Supplementary Materials for "Online Social Networks and Offline Protest"

November 3, 2015

Data

The protest measures in our analysis come from the Global Dataset on Events, Location, and Tone (GDELT) [1]. GDELT is a machine-coded events dataset that extracts dyadic events, including protests, from publicly available news reports (there are 20 main categories of events — for a full description of each category, see the codebook for the Conflict and Mediation Event Observations dataset [2].) Each GDELT row records a primary actor, the primary actor's action (the event), and the actor receiving the action, in addition to metadata such as date, GPS coordinates of the actors and events, the tone of the article, and how many articles write about the event. For example, a row might show that Egyptian police (primary actor) in Alexandria (location) increased their alert status (event) in response to protestors (actor receiving action). Multiple reports are aggregated into one event so that each entry records a unique event; the number of news stories and the number of news organizations writing about the event are separately recorded.

We extract all protests from Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Libya, Lebanon, Morocco, Oman, Qatar, Saudi Arabia, Syria, Tunisia, the United Arab Emirates, and Yemen between November 1^{st}, 2010 and December 31^{st}, 2011. GDELT has identified over 120,000 unique protests during the period of this study, and only 25% of our country-days contain no protests. One GDELT event category is "Protest", so we selected all protest events that occurred in one of the 16 countries during our study. Protests did not start in Tunisia until the middle of December 2010, so the month of November provides a baseline against which to measure subsequent events. The ending date was chosen because our source of geolocated Twitter data ends on that day. We aggregate all data to the daily level for each country for a total of 6,816 observations.

The tweets involved in our analysis represents a sample of about 10% of the public activity within the microblogging platform. There are two ways in which we can identify the country of origin. First, Twitter users have the option to activate location sharing, in which case the tweet would come accompanied by GPS coordinates. As a first step we grab the ones that satisfy this condition and that are indeed in one of the countries of interest. Second, tweets in the raw data stream may be accompanied by a publicly shared location. All tweets without GPS are thus parsed for publicly declared location. We then compare the provided location to a dictionary of cities and country names (and their variants and abbreviations) to assign each tweet to a city or country.

Measurements

Equation 1 shows how we calculate our measure of coordination, a Gini coefficient applied to the distribution of hashtags. For each day i in each country k, we observe n unique hashtags. We count the number of times y each hashtag j occurs and use those counts to calculate a measure of hashtag coordination:

$$Coordination_{i,t} = \sum_{i=1}^{16} \sum_{t=1}^{426} \frac{2\sum_{j=1}^{n} jy_j}{n\sum_{j=1}^{n} y_j} - \frac{n+1}{n} \tag{1}$$

Equation 2 shows how we calculate the hashtag percent measure. For each day i, we count the number of tweets, t, and tweets with hashtags, h. Dividing the latter by former gives the percent of tweets for that day with a hashtag. We repeat this calculation for each of the 426 days for each of the 16 countries.

$$HashtagPercent_{i,t} = \sum_{i=1}^{16} \sum_{t=1}^{426} \frac{h_i}{t_t} \tag{2}$$

Equations 3 through 5 show how we calculate the other possible measures of coordination. Equation 3 replaces h, the numerator of Equation 2, with r, a count of tweets that are retweets. Equation 4 does the same for external links to urls (l); 5 for tweets that mention another Twitter user (m).

$$RetweetPercent_{i,t} = \sum_{i=1}^{16} \sum_{t=1}^{426} \frac{r_i}{t_t} \tag{3}$$

$$LinkPercent_{i,t} = \sum_{i=1}^{16} \sum_{t=1}^{426} \frac{l_i}{t_t} \tag{4}$$

$$MentionPercent_{i,t} = \sum_{i=1}^{16} \sum_{t=1}^{426} \frac{m_i}{t_t} \tag{5}$$

We present summary statistics for all these measures along with the protest measure in Table 1. Table 2 shows the matrix of Pearson correlations for measures taken from the Twitter data and protest measures on the subsequent day. Notice in particular that the coordination measure exhibits the strongest correlation with protest. Finally, Table 3 shows counts and mean values for each of the main variables, by country.

Estimation Strategy

Our base model is:

$$g(E[Protests_{i,t}]) = \boldsymbol{\beta}\mathbf{X} + \gamma Protests_{i,t-1} + \alpha_i + \theta_t \tag{6}$$

where where $g()$ is a link function determined by the form of the dependent variable. Because the dependent variable is a count of protests, it is an integer always greater than or equal to 0. Below, we use a negative binomial model since the dependent variable is a count of protests and the distribution of this variable tends to be overdispersed relative to a poisson distribution. However, we also present models where we apply a simple ordinary least squares (OLS) model to the data after logging the dependent variable with an offset of +1 (results are similar for both specifications). The values $\mathbf{X}$ are independent variables and controls, with estimated linear effects $\boldsymbol{\beta}$. We also control for $Protests_{i,t-1}$ to address the problem of serial correlation, and we include country fixed effects (α) and day fixed effects (θ) to control for unobserved differences between countries (population, history, and so on) and over time (muslim holidays, sporting events, and so on) that may contribute to variation in protest. Finally, to further address the problem of repeat observations within each country, we cluster error terms by country.

We applied a number of tests to the OLS version of our model to diagnose problems related to serial correlation. A Durbin-Watson test for serial correlation returns a test statistic of 2.19, suggesting no positive correlation of the errors but perhaps a slight negative correlation. The Dickey-Fuller coefficient is -10.99 and has a p-value less than 0.01, so we do not need to worry about nonstationarity of the dependent variable. Finally, a Lagrange-Multiplier test for two-way fixed effects returns a chi-square value of 235, suggesting it is appropriate to include country and day fixed effects in each of our models.

Results and Robustness Checks

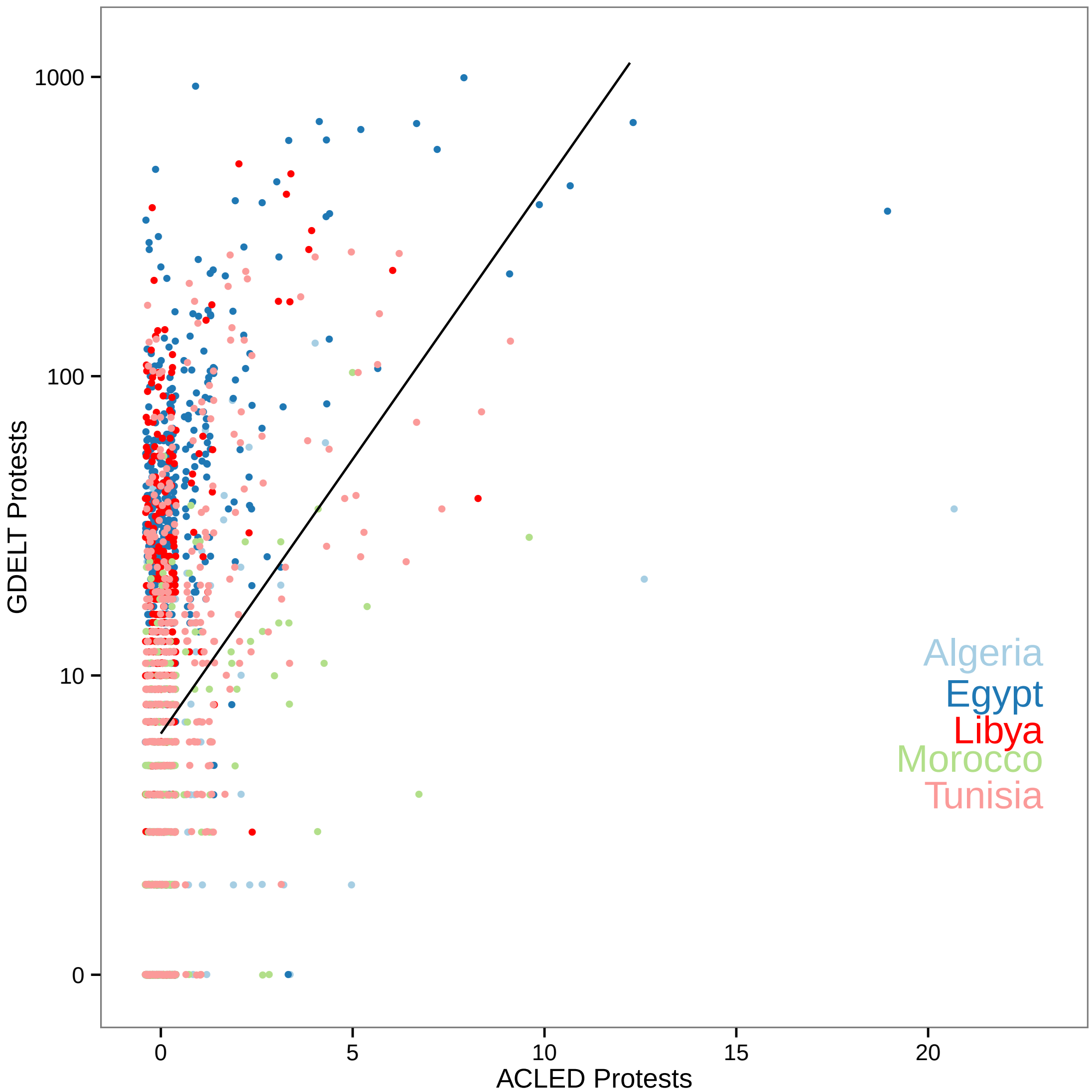
Table 4 shows that hashtag coordination is strongly associated with protests the following day (Model 1). The only other variable that correlates with protests is the percent of tweets containing a hashtag (Model 2), but that relationship disappears when we include the coordination measure in the model (Column 3).

Tables 5 and 6 address two primary threats to validity. First, the main results rely on a machine-coded dataset that itself relies on news reports. Given well-known biases in news coverage [3, 4, 5] and concerns about GDELT's data [6, 7], it is possible that event duplication and reliance on machine-coded data drive our results. Second, our models may be misspecified. In the second set of robustness checks, we therefore use a different estimator, control for Twitter accounts from news media, control for more time lags, and confirm that the results are not driven by local actors attracting international attention. In both sets of checks, coordination is still significantly associated with protest the following day.

Table 5 demonstrates that our main results are not driven by duplication. In Model 1 we rerun the full model from Table 4 using data from the Armed Conflict Location and Event Dataset (ACLED) [8] in lieu of the GDELT data. In addition to more traditional conflictual events such as battles, ACLED also codes for riots and protests. ACLED has events for Algeria, Egypt, Libya, Morocco, and Tunisia for 2010 and 2011, so we took all riots and protests (they are coded as one category) from those countries from November 1^{st}, 2010 through December 31^{st}, 2011. We chose ACLED because it provides greater event granularity than the other main events dataset which contains protests, the Social Conflict in Africa Dataset [9], and is hand-coded so as to avoid duplication of events. Figure 1 shows the correlation between GDELT's record of protest and ACLED's.

In Model 2 of Table 5 we change the dependent variable to $\textit{Protest Rate}_{i,t}$, which is the number of protests on a country-day divided by all the GDELT events in that country that day. If the GDELT algorithm duplicates protests, it should duplicate all of its event categories; dividing protests by events therefore controls for this duplication. Now that the dependent variable ranges from [0,1], a log transform is performed to normalize the data, and ordinary least squares is used in the resulting models. Model 3 uses the coding of Lotan. et al to identify Twitter accounts in Tunisia and Egypt belonging to media organizations, employees of media organizations, digital activists, and spammers [10].

Figure 1: Correlation of GDELT and ACLED Protests



These accounts are heavy users of hashtags and frequent tweeters but they do not reflect how normal people use social media, so we confirm that they are not the accounts coordinating events or creating spurious correlation. Model 4 is same as Model 1 except it ignores any country whose protests are in the top quartile for that country. The results show that high-protest days do not drive our statistical power: small-scale protests increase as coordination increases, not just as coordination for major protests occurs.

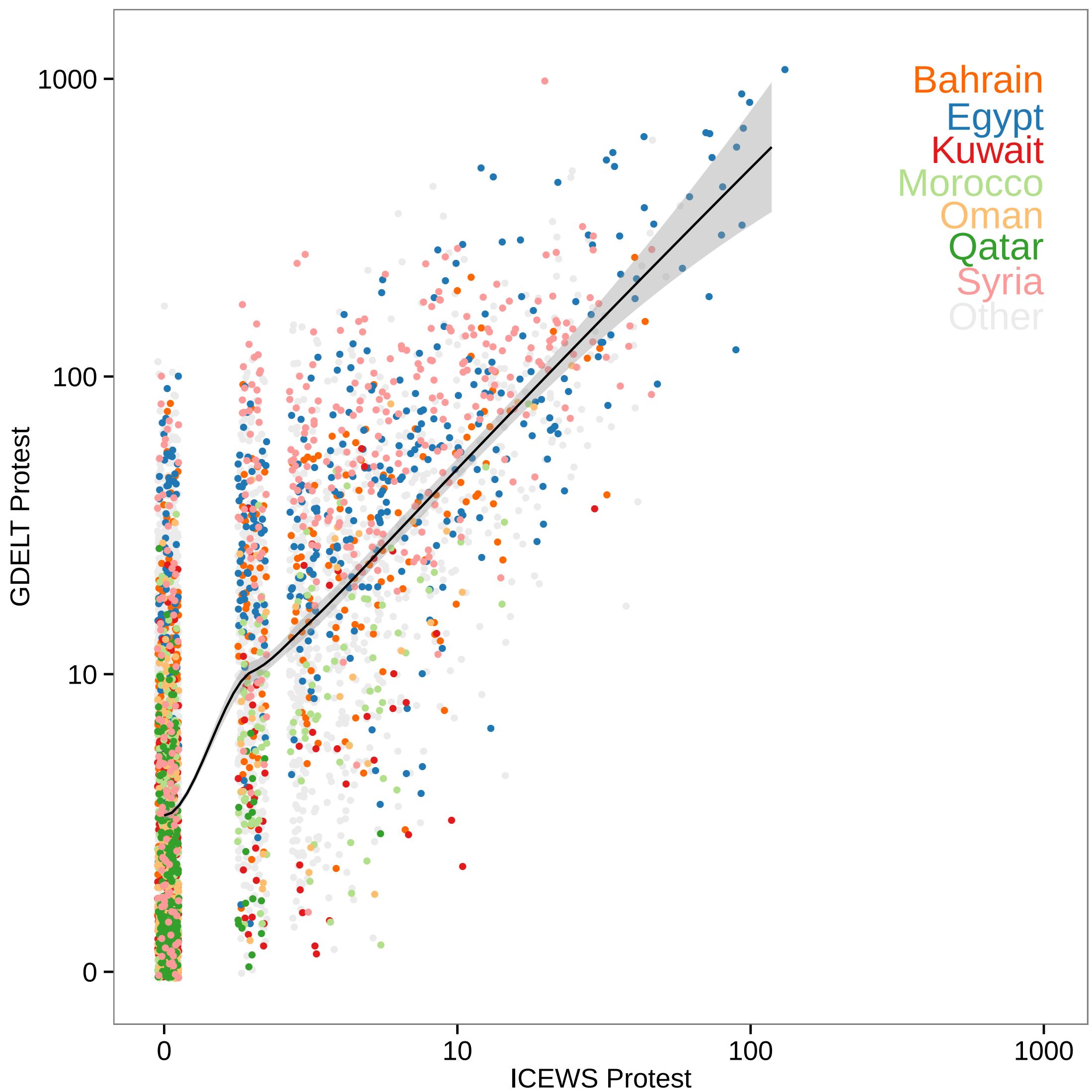
Model 1 of Table 6 replicates the full model from Table 1 but with an ordinary least squares estimator. Models 2-4 revert back to the negative binomial to test different specifications. To check if coordination occurs multiple days before an event and whether we simply capture generic increases in use of a specific hashtag, we include lags up to 3 days; the results in Model 2 show that the main effects do not change. In Model 3, we control for any tweet that has 3 or more hashtags because those are often spam [11]. That model also controls for the production of users who are in the top 5% of the followers' distribution. Since the measure of coordination continues to be strongly associated with protest, it suggests that coordination is a diffuse, self-organizing process that is not controlled by a few central individuals.

Model 4 demonstrates that these results are not driven by elites (English speakers). Much of the research on social media and protest suggests that sites like Twitter are used to attract international attention by spreading information outside of users' countries [10, 12, 13]. Since our data contain English and Arabic tweets and few individuals in these countries will acquire their information through English sources, it is possible that our results are driven by international attention-seeking and not domestic common knowledge and influence. To rule out this possibility, we replicate our main model using only tweets that are in Arabic. (We used a Python implementation of Google's Compact Language Detector to identify each tweet's language.) The results suggest that English speakers are not driving the relationship between coordination and protest the following day.

We also compare GDELT to a different machine-coded dataset. The Integrated

Conflict Early Warning System (ICEWS) is a Lockheed Martin project that uses GDELT's coding scheme with its own software that parses more newspapers, including foreign language ones, than GDELT; Figure 2 shows that GDELT and ICEWS correlate quite closely. The high levels of correlation provide strong evidence that our results are not driven by possible errors in the GDELT dataset.

Figure 2: Correlation of GDELT and ICEWS Protest



High-Coordination Hashtags Contain Words About Protest

Here we analyze the words that are used in conjunction with high-coordination hashtags to ascertain whether they are, indeed, about coordinating protest. We take 2 of the most common hashtags from Egypt and Bahrain and examine the 100 most common words that appear in tweets containing those hashtags. We separated English and Arabic tweets and show here the most common words for the Arabic tweets, since those are the ones most likely to matter for potential protestors in these countries.

Clear differences in language emerge. For Egypt, both sets of hashtags contain overlapping language. Words such as “mubarak”, “#egyelections”, “Allah”, and “security” occur in both. “#tahrir” contains words such as “#occupycabinet”, “rebels”, “#sep9”, “tear gas”, and “#tahrirtv” which reveal how that hashtag is often used for very particular events. On the other hand, words that are associated with “#egypt” include “#syria”, “#libya”, and “#copts” . These words are much more general than those for “#tahrir”, and reveal how that hashtag is used for more general events than “#tahrir”.

Bahrain shows similar separation in knowledge generation based on hashtag. “#lulu” contains words such as “Arrest”, “Roundabout”, “#humanrights”, “#notoexecution”, “Revolution”, and “#manama”. “#Bahrain” contains words such as “King”, “President”, “#syria”, “Security”, and “world”. Words common to both include “#gcc”, “Allah”, “regime”, “people”, “and youth”. Once again, while both hashtags have similar topics, they also differ in key ways, with “#lulu” overall having more focus on specific events and places while “#bahrain” contains more broadly focused information.

- **Egypt**
  - *Words Common to #egypt, #jan25, and #tahrir*
    - Arabic (52)

#egypt, #jan25, #tahrir, on, #25jan, egypt, #noscaf, #news, mubarak, after, ..., #scaf, revolution, tahrir, council, to, today, people, neither, allah, army, between, was, one, egyptian, president, minister, slave, what, other than, mohammed, because, via, #mubarak, police, square, before, now, #ikhwan, #salafi, honour, when, #jul8, the protesters, why, security, youth, until, #ontveg, #may27

    * English (62)

      #egypt, #jan25, #tahrir, time, will, egypt, now, egyptian, #cairo, protest, peopl, say, just, new, #scaf, one, can, get, via, revolut, #mubarak, like, need, day, mubarak, armi, militari, today, know, ..., make, #25jan, want, polic, think, good, pleas, check, love, call, live, see, still, #noscaf, tahrir, come, take, back, right, video, use, watch, start, support, happen, ppl, hope, squar, look, give, help, attack

- *Words Unique to #egypt*

    * Arabic (18)

      someone, #egynews, #ahram, i am, direct, network, #zamalek, #enn, #7urreya, in it, #6april, #syria, #cairo,#ahly, #libya, like that, #egyfootball, prayer

    * English (24)

      2011, current, city, pray, prayer, #egyptian, service, may, #libya, #egyelect, #egypt., elect, news, beta, country, #egypt:, wednesday, arab, sunday,, #israel, saturday,, follow, tuesday,, year

- *Words Unique to #jan25*

    * Arabic (11)

      #amndawla, january, there are, seventh, #dostor2011, arabian, he said, constitution, shafiq, ahmed, through

        * English (17)

          secure, sinc, regime, preside, talk, #amndawla, force, twitter, must, thug, said, freedom, thing, govern, #egyarmi, monist

    - *Words Unique to #tahrir*

        * Arabic (20)

          #tahrirtv, mahmoud, #nov18, we, #occupycabinet, inside, #tahrir, #alex, rebels, with, #fuckscaf, live, field marshal, #sep9, tear gas, who, #suez, street, down with, on him

        * English (27)

          #tahrir., gas, #jul8, tear, #tahrir,, chant, moham, #may27, mahmoud, number, supply, join, anyone, guy, around, plz, #occupycabinet, #july8, scar, tell, mani, now., fire, #tahrirne, got, tweet, show

- **Bahrain**

    - *Words Common to #bahrain, #feb14, and #lulu*

        * Arabic (48)

            · #bahrain, on, #14feb, bahrain, #feb14, #lulu, allah, #gcc, #kuwait, #ksa, today, to, #alwefaq, this, neither, people, before, #egypt, regime, agreement, sheikh, now, hamad, day, was, #bh, salman, until, mohammed, #uae, between, martyr, troops, channel, march, home, report, other than, picture, #q8, village, youth, bahraini, video, jacket, eisa, rajab, noble

        * English (48)

          protest, will, police, now, people, can, attack, ppl, just, right, get, one, call, want, like, see, say, day, know, today, use, arrest, love, need, video, via, think, make, new, back, support, live, still, come, stop, time, good, pleas, tri, ..., happen, start, take, head, watch, thank, look, hope

- *Words Unique to #bahrain*
  - Arabic (27)

    after, king, security, but, iran, basyuni, by, world, minister, president, only, #syria, men, all, one, because, someone, ministry, and from, not, and allah, interior, then, #jun1, inside it
  - English (16)

    make, good, #bahraindeb, country, ask, world, year, work, gas, never, must, king, look, even, old, hope
- *Words Unique to #feb14*
  - Arabic (14)

    mercenary, #news, #ajstream, #world, #bnn, #un..., #feb14..., near, #june1, cinder, hasan, street, high, square
  - English (21)

    #arabspr, #bhn, #eu, #bnn, secure, hospit, ali, death, crackdown, tortur, demonstr, tear, die, anoth, student, activist, medic, shot, brutal, shoot
- *Words Unique to #lulu*
  - Arabic (13)

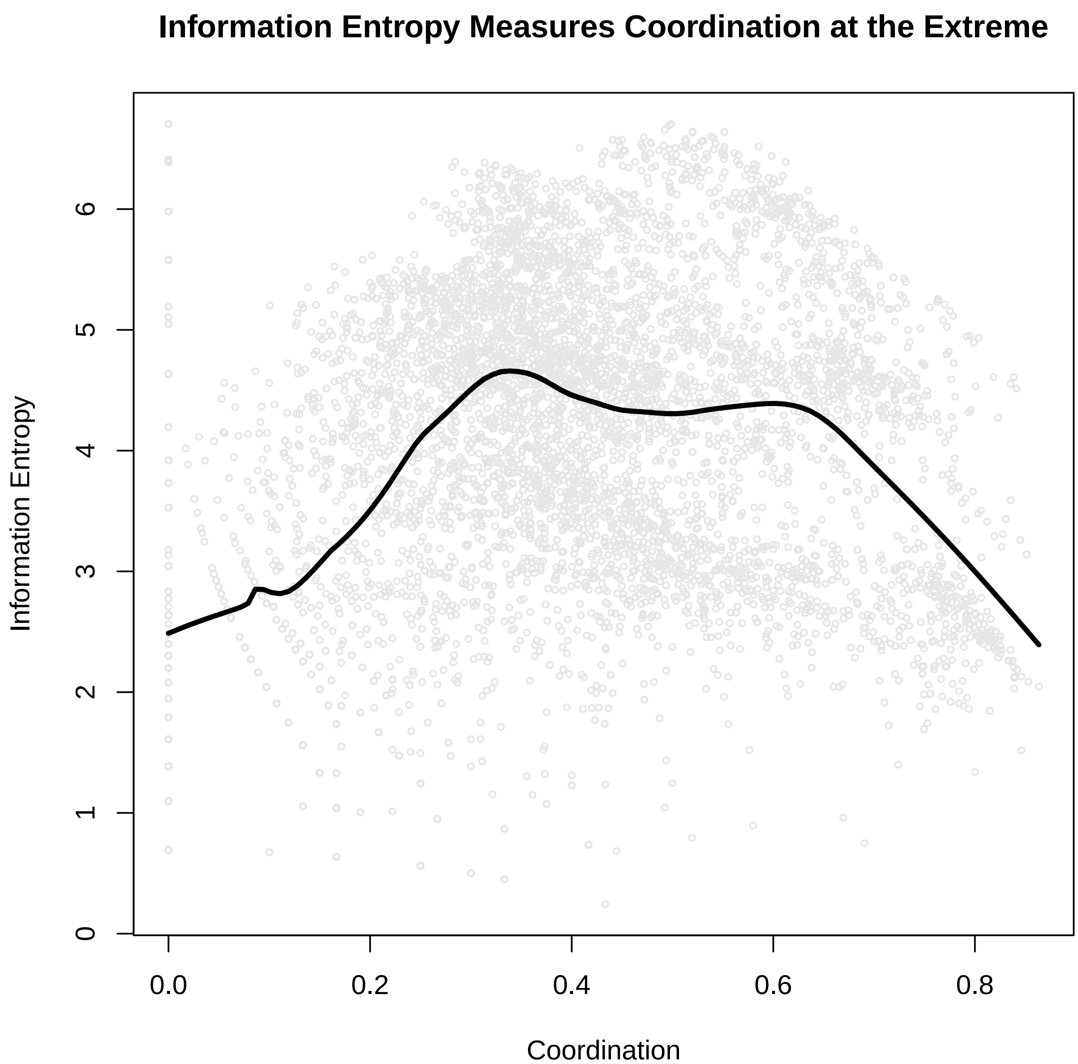
    #arab, village, #lulureturn, that, #usa, #manama, ..., #wefaq, #btv, #lulu..., pearl roundabout, #saudi
  - English (16)

    #lulureturn, roundabout, #freedom, thug, car, #arab, #ksa, tri, #kuwait, demand, confirm, #bh, #lulu,, pearl, thank

Each Day's Most Common Hashtag

Figure 3 shows the correlation between $Coordination_{i,t-1}$ and $Protest_{i,t}$ for Egypt, Qatar, Bahrain, and Morocco. Each dot is colored for the most common hashtag that day, with each country's top 4 hashtags receiving their own color and all other hashtags labeled "Other".

Figure 3: Information Entropy and Coordination



Tables

Table 1: Summary Statistics of Key Variables

Statistic	N	Mean	St. Dev.	Min	Max
Protests	6,816	17.626	45.226	0	994
Tweets	6,816	2,018	3,980	0	32,750
Coordination	6,816	0.242	0.183	0.000	0.790
Hashtag %	6,816	0.192	0.117	0.000	1.000
Retweet %	6,816	0.061	0.061	0.000	1.000
Link %	6,816	0.322	0.192	0.000	1.000
Mention %	6,816	0.388	0.144	0.000	1.000

Table 2: Correlation of Variables

	Protest$_{i,t}$	Coordination$_{i,t-1}$	Hashtag %$_{i,t-1}$	Retweet %$_{i,t-1}$	Link %$_{i,t-1}$	Mention %$_{i,t-1}$	Protest$_{i,t-1}$
Protest$_{i,t}$	-	0.41	0.31	-0.01	0.14	-0.17	0.86
Coordination$_{i,t-1}$		-	0.63	-0.04	0.14	-0.19	0.42
Hashtag %$_{i,t-1}$			-	0.20	0.43	-0.36	0.31
Retweet %$_{i,t-1}$				-	0.11	-0.21	-0.01
Link %$_{i,t-1}$					-	-0.57	0.13
Mention %$_{i,t-1}$						-	-0.16
Protest$_{i,t-1}$							-

Table 3: Main Variables by Country

Country	Protests	Tweets	Coordination	Hashtag %	Retweet %	Link %	Mention %	Population
Egypt	29,035	3,742,648	0.45	0.21	0.04	0.29	0.42	78,075,705
Syria	24,684	229,476	0.49	0.34	0.03	0.61	0.22	21,532,647
Yemen	12,977	61,517	0.28	0.25	0.05	0.61	0.19	22,763,008
Libya	11,146	84,991	0.25	0.24	0.11	0.34	0.32	6,040,612
Tunisia	9,974	228,554	0.24	0.25	0.07	0.43	0.40	10,549,100
Bahrain	7,313	1,056,990	0.41	0.22	0.06	0.16	0.42	1,251,513
Iraq	5,525	146,113	0.23	0.18	0.09	0.28	0.35	30,962,380
SaudiArabia	5,321	4,425,797	0.34	0.13	0.04	0.16	0.51	27,258,387
Jordan	3,830	273,227	0.26	0.21	0.06	0.45	0.36	6,046,000
Lebanon	3,212	522,891	0.23	0.21	0.06	0.28	0.40	4,341,092
Algeria	1,723	7,474	0.05	0.21	0.11	0.44	0.34	37,062,820
Morocco	1,722	300,454	0.15	0.18	0.07	0.34	0.43	31,642,360
Oman	1,243	8,509	0.01	0.09	0.05	0.31	0.48	2,802,768
Kuwait	1,053	29,838	0.05	0.09	0.02	0.13	0.52	2,991,580
UnitedArabEmirates	833	1,531,524	0.22	0.14	0.07	0.20	0.43	8,441,537
Qatar	546	1,104,995	0.21	0.12	0.06	0.10	0.43	1,749,713

Table 4: Negative Binomial Regression of Protest on Twitter Measures from Previous Day

	DV: $\text{Protests}_{i,t}$		
	Protest		
	(1)	(2)	(3)
$\text{Coordination}_{i,t-1}$	2.239***		2.307***
	(0.569)		(0.602)
Hashtag $\%_{i,t-1}$		1.898***	0.194
		(0.686)	(0.540)
Retweet $\%_{i,t-1}$		0.291	0.435
		(0.681)	(0.686)
Link $\%_{i,t-1}$		0.475	0.287
		(0.330)	(0.264)
Mentions $\%_{i,t-1}$		−0.020	0.003
		(0.107)	(0.103)
$Protests_{i,t-1}$	0.011***	0.013***	0.011***
	(0.001)	(0.001)	(0.001)
Intercept	0.126	−0.660*	−0.078
	(0.257)	(0.393)	(0.298)
Day FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Observations	6,800	6,620	6,620
Log Likelihood	−20,054.550	−19,708.050	−19,628.450
AIC	40,993.100	40,286.100	40,128.900
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5: Duplication, Bias, Media, and Extreme Events do not Drive Results

	$Protest_{i,t}$	$log(Protest\ Rate_{i,t})$	$Protest_{i,t}$	$Protest_{i,t}$
	Neg. Bin	*OLS*	*Neg. Bin*	*Neg. Bin*
	ACLED	*GDELT*	*GDELT*	*GDELT*
	(1)	(2)	(3)	(4)
$Coordination_{i,t-1}$	4.206^{***}	1.344^{***}	1.323^{***}	1.372^{**}
	(0.964)	(0.480)	(0.019)	(0.550)
$Hashtag\ \%_{i,t-1}$	0.575	0.445	1.009^{***}	0.657
	(1.602)	(0.339)	(0.276)	(0.640)
$Retweet\ \%_{i,t-1}$	-0.663	-0.674	3.123^{***}	-0.337
	(1.716)	(0.521)	(0.223)	(0.813)
$Link\ \%_{i,t-1}$	1.313	0.468^{***}	-3.884^{***}	0.651^{*}
	(1.040)	(0.163)	(0.098)	(0.376)
$Mentions\ \%_{i,t-1}$	0.709	0.126	-0.523^{***}	0.160
	(0.594)	(0.099)	(0.129)	(0.117)
$Protests_{i,t-1}$	0.206^{***}		0.002^{***}	0.015^{***}
	(0.033)		(0.0001)	(0.002)
$Protest\ Rate_{i,t-1}$		13.648^{***}		
		(1.097)		
$Media\ Org.\ \%_{i,t-1}$			21.950^{***}	
			(0.830)	
$Media\ Indiv.\ \%_{i,t-1}$			17.462^{*}	
			(10.400)	
$Digital\ Activist\ \%_{i,t-1}$			-5.232^{***}	
			(0.088)	
$Spam\ \%_{i,t-1}$			7.448^{***}	
			(0.876)	
Intercept	-24.796	-6.451^{***}	3.343^{***}	-1.706^{***}
		(0.311)	(1.008)	(0.406)
Day FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Observations	2,069	6,617	830	4,849
R^2		0.515		
Adjusted R^2		0.481		
Log Likelihood	-983.332		$-2{,}977.239$	$-10{,}334.300$
AIC	2,816.664		6,806.478	21,540.600

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 6: Robust to Model Misspecification, Popular Users, and Elites

	$Protest_{i,t}$			
	OLS	*Neg. Bin*		
	(1)	(2)	(3)	(4)
$Coordination_{i,t-1}$	19.008***	1.407***	2.231***	2.336***
	(7.147)	(0.414)	(0.628)	(0.519)
$Coordination_{i,t-2}$		1.030**		
		(0.455)		
$Coordination_{i,t-3}$		0.425*		
		(0.242)		
Hashtag $\%_{i,t-1}$	1.651	0.032	−0.222	0.101
	(3.572)	(0.380)	(0.475)	(0.431)
Hashtag $\%_{i,t-2}$		−0.155		
		(0.514)		
Hashtag $\%_{i,t-3}$		0.075		
		(0.429)		
Retweet $\%_{i,t-1}$	0.932	0.252	0.262	0.868***
	(5.223)	(0.510)	(0.598)	(0.280)
Retweet $\%_{i,t-2}$		0.244		
		(0.645)		
Retweet $\%_{i,t-3}$		0.464		
		(0.299)		
Link $\%_{i,t-1}$	−7.120**	0.004	0.152	0.347
	(3.021)	(0.250)	(0.311)	(0.283)
Link $\%_{i,t-2}$		0.247*		
		(0.138)		
Link $\%_{i,t-3}$		0.322		
		(0.234)		
Mentions $\%_{i,t-1}$	−1.595	0.007	−0.068	0.016
	(1.332)	(0.125)	(0.137)	(0.133)
Mentions $\%_{i,t-2}$		0.225**		
		(0.109)		
Mentions $\%_{i,t-3}$				
Spam $\%_{i,t-1}$			−1.477	
			(1.074)	
Top 5% User$_{i,t-1}$			0.584	
			(0.414)	
Top 5% User Hashtag $\%_{i,t-1}$			0.629	
			(0.982)	
$Protests_{i,t-1}$	0.800***	0.011***	0.011***	0.011***
	(0.042)	(0.001)	(0.001)	(0.001)
$Protests_{i,t-2}$		−0.001		
		(0.0005)		
$Protests_{i,t-3}$		0.001		
		(0.001)		
Intercept	4.478**	−1.227***	−0.223	0.487
	(2.271)	(0.427)	(0.288)	(0.384)
Day FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Observations	6,620	6,468	6,620	6,065
R^2	0.784			
Adjusted R^2	0.769			
Log Likelihood		−19,262.320	−19,611.950	−18,409.000
AIC		39,402.630	40,101.900	37,687.990
Note:	*p<0.1; **p<0.05; ***p<0.01			

References

[1] Leetaru, K. & Schrodt, P. GDELT: Global Data on Events, Language, and Tone, 1979-2012 (2013).

[2] Gerner, D. J., Schrodt, P. A., Abu-Jabr, R. & Yilmaz, O. Conflict and Mediation Event Observations (CAMEO): A New Event Data Framework for the Analysis of Foreign Policy Interactions. In *Annual Meeting of the International Studies Association* (2002).

[3] Davenport, C. & Ball, P. Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995. *Journal of Conflict Resolution* **46**, 427–450 (2002).

[4] Herkenrath, M. & Knoll, A. Protest events in international press coverage: An empirical critique of cross-national conflict databases. *International Journal of Comparative Sociology* **52**, 163–180 (2011).

[5] Eck, K. In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation and Conflict* **47**, 124–141 (2012).

[6] Caren, N. It is time to get rid of the E in GDELT (2014). URL `http://badhessian.org/2014/05/it-is-time-to-get-rid-of-the-e-in-gdelt/`.

[7] Hammond, J. & Weidmann, N. B. Using machine-coded event data for the micro-level study of political violence. *Research & Politics* **1**, 1–8 (2014).

[8] Raleigh, C., Linke, A., Hegre, H. & Karlsen, J. Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature. *Journal of Peace Research* **47**, 651–660 (2010).

[9] Hendrix, C. *et al.* Social Conflict in Africa: A New Database. *International Interactions* **38**, 503–511 (2012).

[10] Lotan, G. *et al.* The Revolutions Were Tweeted: Information Flows During the 2011 Tunisian and Egyptian Revolutions Web. *International Journal of Communications* **5**, 1375–1406 (2011).

[11] Kwak, H., Lee, C., Park, H. & Moon, S. What is Twitter, a Social Network or a News Media? In *International World Wide Conference*, 591–600 (2010).

[12] Aday, S., Freelon, D., Farrell, H., Lynch, M. & Sides, J. New Media and Conflict After the Arab Spring. Tech. Rep., United States Institute of Peace, Washington D.C. (2012).

[13] Starbird, K. & Palen, L. (How) Will the Revolution be Retweeted? Information Diffusion and the 2011 Egyptian Uprising. In *CSCW* (Seattle, 2012).