, they reflect some underlying commonality in the way information about and from the these countries is presented across a diverse media landscape. On the one hand, these clusters show which countries have a common media profile and so, to an extent, a common pattern of media usage by the population. On the other hand, these clusters may indicate how countries might be similarly influenced by a media campaign.

VII. Conclusion

In the present work, we present a preliminary comparison of the content of news media and Twitter data over a variety of themes, countries and time periods from data on the Arab Spring. In order to control for biases in the data, we utilized a latent vector autoregressive model first introduced by Eisenstein et al. (2014). Our work focuses on two research questions. First, we considered the extent to which Twitter and news were focused on the same themes at the same time with respect to the same countries. With respect to RQ1, we find the following:

- There is a moderate but significant correlation in the focus of Twitter and news media over time and country across a variety of themes
- The level of correlation is generally stronger for themes related to the concept of revolution
- The level of correlation is strongest with respect to the discussion of Ethnic identities in the region

We then considered how well significant changes in thematic content in Twitter and news media signaled important events occurring on the ground. With respect to RQ2, we find the following:

- The largest jumps in thematic focus in news data tended to occur in content related to war and protest when large-scale protests occurred
- The largest jumps in thematic focus in Twitter data tended to occur in content related to themes of adaptation and change before or during election periods
- In both news and Twitter data, results were biased by the use of Western-centric, English-only data

Finally, we considered a case study that took a wider view of the relationship between countries and content. We found that when grouping countries in terms of average focus across themes, the resulting clusters were roughly organized by the level of social change that occurred in the country. This did not hold true in all cases, particularly for Jordan and Lebanon. Future work is required to appreciate both the underlying social factors influencing why countries clustered in the way they did, and why Jordan and Lebanon appear to be outliers. Our case study also considered how the data roughly supported Cottle's (2011) perspective of the way in which news media and Twitter were used in the discussion of protests in countries where significant social change occurred. Again, future work is needed to more rigorously test these findings.

In general, we find that our work presents a favorable outlook of the opportunities for broadly scoped study of news and Twitter data and their potential for supporting or refuting the host of qualitative work that currently exists on the Arab Spring. However, several limitations much be kept in mind. Beyond a focus on Western-centric, English only data, myriad other methodological trapdoors exist in studying social media data Morstatter et al. (2013); Ruths and Pfeffer (2014); Tufekci (2014). Although our model does well at dealing with some of these biases, results with our Stopword-based themes show that further work is necessary in this area.

Additionally, it is very difficult to truly assess how well models of social change are actually capturing important events on the ground. This is so for at least two reasons. First, a comprehensive, quantitative measure of ground truth events during revolutions is very difficult, if at all possible, to come by. Event-level datasets, particularly during the Arab Spring, are consequently drawn from newspaper data, and therefore cannot be used as a means to assess another model based on the same data. Second, during periods of unrest, important events are rampant, and thus our observations here are important only in so far as we can state that the model captures particular sorts of events (e.g. elections) with a higher frequency than others. Thus, work on creating event datasets using mixed-method and human-in-the-loop approaches is vital in furthering our ability to develop statistical models of text content that we can compare to actual events on the ground during periods of revolution (Hanna, 2014).

References

Anderson, J. R., Matessa, M., and Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188.

Battera, F. (2014). Perspectives for change in tunisia, egypt and syria: the military factor and implications of previous authoritarian regimes. *Contemporary Arab Affairs*, 7(4):544–564.

Baym, N. K. (2010). *Personal Connections in the Digital Age*. Polity.

Bellin, E. (2012). Reconsidering the robustness of authoritarianism in the middle east: Lessons from the arab spring. *Comparative Politics*, 44(2):127–149.

Bishop, C. M. and others (2006). *Pattern recognition and machine learning*, volume 1. springer New York.

Borge-Holthoefer, J., Magdy, W., Darwish, K., and Weber, I. (2014). Content and network dynamics behind egyptian political polarization on twitter. *arXiv:1410.3097 [physics]*. arXiv: 1410.3097.

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.

Comunello, F. and Anzera, G. (2012). Will the revolution be tweeted? a conceptual framework for understanding the social media and the arab spring. *Islam and Christian–Muslim Relations*, 23(4):453–470.

Cottle, S. (2011). Media and the Arab uprisings of 2011: Research notes. *Journalism*, 12(5):647–659.

Dewey, T., Kaden, J., Marks, M., Matsushima, S., and Zhu, B. (2012). The impact of social media on social unrest in the arab spring. *International Policy Program*.

Eisenstein, J., O'Connor, B., Smith, N. A., and Xing, E. P. (2014). Diffusion of Lexical Change in Social Media. *PloS one*, 9(11):e113114.

Gallé, M., Renders, J.-M., and Karstens, E. (2013). Who broke the news?: an analysis on first reports of news events. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 855–862. International World Wide Web Conferences Steering Committee.

Gelvin, J. (2015). *The Arab uprisings: what everyone needs to know*. Oxford University Press, 2nd edition.

Godsill, S. J., Doucet, A., and West, M. (2004). Monte Carlo Smoothing for Nonlinear Time Series. *Journal of the American Statistical Association*, 99(465):156–168.

Goldstone, J. A. (2011). Cross-class coalitions and the making of the arab revolts of 2011. *Swiss Political Science Review*, 17(4):457–462.

Goldstone, J. A. (2013). Bringing regimes back in–explaining success and failure in the middle east revolts of 2011. *Available at SSRN 2283655*.

Hanna, A. (2014). Developing a System for the Automated Coding of Protest Event Data. *Available at SSRN 2425232*.

Hussain, M. M. and Howard, P. N. (2013). What best explains successful protest cascades? ICTs and the fuzzy causes of the arab spring. *International Studies Review*, 15(1):48–66.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., and Boyd, D. (2011). The revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International Journal of Communication*, 5:1375–1405.

Morstatter, F., Pfeffer, J., Liu, H., and Carley, K. M. (2013). Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. *Proceedings of ICWSM*.

Ruths, D. and Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346(6213):1063–1064.

Sick, G. (1985). *All fall down: America's tragic encounter with Iran*. Random House New York.

Starbird, K. and Palen, L. (2012). (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pages 7–16. ACM.

Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*.

Tufekci, Z. and Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from tahrir square. *Journal of Communication*, 62(2):363–379.

Wasserman, L. (2003). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

Wolfsfeld, G., Segev, E., and Sheafer, T. (2013). Social media and the arab spring politics comes first. *The International Journal of Press/Politics*, 18(2):115–137.

Yang, J. and Leskovec, J. (2011). Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 177–186. ACM.

Yuce, S., Agarwal, N., and Wigand, R. T. (2015). Women's Right to Drive: Spillover of Brokers, Mobilization, and Cyberactivism. In Agarwal, N., Xu, K., and Osgood, N., editors, *Social Computing, Behavioral-Cultural Modeling, and Prediction*, number 9021 in Lecture Notes in Computer Science, pages 232–242. Springer International Publishing.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological methods*, 12(4):399.