



# Online social networks and offline protest

Zachary C Steinert-Threlkeld<sup>1\*</sup>, Delia Mocanu<sup>2</sup>, Alessandro Vespignani<sup>2</sup> and James Fowler<sup>1,3</sup>

\*Correspondence:  
zsteiner@ucsd.edu

<sup>1</sup>Department of Political Science,  
University of California - San Diego,  
9500 Gilman Drive #0521, San  
Diego, CA 92093, USA

Full list of author information is  
available at the end of the article

## Abstract

Large-scale protests occur frequently and sometimes overthrow entire political systems. Meanwhile, online social networks have become an increasingly common component of people's lives. We present a large-scale longitudinal study that connects online social media behaviors to offline protest. Using almost 14 million geolocated tweets and data on protests from 16 countries during the Arab Spring, we show that increased coordination of messages on Twitter using specific hashtags is associated with increased protests the following day. The results also show that traditional actors like the media and elites are not driving the results. These results indicate social media activity correlates with subsequent large-scale decentralized coordination of protests, with important implications for the future balance of power between citizens and their states.

**Keywords:** social media; Twitter; protests; Arab Spring; collective action

## 1 Introduction

In the last two decades, large-scale protests have increased in frequency; these protests often lead to mass casualties, change countries' political systems, and sometimes lead to war, as is the case in Iran in 1979 or Libya and Syria in 2011. Here we present results from a large-scale longitudinal study that connects these two worlds, showing that coordination on social networks correlates with protest offline. Online social networks have become an increasingly common component of people's lives and are used to predict box office returns [1], stock market fluctuations [2], social collective phenomena [3–5], and even private traits such as sexual orientation and political ideology [6]. They can also be harnessed for crowd sourced searching [7] and get-out-the-vote campaigns [8]. Using 13.8 million geolocated tweets and machine-coded data on protests from 16 countries during the Arab Spring, we show that coordination of messages on Twitter is associated with increased protests the following day. The large-scale, systematic study we present here provides extensive evidence that social media may help decentralized groups coordinate online to organize protests offline. These results suggest that individuals in a wide range of countries use social media to organize large-scale protests and not only at the height of protest events; coordination is a perpetual, systematic activity. The results therefore extends existing work which focuses on protest events in one country [9–12].

The Arab Spring protests started in Tunisia on December 14th, 2010, and eventually led to the resignation of that country's president on January 14th, 2011. They spread to neighboring countries over the following weeks, inspiring massive turnout in Egypt that caused President Hosni Mubarak to resign on February 11th, 2011. In the aftermath, many

other countries experienced protest, with some, such as Libya and Syria, quickly turning into civil wars.

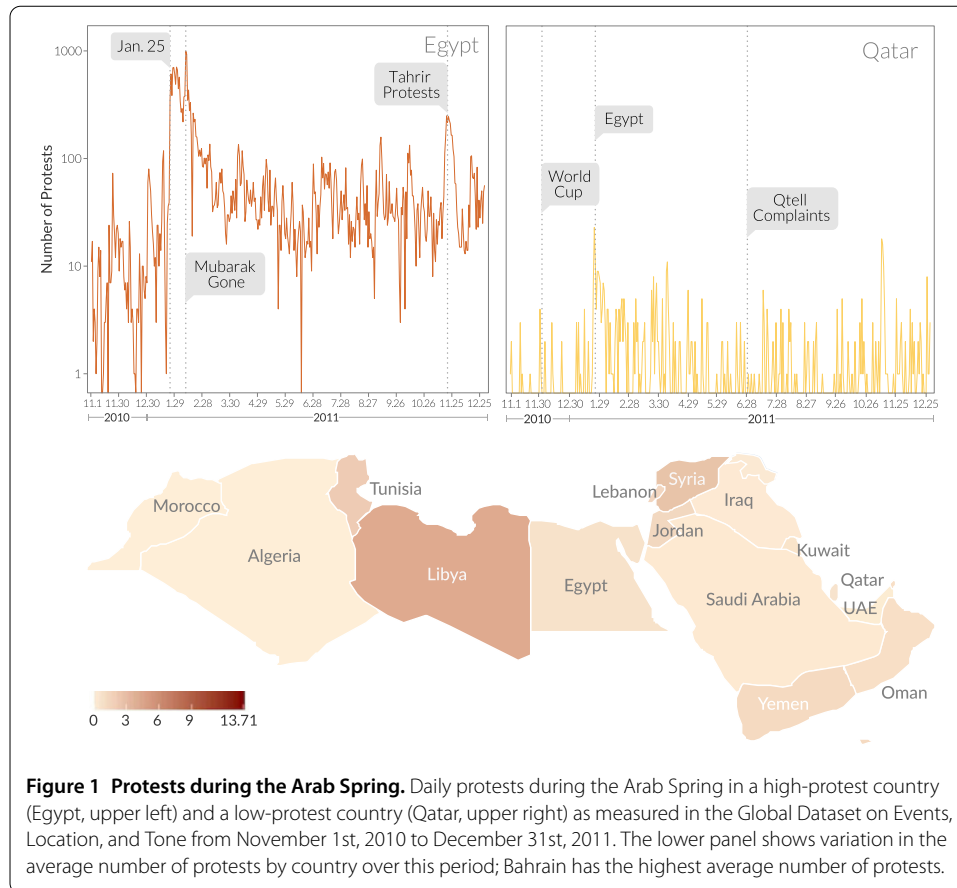
Accounts of the Arab Spring have frequently credited Twitter, Facebook, and other social network sites with helping the protests self-organize [13]. Indeed, there are many reasons to expect that social media did play a role. When deciding whether or not to protest, individuals have to estimate the risk of arrest or violent state repression, and they have to weigh that cost against the potential benefits of any change in policy that may result. They are primarily interested in how many other individuals are going to protest and whether those participants are first-time protesters or not [14]. Protesting *en masse* decreases the chance an individual will personally experience state violence and increases the probability of achieving policy change. Individuals wanting to protest are therefore strongly incentivized to coordinate their protest action with others.

## 2 Data

Social media can make it easier to protest because it lowers the barriers required to coordinate, making it easier to know if others will protest and whether or not they are habitual protesters. Joining a social media platform requires many fewer resources than establishing a newspaper, running a television station, or opening a civil society organization [15], meaning many more people can produce information seen by others. For example, tweets can be composed and read using a basic mobile phone, and blogs only require access to an internet connection. Social media also facilitates connections between people who otherwise would not come into contact [16], greatly increasing the number of people that know about an event. While governments can also use social media to monitor and repress their citizens [17, 18], the lower barriers to entry and broadcast-like features of social media give individuals more power to coordinate than they have without online social networks.

Here, we quantitatively test the hypothesis that social media usage correlates with subsequent protests, using longitudinal data from the Arab Spring. Though individuals use many online networks to organize protest, we focus on the social media site Twitter for three reasons. First, it has become a tool for citizens to gather and disseminate information in information-scarce environments such as authoritarian regimes. It therefore is a critical component of many protest movements, starting with the 2009 Iran election protests and continuing through the Ukraine civil war. Second, it provides some of the best temporal resolution of any data source. It is therefore one of the few sources available to researchers interested in quickly-changing processes such as protests. Third, state actors belatedly realized the power of social media, making social media an attractive tool for anyone seeking independent information; the content contained in social media therefore more closely reflected the offline world than did official news sources [19].

To determine that social media were used to coordinate protest, we measure the daily number of protests across 16 countries in the Middle East and North Africa from November 1st, 2010 through December 31st, 2011. These data come from a publicly-available machine-coded events dataset, the Global Database of Events, Location, and Tone (GDELT). GDELT machine-reads American and foreign newspapers and extracts the events each article is about [20]. Figure 1 shows the daily measure of protests in a high-protest country (Egypt) and low-protest one (Qatar) as well as the average level of protest per country in our study. In the Supplementary Materials (Additional file 1), we show that this measure correlates with hand-coded datasets of protest; others have shown that it correlates with other machine-coded data [21].



To measure social media behavior over the same period, we collected almost 14 million geolocated tweets that originated in these countries [22]. Unlike previous studies [23, 24], we did not select tweets based on protest-related topic words. This gives us a representative sample of online conversations that includes tweets that were not about protests or surrounding events in these countries (see the Supplementary Materials for a discussion of the advantages of this approach). Our inferences therefore do not risk being caused by selecting on the dependent variable. Overall, 1.9% of tweets in our sample have GPS coordinates, the rest have a user reported location.

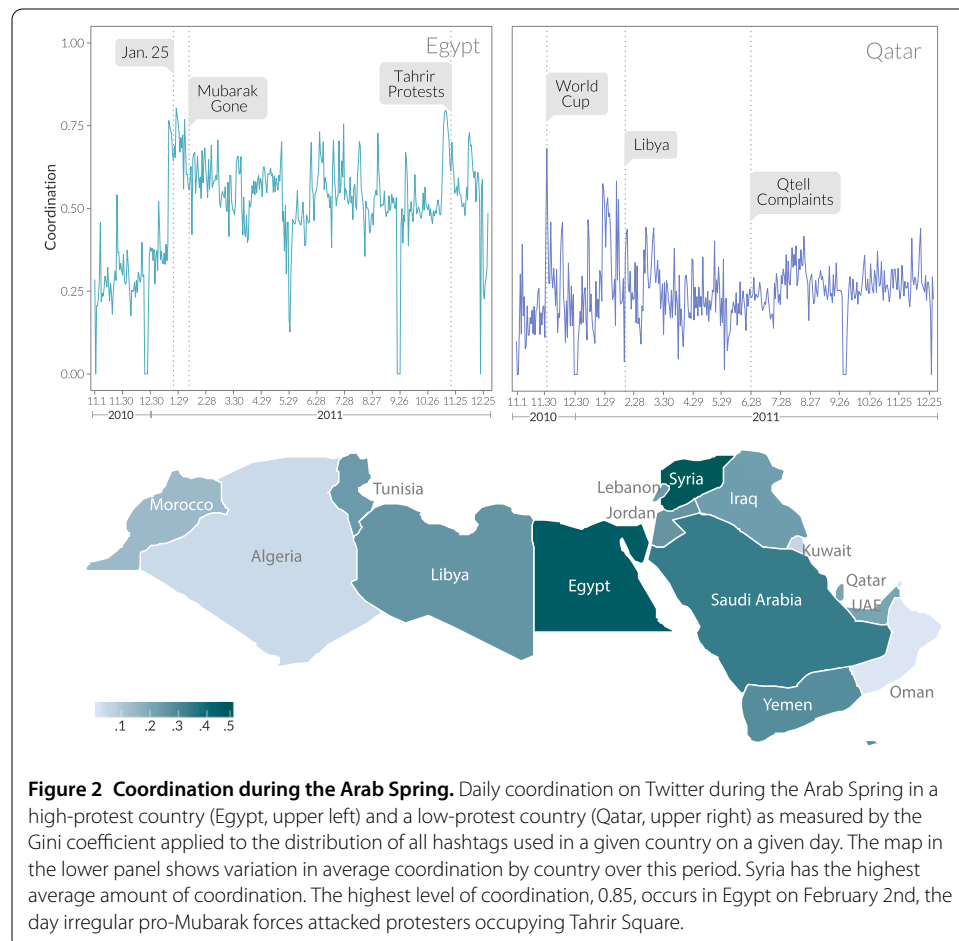
We identified topics using hashtags, self-categorized topic words preceded by the # sign. Hashtags associate the message that contains them with a larger discourse. For example, the tweet ‘You guys! We’re about to head to a meeting point in front of Marie Louis store in batal ahmed street #jan25’ contains the ‘#jan25’ hashtag, a common hashtag used to talk about protests in Egypt, since January 25th was the first day of major protests in that country. Anyone can search Twitter for a hashtag, and Twitter will return all tweets containing that hashtag; one can also click a hashtag seen in a tweet and instantaneously see the tweets using that hashtag. One does not need an account with Twitter to see tweets or hashtags.

Users quickly coordinate on a few hashtags to use for an event, whether that event is a protest, sporting event, or meme [25]. As a result, the set of unique hashtags people use may decline when hashtags are being used most frequently, even if the total number of tweets with hashtags increases. Converging to a few hashtags and using them intensively

is what we call coordination. We call this coordination because individuals are more likely to protest when they know many others will protest, and using a few hashtags repeatedly signals that there exists a latent demand for protest [26]. As the communication about upcoming protest occupies more of a country's total communication, individuals should have higher confidence that others will protest.

We measure the extent of this coordination with the Gini coefficient, which indicates the amount of inequality in a distribution. This coefficient ranges from a value of 0, which indicates complete equality (no coordination: all hashtags are used the same number of times), to 1, which indicates complete inequality (perfect coordination: everyone uses a single hashtag and no other hashtags are used). A high hashtag Gini coefficient therefore means that individuals are coordinating on the event that that hashtag represents. We do not use the percent of tweets containing a hashtag because that measure is agnostic to the hashtag(s) used in a tweet, so a day with a high percentage of hashtagged tweets may have little coordination. We also do not use an information entropy measure because low entropy could correspond to a day where all hashtags are the same or a day where no hashtags are used; this point is explored more in the Supplementary Materials.

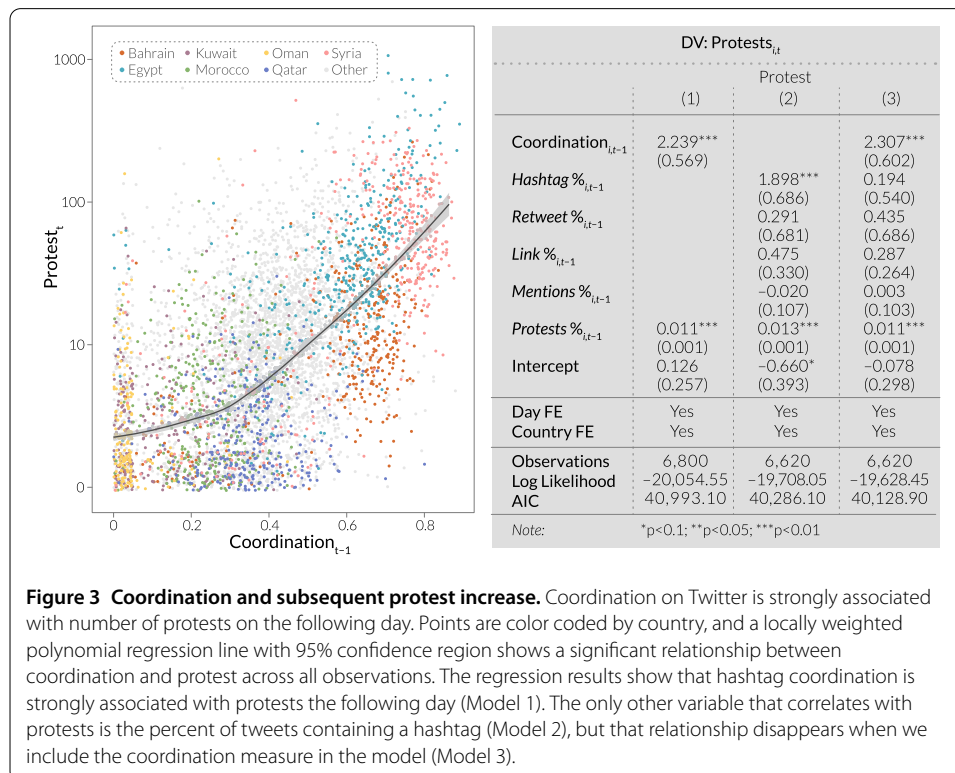
The measure of coordination captures when people on Twitter coordinate protests. Figure 2 shows how the coordination measure changes over time in Egypt and Qatar as well as the average level of coordination in the 16 countries. The correspondence between coordination and events is noticeable, and countries with more protest also had higher av-



erage levels of coordination. Egypt’s average level of coordination is 0.59, Syria’s 0.6. At the other extreme, Kuwait’s level of coordination is 0.09, Oman’s 0.02; the lowest income Gini is Sweden’s, at 0.25.

### 3 Results

Figure 3 reveals a robust statistical relationship between yesterday’s level of coordination and today’s number of protests. To reach this conclusion, we use a negative binomial regression model where the dependent variable is the count of protests at time  $t$  and the primary independent variables is the coordination at time  $t - 1$ . The coordination measure has a  $p$ -value less than 0.01 (coefficient of 2.239 and standard error of 0.569), and a one standard deviation increase in the coordination measure is associated with a 25.4% increase in the number of protests on the following day. Figure 3 also shows that the result holds when a number of potential other methods of coordination are modeled. The percent of tweets that contain a hashtag indicates the extent to which people are contributing to a tagged conversation, in some cases about protests. When a high percent of tweets are retweeted, a smaller number of original tweets drive the conversation. The percent of tweets that contain links indicates that more people are referencing an important blog or news item. And the percent of tweets that mention other users indicates that more direct communication is happening between people on Twitter. Retweets and hashtags are especially conducive to spreading information that may have a coordinating effect. None of these measures correlate as strongly with protests on the following day as coordination, and none of them are significantly associated with protests after controlling for coordination. A more detailed presentation of the model and controls is reported in the Supplementary Materials.

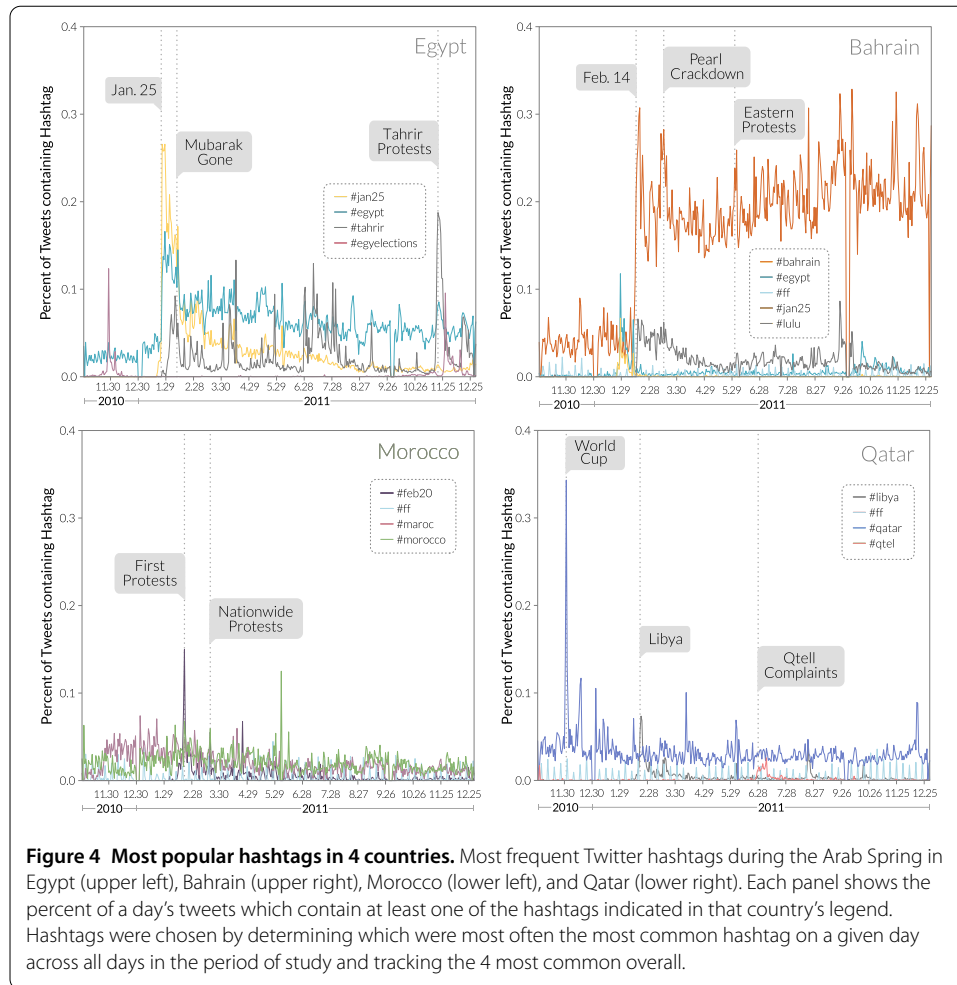


We also show in the Supplementary Materials that these results are robust to model type (linear vs. negative binomial regression), protest data source (a handcoded dataset and a different machine-coded one deliver the same results), and serial correlation in errors (multiple tests suggest that including a lagged dependent variable is sufficient to deal with the problem, and including several extra lags does not change the results). All models account for unobserved between-country differences with country fixed effects (accounting for stable characteristics like country size and political history), and they control for unobserved sources of variation over time with day fixed effects (helping to control for day-of-week variation in Twitter activity or special events like Ramadan that affect all countries in the study). Moreover, a model that drops high protest days also shows a significant relationship, suggesting that the association is not driven by a few large events. Finally, to ensure that the results are not driven by individuals trying to draw international attention to the protests [23], we drop all English tweets; the results remain the same.

An important question about the potential effect of social media coordination is its source - is it decentralized or does it come from traditional sources like news media and political activists? In the Supplementary Materials, we add to our models a number of measures of Twitter activity from these traditional sources [27], including the percent of tweets on a given day that were sent by media organizations, media personalities, digital activists, and the top 5% most active Twitter users. However, these models with controls continue to show a similar association with hashtag coordination, with an effect for media organizations and no effect for digital activists. Coordination therefore appears to come through decentralized activity of many individuals, not the frequent tweeting of a few specialized actors.

Figure 4 shows the prevalence of the hashtags in Bahrain, Egypt, Morocco, and Qatar most often used for coordination during the study period. They were chosen by observing the most common hashtag in a country on a given day, counting how many times that hashtag was the most common during the study period, and keeping the 4 most common. In Egypt, three features stand out. First, there is little coordination on any one hashtag before January 25th. The hashtag '#egypt' is barely used, the first appearance of '#jan25' is not until January 19th (and then accounts for only 0.1% of tweets), and '#tahrir' does not appear until January 25th (and it does not appear in large numbers until just before the resignation of Mubarak). Second, which hashtag is most prevalent depends on the type of upcoming event. During the 18 days of initial protest, '#jan25' dominates, as this was the focal date for the protests. Though the largest protests while Mubarak was in power took place in Tahrir Square, they were contentious, which is why more general hashtags such as '#jan25' and '#egypt' dominate. Moreover, '#jan25' consistently declines in usage after the 18 days of protest. By the middle of March, it will never be the most common of the three hashtags again, and it ceases to correlate with the other two. Third, overall levels of coordination decrease after the first 18 days. They first decrease sharply after President Mubarak's resignation, and their average prevalence gradually continues to decline. '#tahrir' might be an exception, as it is used much more narrowly to coordinate specific, Tahrir Square-centric events, but the frequency of its usage declines as well.

In Bahrain, there is a dramatic spike in coordination almost immediately after the start of protests on February 14th. This coordination is then sustained throughout the year, with increases and spikes around important events. We include in Figure 4 the prevalence of the hashtag #lulu - though it is never the most common hashtag, it is the one used to



discuss events concerning the Pearl Roundabout, the main focus of protestor activity in Bahrain. #lulu is not in our dataset before February 14th, and its use varies over time in more specific patterns than #Bahrain. Finally, note the focus on events in Egypt at the end of the Egyptian protests and starting again in mid-October.

Morocco experienced much less protest, as shown in Figure 1, and Figure 4 shows they also experienced less coordination. The initial protests in Morocco occurred on February 20th, when coordination peaked, but otherwise the level of coordination is not on the same scale seen in more contentious countries. Note also that Moroccans did not appear to take as much interest in events occurring elsewhere in the Middle East and North Africa, suggesting that they may have felt less affinity towards protestors in other countries. Similarly, Qatar exhibits low levels of coordination and no attempts at organizing protests. The day with the highest level of hashtag coordination is December 2nd, 2010, when Qatar learned it was chosen to host the 2022 World Cup. However, people in Qatar paid close attention to the events in #Libya, the third most common hashtag (and, not shown, hashtags about the Egypt protests ranked 6th and 7th).

#### 4 Discussion

We have used a large dataset that covers thousands of protests in many countries to reveal a relationship between people coordinating on social media and increased protests the

following day. Past work suggests that individuals are more likely to protest when their close friends or neighbors are protesting [28] and when they have prior experience with protest [29], but the results here suggest that weak ties can also facilitate mobilization [16]. They do so by exposing individuals to information about participation from outside of their local, strong-tie social network, allowing those who are on the fence about protesting to approximate how widespread, and therefore safe, the protests are. In this sense, social media helps people not only to learn about protests, but also to see that others are learning about them, aiding in coordination [30].

These results do not mean that there will always exist a relationship between coordination and subsequent events. For example, in democracies, where the cost of protesting is lower, individuals may not need to coordinate with each other to protest. Or in countries with free media, coordinating may be more likely to come from media sources than individuals. Future work should investigate these boundaries.

A growing literature suggests that new technology lowers violence [31] and decreases a state's control of information [15]. But as states learn how to control information and communication technology [32], they have also learned how to use it to track individual participants in protest [18] and activities the state has defined to be illegal [33]. By making coordination visible to a broad audience, activists can facilitate protest, but they also make their coordination visible to the state, enabling targeted repression. Thus, online social networks may help individuals protest, but they also help states repress protesting individuals. Future research should try to discern which of these effects prevails.

## Additional material

[Additional file 1: Supplementary Materials \(pdf\)](#)

## Competing interests

The authors declare that they have no competing financial interests.

## Authors' contributions

AV and DM collected the Twitter data; ZST collected the data on protests. DM and ZST cleaned the Twitter data. All authors contributed to model design and testing. ZST drafted the initial report, and all authors edited it. ZST and AV created the figures.

## Author details

<sup>1</sup>Department of Political Science, University of California - San Diego, 9500 Gilman Drive #0521, San Diego, CA 92093, USA.

<sup>2</sup>Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA 02115, USA.

<sup>3</sup>Department of Medicine, University of California - San Diego, 9500 Gilman Drive #0521, San Diego, CA 92093, USA.

## Acknowledgements

AV acknowledges the support by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (Dol/NBC) contract number D12PC00285. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, Dol/NBE, or the United States Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Received: 20 August 2015 Accepted: 28 October 2015 Published online: 09 November 2015

## References

1. Asur S, Huberman BA (2010) Predicting the future with social media. In: 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, pp 492-499
2. Curme C, Preis T, Stanley HE, Moat HS (2014) Quantifying the semantics of search behavior before stock market moves. *Proc Natl Acad Sci USA* 111:11600-11605
3. Xu J, Lu T-C, Compton R, Allen D (2014) Civil unrest prediction: a Tumblr-based exploration. In: Kennedy W, Agarwal N, Yang S (eds) *Social computing, behavioral-cultural modeling and prediction*. Lecture notes in computer science, vol 8393. Springer, Berlin, pp 403-411. doi:10.1007/978-3-319-05579-4\_49



4. Ramakrishnan N, Butler P, Muthiah S, Self N, Khandpur R, Saraf P, Wang W, Cadena J, Vullikanti A, Korkmaz G, Kuhlman C, Marathe A, Zhao L, Hua T, Chen F, Lu CT, Huang B, Srinivasan A, Trinh K, Getoor L, Katz G, Doyle A, Ackermann C, Zavorin I, Ford J, Summers K, Fayed Y, Arredondo J, Gupta D, Mares D (2014) "Beating the news" with EMBERS: forecasting civil unrest using open source indicators. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '14. ACM, New York, pp 1799-1808. doi:10.1145/2623330.2623373
5. Boecking B, Hall M, Schneider J (2015) Event prediction with learning algorithms - a study of events surrounding the Egyptian revolution of 2011 on the basis of micro blog data. *Policy Internet* 7:159-184
6. Kosinski M, Stillwell D, Graepel T (2013) Private traits and attributes are predictable from digital records of human behavior. *Proc Natl Acad Sci USA* 110:5802-5805
7. Pickard G, Pan W, Rahwan I, Cebrian M, Crane R, Madan A, Pentland A (2011) Time-critical social mobilization. *Science* 334:509-512
8. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489:295-298
9. Conover MD, Ferrara E, Menczer F, Flammini A (2013) The digital evolution of occupy Wall Street. *PLoS ONE* 8:e64679
10. Alvarez R, Garcia D, Moreno Y, Schweitzer F (2015) Sentiment cascades in the 15M movement. *EPJ Data Sci* 4:6. <http://www.epjdatascience.com/content/4/1/6>
11. Budak C, Watts D (2015) Dissecting the spirit of Gezi: influence vs. selection in the occupy Gezi movement. *Sociol Sci* 2:370-397. <https://www.sociologicalscience.com/articles-v2-18-370>
12. Gruzd A, Tsyganova K (2015) Information wars and online activism during the 2013/2014 crisis in Ukraine: examining the social structures of pro- and anti-Maidan groups. *Policy Internet* 7:121-158. <http://doi.wiley.com/10.1002/poi3.91>
13. Pollock J (2011) Streetbook: how Egyptians and Tunisian youth hacked the Arab Spring. *MIT Technol Rev*: 70
14. Lohmann S (1994) The dynamics of informational cascades: the Monday demonstrations in Leipzig, East Germany, 1989-91. *World Polit* 47:42-101
15. Edmond C (2013) Information manipulation, coordination, and regime change. *Rev Econ Stud* 80:1422-1458
16. Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78:1360-1380
17. Egorov G, Guriev S, Sonin K (2009) Why resource-poor dictators allow freer media: a theory and evidence from panel data. *Am Polit Sci Rev* 103:645-668
18. Bassiouni MC, Rodley N, Al-Awadhi B, Kirsch P, Arsanjani MH (2011) Report of the Bahrain Independent Commission of Inquiry. Technical report, Bahrain Independent Commission of Inquiry, Manama, Bahrain
19. Hamdy N, Gomaa EH (2012) Framing the Egyptian uprising in Arabic language newspapers and social media. *J Commun* 62:195-211
20. Leetaru K, Schrodt P (2013) GDELT: Global Data of Events, Language, and Tone, 1979-2012
21. Ward MD, Beger A, Cutler J, Dorff C, Radford B (2013) Comparing GDELT and ICEWS event data
22. Mocanu D, Baronchelli A, Perra N, Vespignani A, Goncalves B, Zhang Q (2013) The Twitter of Babel: mapping world languages through microblogging platforms. *PLoS ONE* 8:e61981
23. Aday S, Freelon D, Farrell H, Lynch M, Sides J (2012) New media and conflict after the Arab Spring. Technical report, United States Institute of Peace, Washington
24. Bruns A, Highfield T, Burgess J (2013) The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. *Am Behav Sci* 57:871-898
25. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C (2012) Dynamical classes of collective attention in Twitter. In: Proceedings of the 21st international conference on world wide web - WWW '12, pp 251-260
26. Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. In: International world wide web conference, pp 695-704
27. Lotan G, Ananny M, Gaffney D, Boyd D, Pearce I, Graeff E (2011) The revolutions were tweeted: information flows during the 2011 Tunisian and Egyptian revolutions web. *Int J Commun* 5:1375-1406
28. Gould RV (1991) Multiple networks and mobilization in the Paris Commune, 1871. *Am Sociol Rev* 56:716-729
29. McAdam D (1986) Recruitment to high-risk activism: the case of freedom summer. *Am J Sociol* 92:64-90
30. Chwe MS-Y (2000) Communication and coordination in social networks. *Rev Econ Stud* 67:1-16
31. Pierskalla JH, Hollenbach FM (2013) Technology and collective action: the effect of cell phone coverage on political violence in Africa. *Am Polit Sci Rev* 107:207-224
32. Kalathil S, Boas TC (2003) Open networks, closed regimes: the impact of the Internet on authoritarian rule. Carnegie Endowment for International Peace, Washington
33. Gerber MS (2014) Predicting crime using Twitter and kernel density estimation. *Decis Support Syst* 61:115-125

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---