Social media manipulation during the Gulf Crisis (2017-present)

Andrew Leber and Alexei Abrahams

Social media platforms like Facebook and Twitter were heralded circa 2009-2011 as 'liberation technology' that would facilitate mass mobilization against Middle Eastern authoritarians. In this article, however, we present evidence from the ongoing Gulf Crisis (2017-present) that regimes are now capably exploiting Twitter as a vector of political propaganda and social polarization. Drawing in part on novel data collected by the authors, we show that certain state officials act to manipulate discourse on Twitter through direct dissemination, offline coercion or co-optation of existing social-media "influencers," and the mass production of online statements via automated, 'fake' accounts. We present evidence of state actors manufacturing online discourses to make claims about public opinion at home and abroad, while highlighting limits to resonance of these claims with Arab publics.

Introduction

Circa 2009-2011, social media platforms like Twitter and Facebook were celebrated as 'liberation technology'[1] that would facilitate mass mobilization against authoritarian regimes of the Middle East and North Africa (MENA). After the 2009 Iranian election protests and 2010-12 Arab Spring uprisings earned the moniker "Twitter revolutions," numerous studies have sought to articulate how exactly social media facilitated the coordination of contentious political action across MENA, even in the face of brutal repression.[2] After this initial wave of euphoria, however, examination of the cumulative impact of day-to-day social media usage has highlighted the potential for online communities to develop homophilic 'echo chambers', accelerating ideological or sectarian polarization in societies.[3] Social media, it was realized, have a darker side.

Thus far, however, even these identified vices of social media in the MENA region have at least enjoyed the redeeming quality of being 'organic' or in some sense 'self-inflicted'. Unfortunately, though perhaps unsurprisingly, we present evidence from the Gulf that the situation has changed. From their initially censorial, defensive, and reactive posture during the Arab Spring, regimes have now gone on the offensive, exploiting Twitter as a vector for political propaganda by which to manufacture the perception of support for themselves and their policies while dividing and discouraging their opposition. In line with sophisticated efforts underway

---

[1] Diamond, Larry, and Marc F. Plattner, eds. *Liberation technology: Social media and the struggle for democracy*. JHU Press, 2012.

[2] For the Arab Spring in general, see Steinert-Threlkeld, Zachary, "Spontaneous collective action: Peripheral mobilization during the Arab Spring," *American Political Science Review* 111, no. 2 (2017), pp. 379–403; — (2017c). Steinert-Threlkeld, Zachary, "Twitter as Data," Working Paper, 2017; Steinert-Threlkeld, Zachary C et al. "Online social networks and offline protest," *EPJ Data Science* 4, no.1, 2017. p. 19. For Tunisia's 'Jasmine Revolution', see Breuer, Anita, Todd Landman, and Dorothea Farquhar (2015). "Social media and protest mobilization: Evidence from the Tunisian revolution," *Democratization* 22, no.4 (2015), pp. 764–792. For Egypt's 'January 25' revolution, see Tufekci, Zeynep and Christopher Wilson, "Social media and the decision to participate in political protest: Observations from Tahrir Square," *Journal of Communication*, 62, no.2 (2012), pp. 363–379 or Acemoglu, Daron, Tarek A Hassan, and Ahmed Tahoun, "The power of the street: Evidence from Egypt's Arab Spring," NBER Working Paper (2014).

[3] Lynch, Marc, Deen Freelon, and Sean Aday, "Online clustering, fear and uncertainty in Egypt's transition," *Democratization* 24, no. 6 (2017), pp. 1–19."

in China to monitor and manipulate citizens' online experiences,[4] Gulf governments have sought to refashion online spaces to their own ends rather than silencing them outright. Shutting down or blocking online platforms sends an undeniable signal of censorship that can generate protests of its own accord (the so-called 'Streisand' effect). Careful enforcement of red lines, however, can keep online organizing from getting out of hand while preserving a circumscribed channel for publics to voice selective criticism. Moreover, within these boundaries, social media offer regimes the opportunity for a particular kind of claim-making, encouraging citizens to believe that official statements or narratives enjoy widespread popular support due to the volume of online engagement that authoritarian rulers can generate – whether through the cultivation of "organic" online personalities around state elites, or through the automated simulation of online activity using so-called 'bot' accounts.

This article examines social media manipulation in the early months of the intra-GCC crisis, in which Saudi Arabia, the United Arab Emirates, and Bahrain have mounted an effective blockade of neighboring Qatar over opposition to aspects of Qatari foreign policy, most notably its regional support for various Islamist movements (especially in the wake of the Arab Spring). The events of the crisis present ideal ground to explore online communication and manipulation in the region. Internet and social-media penetration for the Gulf countries are among the highest in the world, while the cross-border war of words forces governments to rely on less direct methods of influencing online narratives. While state manipulation of social media activity is hard to quantify in aggregate, we find suggestive evidence that dedicated institutions within ruling regimes support state claim-making by generating or drowning out specific online narratives. Additionally, we provide evidence that not all top-down narratives resonate equally with online audiences – the more that claims clash with pre-existing beliefs about actors and intentions, the less likely "real" users will relay claims within their own social networks.

By examining the strategies employed by identifiable state elites and institutions within one Gulf country – Saudi Arabia – and providing evidence of these strategies' ability to warp online discourse to regimes' advantage, this article encourages greater consideration of the ways in which new media technologies may empower state surveillance as much if not more than it serves as a catalyst for mass mobilization. Engaging with recent work in political science on strategies of online censorship, it argues that open-access social media platforms such as Twitter are particularly vulnerable to state manipulation, where users have limited resort to understand whether the "trends" they see reflect the aggregate views of real, verified users.

---

[4] Molly Roberts, *Censored: Distraction and Diversion Inside China`s Great Firewall*, Princeton University Press (2018).

Liberating what?

Social media platforms like Twitter, Facebook, their peers and predecessors have facilitated new means of communication and organization over the past decade and a half. Users can post thoughts, article links, content and event information to personal profiles or online groups, potentially building new communities or developing common knowledge even in the face of governments opposed to independent civil society and the free flow of information. The popular association of these platforms with both the 2009 Iranian election protests and the 2010-12 Arab Spring uprisings have motivated a burgeoning research agenda on the interplay between social media use and contentious politics, particularly in the Middle East and North Africa. Infused with revolutionary fervor, one major body of research has focused on the protests themselves, demonstrating empirically how social media help activists solve coordination problems, facilitating mass mobilization even in the face of notoriously repressive regimes.[5] Scholars have also used social media as a prism though which to understand broader patterns of organization and mobilization in society. Zachary Steinart-Threkeld uses online coordination around hashtags during the Arab Spring protests to argue that contentious politics are more likely when there is greater coordination among the periphery of politically active Arab publics, rather than just among a core group of activists.[6] Adria Lawrence, meanwhile, surveys a key Facebook group of Moroccan activists to argue that histories of repression can increase individual willingness to undertake risky political action.[7]

These claims have been subject to repeated critiques, from arguments that social media platforms on their own rarely play an independent role in mass mobilization, to Gholam Khiabany's biting critique of Western narratives' focus on the role of information technology as a way to claim credit for mass mobilization that Western governments themselves have militated against.[8] Likewise, others have noted that beyond intense episodes of contentious politics, and outside of these 'weeks where decades happen',[9] social media platforms are used on a more mundane, day-to-day basis to communicate political opinions. Though less dramatic, these lull periods are where citizens' perceptions of each other, of their rulers, and of their societies are shaped and molded, helping to determine their willingness or reluctance to mobilize when 'critical junctures' eventually arise. In contrast to generally upbeat takeaways of mobilization studies, research on social media's role during these lull periods has generally been gloomier. Tamar Mitts, for example, has substantiated widespread accounts of extremist groups such as al-Qaeda or the Islamic State of Iraq and al-Sham using social-media forums to recruit new members, demonstrating that ISIS-linked accounts attract more engagement from followers in the wake of publicized anti-Muslim acts in the West.[10]

Even beyond organized recruitment tactics, decentralized political networks on social media appear to exhibit significant homophily or cliquishness,[11] mirroring and exacerbating polarization within their societies. Regional application of this work documents Egypt's 2011-2013 descent into ideological polarization between Islamists

---

[5] In addition to those cited above, for an early example see Breuer, Anita, "The Role of Social Media in Mobilizing Political Protest: Evidence from the Tunisian Revolution," Discussion Paper, Deutsches Institut für Entwicklungspolitik 2012; for an early conceptual piece, see Lynch, Marc, "After Egypt: The Limits and Promise of Online Challenges to the Authoritarian Arab State," *Perspectives on Politics* 9, no. 2 (2011), pp. 301-310.

[6] Steinart-Threlkeld 2017a.

[7] Lawrence, Adria. "Repression and Activism among the Arab Spring's First Movers: Morocco's (Almost) Revolutionaries," APSA 2013 Annual Meeting Paper, American Political Science Association 2013 Annual Meeting, 2013.

[8] Gholam Khiabany, "Technologies of Liberation and/or Otherwise," IJMES 47 (2015), 348-353.

[9] Holmes, Amy (2012). "There are weeks when decades happen: structure and strategy in the Egyptian revolution," In: Mobilization: An International Quarterly 17.4, pp. 391–410.

[10] Mitts, Tamar. (2017). From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West, Working Paper.

[11] Halberstam, Yosh and Brian Knight (2013). "Are Social Media more Social than Media? Measuring Ideological Homophily and Segregation on Twitter." Working Paper.

and secularists, showing how rumor mills and echo chambers paved the way for the June 2013 military coup and subsequent democratic deconsolidation.[12] More hopefully, Alexandra Siegel finds that Egyptians following a more ideologically diverse set of elite tweeters tended not to succumb to polarization.[13] In summary, outside of the mobilization literature, political research on social media has focused on the question of polarization, with the weight of evidence leaning towards the thesis that social media reinforce political polarization.

Comparatively fewer studies have considered the potential for governments to shape their publics' use of social media in ways that go beyond isolated arrests or offline harassment of individual bloggers. This is surprising given that political-science framings of the Middle East and North Africa have frequently positioned robust security apparatuses as one of the defining characteristics of the region's persistent authoritarianism.[14] While the coordinating power of social media platforms no doubt took many such agencies by surprise at the outset of the Arab Spring – prompting panicked reactions such as the wholesale shutdown of access to the internet in Egypt – many of these agencies have the resources to acquire and deploy the latest digital techniques in overseeing and censoring digital speech. In a survey of the state-of-play, Seva Gunitsky outlines the ways in which autocratic governments have moved to co-opt the disruptive potential of social media for their own purposes: mobilizing supporters in the face of demonstrations and shaping narratives in ways that "can more easily avoid the appearance of artifice."[15] Research on China has produced some of the most detailed profiles of online censorship strategies, with Molly Roberts documenting the Chinese government's various choices employs to employ tactics of striking fear into individual activists, restricting access to critical views or flooding online platforms with distracting information in overseeing its self-contained online platforms.[16]

Gulf Media and the Gulf Crisis

Regimes in the MENA region – and especially the GCC, where social media usage rates are among the highest in the world – cannot hope to have the same degree of control over open-access platforms where their citizens share photos, message friends, and open informal businesses. Nearly half of survey respondents in Qatar, Saudi Arabia and the United Arab Emirates reported using Twitter accounts in 2017, with few respondents of any age category reporting no social-media use whatsoever.[17] By now, smart phones are ubiquitous across the Arab Gulf states, as are Snapchat videos, Instagram handles, and official Twitter profiles for ministries and ministers alike.

Banning platforms outright, as Turkey has done periodically with Facebook and Whatsapp and Iran has done with Twitter and Facebook, eliminates the potential for social media posts to serve as an "advance warning system" for rulers interested in hearing of specific grievances as they arrive and addressing them in piecemeal fashion to present the image – if not the reality – of government responsiveness to popular concerns. Qatari authorities, for example, have repeatedly backed away from fee hikes or cancelled Western-style concerts in the face of criticism on Twitter, while online grumbling appears to have cued Saudi policymakers to delay or alter

[12] Borge-Holthoefer, Javier et al. (2015). "Content and network dynamics behind Egyptian political polarization on Twitter". In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, pp. 700–711; Lynch, Marc, Deen Freelon, and Sean Aday (2017). "Online clustering, fear and uncertainty in Egypt's transition". *Democratization*, pp. 1–19.

[13] Siegel, Alexandra et al. (2015). "Tweeting beyond tahrir: Ideological diversity and political tolerance in Egyptian twitter networks," SMaPP Lab Papers in Progress.

[14] Eva Bellin. (2004). "The robustness of authoritarianism in the Middle East: Exceptionalism in comparative perspective," *Comparative politics*, 139-157.

[15] Seva Gunitsky, "Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability," *Perspectives on Politics* 1 3, no. 1, March 2015.

[16] Roberts (2018).

[17] "% Who Use the Following Social Media Platforms," *Media Use in the Middle East*, Northwestern University in Qatar, 2017. < http://www.mideastmedia.org/survey/2017/>

aspects.[18] As a result, more upbeat accounts once focused on the new opportunities afforded Gulf citizens to express dissent online; "Twitter for us is like a parliament," one cheerful interviewee for the *New York Times* noted in 2012.[19]

More recently, though, journalistic accounts and anecdotal evidence have highlighted the potential for unaccountable rulers to hire or develop sophisticated tools for warping social media to their own ends. The United Arab Emirates and other Gulf countries have been a frequent setting in more recent years for stories of cyber-intrigue, ranging from expatriate employees jailed for social-media rants to the use of sophisticated, Western-developed spyware to track and target potential activists.[20] Crossing unspoken red-lines – typically deemed insulting to religion or to the monarchies' ruling families – has often resulted in offline efforts at prosecution when Gulf citizens post online under their real identities, even in slightly more liberal Kuwait.[21]

Only recently, though, has evidence emerged of Gulf efforts that extend beyond piecemeal crackdowns to suggest systematic policing and shaping of online spheres. In pioneering work, Marc Owen Jones has documented how online "bots" (automated social media accounts) have been programmed to accomplish a range of political tasks in Bahrain and elsewhere across the Arab Gulf. For example, in one incident around June 21, 2016, Jones documents how a Twitter conversation among Bahraini activists and opposition sympathizers was suddenly inundated with sectarian hate speech in a coordinated flooding attack by numerous 'bot' accounts. In May, 2017, Jones finds evidence of bots retweeting President Donald Trump as he issued favorable remarks about Saudi Arabia during his visit to the Kingdom.

Social Media Moguls

That Gulf regimes should seek to manipulate their citizens' political perceptions via social media is unsurprising. Pinning down the role of any particular set of officials or agencies, however, is a more difficult task, given that such efforts are typically shrouded in a degree of secrecy and may be confounded by the social-media equivalent of "patriotic hacking" – citizens taking to cyberspace to express their support for government actions of their own accord, with organized government efforts taking a backseat role.

The Gulf Crisis has seen no shortage of accusations that state agencies are ramping up online efforts. Saudi and Emirati commentators in both social and traditional media, for example, have repeatedly accused the Qatari government of a protracted anti-Saudi and anti-UAE media campaign, one waged (among other fronts) via social media accounts and hacking efforts.[22] The catch-all term for Qatar's electronic foot-soldiers in these accounts is *khilaya 'Azmi* ("Azmi cells"), a reference to the purported role of Azmi Bishara – Palestinian

---

[18] Leber, Andrew and Charlotte Lysa (2016) "The Banality of Protest? Twitter Campaigns in Qatar," *Gulf Affairs*, OxGAPS 2016; Justin Gengler, "Collective Frustration but No Collective Action in Qatar," Middle East Reporting Project, 2013. <http://www.merip.org/mero/mero120713>

[19] Robert Worth, "Twitter Gives Saudi Arabia a Revolution of its Own," *New York Times*, 20 October 2012. <https://www.nytimes.com/2012/10/21/world/middleeast/twitter-gives-saudi-arabia-a-revolution-of-its-own.html>

[20] Elias Groll, "The UAE Spends Big on Israeli Spyware to Listen In on a Dissident," *Foreign Policy*, 25 August 2016; "Facebook rant lands US Man in Emirati Jail,"BBC, 5 March 2015. <http://www.bbc.com/news/technology-31692914>

[21] "Kuwait: Teacher Faces Jail over Twitter Comments," Human Rights Watch, 20 July 2013. <https://www.hrw.org/news/2013/07/20/kuwait-teacher-faces-jail-over-twitter-comments>

[22] See, for example, Hani al-Dhahiri, "Hal matat khilaya 'azmi am ghayarat mulabisiha?" ["Have the Azmi cells died off or changed their appearance?], *Okaz*, 11 March 2018. <https://www.okaz.com.sa/article/1622491?rss=1; @abonaif00777, "Man hum Khilaya 'azmi?" ["Who are the 'Azmi Cells?"], Twitter moment, December 19 2017. <https://twitter.com/i/moments/943018598485905408>

intellectual, Qatar resident, and director of the Doha-based Arab Center for Research and Policy Studies – in directing the overall media campaign and especially social-media tactics.

The most well-documented allegations of Gulf government efforts to warp the social media environment, however, come from Saudi Arabia, where enough information has been made available to at least sketch out (however roughly) the role played by a key Saudi official in the social-media front of the ongoing Gulf Crisis. Saud al-Qahtani, an advisor to the Royal Court of Saudi Arabia and minister-rank General Supervisor of the Center for Studies and Information Affairs, has been one of the most pronounced official Saudi voices on Twitter since the outset of the Gulf Crisis. Over the course of the past year he has frequently Tweeted out photos or statements mocking or demeaning Qatar and its rulers – often referencing Qatar's small size as a country or hyping anti-Qatar conspiracy.[23] While Qahtani's Twitter profile includes the disclaimer that @saudq1978 is a "personal account that does not represent official views," he has suggested that his online statements *do* reflect official policy. "I am an employee faithfully carrying out the orders of my lord the King and His Excellency the Crown Prince," he replied to one critic on Twitter, while at other times he has tweeted out what appear to be behind-the-scenes information on Saudi government spending.[24]

Qahtani's rise to prominence has been associated with the growing power and influence of Crown Prince Muhammad bin Salman, as Qahtani has frequently joined MBS' entourage in trips abroad over the past several years. A glowing account of MBS by Qahtani himself places their first meeting in early 2015, shortly before King Salman appointed Qahtani to Minister-level rank as head of the Center for Studies and Media Affairs within the Saudi Royal Court.[25] Yet Qahtani has served in the Royal Court in some capacity since at least 2004, with his professional engagement with online media dating back some time. Email exchanges from 2012 through to 2015 between him and the staff of Milan-based cyber-security and surveillance firm Hacking Team surfaced in leaked emails from the company, providing scattered references to efforts by Qahtani to secure the company's services on behalf of the Kingdom and his office within the Royal Court.[26]

Evidence suggests that Qahtani, and the office he leads, engages multiple means of shaping online conversations. As noted, Qahtani has a prominent Twitter presence in a personal capacity. He can drive Twitter traffic directly, rallying online followers as any social media "influencer" might – early in the Gulf Crisis, for example, he created a hashtag aimed at building a ``black list" of Saudi citizens or other individuals in the Gulf who might sympathize with Qatar during the crisis.[27] Accounts of Saudi Arabia's media output on the Gulf Crisis likewise suggests efforts at dragooning other prominent social media personalities into relaying regime claims. Anonymous Twitter account @mujtahidd claimed that MBS had given Qahtani a mandate to develop online surveillance and manipulation capabilities, additionally empowering him to supersede the Minister of Culture and Information in regulating the Kingdom's journalistic output.[28]

Of particular interest is whether Qahtani is able to "signal-boost" beyond relying on his immediate followers or those of potential surrogates. Early in 2017, Saudi journalist Turki al-Ruqi posted an extensive attack on Qahtani's character and media activities, accusing him of building up an "army" of hackers and "hacking tools"

---

[23] One such theory included an alleged Saudi plan to drive a canal across the base of the Qatari peninsula, rendering Qatar an island. @saudq1978, Twitter post, 15 April 2018. <https://twitter.com/saudq1978/status/985533292366704640>

[24] @saudq1978, Twitter reply, 17 August 2017. <https://twitter.com/saudq1978/status/898273541367451648?lang=en>

[25] "Saud al-Qahtani yaktub: Kayfa t'aml m'a al-'amir Muhammad bin Salman?," ["Saud al-Qahtani writes: What's it like to work with Prince Muhammad bin Salman?"], *al-Riyadh*, 4 April 2018, <http://www.alriyadh.com/1673253>

[26] Inter alia, "RE: Requist [sic]," Hacking Team Archive, Wikileaks, 29 June 2015. <https://wikileaks.org/hackingteam/emails/emailid/1144691>

[27] @saudq1978, Twitter post, 17 August 2017, <https://twitter.com/saudq1978/status/898257245183463424>

[28] @mujtahidd, Twitter post, 30 May 2017. <https://twitter.com/mujtahidd/status/869726907620831232>

to harass critical media outlets within the Kingdom.[29] Later, in August, another anonymous Twitter account with the name *Tarikh wa Dhikriyat* (History and Memories) tweeted out a series of screenshots purporting to show a user associated with Saud al-Qahtani's screenname and email address (saudq1978) requesting account-access and -creation tools on the website Hack Forums. Qahtani and his office certainly have access to sophisticated tools of social media analysis – he periodically periodically disseminates reports on Twitter traffic in the region from the center he leads within the Royal Court, usually to substantiate claims of Qatari social-media manipulation (i.e. "Azmi cells").[30]

Speculation about Qahtani's role can be hard to substantiate, especially as he is only the public face of much larger state efforts to monitor and censor media outlets in the Kingdom, online or off. Offline efforts at directing prominent individuals' online commentary, as suggested by al-Ruqi and other accounts of Saudi Arabia's media environment, would be hard to detect short of a strict party line being enforced across a range of personalities.[31] Still, available data on Twitter "hashtag-campaigns," especially those involving Qahtani and prominent media outlets, can provide some initial clues as to the shape of the online "war of position" in the Gulf. Does Qahtani's social media presence function merely as an outsize "influencer," his associations with the halls of power a signal as to where patriotic Saudi citizens should direct their online activity? Or do the digital tools allegedly under his control lend his office additional leveraging in generating what might otherwise appear to be waves of "organic" social media activity?

---

[29] @turkialroqi, Twitter post, 26 February 2017. <https://twitter.com/turkialroqi/status/835869428675915776>

[30] "Qahtani: 23 thousand dragooned by Qatar to attack Saudi Arabia," Al Arabiya, 7 July 2017. https://www.alarabiya.net/ar/saudi-today/2017/07/06/القحطاني-23-ألف-حساب-جندتها-قطر-لمهاجمة-السعودية.html; Marc Owen Jones has raised questions about the accuracy of Qahtani's analysis. Marc Owen Jones, "Saudi Royal Court Advisor Saud Al-Qahtani is Using Bad Science to Inflame Tensions With Qatar," Personal website, 23 August 2017. <https://marcowenjones.wordpress.com/2017/08/23/saudi-royal-court-advisor-saud-al-qahtani-is-using-bad-science-to-inflame-tensions-with-qatar/>

[31] Anonymous comments have suggested that the Saudi government communicates key points of desired narratives through private messaging apps. Ahmed al-Omran, "Gulf media unleashes war of words with Qatar," *Financial Times*, 3 August 2017. <https://www.ft.com/content/36f8ceca-76d2-11e7-90c0-90a9d1bc9691>

Tracing Twitter Interventions

One spate of online activity from August 2017 allows us to provide some initial answers to these questions. In the afternoon of August 21, the hashtag إرحل_يا_تميم# ("Get Out Tamim!") began to circulate among some Gulf-based Twitter accounts, initially reacting to a news article that claimed some Qatari citizens had scrawled "Get Out, Tamim!" on a mural of the Qatari Emir in Doha. Soon after, a competing hashtag – *Tamim fi Kul Makan* ("[Emir] Tamim is everywhere") - arose to counter this narrative, with ostensibly Qatar-linked users tweeting their support for the Emir and seeking to deny that there was significant political dissent within the country.
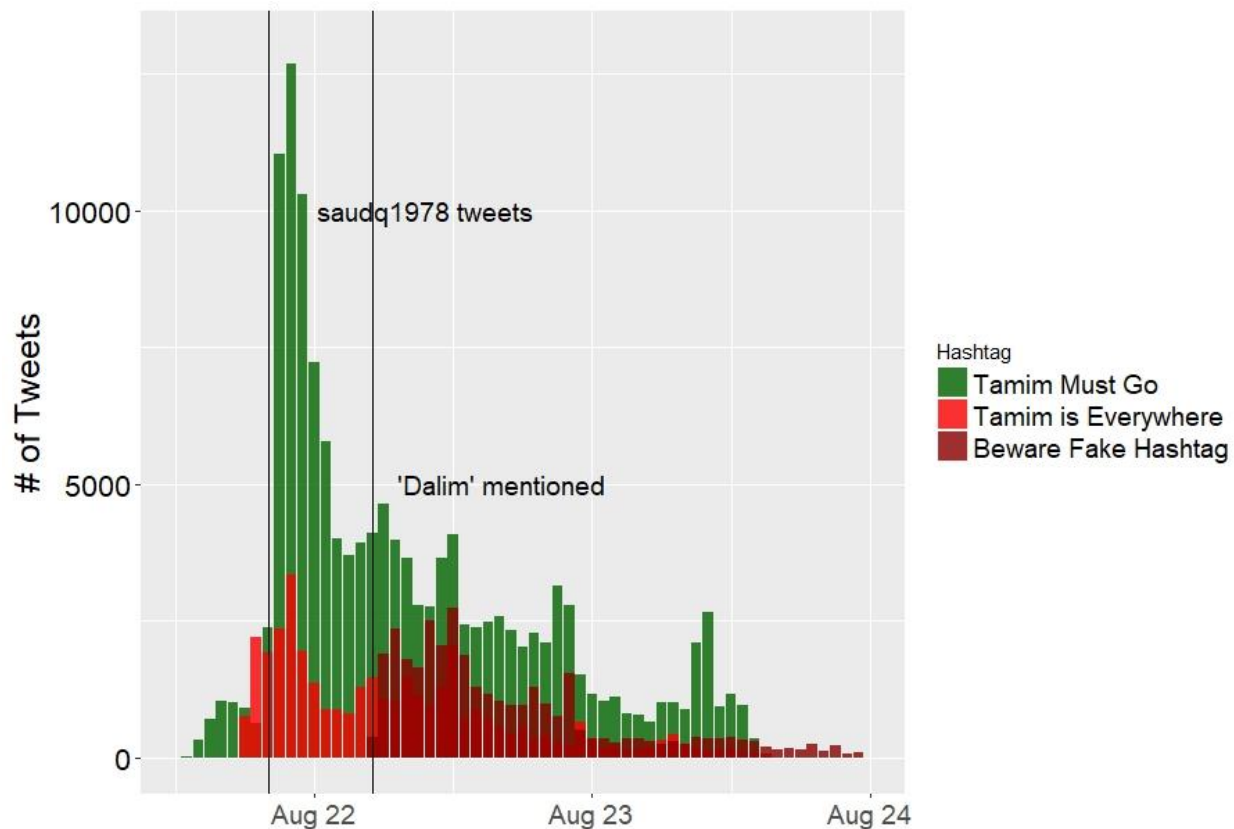

Figure 1: Hashtags around "Get Out, Tamim!" and notable interventions.

Notably, Saud al-Qahtani made a single yet seemingly influential intervention into this conversation at around 11:00pm Riyadh time (see Figure 1). In a single Twitter thread, he claimed that إرحل_يا_تميم# was a "trending topic" inside Qatar – indicating widespread activity on the hashtag within the country – citing Twitter's own list of "trending topics" in Qatar to make this claim. His intervention was followed by a dramatic spike in activity on the إرحل_يا_تميم# hashtag. This in turn provoked a hashtag response by prominent Qatari-affiliated accounts, *La Musharika fi Hashtagat Mashbuha* ("No participation in suspicious hashtags"), in which Qatari users explicitly referenced Qahtani (by his nickname "Dalim") and attempts at social-media manipulation.[32]

---

[32] See, for example, the explanation offered in @3issaqtr, Twitter post, 22 August 2017. <https://twitter.com/3issaqtr/status/899864594230853632>. Marc Owen Jones, analyzing Qahtani's activity in the blog post above, speculates that a variety of factors led to "Get Out, Tamim!" being listed as a trending topic inside Qatar beyond this being an accurate reflection of Qatari public opinion. In particular, he finds that a substantial portion of Twitter accounts associated with the hashtag campaign originated in Saudi Arabia.

We downloaded tweets associated with all three hashtag campaigns, given that each entailed a broad claim or counter-claim about public opinion in Qatar towards the ruling family. Statistics on each of the three hashtags are presented in Table 1. The Twitter REST API allows us to download only a fraction of

Table 1:

Hashtag	Total participants in sample	Tweets in sample	% retweets	% traffic driven by *influencers**	% of participants who are influencers	% traffic driven by *core influencers*
#إرحل_يا_تميم	52,445	258,045	89.4	75.4	0.45%	55.1
#تميم_في_كل_مكان	14,454	61,292	84.3	61.2	0.66%	36.9
لا_مشاركة_في_هشت #جات_مشبوهة	14,300	55,412	87.7	67.2	0.45%	34.8

*By *influencers* we mean the top two percentiles of retweeted accounts within the hashtag. Since retweets exactly repeat verbatim the original tweet, these accounts may be understood as the most promoted or amplified by others.

the tweets generated per hashtag.[33] Having downloaded these hashtags across the entire duration of the episode, however, the differences in sample sizes are reflective of the differences in overall traffic generated by each hashtag. We can see that the Qahtani-promoted hashtag #إرحل_يا_تميم generated over a quarter-million tweets, compared to the two response hashtags, whose combined total of tweets was less than half that sum. Likewise, #إرحل_يا_تميم involved 52,445 participants, dwarfing the turnout for both response hashtags. Evidently a 'manufactured' or misleading hashtag can generate far more attention than hashtags seeking to correct them.

Table 1 further reports the percentage of tweets that were in fact simply the verbatim repetition of tweets by others (*retweets*). Retweets make up the vast majority of tweets, ranging from 84% to 89% of traffic across these samples. Indeed, per hashtag most participants limit themselves only to retweeting what others say, with only 22-32% of participants composing original tweets of their own.

Among this minority, moreover, the rate at which original tweets are retweeted is profoundly skewed. Indeed, we find for #إرحل_يا_تميم that the top 2 percentiles of retweeted accounts generated over three quarters of total traffic. In other words, 194,566 tweets in the #إرحل_يا_تميم sample can be traced back to just 238 participants, by virtue of the fact that whatever they said was retweeted at such a high rate. The two response hashtags were similarly dominated by a small set of 'influencers'. We are not aware of any study thus far that has highlighted so clearly the incredible inequality with which voices on social media are amplified.

The consequences of this skewness for the project of manipulating political conversations are profound. If participants' voices are equally amplified, a regime wishing to threaten or co-opt participants to influence a conversation would have to track down tens of thousands of people. When just a few hundred participants drive the discourse, however, a regime seeking to manipulate the conversation need only threaten or co-opt these highly influential actors – a much more feasible task.

These 'influencers', moreover, are not isolated actors. Since we know from our data who retweeted whom, we can calculate a retweet network graph of the #إرحل_يا_تميم conversation. Using standard community detection algorithms that exploit variation in the density of retweeting,[34] we identify communities of 'influencers' participating in the #إرحل_يا_تميم conversation. We plot all such communities that were at least 10% the size of the largest detected community, corresponding to 141 of 238 influencers (Figure 2). Reviewing a random

---

[33] See Steinert-Threlkeld, Zachary, "Twitter as Data," Working Paper, 2017. Studies have shown that samples obtained via the REST API are non-random in certain ways, though the selection rule is unknown.

[34] See Lynch et al. (2017) for another example of community detection using retweet network data from Twitter.

sample of tweets from each of these communities, we are able to identify the 'theme', ideological or otherwise, that distinguishes each community. Firstly, there is a community of advertisement bots (highlighted in black), which are apolitical automated accounts that detect trending hashtags and hijack them to advertise products or services. The remaining three communities are all political, but of different varieties. The largest of the three, highlighted in yellow, generally expresses distaste for Emir Tamim, but sympathy for the Qatari people. The second largest of the three, highlighted in light green, expresses contempt for both Emir Tamim and Qatar generally. The third and smallest of the three, highlighted in dark green, is a small network of accounts associated with the pro-Saudi news agency Okaz. Note that none of the influencer communities takes a pro-Tamim stance, suggesting that the hashtag as a whole was dominated by anti-Tamim sentiment, and constituted variations on the theme of Saudi Arabia's official stance against Qatar as of August, 2017.
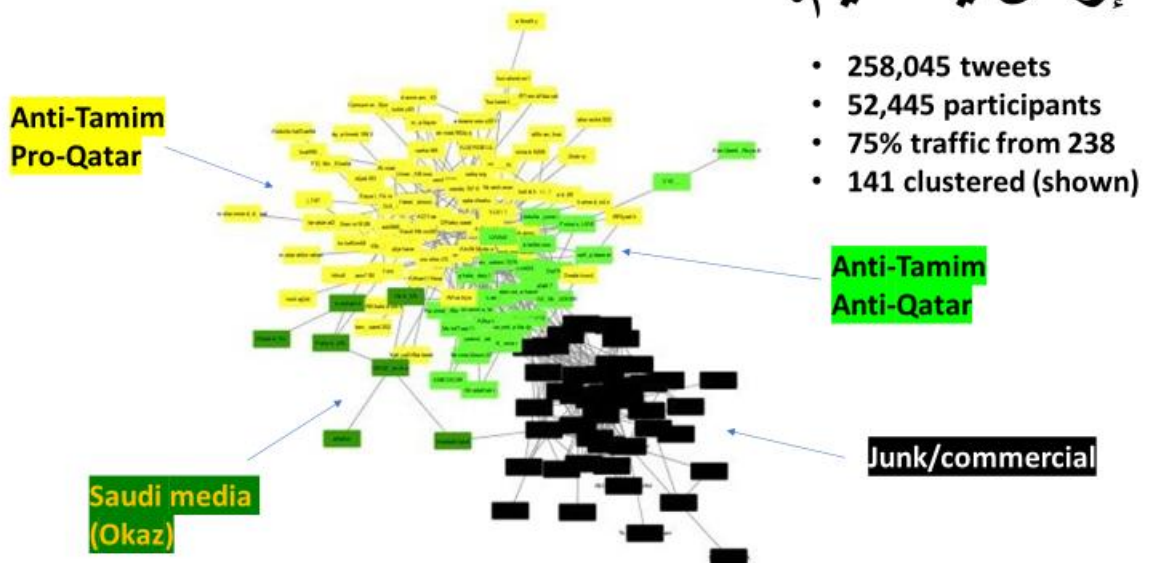

Figure 2: Influencer communities active on ارحل يا تميم # in August, 2017. Tweet data were downloaded and analyzed by the authors via Twitter's REST API.

Finally, the degree to which influencer communities are consolidated as a 'core' appears to be higher for ارحل_يا_تميم# than the two response hashtags. We calculate the 'core' of each hashtag as a group of influencers who enjoy the additional affirmation of having been retweeted by at least one other influencer. In this sense, the 'core' of each hashtag is a kind of elite club of social media accounts whose members are both widely retweeted by the public, and vouched for by at least one other elite club member. Within the manufactured hashtag إرحل_يا_تميم#, more than half (55%) of influencers belong to the 'core'. By comparison, the cores of the two response hashtags constitutes 34-36% of the influencer community, suggesting these influencers were somewhat less orchestrated, and more decentralized, in their behavior.

So far we have pointed to several suspicious features of the إرحل_يا_تميم# conversation that raise concern of manipulation. The conversation appears to be led by a small set of highly influential social media accounts, and many of these accounts seem to be ideologically related to each other in the sense that they retweet each other's

messages. As further evidence of potential manipulation, we can exploit what we know about the timing of tweets. Figure 1 depicts the number of tweets tweeted per minute, per hashtag, throughout the episode. The spike in activity that follows Qahtani's brief engagement with the hashtag conversation is certainly dramatic, and the 6,300 or so retweets of his statement alone would barely account for half of the initial spike. Still, Qahtani's online popularity following could simply mean that his engagement sparked a chain-reaction of sympathetic tweets using the same hashtag – further analysis is needed to compare the profiles of those tweeting to known profiles of "bot" accounts.

The Limits of Manipulation?

At the same time, the agency of social media users cannot be assumed away – elite attempts at framing online discourse can only succeed so far as claims made resonate with pre-existing narratives about the claim in question. Here, as an example, we examine a pair of hashtags (totaling ~560,000 tweets) that appeared in early July following news of Saudi Arabia, the UAE, Egypt and Bahrain adding a series of individuals and organizations to their respective terror watch lists – alleging direct or indirect Qatari ties in either case. This in turn gave rise to the hashtags *Qatar Tumawal al-Irhab* ("Qatar funds terrorism") and *Tamweel Qatar lil-Irhab* ("Qatar's funding of terrorism"). While a major limitation of Twitter data is the fact that each Tweet can only carry but so much ``content" to analyze in 140 characters, several of the tweets bearing these hashtags provide us with a measure of useful information: exactly which alleged terror groups or individuals various news outlets and social media users associated with Qatari support. We can then examine which kinds of tweets were then retweeted by other users, which we interpret as evidence that the claimed connection resonated with users as at least plausible.
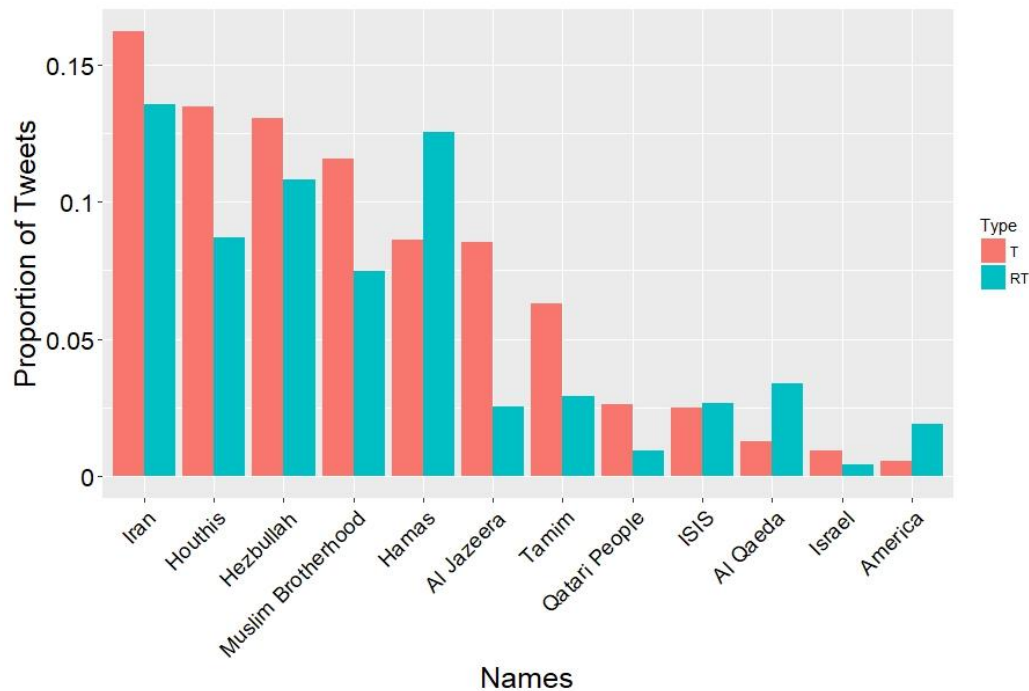

Figure 3: Proportion of "Qatar supports terror" tweets by topic mentioned. T is for original tweets, RT is for retweeted content.

By combing through the tweets through a measure of sampling and –a list of potential "terror affiliates" was created, along with related subjects included in the tweets. This resulted in a distribution of tweets and retweets according to subjects mentioned (some tweets mentioned more than one subject). We anticipated that some claims about Qatar's foreign policy associations might "resonate" more with Twitter users than others as a sign of "support for terror." In particular, a number of initial Tweets cited Qatar's ties to Palestinian militant group

and de facto ruling power in the Gaza Strip Hamas – an organization listed as a Foreign Terrorist Organization by the United States from 1997, but one with at least a measure of popular legitimacy in some quarters of the Arab world as a source of armed resistance to Israel.[35] With the potential for dissonance between popular views of Hamas and elite online claims that Qatar's contacts with Hamas and its leaders amounted to "support for terrorism," we generated 95% confidence intervals for the ratio of Tweets to Retweets to determine whether we observed a significantly *higher* or *lower* level of retweeting of various topics relative to random chance (Figure 4).[36] In line with our prior beliefs, there was only around 1 retweet of a Hamas-related tweet for every 10 original tweets – fewer than we would expect from random patterns of retweeting. In all, this provides suggestive evidence that online discourse is not infinitely malleable – elite cues that seek to shape social-media discourse must navigate existing frames of reference and associations rather than making them up out of whole cloth.
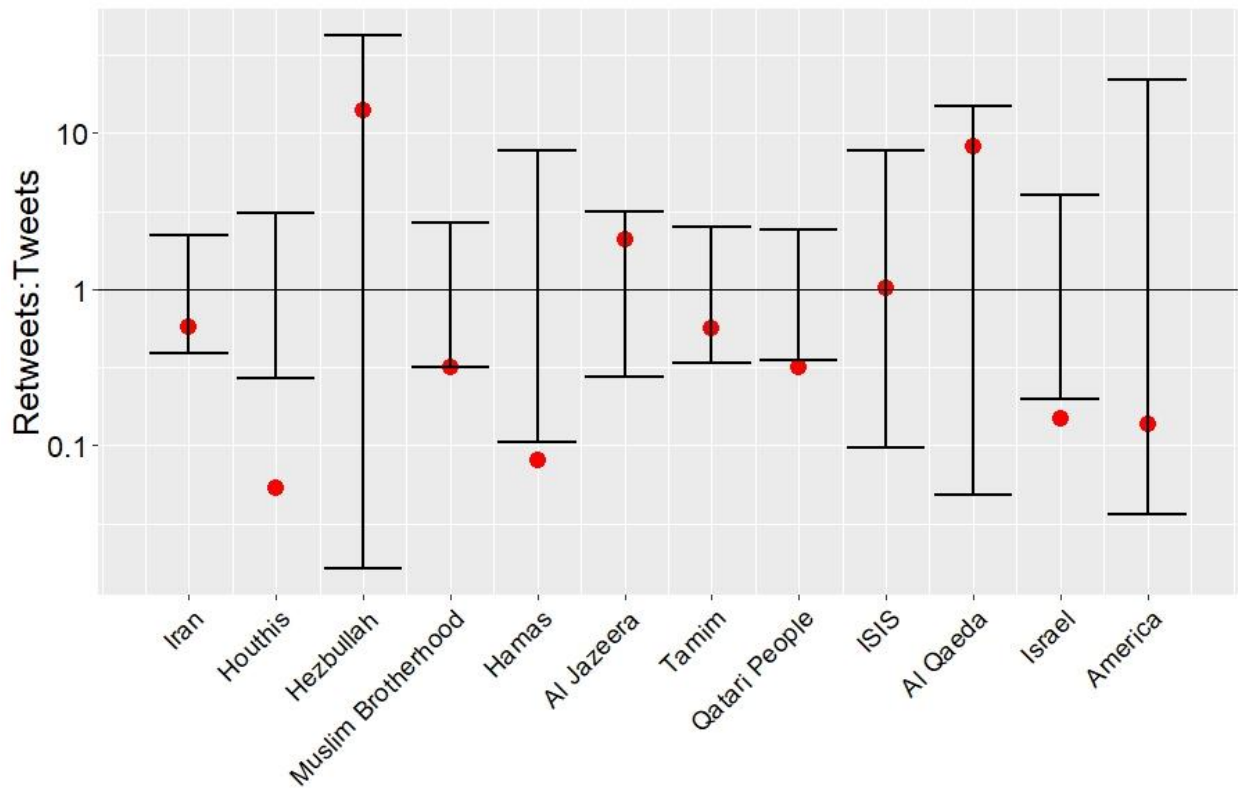

Figure 4: Actual Tweet-to-Retweet ratios for various topics given in red. 95% confidence intervals displayed in black.

Conclusion

Despite the initial promise of social media platforms as the basis for diffuse, even democratic discussion of an unending range of issues, developments over the past decade suggest that even platforms made possible by "liberation technology" risk falling victim to Michel's iron law of oligarchy. Entities with significant offline resources and organizational power – especially authoritarian states – have both capacity and incentive to try

---

[35] "Spring 2012 Survey," *Global Attitudes Project*, Pew Research Center, July 2012.
[36] We took 2000 random samples from our pool of original tweets, with replacement, and calculated the values that bounded 95% of the resulting tweet-to-retweet ratios for each topic.

to shape online discussions in their favor, biasing online free expression towards mindless repetition and careful self-censorship. Drawing on recent research on Chinese censorship strategies as well as our own analysis of Twitter activity concerning the Gulf crisis, we propose several channels along which Gulf governments may be seeking to steer online discourse: building a direct social-media presence for key state officials or political leaders; engaging in offline coordination between traditional media figures and other social-media "elites" to elevate the prominence of certain narratives online; and investing in tools of automated manipulation such as "bot" accounts that can repeat, retweet, or drown out certain voices as needed.

While we provide suggestive evidence as to influence of one particular Gulf official – Saudi Advisor to the Royal Court Saud al-Qahtani – further evidence is needed to understand the comparative capability and willingness of Gulf governments to warp social-media environments, either to influence public opinion towards' citizens own governments or towards potential rivals. Given ongoing incentives to instrumentalize social-media discourse in this fashion, the next few years may witness a small-scale tragedy of the digital commons as otherwise engaging and insightful users depart the increasingly uncertain digital reality of platforms such as Twitter for smaller circles (such as WhatsApp groups) where the range of views present may be smaller but the provenance of those speaking more certain.