

Oxford Handbooks Online

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

Javier Borge-Holthoefer, Muzammil M. Hussain, and Ingmar Weber

The Oxford Handbook of Networked Communication

Edited by Brooke Foucault Welles and Sandra González-Bailón

Subject: Political Science, Political Behavior, Regional Studies

Online Publication Date: Dec 2018 DOI: 10.1093/oxfordhb/9780190460518.013.24

Abstract and Keywords

Digital infrastructure has been rapidly embraced in the Arab Middle East and North Africa in the last decade, opening a unique window for computational social science and network data science scholars. However, there are currently two coexisting social and economic realities in the region, which result in very different usages and dynamics of networked communication: countries with chronic civil unrest in which digital media have largely served as mobilization tools (e.g., Tunisia, Egypt), and relatively stable and wealthy societies that face social change and economic hyper-development (e.g., Qatar, Kuwait). Given such diversity across the region, how and why should social scientists study digital networks in the Middle East? What can digital networks teach us about the social and political aspects of the modern Middle East? In sum, while claims about digital technologies' impacts across the region have been critiqued for being speculative and overblown, we suggest that digital technologies have instead broadened our ability to understand ongoing transformations among Arab states and societies.

Keywords: computational social science, social media, social disrupter, social observatory, collective action

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

This chapter focuses on a complex and diverse region: the Arab Middle East and North Africa. Digital infrastructure has been rapidly embraced in the last decade, opening a unique window for computational social science and network data science scholars: data from social media already constitute our best observatory for societies in which traditional data sources are not as abundant as in other parts of the world. However, there are currently two coexisting social and economic realities in the region, which result in very different usages and dynamics of networked communication overall. On the one hand are countries with chronic civil unrest, where digital media have largely served as mobilization avenues and coordination tools for political purposes (e.g., Tunisia, Libya, and Egypt). On the other hand are relatively peaceful, stable, and wealthy societies that nonetheless face serious social challenges emanating from social change and economic hyper-development: a progressive minoritization of the local population, a rapid change to urban and cultural spaces, and sometimes incongruous traditional and modern lifestyles (e.g., Qatar, Bahrain, and Kuwait). Given such complexity and diversity across the region, how and why should social scientists study digital networks in the Middle East? How have digital networks changed the Middle East? And what can digital networks teach us about the social and political aspects of the modern Middle East? In sum, while claims about digital technologies' impacts across the region have been criticized for being both speculative and overblown, we suggest that digital technologies have in fact broadened and deepened our ability to understand ongoing transformations between Arab states and societies.

1. Introduction

In this chapter we approach the Arab Middle East from a dual perspective, analyzing the current sociopolitical tensions in the region and the emergent social scientific opportunities for examining them. Digital technologies have been rapidly embraced in the last decade across populations and institutions, opening a unique window to social scientific inquiry in the region. Metadata curated from digital media now constitute observatories of societies and social life and make observable these trends and trajectories on a scale and at a depth hitherto unparalleled from the "Arab street" (Regier and Khalidi 2009). Since the diffusion of digital media, how have social scientists investigated social and political life in the region? How are new research practices from the computational social sciences contributing to and changing this paradigm? How might new sociotechnical transformations in the region further enable and challenge this work?

To address these questions, we review the research and findings on digital media and the Middle East from the early period of Internet diffusion to the aftermath of the Arab Spring, with special attention to the new influx of work stemming from scholarship on complex networks and computer science. We argue that there are two presiding and coexisting social and economic realities in the region, which result in very different

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

usages and dynamics of networked communication: fragile societies and prospering states. On the one hand, fragile societies (e.g., countries with recent civil unrest and lacking stable governance infrastructure) have often co-opted digital media spaces as avenues for mobilization and tools for coordinating social change. Prominent cases include Jordan, Tunisia, Lebanon, Algeria, and Egypt, which have been featured in political studies after the Arab Spring mobilizations of 2011–2012. On the other hand, even prospering and stable wealthy states, which nonetheless face their own unique challenges (e.g., rapid societal transformations stemming from economic hyper-development), have counter-exemplified ways in which digital communication and social networking patterns replicate “offline” realities and challenges. Prominent cases in which rapid digital diffusion has worked hand in hand with state-sanctioned hyper-development include Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates (UAE) and have been prominently featured in complexity studies. By comparing and contrasting these research paradigms on sociopolitical versus socioeconomic phenomena and analyzing their underlying research perspectives (i.e., political studies versus complexity analyses), we find that in both examples social science scholars increasingly rely on and struggle with shared opportunities and pitfalls of data abundance and verification norms.

Our overarching perspective hinges on the idea of moving past concerns about access (via technologies and skills) to the observation of sociopolitical modularity. While past scholarship has rested on the limits of user access to digital technologies in the Middle East juxtaposed with the recent debates between cyber-optimists and cyber-pessimists reflecting on the Arab Spring, we propose that the future of these networked societies is far less coherent and much more chaotic than existing frames of analysis allow us to investigate. In contrast, with data from social computing, computational social science scholars have begun to observe how societies in the Middle East have restructured in unique social networks and community structures that form and recombine based on opposing political interests, segregated social groups, and diverse cultural and linguistic divides across a region that is often depicted as being far more coherent and resistant to change than is true. We suggest that scholarship on networked communication in the region move beyond simply debating whether and to what extent information communication technologies (ICTs) have disrupted Middle Eastern societies, and instead take more observational approaches to unpacking Middle Eastern societies’ increasing modularity, fragility, and complexity.

2. Social Media as Social Disrupter

How have digital networks changed the Middle East?

The study of new ICT infrastructure adoption and use and the consequences it has had for institutions, groups, and individuals is not limited to the narrow confines of the Arab world. However, for over half a decade no single geopolitical region has been under more academic and policy scrutiny than the nation-states of North Africa and the Middle East when considering the role of ICTs and digital media. Before the past five years, the broader intersections of work on these topics were situated in the contexts and experiences of industrialized democracies in North America and Western Europe. Since 2011 the events of the Arab Spring have introduced a complex disruption into these intellectual trajectories and the interpretive toolkits available for social scientific analysis being applied to this region. In this section we provide a summary of these disruptions to established areas of work in communication and political studies and offer a brief synthesis of how recent studies have begun to address the core new issues, observations, and questions raised since the Arab Spring of 2011 with research frameworks and methods appropriated from computational social sciences.

The bulk of new scholarship in communication and political studies published since 2011 has pursued two general and complementary strategies while employing computational data and analytics based on availability and convenience. First, the focus of most work has been on advancing new interpretive frameworks developed from in-depth case studies drawn during and in the aftermath of the Arab Spring; second, studies have also tested generalizable processes through computational analyses. It is this latter body of work that reflects the spaces where new crosscutting collaborations are taking place by marrying traditional social scientific research questions with methods and data arising from data sciences and social computing. Between these two trajectories of work there has been notable progress and advancement, both in debating the priority of research questions concerning social and political issues in the region and in innovating and testing the limits of new strategies for empirically evaluating these claims through new sources of data.

In contrast, before the Arab Spring activities of 2011, much of the scholarship in the region was concerned with the problems of limited access to ICT infrastructure for most citizens in the region. While even the earliest analysis of the Arab world's information revolution forecasted possibilities of "major political, social and or economic transformation" because of new information technology (IT) systems, public access at the turn of the century was limited to fewer than two million users (Ghareeb 2000, pp 396). However, within a decade, by the eve of the Arab Spring protests, the fabric and structure of Arab politics, society, and governance had undergone incremental and rapid transformation through the adoption of unequal yet widespread ICT infrastructure. By

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

2010 the overlaying of new digital networks, in addition to the transformations in the existing centralized broadcast communication systems, had foundational impacts for Arab political communication systems (Anderson 2013).

Just as Web 2.0 ecologies had rewired political communities and media networks in American and European societies, many Arab political societies had been fundamentally rewired through blogging networks that connected political factions within and across the region (Etling et al. 2010). The expansion of mediated communication spaces supported by ICT diffusion has had a permanent impact on news work and the media landscape. For example, during the peak of the protests in 2011, in a media system previously tightly controlled by Egypt's military regime before Internet diffusion, state-controlled newspapers were unsurprisingly ineffective in supplying the public with critical coverage of the uprisings. Similarly, independent newspapers were only slightly more effective in diversifying the public's media coverage with alternative media frames. However, social media outlets supplied the most prescient interpretive frames of "revolution" as event, "dictatorship" as cause, and "resignation" as solution for Egyptian and global citizens seeking understanding of on-the-ground events (Hamdy and Gomaa 2012). Social media outlets also dared to break the censored news agenda in Egypt: gatekeeping, the process by which media experts traditionally construct and shape the public agenda, had been flipped into a networked activity because a digitally enabled user base had taken root in Egypt and many Arab societies. This had created the combined disruption in which broadcast frameworks were notably influenced and co-constructed with the participation of large masses of social media users actively curating their interests and interpretations about the protests in a media system previously censored in the supply and production of media frames (Meraz and Papacharissi 2013).

These changes in how news work was conducted and how the media landscape had evolved suggest conclusively that the diffusion of Internet infrastructure to educated urban elites in the region has been a profoundly transformational process that it is difficult and unproductive to reduce to a few set levels. For example, moving beyond news studies, network diffusion to users has also provided more mechanisms for interlinking communities across and beyond the region. By tracing the volume and structure of communication in Twitter, Bruns et al. (2013) find that in both Egypt and Libya, local activists and global citizens functioned as significant bridges, connecting Arab communities with English audiences around the world. This suggests the possibility not only that social media spaces have altered the official narratives produced by media organizations locally, but also that online media activism itself is now a crucial component of the larger mosaic of political participation available to activist citizens even in closed political systems.

Indeed, numerous studies illustrate this possibility through triangulation and cross-case comparison. During the Arab protests, the design and policies governing social media platforms critically influenced the ability of activist users to accomplish their goals (Youmans and York 2012). In her comparative case analysis of the Iranian Green revolution, the Moldovan protests, and the G-20 summit in Pittsburgh, Ems (2014)

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

demonstrates that social media technologies functioned as “socio-technical assemblages” affording both opportunities and risks to users, helping to advance the intellectual project beyond vehement calls against digital optimism or cyber-pessimism. Similarly, in their historical case analyses of Belarus, Moldova, Russia, and Ukraine, Lysenko and Desouza (2014) convincingly trace the back and forth tussle over ICT infrastructure between authorities and dissenters over a period of fourteen years. In fact, through both contextual and historical analyses, these examinations consistently find that new communication technologies *do not have an inherently democratic or antidemocratic utility to users or elites*. Rather, the features and uses afforded by new technologies are something actively constructed and contested by everyday users and political elites. While this should give support to the idea that scholars must move past polemical debates segregated along cyber-optimist and cyber-pessimist lines, it also suggests that the eventual democratic or antidemocratic impact of new communication infrastructure is a shared outcome of the assemblages of tools, affordances, uses, and governing policies surrounding the imaginaries of ICTs.

So how have communication technologies and digital ecologies impacted political and social life in Arab societies? First, the impact has been both complex and dramatic. On average, the Arab citizen in today’s media environment has more capacity to participate in and organize collective action, especially by reaching out internationally (Lynch 2011), than ever before. Second, Arab regimes now also have increased capacities for accessing and observing their publics and opposition parties (Bellin 2012), to monitor, manage, or coerce them. In fact, the disruption caused by digital media in empowering the average citizen may have further encouraged regimes to resort to stricter and more drastic measures of culling the capacity of civil society. Third, whether Arab regimes take a guarded or more open approach to managing online activities, it is difficult for states to unwire their newly wired societies. With more globally connected users, we must expect that networked communities have more capacity to mobilize and coordinate action among their peers—whether that coordination and mobilization are oriented toward democratic or antidemocratic outcomes—than before their societies were wired.

In this final space of scholarship on social movements and collective action networks, scholars from social computing and computational social sciences have been well equipped and perceptive in focusing their analysis on understanding how connected individuals and groups coordinate action, with or without the support of established institutions (Bennett and Segerberg 2012). For example, the most successful movements in the region (e.g., Kefaya and April 6th Youth) relied heavily on social media infrastructure to coordinate sympathetic crowds to expand their networks globally, shaped media framing sympathetically to their goals, and strengthened core group ties into dense communication networks (Lim 2012). While the language of crowds and groups may at first seem conceptually barren, network scholars have developed definitions and meanings for these rich structural components of dense group activity by inductively and deductively analyzing their communication and coordination patterns.

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

Contemporary collective action coordinated with network communication infrastructure in the Arab Spring and related movements took place with the involvement of hidden influentials, who defined the movement's identity and framing and worked closely with the influentials who projected the message; broadcasters, who gave dense crowds their structure; and common users, who gave the movements their mass effect (González-Bailón, Borge-Holthoefer, and Moreno 2013). More recent mass participation events, like Kony 2012, Occupy Wall Street, and the Arab Spring, have reflected the capacity of crowds to form and mobilize around shared grievances with seemingly little or no coordination, but network scholars and data science methods are unpacking the order and structure of what resembles chaos and confusion to the untrained observer. The next section engages with the second core question of our analysis: *What can digital networks teach us about the Middle East?* Answering this demands a focused elucidation of recent and necessary developments around the *formal methods* that network analysts and data scientists exploit at the intersection of computer science, complex networks, and big data.

3. Social Media as Social Observatory

A key feature that explains the rapid emergence of computational social science is that it draws on the productive interaction of previously divergent academic avenues. Social media typically come in the form of networked communication systems; as such, our attempts to improve and quantify our sociological insights can appropriate the rapid development of complex network theory and the availability of big data sources. In the Arab world this translates into the application of network methodology to empirical data to answer questions about political mobilization and the integration/segregation dynamics in some wealthy Arab societies. The case studies that follow focus precisely on these two questions, illustrating the strength of combining these areas. Both case studies are presented in an accessible way. That is, technical details about the methods have been omitted (further details and overviews on the literature about the structure and dynamics of complex networks are available in Boccaletti et al. 2006; Kadushin 2012; and González-Bailón et al. 2014a). But first, as the guiding thread of our case studies is an exploration of the sociopolitical modularity reflecting the region, we now demonstrate how complex network theory provides a quantitative proxy for observing it. In particular, we briefly review the basics of community detection in networks.

In network studies, *community* refers to a set of nodes in a network such that, intuitively, the nodes within one community are densely connected, whereas the connections between different communities are sparse. Note that implicitly, the use of *dense* and *sparse* implies the existence of a null model that defines an expectation of the level of density created by random chance. In 2002, Girvan and Newman devised a formalization

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

of this intuition in the form of the following equation, which quantifies the quality of a given partition of nodes:

$$Q = \frac{1}{L} \sum_{ij} (a_{ij} - p_{ij}) \delta(C_i, C_j)$$

Here the sum runs over all the connections between pairs of nodes ($\{a_{ij}\}$) in the network; $\{p_{ij}\}$ represents the probability of an edge existing between nodes i and j in some null model; L is the total number of edges in the graph; and $\delta(C_i, C_j)$ is the Kronecker delta, valued 1 if nodes i and j belong to the same community ($C_i = C_j$) and 0 otherwise.

This mathematical expression merely implies that for a given classification of nodes in groups, one can measure the quality of the communities, that is, the partition found with the coefficient Q (known as *modularity*). In doing so, the door is opened for researchers to propose heuristics that explore as many candidate partitions as possible. Solutions to this challenge are abundant (Fortunato 2010), and social network scholars have especially benefited from efficient algorithms that can handle very large structures (from thousands to millions of nodes). These algorithmic developments have opened a fruitful research avenue to investigate the determinants and effects of modularity in networks. Beyond political and opinion polarization (which we overview later in the chapter), community detection has been used to understand the characteristics of online collegiate social networks (Traud et al. 2011), the structure of protest campaigns in social media (González-Bailón and Wang 2016), and the fragmentation in the House of Representatives of the US Congress (Porter et al. 2005). In all these case studies, community detection has helped to identify structural holes in networks (i.e., areas where information is less likely to travel because there are relatively fewer channels for its dissemination).

4. Case 1: Networks of Political Polarization in Egypt

Now that scholars and policymakers are re-evaluating the costs and benefits of the Arab Spring,¹ it seems appropriate to re-examine the political dimensions of communication networks in the Middle East. These networks of conflict emerge worldwide, as an extension of the natural political discussion happening everywhere, anytime, through any available communication channel. What makes online platforms and channels particularly useful for the study of political conflict is the availability of digital traces that allow scholars to track, for thousands—often millions—of individuals, *who* said *what*, to *whom*, and *when*. The quantity and quality of such detailed traces of human expressions and behaviors give room to exploit the full capabilities of data science, whose outputs can then be appropriately interpreted through a sociological lens (with important caveats and limitations; for more, see González-Bailón et al. 2014b).

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

These applications of data sciences are of particular importance in the Arab world, where chronic political unrest is combined with a swift adoption of online technologies.

Furthermore, the rapid growth of social networking sites since 2011 has shifted away from typical social and entertainment uses of online media toward those that are more political and civic.² Indeed, recent turmoil in Middle East and North Africa (MENA) countries has provided useful case studies of online interaction: Iran (2009), Tunisia, Libya (currently a failed state), Egypt and Bahrain (2011), Palestine and Israel (ongoing Gaza conflict), and Syria (2011 to present) have witnessed social revolts, violence, and in some cases, regime changes. In all of these cases, it has been claimed that online social networks have played a key role, so the possibility of observing their processes vastly and intimately is most opportune.

One of the most outstanding features shaping online political discussion is polarization. The obvious expression of such polarization is in the content (what is being said), but it can also be tracked from interaction patterns; in communication networks, structure has a meaning too. Opinion-aligned peers tend to flock together, which translates into the formation of clear-cut modular topologies (i.e., networks in which the main feature is the presence of communities or modules, as previously discussed). These two aspects of polarization have encouraged researchers to exploit *natural language processing* tools, such as opinion mining (Pang and Lee 2007), and *complex networks methods*, such as community detection heuristics (Fortunato 2010), to measure and characterize to what extent polarization emerges around a certain topic. Furthermore, content and structure are studied as they evolve in time, given the availability of temporally resolved data.

In our case study, we pay attention to polarized content and structure in time, as observed in Egypt at different stages of the political turmoil that has dominated the country since 2011. Our study builds on useful precedents that focused on particular aspects of online political conflicts. One of the first approaches to political polarization in social networks is that of Adamic and Glance (2005) and Conover et al. (2011), who reported a one-dimensional, left-to-right schema for US politics. Regarding the temporal dimension, some works provide longitudinal accounts of political (protest) events: the work by Borge-Holthoefer et al. (2011), devoted to the growth of the “Indignados” movement in Spain, or of González-Bailón et al. (2011), devoted to studying the recruitment processes in the 2011 protests in Spain. Closer to the Middle East, there is a large body of work on the Arab Spring revolts, some of which focuses on the Egyptian Revolution in particular. These depart from this chapter insofar as they typically include some discussion of the active role of social media during the protests and the revolution but do not use data from social networks (Twitter in particular) to study polarization or major events. Concerning conflictive situations, we find a plethora of works devoted to the Arab Spring, as the expression “Twitter Revolution” was coined and settled in mass media (e.g., Attia et al. 2011; Azab 2012; Khamis 2011; Lim 2012; Oh et al. 2012), often from a qualitative point of view or relying on surveys. Using data from blogs, Al-Ani et al. (2012) explore alternative news sources, beyond the government-supplied versions of events. However, these studies do not quantify online polarization or provide a longitudinal point of tracking and comparison. Other countries and conflicts have also

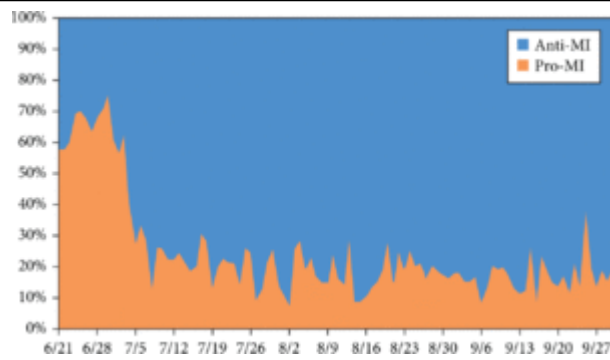


Figure 2. Fraction of pro- and anti-MI hashtags during the three-month period before, during, and after the military intervention in Egypt, summer 2013.

4.2. Structural Polarization

As stated previously, networks of conflict typically display polarization under the form of modular structures. In the case of Twitter, this can be detected at various levels: in networks of mentions, in networks of retweets, or simply in the follower-friend connections pattern. Here we report the evolution of the Egyptian retweet network during the same period of time as for the content analysis in the preceding section (Borge-Holthoefer et al. 2015). In particular, a weighted directed link (i, j, w) is laid between two nodes if user i retweeted a previous tweet by j a number of times w . A three-day overlapping sliding window scheme is used, in which the network on day t is obtained by aggregating three days of activity, that is, retweets from $t - 1$ to $t + 1$. In general, the sequence of retweet-reconstructed networks exhibits fairly constant global topological properties: density ($\sim 10^{-4}$), average degree ($2 \leq \langle k \rangle_t \leq 5, \forall t$), or clustering ($\sim 10^{-3}$) display small differences across time. The networks are thus very sparse and highly disassortative, in the sense that a few Twitter users get most of the retweeting activity, with hardly any presence of triadic closure. A modularity optimization heuristic is applied to these evolving structures (Fortunato 2010; Borge-Holthoefer and González-Bailón 2016) to obtain a two-community partition, since we track for a bipolar political scenario in Egypt.

In accordance with content-based results (figure 2), the structure of the Egyptian retweet network shows that polarization is quite stable (see figure 3): each camp accounts for almost half of the Twitter users in Egypt, with some variations over time as violence emerged during the conflict.

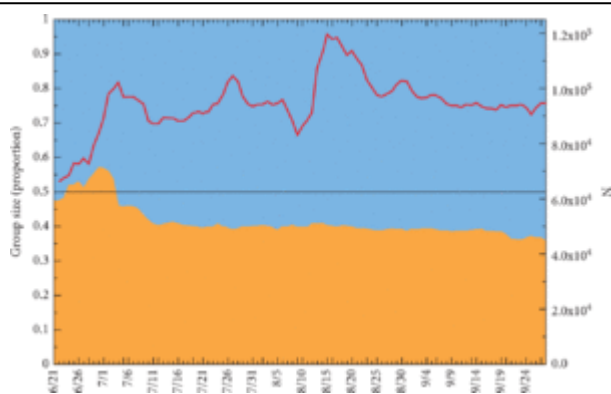


Figure 3. Content-blind community detection results for the Egyptian evolving retweet networks: fraction of users classified as pro-MI (orange) and anti-MI (blue), during the three-month period before, during, and after the military intervention in Egypt, summer 2013.

The examination of tweets from the perspective of pro- and anti-military intervention provides a solid example of what is meant by the expression *networks of conflict*. In the case of the Middle East, this phenomenon is also well expressed by the secularist versus Islamist dichotomy, which underlies the political instability observed for the past five years.³ This clear-cut categorization (which

admittedly simplifies the complexity of political conflict in the MENA region and elsewhere) is helpful to quantify disagreement in communication networks under the form of polarization.

In our particular case study, to reach the reported conclusions we have pivoted on two methods, which rely on different aspects of a given data set: the semantic and the structural dimensions. It is noteworthy that the trends observed from these nonoverlapping methodologies converge, giving evidence that both aspects of online platforms are indicative of the underlying social trends.

5. Case 2: Networks of Social Fragmentation in Qatar

In this section we describe two recent studies on digital communication in Qatar. We chose Qatar as representative of a set of politically stable and well-developed Gulf monarchies. During the last decade, Qatar has seen an enormous population growth, almost tripling, from 0.8 million in 2005 to 2.3 million in 2014.⁴ Fewer than 0.3 million people are Qatari nationals, and the explosive growth has been driven by the importation of foreign labor,⁵ predominantly from countries such as India (0.5 million), Nepal (0.4 million), Philippines (0.2 million), and Egypt (0.2 million). Ethnic Qataris constitute only 12% of the total population. In sum, the social networks and communities in Qatar are new, nimble, and in a period of formation.

This unique context of a ruling population in juxtaposition to rapidly imported populations comes with a number of social challenges, including maintaining social cohesion and ensuring political stability, tensions made observable in new ways through social media

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

data. We present summaries of recent studies that use social media data from Qatar to study these issues by looking at how Twitter users in Qatar connect across languages and how politicized migrants in Qatar are online.

5.1. Interlingual Social Networks

Maintaining a cohesive society without excluding minorities is a challenge for any society, but it is even more pressing for Qatar because of the exceptionally high percentage of foreigners among its population. Unlike most other countries in the world, Qatar does not award its foreign residents Qatari citizenship, irrespective of how much time they have spent in the country or of whether they were born there. The highly skewed nationality demographics and the interest in preserving a sense of national identity have led to a general approach in which all foreign workers are treated as welcomed guests, who are staying for a limited amount of time.

This setup creates inherent limitations on how tightly connected Qatar's overall population can ultimately be. One obvious obstacle to strong cohesion is the fact that the majority of the population does not speak Arabic and that their native languages include a mix of Indian languages; Tagalog, also referred to as Filipino; European languages; and more. Apart from potentially limiting social capital due to the difficulty of creating and maintaining cross-lingual social ties, this also puts pressure on the status of the Arabic language. Online content producers, including regular users, might want to maximize their potential audience reach by creating content solely in English, which is understood by almost everyone, rather than in Arabic, the reach of which is more limited. The potential power of social media to connect expats of different nationalities with Qataris and other Arabs could therefore, conceptually, lead to the demise of the Arabic language, as English is becoming the lingua franca of the multinational online world. In the following paragraphs we report on a recent study that used data from Twitter to study the digital communication network in Qatar through the lens of language use (Kim et al. 2014).

In their study, Kim et al. used data from forty-two thousand Twitter users with a self-declared location in Qatar to study how they connect to each other. The "search bio" feature of FollowerWonk,⁶ which allows the retrieval of users with a particular string in their location field, was used to obtain this set of users. For each user, up to the most recent thirty-two hundred tweets were obtained using Twitter's REST API.⁷

Arguably the most crucial part in the researchers' data acquisition is the detection of the language or set of languages a user tweets in. To infer the language for text written by a Twitter user, the Compact Language Identifier tool⁸ was used. Language inference at the level of individual tweets is, however, a difficult task due to their extreme brevity (with a maximum of 140 characters) and the highly nonstandard type of language or "lingo" being used. To sidestep many of these issues, Kim et al. infer the language at the user level, not the tweet level. For this, all of a user's tweets are concatenated into a single

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

document. The Compact Language Identifier tool then outputs the percentage of bytes written in a given language. As this output still contains a certain amount of noise (e.g., due to terms appearing in many languages such as “de”), only languages with more than 15% of bytes were considered as languages spoken by the user.

For 55% (23,000) of the users, Arabic was inferred as their sole language used on Twitter. For 25% (10,000) it was only English, 13% (6,000) had both English and Arabic content, and 2% (1,000) had both English and Tagalog content. A long tail of very small languages was dropped from the analysis. Figure 4 shows a network visualization for the following networks of the users in these four language groups. This graph was created using the Gephi software⁹ for plotting with the Yifan Hu graph layout algorithm. This algorithm attempts to keep users who follow each other close in the resulting graph. The graph indicates a high degree of separation between the Arabic-only (middle shade, upper left part) and English-only (darkest shade, lower right part) users. Arabic-English bilinguals (lightest shade, middle part) form a bridge between these two groups, as can also be seen in Figure 5. The Tagalog-English speakers are furthest from the Arabic-only speakers (light shade, small cluster on the right edge). Notably, modularity for this language-based partition is $Q \sim 0.4$, which doubles the average modularity ($Q \sim 0.19$) obtained for randomized versions of the same network. This is interesting because it reflects a fundamental fact about networked systems—that is, that dynamics (in this case, communication in a certain language) and structure (following relations here) are mutually shaped and constrained in a feedback loop.

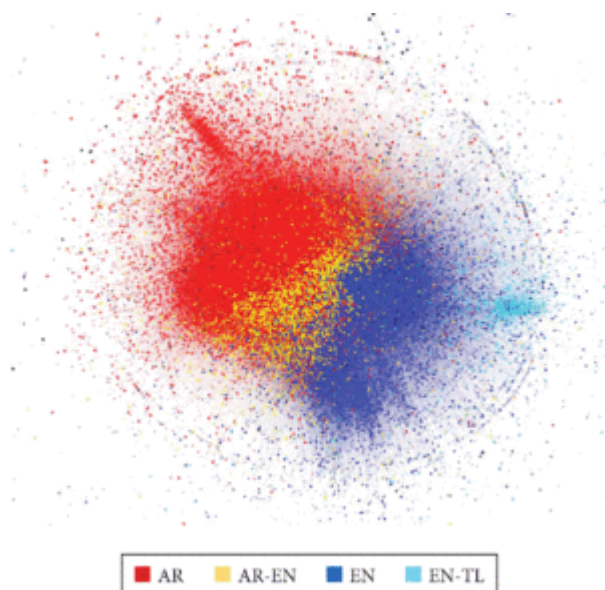


Figure 4. Visualization of Qatar Twitter network. Each node represents a user, and each edge is a “following” link in the Twitter network. Figure 5 summarizes the user-level network into a language-group-level network with four nodes. It shows that Arabic-only users have a very low propensity to follow non-Arabic speakers. Similarly, Tagalog-English speakers rarely connect to Arabic-only users.

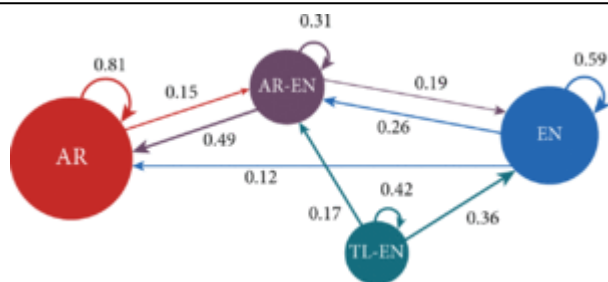


Figure 5. Following edges between the different language groups.

As one might expect, there is a strong within-group linkage preference. It also shows the expected result that languages of less-privileged groups of migrants, in particular those from the Indian subcontinent, are largely absent on Twitter. A study

by Huang et al. (2014) estimated that users from the Indian subcontinent make up only 3% of Twitter users in Qatar. They are, however, connected to the Internet and rely more on Facebook as a social networking site (Kamrava and Babar 2012).

5.2. Political Engagement

Apart from the challenge of communicating effectively in a multilingual society while preserving non-English languages, the large number of foreign workers could potentially pose a risk for political stability in Qatar and other Gulf Cooperation Council (GCC) monarchies. It is assumed that in particular Arab migrants could introduce externally driven politically, socially, and religiously salient ideologies due to their linked identity with the Gulf nationals and their inherent embeddedness in the geopolitical atmosphere of the Middle East. Here we report on a study that used Twitter data to test the premise that migrants from Arab countries are engaging more with Qatar's domestic politics online than are migrants from other countries.

This study used data from sixty-eight thousand Twitter users with a self-declared location in Qatar. Whereas language was a crucial variable for the previous study, this study relied on having a "nationality group" label for each user. Unfortunately Twitter does not provide information on the nationality of a user; hence this information has to be inferred. Akbar et al. employed the crowdsourcing platform CrowdFlower¹⁰ to infer users' most likely nationality.

Crowd workers were presented with a link to the Twitter profile, as well as a summary of users' language distribution across their tweets and the self-declared locations, grouped by country, of their friends and followers. Huang et al. (2014) had previously observed that the location of a user's followers is a strong indicator of that person's nationality. Based on this combined information of name, profile picture, language use, and friends' and followers' locations, crowd workers were then tasked with labeling the Twitter users according to their most likely nationality group, such as "North American," "Eastern European," or "Southeast Asian." To ensure an appropriate level of understanding and effort of the workers, a "quiz" had to be passed before they were admitted to the job. Workers with too many incorrect answers on the quiz were blocked from the job. Workers

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

who later performed poorly on occasional “hidden test questions” (called “gold units” in CrowdFlower terms) were also excluded.

To quantify the degree of political engagement of the Twitter users in Qatar, certain “markers” were used. These markers included (1) following certain accounts pre-labeled as being related to domestic Qatari politics, (2) retweeting tweets from such accounts, and (3) using hashtags related to domestic Qatari politics.

Regardless of which metric was used, the overall level of political engagement found on Twitter in Qatar was very low. For example, for all nationality groups except Australian, there was on average less than one tweet per user with a political hashtag. For Australian Twitter users in Qatar the average was just under 3.0, though this was due to a relatively low user count for this group and overproportionate use of the hashtag *#freeAJStaff*. Among all nationality groups, users from the Arabian Gulf were most likely to use political hashtags. More than half of the hashtags used were about Mahmoud Al-Jaidah, a Qatari imprisoned by the UAE on charges of cooperating with a reformist movement. A large portion of the debate on this topic was in Arabic, which, given the lack of strong interlingual ties (see the previous sections), means that Western expats would be less likely to be exposed to such content.

Still, despite the heavy use of Twitter during the Arab Spring (see the discussion in this chapter’s introduction), in Qatar Twitter does not seem to be heavily used to share information about domestic politics. On the other hand, Facebook is used by 41% of Arab youth to share news articles with friends and family, according to a recent Arab Youth Survey.¹¹ This survey did not, however, provide a breakdown for Qatar. Apart from Twitter not being the medium of choice for politics, the presented results could also indicate that many expats simply do not engage with domestic politics, at least not publicly.

6. Conclusion

Social science scholars seeking to expand their cases of study beyond advanced industrialized democracies have often encountered the challenges of data sparsity and the inability to extend their conceptual frameworks neatly to different contexts. In this way, the challenges and opportunities we have outlined in this chapter are by no means new or unique only to the Arab Middle East. However, we have organized this discussion to address two broad questions: How have scholars approached the study of newly enmeshed digital networks in the Arab states and societies? How else might they approach these new realities with regard to existing scholarly agendas? We have suggested that, traditionally, the issue of digital networks has been approached as a “disruptor” to social formation, political organizing, and the established analytical frameworks used to study them. However, beyond seeing digital networks as disruptors, computational social scientists and data science scholars have approached digital

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

networks as an important new opportunity to *observe* the current social and political transformations in the region with new strategies.

In this chapter we have presented two original case studies on how digital communication via Twitter data can be used to study sociopolitical modularity in both Egypt and Qatar, two very distinct countries representing both fragile and wealthy countries' experiences in the Arab world. These case studies were not limited to the current focus of the field (i.e., how digital communication *affects* modularity in the society), but rather focused on how it *reflects* this modularity and can serve as a social observatory to unpack these ongoing transformations. Computational social science methods provide the tools for using data from digital communication to quantify concepts such as modularity, polarization, and political engagement. These measurements and comparisons can provide valuable insights into societies' complex realities that traditional social science methods, like interviews and surveys, may not be as adept at obtaining from within developing regions and fragile states.

Despite the enormous potential as a social observatory that open digital communication data hold and the range of existing studies now using data from Twitter in particular (Mejova et al. 2015), there are limitations to using such data for population-level studies and generalizations. There is a robust and important sub-area of work in computational social science methodology that is ongoing, so we highlight these limitations as important avenues for interested readers and practitioners to follow up with regularly. First and foremost is selection bias: not everyone in a country or region is on Twitter, and the important platforms of users and associated data change as competing platforms gain or lose market share. For example, the population from the Indian subcontinent—the single largest community in Qatar—was largely absent from recent Twitter data archives. More qualitative work needs to be done to understand why this is. However, for the purpose of using digital networks as observatories of social life in the region, even this lack of data can, to a certain degree, be an important indicator of offline realities.

Second, there is always a level of residual noise in digital-network-borne data, because variables such as the languages used, the users' nationalities, their political leanings, and their locations must be inferred. Such inference often comes with a trade-off between precision and recall: one can either find a small set of users in which the variable of interest is easy to identify (i.e., high precision, low recall) or can infer the variable for a large set of users but with lower confidence (i.e., high recall, low precision). The standard approach in data sciences is that a bigger sample size n , even if it comes with significant noise and variance, is preferable to a much smaller but cleaner or more precise data set. This is because, mathematically, the Law of Large Numbers guarantees that despite noise, any desired level of statistical significance can be achieved, provided that n is large enough. So rather than trying to avoid any type of noise altogether, noise is accepted as part of the paradigmatic setup and taken into account in the analysis. Or stated more simply: more data beats better algorithms (Pereira, Norvig, and Halevy 2009). Finally, while much of the existing mathematical logic undergirding standard social science statistics is based on parametric statistics due to social science's undergirding

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

assumption that generalized social phenomena are distributed “normally” in the natural world, the opportunities presented in large-n “big data” archives require nonparametric statistics to parse digitally mediated social phenomena that are more often distributed logarithmically.

In sum, while these issues we have highlighted are often labeled as validity and reliability limits and concerns present in scholarly efforts to expand social science research to include nontraditional sources of data (boyd and Crawford 2012), computational social science scholars in turn view the very same concerns as opportunities for innovation and expansion of the social scientific enterprise. To overcome these limitations and make fuller use of the promise that studying digital communication holds, a number of challenges are being addressed. First, a tighter integration of different expertise and more truly interdisciplinary work is needed, in particular when it comes to the mutual exchange of methodologies. For example, it is not uncommon for research teams to include multiple statisticians and data architects alongside regional and linguistic experts. To illustrate, in computer science “classifiers” are thoroughly evaluated along dimensions such as precision or accuracy, but the process of establishing a construct’s validity is less thoroughly conducted. To address this, recent studies have included construct validity testing as a more integral part of the data curation and analysis process (see Laufer et al. 2015 for an example of validating measures of cultural similarity derived from Wikipedia). Another related challenge stems from the complexity of automating analysis of non-Latin characters, such as Arabic. For example, whereas there are dozens of ready-to-use tools and libraries for sentiment analysis in English, one is hard pressed to find as many robust text libraries for Arabic, an especially challenging issue given the sheer diversity of mutually intelligible but textually varied Arabic dialects.

Finally, given the importance of ethics and human subject protection in social science research generally and the added sensitivities of the region’s geo-cultural politics, the issue of *privacy* is of foremost importance. In particular, women in the Middle East often make conscious efforts to maintain an image of modesty. This includes more carefully limiting who can view the content they share online (Vieweg and Hodges 2016). Similarly, issues of state surveillance and control are also substantively more pronounced here than in other parts of the world, creating additional elements of privacy. Together, not only are questions of data completion and accuracy implicated by these cultural and political contexts, but they also present opportunities for research practitioners to rethink their relationships and impacts on the human subjects from whom these data originate. Overall, while the events of the Arab Spring and digital diffusion more broadly have narrowed the intellectual project to questions of impacts and disruptions, we suggest that the same tools and technologies have expanded and deepened opportunities to observe and understand social transformations both due to and beyond the digital technologies themselves. In the Arab Middle East, a region often touted as critically important to understand for the stability of the global system because of its ethnolinguistic diversity, pronounced political fragility, and burgeoning youth bulge, the recent and rapid diffusion

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

of digital media infrastructure has opened to scholars new methodological avenues by which to observe these changes. Much work has looked at identifying the mechanisms and consequences of social media on events like the Arab Spring, but not very much work has seriously explored social media as a source of data to make sense of Arab societies.

References

- Adamic, L. A., and Glance, N. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD@KDD*, 36–43 (2005).
- Al-Ani, B., Mark, G., Chung, J., and Jones, J. The Egyptian blogosphere: a counter-narrative of the revolution. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, ACM, 17–26 (2012).
- Attia, A. M., Aziz, N., Friedman, B., and Elhusseiny, M. F. Commentary: The impact of social networking tools on political change in Egypt's "revolution 2.0". *ECRA* 10, 369–374 (2011).
- Azab, N. A. The role of the internet in shaping the political process in Egypt. *IJEP* 3, 31–51 (2012).
- Anderson, Jon V. "Is Informationalization Good for the Middle East?" *Arab Media and Society* 18 (2013):.
- Bellin, Eva: "Reconsidering the Robustness of Authoritarianism in the Middle East Lessons from the Arab Spring." *Comparative Politics* 44 (2) (2012): 127+.
- Bennett, W. Lance, and Alexandra Segerberg. "The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics." *Information Communication & Society* 15 (5) (2012): 739–768. doi:10.1080/1369118X.2012.670661.
- Boccaletti S., Latora V., Moreno Y., Chavez M. and Hwang D. Complex networks: structure and dynamics. *Physics Reports*, 424, 175–308 (2006).
- Borge-Holthoefer, J., Rivero, A., García, I., Cauhé, E., Ferrer, A., Ferrer, D., Francos, D., Iñiguez, D., Pérez, M., Ruiz, G., et al. Structural and dynamical patterns on online social networks: the Spanish may 15th movement as a case study. *PLoS One* 6, 8, e23883 (2011).
- Borge-Holthoefer J., Magdy W., Darwish K. and Weber I. Content and Network Dynamics Behind Egyptian Political Polarization on Twitter. *Proceedings of the 18th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)* (2015)

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

Borge-Holthoefer, J. and González-Bailón, S. Scale, Time, and Activity Patterns: Advanced Methods for the Analysis of Online Networks. In Fielding, N., Lee, R., and Blank, G. (eds). *Handbook of Online Research Methods*, 2nd edition, Thousand Oaks: Sage (2016)

boyd, danah, and Kate Crawford. "Critical Questions for Big Data—Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15 (5) (2012): 662–679.

Bruns, Axel, Tim Highfield, and Jean Burgess. "The Arab Spring and Social Media Audiences: English and Arabic Twitter Users and Their Networks." *American Behavioral Scientist* 57 (7) (2013): 871–898. doi:10.1177/0002764213479374.

Choudhary, A., Hendrix, W., Lee, K., Palsetia, D., and Liao, W.-K. Social media evolution of the Egyptian revolution. *CACM* 55, 5, 74–80 (2012).

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., and Menczer, F. Political polarization on Twitter. In *ICWSM* (2011).

Ems, Lindsay. "Twitter's Place in the Tussle: How Old Power Struggles Play out on a New Stage." *Media Culture & Society* 36 (5) (2014): 720–731. doi:10.1177/0163443714529070.

Etling, Bruce, John Kelly, Robert Faris, and John Palfrey. "Mapping the Arabic Blogosphere: Politics and Dissent Online." *New Media & Society* 12 (8) (2010): 1225–1243. doi:10.1177/1461444810385096.

Fortunato, S. Community detection in graphs. *Physics Reports*, 486(3–5), 75–174 (2010).

Ghareeb, E. "New Media and the Information Revolution in the Arab World: An Assessment." *Middle East Journal* 54 (3) (2000): 395–418.

Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *PNAS*, 99(12), 7821–7826 (2002).

González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. The dynamics of protest recruitment through an online network. *Scientific Reports* 1, 197 (2011).

González-Bailón, Sandra, Javier Borge-Holthoefer, and Yamir Moreno. "Broadcasters and Hidden Influentials in Online Protest Diffusion." *American Behavioral Scientist* 57 (7) (2013): 943–965. doi:10.1177/0002764213479371.

González-Bailón S., Wang N., Rivero A., Borge-Holthoefer J. and Moreno Y. Assessing the bias in communication networks sampled from twitter. *Social Networks* 38, 16–27 (2014a).

González-Bailón S., Borge-Holthoefer J. and Moreno Y. Online Networks and the Diffusion of Protests. In G. Manzo (ed), *Analytical Sociology: Actions and Networks*, Wiley (2014b).

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

González-Bailón, S., & Wang, N. Networked Discontent: The Anatomy of Protest Campaigns in Social Media. *Social Networks*, 44, 95–104 (2016).

Hamdy, Naila, and Ehab H. Gomaa. "Framing the Egyptian Uprising in Arabic Language Newspapers and Social Media." *Journal of Communication* 62 (2) (2012): 195–211. doi: 10.1111/j.1460-2466.2012.01637.x.

Huang, Wenyi, Ingmar Weber, and Sarah Vieweg. "Inferring Nationalities of Twitter Users and Studying Inter-National Linking." *HyperText* (2014): 237–242.

Kadushin, C. *Understanding Social Networks: Theories, Concepts, and Findings*. New York: Oxford University Press, 2012.

Kamrava, Mehran, and Zahra Babar (Eds.). *Migrant Labor in the Persian Gulf*. London: Hurst & Company, 2012.

Khamis, S. The transformative Egyptian media landscape: Changes, challenges and comparative perspectives. *IJOC* 5, 1159–1177 (2011).

Kim, Suin, Ingmar Weber, Li Wei, and Alice Oh. "Sociolinguistic Analysis of Twitter in Multilingual Societies." *HyperText* (2014): 243–248.

Laufer, Paul, Claudia Wagner, Fabian Flöck, and Markus Strohmaier. "Mining Cross-Cultural Relations from Wikipedia: A Study of 31 European Food Cultures." In *Proceedings of the ACM Web Science Conference (WebSci '15)*, 2015. doi: <http://dx.doi.org/10.1145/2786451.2786452>.

Lim, Merlyna. "Clicks, Cabs, and Coffee Houses: Social Media and Oppositional Movements in Egypt, 2004–2011." *Journal of Communication* 62 (2) (2012): 231–248. doi: 10.1111/j.1460-2466.2012.01628.x.

Lynch, Marc. "After Egypt: The Limits and Promise of Online Challenges to the Authoritarian Arab State." *Perspectives on Politics* 9 (2) (2011): 301–310. doi:10.1017/S1537592711000910.

Lysenko, Volodymyr V., and Kevin C. Desouza. "Charting the Coevolution of Cyberprotest and Counteraction: The Case of Former Soviet Union States from 1997 to 2011." *Convergence—The International Journal of Research into New Media Technologies* 20 (2) (2014): 176–200. doi:10.1177/1354856512459716.

Mejova, Yelena, Ingmar Weber, and Michael Macy (Eds.). *Twitter—A Digital Socioscope*. Cambridge University Press, Cambridge, UK, 2015.

Meraz, Sharon, and Zizi Papacharissi. "Networked Gatekeeping and Networked Framing on #Egypt." *International Journal of Press-Politics* 18 (2) (2013): 138–166. doi: 10.1177/1940161212474472.

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

Oh, O., Eom, C., and Rao, H. Collective sense-making through the Twitter service during the 2011 Egypt revolution. In ICIS (2012).

Pang B. and Lee L. Opinion mining and sentiment analysis, FTIR, vol. 2, no. 1-2, pp. 1-135 (2007).

Pereira, Fernando, Peter Norvig, and Alon Halevy. "The Unreasonable Effectiveness of Data." *IEEE Intelligent Systems* 2 (24) (2009): 8-12.

Porter, M. A., Mucha, P. J., Newman, M. E. J. & Warmbrand, C. M. A network analysis of committees in the U.S. House of Representatives. *PNAS*, 102(20), 7057-7062 (2005).

Regier, Terry, and Muhammad Ali Khalidi. "The Arab Street: Tracking a Political Metaphor." *Middle East Journal* 63 (1) (2009): 11-29 doi: 10.3751.63.1.11.

Traud, A., Kelsic, E., Mucha, P. & Porter, M. Comparing Community Structure to Characteristics in Online Collegiate Social Networks. *SIAM Review*, 53(3), 526-543 (2011).

Vieweg, Sarah, and Adam Hodges. "Surveillance & Modesty on Social Media: How Qataris Navigate Modernity and Maintain Tradition." *CSCW* (2016): 527-538.

Weber, I., and Garimella, K. R. K. #Egypt: visualizing Islamist vs. secular tension on Twitter. In *ASONAM*, 1100-1101 (2013a).

Weber, I., Garimella, V. R. K., and Batayneh, A. Secular vs. islamist polarization in Egypt on Twitter. In *ASONAM*, 290-297 (2013b).

Wulf, V., Aal, K., Abu Kteish, I., Atam, M., Schubert, K., Rohde, M., Yerosis, G. P., and Randall, D. Fighting against the wall: Social media use by political activists in a Palestinian village. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1979-1988 (2013a).

Wulf, V., Misaki, K., Atam, M., Randall, D., and Rohde, M. 'On the ground' in Sidi Bouzid: investigating social media use during the Tunisian revolution. In *Proceedings of the 2013 conference on Computer Supported Cooperative Work*, ACM, 1409-1418 (2013b).

Youmans, William Lafi, and Jillian C. York. "Social Media and the Activist Toolkit: User Agreements, Corporate Interests, and the Information Infrastructure of Modern Social Movements." *Journal of Communication* 62 (2) (2012): 315-329. doi:10.1111/j.1460-2466.2012.01636.x.

Notes:

(¹) "I was terribly wrong"—writers look back at the Arab spring five years on. *The Guardian*, January 23, 2016, <https://www.theguardian.com/books/2016/jan/23/arab-spring-five-years-on-writers-look-back>.

Studying Networked Communication in the Middle East: Social Disrupter and Social Observatory

(²) Arab social media report, Dubai School of Government's Governance and Innovation Program, vol. 1, issue 2 (2011), <http://www.arabsocialmediareport.com/>.

(³) The Arab Spring: 5 years on. *Middle East Monitor*, 2016, <https://www.middleeastmonitor.com/specials/arabspring/>.

(⁴) World Bank data portal: Population, <http://data.worldbank.org/indicator/SP.POP.TOTL>.

(⁵) Qatar population, *Business in Qatar Magazine*, 2013, <http://www.bq-magazine.com/economy/2013/12/population-qatar>.

(⁶) FollowerWonk, <https://moz.com/followerwonk/bio>.

(⁷) Twitter API: specific user's timeline, https://dev.twitter.com/rest/reference/get/statuses/user_timeline.

(⁸) Compact language detector, <https://github.com/CLD2Owners/cld2>.

(⁹) Gephi: The Open Graph Viz Platform, <https://gephi.org/>.

(¹⁰) CrowdFlower, <https://www.crowdfunder.com/>.

(¹¹) Arab youth survey, <http://www.arabyouthsurvey.com/en/home>.

Javier Borge-Holthoefer

Javier Borge-Holthoefer, Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona and Qatar Computing Research Institute (QCRI), Hamad bin Khalifa University, Doha

Muzammil M. Hussain

Muzammil M. Hussain, University of Michigan, Ann Arbor

Ingmar Weber

Ingmar Weber, Qatar Computing Research Institute (QCRI), Hamad bin Khalifa University, Doha

