

# Detecting Mass Protest through Social Media

Steven Lloyd Wilson

## *Abstract*

Building on the existing understanding of the dynamics of mass protest, this paper argues that social media data can be used to detect the occurrence of such protests. It outlines a theoretical framework arguing that during times of mass mobilization, network central actors will geographically converge upon city centers, and that the relative magnitudes of social media activity in the center and periphery of the state can be used to detect the occurrence of protests. This article presents a new dataset of 2.2 million geocoded Ukrainian tweets, which are used to empirically test the theory against the observed 2014 Euromaidan protests in Ukraine. By relying completely on count and network data rather than keywords, hashtags, or other contextual clues from the content, this technique is portable across language barriers and national borders.

---

*Dr. Steven Lloyd Wilson is an Assistant Professor in the Political Science Department at the University of Nevada at Reno. Correspondence can be directed to [stevenwilson@unr.edu](mailto:stevenwilson@unr.edu).*

**F**rom February 18 to 20 of 2014, Ukrainian citizens entered the streets of Kiev in numbers not seen since the Orange Revolution of 2005. The result was eerily similar: Viktor Yanukovich driven from the capital and a pro-western regime installed in his place. The revolution reflected the ubiquity of new communication technologies: streamed live in video, picture, and text over social media. As became evident during the Arab Spring, the Internet's spread of universal, instantaneous, and mobile mass communication is transforming the way that protest and mass social action occurs (Anderson, 2011; Farrell, 2012; Sabadello, 2012; Stepanova, 2011; Tucker, Barberá, & Metzger, 2013; Zhuo, Wellman, & Yu, 2011).

While such work has probed the use of new communication technologies for the organization of mass protest in the context of theories of collective action and social movements, little work has focused on effects in the opposite causal direction. That is, while we have analyzed the effects of social media on the development of mass protest, we have little understanding of how mass events in the real world affect activity on social networks. Thus this project's exploratory research question can be formulated: how does mass protest geospatially affect social media activity?

This exploration has the additional benefit of providing a new tool for the measurement of mass protest. Our capacity for measuring the occurrence of protest still remains grounded in manual efforts by those on the ground, such as reporters or nongovernment organizations (NGO) observers estimating how many people are in a particular capital square on a particular day. This leaves

enormous gaps in our measurement of protest because it relies on experts being in place on the ground as events are occurring, and upon the ability of free media to publish such reports. Data from Ukraine during the several months of waxing and waning protests in 2013 and 2014 are especially telling in this regard. News reports tended to report days of truly massive turnout (usually on Sundays) but had disjointed and inconsistent reports of turnout on other days. In addition, virtually all coverage was of Kiev, with only occasional reports indicating that protests were occurring in regional capitals as well during much of this period.

This article explores how social media geospatially reacts to the occurrence of mass protest, using a new, massive dataset of geocoded Twitter data originating in Ukraine before, during, and after the events of Euro-*maidan*. It finds strong evidence of mass protest in a state being associated with an increase in social media activity nationwide, an absolute drop in social media activity at the physical location of the protest, but an increase in the *network-weighted* social media activity at the protest's location.

### **The Internet, Social Media, and Mass Protest**

There are two basic schools of thought regarding the effect of the Internet's development upon politics. The first argues that the various technologies afforded by the Internet are at best irrelevant to politics in the grand scheme of things, and at worst have various negative effects such as homophily, diminished institution-building, and empowerment of authoritarian regimes (Faris, & Etling, 2008; Gentzkow, & Shapiro, 2011; Morozov, 2012;

Page, 2008; Sunstein, 2009; Wilson, 2016a; Wojcieszak & Mutz, 2009). The second school of thought argues that the Internet deeply changes political equilibria by virtue of greatly reducing the transaction costs that make collective action difficult to organize (Castells, 2009; Diamond, 2010; Shirky, 2009; Wilson, 2016a).

Social media are of particular interest to this debate given its widespread adoption over the last decade, and specific capacity for allowing individuals to communicate en masse. The vast bulk of the social science literature on the effect of social media on political mobilization thus far has focused on its influence upon the events of the Arab Spring (Alqudsi-ghabra, Al-bannai, & Al-bahrani, 2011; Hofheinz, 2005; Mackell, 2011; Murphy, 2006; Oghia & Indelicato, 2011; Sabadello, 2012; Stepanova, 2011; Zhuo et al., 2011) and Color Revolutions (Bunce & Wolchik, 2010; Chowdhury, 2008; Dyczok, 2005; Goldstein, 2007; Kyj, 2006).

There is also disagreement over the role that social media plays with regard to the development of mass protest in particular. Some scholars argue that social media are used to better solve the classic collective action problem by lowering the transaction costs of organization (Shirky, 2009). In addition, this literature argues that social media benefit from the easy proliferation of weak ties among a population. Weak ties (as opposed to the strong ties formed by ethnic or class identities) are understood to play an important role in mass mobilization. Others argue that the usage of social media is *ex post facto* informational, rather than organizational. That is, discussion of mass events on social media happens after the fact, as descriptive discussion of events rather than as a tool for organiz-

ing those events in the first place. The distinction between these two theories is one that should be relatively simple to empirically test, at least in principle. If social media are organizational, events should be discussed *before* they happen, which should not be the case if usage is merely descriptive. In practice, however, this distinction has been challenging to reliably test. For instance, in the case of the 2011 London riots, the richness of available data has led to various scholars tackling the use of Twitter by various populations, including police, government, rioters, and journalists. However, that extensive literature has been unable to reach any consensus as to the causal role of social media in events that occurred (Baker, 2012; Denef, Bayerl, & Kaptein, 2013; Glasgow, & Fink, 2013; Panagiotopoulos, Bigdeli, & Sams, 2014; Procter, Vis, & Voss, 2013; Tonkin, Pfeiffer, & Tourte, 2012; Vis, 2013).

In addition, the problem of measuring incidents of mass protest is a long-standing one in the social sciences. The challenge is typically met by hand-coding of newspaper accounts from local areas, or by ethnographies gathering accounts of individuals present at events, after the fact. However, automated event identification algorithms have gained traction over the last decade, as increasingly sophisticated pattern detection software and the wide availability of electronic news media have proliferated. These projects, such as the EMBERS project (Ramakrishnan et al., 2014), the Uppsala Conflict Data Program (Sundberg & Melander, 2013), and ICEWS (Boschee et al., 2015), share a basic conceptual basis despite different technical implementations. Computer software examines news reports for keywords (such as “protest” or “violence”) and uses contextualization tech-

niques to identify events along with their location, date, time, and metadata of interest to the project such as number of participants, number of arrests or deaths, demographic group of individuals involved, or issues being protested.

This approach has several limitations. First, these projects are entirely dependent on the availability of timely electronic copies of free media in the geographic locations of interest. This is problematic in the cases of authoritarian states that control the media narrative or areas that have very limited media coverage due to poverty, lack of political power, or other factors. Second, they are language dependent, requiring keywords and the context algorithms to be adapted for every required language. This process does not entail simple dictionary translations, but requires understanding how such information is contextually conveyed in different ways in different languages.

Some of these issues have been taken up by applying event identification algorithms to social media as well. These tools use keyword searches of various levels of sophistication to detect whether or not people are talking about protest (or other events of substantive interest) in a particular geographic area (van Zoonen & van de Meer, 2016; Becker, Naaman, & Gravano, 2010; Chae et al., 2012; Pohl, Bouchachia, & Hellwagner, 2012).

This article explores an understudied element of the intersection between social media and mass protest by using geocoded tweets in order to understand how social media *geographically* reacts to the occurrence of mass protest within a country. It does so by taking a systems approach that is not dependent on either the availability of traditional media or the content of tweets.

**Data: Euromaidan and Geocoded Twitter**

Exploring the effect of mass protest on social media in the proposed way requires two sets of data: geocoded social media data and event data on mass protests. Given their significant size, discrete start and end point, relatively long period of contiguous time, and satisfactory coverage in both the news media and social media, the Euromaidan protests make for an acceptable case study for the project. As such, this study's data collection is bounded from October 1, 2013 to March 31, 2014. This six-month block of time includes a month and a half before and after the protests began and ended, respectively.

The social media dataset was constructed from a custom database infrastructure built to collect tweets en masse, which has been downloading all geocoded tweets (i.e., those that have a GPS-generated latitude and longitude attached to the tweet) originating from the former Soviet world in real-time since the beginning of 2012 (Wilson, 2016b). This particular project uses the subset of this database corresponding to tweets posted from within Ukraine during the six months of the study, amounting to approximately 2.2 million individual tweets.

This data is all geocoded, with the exact latitude and longitude of the originating device recorded at the moment of the tweet. These coordinates are accurate to within two meters, so they can be used for precise subnational analysis impossible with most social science data sources. Custom written GIS code identified the Ukrainian raion (roughly equivalent to county-level, with 490 total in Ukraine) from which each tweet originated. Additional code was written to aggregate the tweets such that the unit of analysis was raion-days, yielding an  $n$  of 181 days.

In addition, the data includes the number of friends that the user who tweeted had at the moment of the tweet, which allows quantification of how important to the overall network the particular user is. On social media, users with more friends or followers can be understood to have a higher number of weak ties, that is, they are more network-central actors. As such, this paper defines the *network-weighted* number of tweets as the multiplication of the number of tweets by the poster's number of friends at the time of the tweet, effectively weighting every tweet by the number of network connections possessed by the tweeter. For instance, if Anna had three friends and tweeted once, while Bob had one friend and tweeted twice, Anna's network-weighted measure would be three, while Bob's would be two.

In order to capture geographic variation made possible by the tweets' geocoding, these tweets are divided into two groups: those originating from within the urban raions of Pecherskyi and Shevchenkivskyi (the two small raions in central Kiev that feature the capitol, government buildings, and Maidan Square), and those originating from the rest of Ukraine. These two groups of tweets are referred to as *core* and *periphery* for ease of reference.

Finally, operationalization of the occurrence of mass protest was accomplished by hand-coding events using mainstream international media sources such as the BBC and Associated Press, in conjunction with event data from ICEWS (Boschee et al., 2015). First, a dichotomous variable was coded on a daily basis over the period of the study, with a one indicating a mass protest occurred in central Kiev, while a zero indicated there was no protest. It was well-documented that protest of some minimal level

was constant during the period of November 21 to February 28, and so those days were coded as one, while days outside that range were coded as a zero. Media on these days would often use qualitative descriptions such as "a large number of protesters," while they tended to provide estimates of exact numbers of participants only for the larger Sunday protests.

In addition, for the three-day peak of protest in which the largest crowds were present, violence erupted, and the government fled the country, hand-coding from media accounts (both Ukrainian and international) established where and when specific violent incidents occurred within central Kiev. These incidents were assigned a timestamp, in addition to as precise a latitude and longitude as possible given the descriptions of news accounts, generally corresponding to either a specific landmark or a street intersection, for instance: "In a renewed assault shortly after 04:00 local time on Wednesday (02:00 GMT), the police tried to move on the protesters' tents near the main monument on the square" (BBC, 2014).

These data provide a rich set of measures in order to explore how social media are impacted by events of mass protest. They allow the distinction between central and peripheral social media activity geographically, and along the dimensions of both absolute and network-weight quantities of tweets. Finally, the mapping of incidents of violence at the peak of protest and the dichotomous coding of whether mass protest occurred in central Kiev on a daily basis provide two different elements of events in the real world that can be analyzed as to their relationship with social media activity.

## Empirics

Given the dichotomous variable of whether mass protest occurred, and interest in its relationship with several social media variables, a logistic regression framework is ideal, especially since the dichotomous variable is one in approximately half of the cases. Regressions were run in R, using the `glm` package (R Core Team, 2017). Code and data files are available upon request. This treats the occurrence of mass protest as the dependent variable and the social media variables as independent variables, with the units of analysis being days, with an  $n$  of 181. Note that as this is an exploratory analysis, the regression design should not imply that the social media variables are *causing* the occurrence of mass protest, but is simply measuring the *correlative* relationship between the two sides of the equation.

Additionally, "time" has been also been used as an independent variable, operationalized simply as the number of days since the start of the sample (i.e. October 1 is "1" and March 31 is "181"). This is to control for the fact that social media activity, all else being equal, is rising over time (from February 2012 to April 2014, the number of tweets originating from Ukraine over time increased in a highly linear manner, with an adjusted  $R^2$  on a univariate Ordinary Least Squares, OLS, of 0.85).

Table 1 reports the results of the logistic regression, with highly statistically significant effects for each of the independent variables, in both intuitive and counter-intuitive directions.

First, note the sign for the absolute quantity of tweets in the periphery. It is positive, which is intuitive: when a mass protest is occurring, the country's eyes are on

**Table 1**

*Logistic Regression of Occurrence of Protest on Social Media Variables (n=181)*

	Coefficient	P
Periphery tweets	$1.880 \times 10^{-3}$	<0.001***
Core tweets	$-5.453 \times 10^{-2}$	<0.001***
Periphery network-weighted tweets	$-3.979 \times 10^{-6}$	<0.001***
Core network-weighted tweets	$1.123 \times 10^{-4}$	<0.001***
Time	$-2.981 \times 10^{-2}$	0.004**
Intercept	-10.32	<0.001***

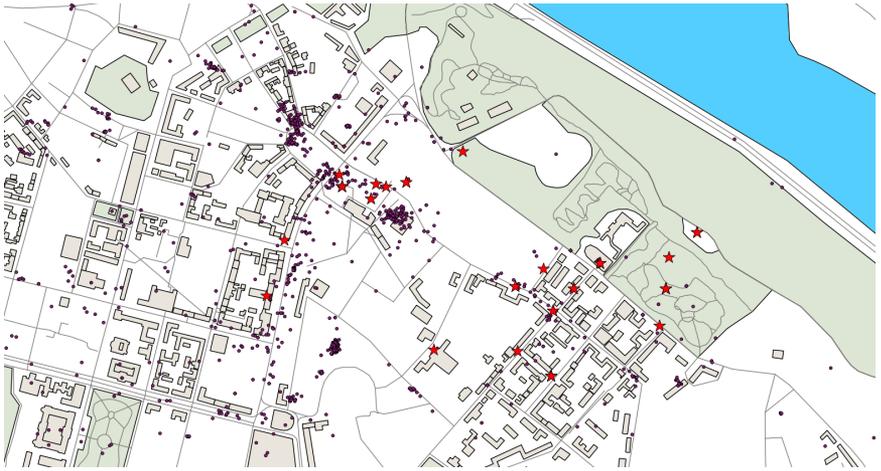
it. Since individuals tend to communicate more during historic events, it would be expected for communication usage of all kinds to surge, which would include increased posting to social media sites.

On the other hand, the absolute quantity of tweets in central Kiev, where the mass protests were predominantly occurring, is actually negatively associated with the occurrence of mass protest. This is highly counter-intuitive in one sense: if there are hundreds of thousands of people packed into the city center, one would expect social media usage in the city center to greatly increase. One explanation would be that those on the ground at the site of protest will not post as much on social media in real time, because they are engaged in other activities. While there are certainly posts to social media being transmitted from the midst of mass protest (recall the famous images and videos of regime violence during the peak of the Arab Spring), all else being equal, active protesters will tweet less than if they are not protesting. This may be evidence that the use

of social media in conjunction with protest may indeed be organizational rather than descriptive.

An examination of the geography of tweets and violence provides some support for this interpretation. Figure 1 is a map of central Kiev, with the small dots indicating an individual tweet posted during the period of February 18-20 (the height of the protests), while the large stars indicate individual instances of violence (hand-coded from newspaper sources, as described in the last section). The bulk of the violence took place on the east side of the map, near the main government buildings. Each of those locations represents multiple instances of violence due to the imprecision of newspaper accounts. By and large, where violence occurred, tweets did not, while most of the twitter activity occurred to the northwest in Maidan Square itself, where speaking platforms and tents were erected. In one sense, this is a microcosm of the organizational versus descriptive debate. Organizational tweets occurred in the northwest, and descriptive tweets (of which there were proportionately much fewer) originated in the east.

Finally, the network-weighted social media activity has a positive sign in the center and a negative sign in the periphery. This relationship follows from the literature on social movements and protests, in which the best predictor of attendance of a protest is not ideology or other substantive factors but whether an individual already knows someone else going to the protest (Schussman & Soule, 2005; Tufekci & Wilson, 2012). The so-called "power of weak ties" suggests that social media's organizational impact on protest may not just be one of easing coordination of action, but simply one of making the participation of weakly tied individuals far more visible to each other. In



*Figure 1. Violence and tweets in Central Kiev Feb. 18-20, 2014.*

operational terms, mass protests should be evinced by the geographic convergence of network-centric individuals, i.e. those with many friends, upon the site of the protest. This would drain network-weighted social media activity out of the periphery and into the center during incidents of mass protest.

Table 2 renders the results of this regression framework as a classification table, showing whether protest was predicted by the model, and how that compared with observed instances of mass protest. The results are quite stark, with the model correctly predicting the outcome in 91% of cases, with only eight false negatives and nine false positives in the 181 days in the sample. Five of the eight false negatives occur during a single week in early February (from the third to the eleventh), which was a period described in news accounts as calm, with the number of protesters in the low thousands.

Five of the nine false positives are either just before or just after the defined period of protest (particularly in

**Table 2**  
*Classification Table of Logistic Regression*

	No Protest (Predicted)	Protest (Predicted)	Sum
No Protest (Observed)	73	9	82
Protest (Observed)	8	91	99
Sum	81	100	181

the first week of March), which may be indicative of protests slowly taking off at the beginning, and tapering off at the end. Of particular interest though is the false positive on October 2. Upon further investigation, this was the day of a large one-day protest unrelated to the subsequent Maidan protests, in which thousands of protesters were turned away from the city council building in central Kiev by police with tear gas. This demonstrates the model in fact picking up on mass protest events outside of the roughly coded dependent variable of the project.

### **Additional Discussion**

This article explores how the occurrence of mass protest is reflected in social media activity, identifying both intuitive and counter-intuitive patterns that lay a foundation for further work.

One avenue of additional research would be pursuing the distinction between the use of social media for organizational and descriptive purposes. This article noted the stark distinction between Twitter activity adjacent to violence and at staging areas several blocks away. This points to a complex combination of social media being utilized for both of the theorized uses, but in close geographic

and temporal proximity that studies with less resolution cannot distinguish between. A content analysis of the full texts of this sort of social media data with a resolution of meters and seconds, could yield a rich understanding of how social media and mass protest are causally connected at a micro-level.

The finding that mass protest in the Euromaidan context was associated with drops in absolute levels of social media activity in the center of Kiev but spikes in network-weighted activity can be exploited in a couple of different ways. First, the existence of this phenomenon can be tested generally in other countries and protest contexts to see if it is a robust pattern, or an idiosyncrasy of the Euromaidan case. Initial work has begun on duplicating the study with geocoded Twitter data from the March 2017 protests in Russia and Belarus. In addition, if this finding is robust, it has the potential to aid in identification of mass protest events in areas besides country capitals. Preliminary work has applied this finding to the same tweets from this paper, except defining the "center" as particular regional capitals instead of Kiev. For instance, reports of mass protest in Lviv were found to coincide with this same pattern of social media usage, although comprehensive testing has not yet been possible due to the scarcity of event data on protest in Lviv or other Ukrainian regional capitals.

However, this also points to one potential application of this paper's findings. The sparse data available on protests outside of capital cities is one major drawback of existing event detection algorithms, and so the capacity of social media usage patterns to suggest when and where mass protests are occurring could supplement such detec-

tion algorithms. Because this is non-keyword or language dependent, it can provide an additional input into such algorithms to indicate places where additional attention should be paid. This could help detect protest in cases where free media is not available, or when protesters are using language not accounted for in keywords. For example, discussion of opposition in authoritarian regimes often utilizes coded speech, avoiding hot button words like "protest" but using euphemisms or alternate terms. The findings of this paper can be used to detect protest independently of predicted terms, either as an end in and of itself, or as a way to identify what keywords are actually being used by the population in question, in order to refine keyword event detection algorithms, or better understand the way that protest is organized and discussed by populations.

Building on our existing understanding of the dynamics of mass protest, this paper explores how social media responds to the occurrence of mass protest, using the Euromaidan protests in Ukraine as a case study. The article finds that social media usage increases nationwide during large mass protests in the capital, and that while absolute social media traffic decreases at the site of mass protest, network-weighted activity increases. By relying completely on count and network data rather than keywords, hashtags, or other contextual clues from the content, its findings point to techniques that should be portable across language barriers and national borders. In addition, the findings are applicable to any geocoded social network, and thus have potential for use in any country for which similar data are available, even in the absence of widespread Twitter usage. This article lays the foundation for

additional work that could apply its findings both to other states and to advancing the fields of detection and analysis of mass protests.

## References

- Alqudsi-ghabra, T. M., Al-Bannai, T., & Al-Bahrani, M. (2011). The internet in the Arab Gulf Cooperation Council (AGCC): vehicle of change. *International Journal of Internet Science*, 6(1), 44-67.
- Anderson, L. (2011). Demystifying the Arab Spring. *Foreign Affairs*, 90(2), 2-7.
- Baker, S. A. (2012). From the criminal crowd to the "mediated crowd": the impact of social media on the 2011 English riots. *Safer Communities*, 11(1), 40-49.
- BBC News. (2014, February 19). Ukraine crisis: Police storm main Kiev 'Maidan' protest camp. Retrieved from <http://www.bbc.com/news/world-europe-26249330>.
- Becker, H., Naaman, M., & Gravano, L. (2010, February). Learning similarity metrics for event identification in social media. In *Proceedings of the third ACM international conference on Web search and data mining* (pp. 291-300). ACM.
- Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, S., Starz, J., & Ward, M. (2015). ICEWS Coded Event Data. *Harvard Dataverse*, V12.
- Bunce, V. J., & Wolchik, S. L. (2010). Defeating dictators: Electoral change and stability in competitive authoritarian regimes. *World Politics*, 62(1), 43-86.
- Castells, M. (2009). *Rise of the Network Society, with a new preface: The Information Age: Economy, society, and culture Volume I*. Hoboken, NJ: Wiley-Blackwell.
- Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D. S., & Ertl, T. (2012, October). Spatiotemporal social media analytics for abnormal event detection and exami-

- nation using seasonal-trend decomposition. In *Visual Analytics Science and Technology (VAST)*, 2012 IEEE Conference on (pp. 143-152). IEEE.
- Chowdhury, M. (2008). The Role of the Internet in Burma's Saffron Revolution. Harvard University, Berkman Center for Internet & Society, 8. Retrieved from [https://cyber.harvard.edu/publications/2008/Role\\_of\\_the\\_Internet\\_in\\_Burmas\\_Saffron\\_Revolution](https://cyber.harvard.edu/publications/2008/Role_of_the_Internet_in_Burmas_Saffron_Revolution)
- Denef, S., Bayerl, P. S., & Kaptein, N. A. (2013, April). Social media and the police: Tweeting practices of British police forces during the August 2011 riots. In *proceedings of the SIGCHI conference on human factors in computing systems* (pp. 3471-3480). ACM.
- Diamond, L. (2010). Liberation technology. *Journal of Democracy*, 21(3), 69–83.
- Dyczok, M. (2005). Breaking through the information blockade: Election and revolution in Ukraine 2004. *Canadian Slavonic Papers*, 47(3-4), 241-264.
- Faris, R., & Etling, B. (2008). Madison and the smart mob: The promise and limitations of the Internet for Democracy. *Fletcher Forum of World Affairs*, 32, 65.
- Farrell, H. (2012). The consequences of the Internet for politics. *Annual Review of Political Science*, 15(1), 35–52.
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4), 1799-1839.
- Glasgow, K., & Fink, C. (2013, April). Hashtag lifespan and social networks during the London riots. In A. M. Greenberg, W. G. Kennedy, & N. D. Bos (Eds.) *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 311-320). Springer Berlin Heidelberg.
- Goldstein, J. (2007). The role of digital networked technologies in the Ukrainian Orange Revolution. In *Berkman Center Research Publication*, 14. Retrieved from <https://cyber.harvard.edu/sites/cyber.harvard.edu/files/>

Goldstein\_Ukraine\_2007.pdf

- Hofheinz, A. (2005). The Internet in the Arab world: Playground for political liberalization. *International Politics and Society*, 3(1), 78-96.
- Kyj, M. J. (2006). Internet use in Ukraine's Orange Revolution. *Business Horizons*, 49(1), 71-80.
- Mackell, A. (2011, May 25). The IMF versus the Arab spring. *The Guardian*, Retrieved from <http://www.theguardian.com>
- Morozov, E. (2012). *The Net delusion: The dark side of Internet freedom*. PublicAffairs.
- Murphy, E. C. (2006). Agency and space: The political impact of information technologies in the Gulf Arab states. *Third World Quarterly*, 27(6), 1059-1083.
- Oghia, M. J., & Indelicato, H. (2011). Ruling the Arab internet: an analysis of internet ownership trends of six Arab countries. *Arab Media & Society*, Summer, 14.
- Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.
- Panagiotopoulos, P., Bigdeli, A. Z., & Sams, S. (2014). Citizen-government collaboration on social media: The case of Twitter in the 2011 riots in England. *Government Information Quarterly*, 31(3), 349-357.
- Pohl, D., Bouchachia, A., & Hellwagner, H. (2012, April). Automatic sub-event detection in emergency management using social media. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 683-686). ACM.
- Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3), 197-214.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., et al. (2014, August). '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* (pp. 1799-1808). ACM.
- Sabadello, M. (2012). The Role of New Media for the Democratization Processes in the Arab World. In Z. Sereghy, S. Bunk, & B. Preiss (Eds.), *The Arab Revolutions: Reflections on the role of civil society, human rights and new media in the transformation processes*. Austrian Study Center for Peace and Conflict Resolution.
- Schussman, A., & Soule, S. A. (2005). Process and protest: Accounting for individual protest participation. *Social Forces*, 84(2), 1083-1108.
- Shirky, C. (2009). *Here comes everybody: The power of organizing without organizations*. New York: Penguin.
- Stepanova, E. (2011). The role of information communication technologies in the "Arab Spring". *Ponars Eurasia*, 15, 1-6.
- Sundberg, R., & Melander, E. (2013). Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, 50(4), 523-532.
- Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.
- Tonkin, E., Pfeiffer, H. D., & Tourte, G. (2012). Twitter, information sharing and the London riots?. *Bulletin of the American Society for Information Science and Technology*, 38(2), 49-57.
- Tucker, J., Barberá, P., & Metzger, M. (2013). A breakout role for Twitter? Extensive use of social media in the absence of traditional media by Turks in Turkish in Taksim Square protests. *The Monkey Cage [blog]*. Retrieved from <http://themonkeycage.org/2013/06/a-breakout-role-for-twitter-extensive-use-of-social-media-in-the-absence-of-traditional-media-by-turks-in-turkish-in-taksim>

square-protests/

- Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication, 62*(2), 363-379.
- van Zoonen, W., & van der Meer, T. G. L. A. (2016). Social media research: The application of supervised machine learning in organizational communication research. *Computers in Human Behavior, 63*, 132-141.
- Vis, F. (2013). Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 UK riots. *Digital Journalism, 1*(1), 27-47.
- Wilson, S. L. (2016a). *Information and revolution* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (No. 10155683)
- Wilson, S. L. (2016b). The Geocoded Twitter Archive for the Social Sciences: A New Dataset. Tech. rep.
- Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement?. *Journal of Communication, 59*(1), 40-56.
- Zhuo, X., Wellman, B., & Yu, J. (2011). Egypt: The first Internet revolt? *Peace Magazine*, (Jul/Sep), 6-10.