International Journal of Communication 15(2021), 1173–1199 1932–8036/20210005

Electronic Armies or Cyber Knights? The Sources of Pro-Authoritarian Discourse on Middle East Twitter

ALEXEI ABRAHAMS
Shorenstein Center, Harvard University, USA

ANDREW LEBER
Harvard University, USA

A decade after commentators hailed social media as "liberation technology," researchers, analysts, and journalists are now likelier to emphasize the "dark side" of social media as offering autocratic regimes a potent toolkit for information control. Although important work has established how social media can advantage authoritarian regimes, we suggest that much proregime content stems from the decentralized efforts of active regime supporters rather than top-down manipulation by a handful of state officials. We demonstrate the plausibility of this claim by finding little evidence of top-down manipulation within several hashtags that align with regime narratives in Saudi Arabia. Furthermore, we find little evidence of bot presence on 279 prominent hashtags from across the Middle East and North Africa. Regimes may lay the groundwork for online displays of support, but they hardly have a monopoly on proauthoritarian rhetoric.

Keywords: social media, bots, Saudi Arabia, misinformation, Middle East

Over the past decade in the Middle East, social media technology has gone from being praised as "liberation technology" (Diamond & Plattner, 2012; Tufekci & Wilson, 2012) to being condemned as a tool of authoritarianism (Bradshaw & Howard, 2019; Howard & Bradshaw, 2018; Howard, Wooley, & Calo, 2018; Marczak, Scott-Railton, McKune, Razzak, & Deibert, 2018; Michael, 2017). Arguably, a substantive driver behind this opinion shift is a body of scholarly and journalistic evidence documenting the prevalence of political "bots" on Middle Eastern social media threads—armies of centrally commanded social media accounts that distort political discourse. Scholars (including the authors) have developed bot detection techniques and documented incidents of bots systematically promoting specific Arabic, Turkish, or Persian hashtags (M. O. Jones, 2019a; M. O. Jones & Abrahams, 2018; Stubbs & Bing, 2018), warping online narratives to the benefit of autocratic rulers. Fear of electronic bot armies has also driven considerable media and tech sector attention, with both Twitter and Facebook periodically announcing purges of suspect accounts (Gleicher, 2019; Twitter, 2019a, 2019b).

Alexei Abrahams: alexei_abrahams@alumni.brown.edu
Andrew Leber: andrewmleber@g.harvard.edu
Date submitted: 2020–06–17

In this article, however, we revisit this perception of bot infestation with new Twitter data and a mixed-methods research design. We begin by performing an in-depth qualitative process trace and statistical analysis of several politically charged hashtags relevant to Saudi Arabia. The Saudi regime has been closely identified with efforts to manipulate Twitter content, going so far as to recruit spies within the company itself (Nakashima & Bensinger, 2019). This renders the Kingdom a "most likely" case for observing manipulation on politically relevant and prominent hashtags. After closely examining three prominent and politically divisive hashtags from 2018 related to Saudi Arabia, however, we fail to reject the null hypothesis that proregime narratives are primarily "organic."

We then run a broad sweep for bots across 279 trending Middle East and North Africa (MENA) hashtags collected by the authors from October 2019 to January 2020, coinciding with a fresh wave of protests across Lebanon, Iraq, and Iran, and a Turkish land invasion of Kurdish Syria. Adapting two credible bot detection methods from the literature, we find in both cases that, on average, roughly just 5–9% of users per hashtag were plausibly bots (and for reasons explained below, this likely constitutes a substantial overestimate of the true rate); smaller, fringier hashtags—not the highest-volume—are "bottier."

These findings lead us to hypothesize, in contrast to recent scholarship, that much of what we observe in the way of proauthoritarian speech on Saudi (or Gulf, or, less cautiously, Middle Eastern) Twitter may be the result of organic activity driven by influential accounts that have built up their own followings by toeing the party line—perhaps voluntarily, or else by responding to pressure from authorities. We encourage scholars and journalists to shift the emphasis of their future investigations from searching for mass electronic armies of bots to investigating instead this rarified class of influential "cyber knights"—their motivations, tactics, and the pressures they respond to online and off as they shape political discourse in the region.

Electronic Armies or Cyber Knights?

Circa 2011, during the height of the Arab Spring, social media platforms like Facebook and Twitter were exploited to great effect by social movement activists to coordinate and mobilize protesters across the Middle East (Diamond & Plattner, 2012; Tufekci & Wilson, 2012). For a moment, social media seemed the darling of Western liberalism, a technological innovation to help deliver the world's final wave of democratization and bring an end to history once and for all. Just a half-decade later, however, in the wake of Brexit, the election of Donald Trump, and the Cambridge Analytica scandal, the euphoria surrounding social media metastasized into an equal and opposite hysteria amid allegations that malicious actors have exploited social media to surveil and mislead citizens with targeted propaganda and political disinformation (Allcott & Gentzkow, 2017; Davies, 2015; Nadler, Crain, & Donovan, 2018; Waltzman, 2017).

Paralleling this disenchantment, a wave of scholarship and journalism focused on the MENA region has documented with growing dismay how authoritarian regimes, having survived the initial onslaught of protests facilitated by social media, increasingly appear to have adapted to its existence, deploying considerable resources to contest and disrupt revolutionary narratives online (Greenberg, 2019; M. O. Jones & Abrahams, 2018; M. O. Jones, 2019b; Leber & Abrahams, 2019; Patin, 2019; Ritzen, 2019). Although scholars have emphasized a nuanced understanding of the Internet's overall effect on authoritarianism (Weidmann & Rød, 2019), recent articles about MENA-region Twitter speculate that the platform's repressive

tendencies now outweigh its liberating potential (Abrahams, 2019). Within this scholarship, research has particularly focused on the role of automated accounts, or "bots," acting in concert to advance particular political agendas. Professor Marc Owen Jones, undoubtedly the original "bladerunner" of Gulf Twitter, has documented as early as 2016 how centrally commanded accounts tweet together to promote authoritarian narratives or drown out human rights conversations (M. O. Jones, 2016, 2019b).

The deployment of bots to manipulate the discourse makes good sense as an act of authoritarian self-preservation. Platforms like Facebook and Twitter, with their public-facing conversation threads, proved particularly threatening to authoritarian regimes during the Arab Spring. Apart from helping activists plan mobilization, the upvoting features common to these platforms allow citizens not only to observe radical political speech, but to discover—by seeing how many likes or retweets a post received—to what degree their political grievances may be shared by others (DiResta, 2016). This disrupts the "preference falsification" by which citizens under authoritarianism display outward support for the regime yet are privately critical (Kuran, 1997; Wedeen, 1999), potentially generating online and offline mobilization (Lohmann, 1994; Weidmann & Rød, 2019).

For authoritarian regimes that do not design and operate their own social media platforms, such as those of the Arab Gulf monarchies, policing online spaces becomes more difficult because regimes generally lack the ability to selectively censor and delete content.[1] These regimes are restricted to flooding existing online conversations, drowning out opposition voices alternative content, or inflicting fear in would-be critics through online harassment or offline punishments.[2] Bots can certainly prove useful on both fronts, whether by "hijacking" an existing hashtag or online narrative with proregime messaging, generating entirely new conversations (with or without inspiring organic participation) to dwarf the scale of opposition narratives, or by harassing activists with demoralizing messages (Leber & Abrahams, 2019; Tucker, Theocharis, Roberts, & Barberá, 2017). These tactics have been documented under a wide range of authoritarian regimes and other repressive contexts, from Russia to Bahrain to Venezuela (Forelle, Howard, Monroy-Hernández, & Savage, 2015; M. O. Jones, 2013; Stukal, Sanovich, Bonneau, & Tucker, 2017).

This notion, however, that Twitter has been "captured" by authoritarian regimes seeking to preclude another wave of protests should be met with skepticism at two levels.

First, it is not out of the question for influential social media users to champion official narratives of their own accord. Although many individuals no doubt conceal their true, critical preferences under authoritarianism, some contingent of individuals may harbor genuinely proregime sentiments, or opportunistically express this sentiment to curry favor with authorities (Gerschewski, 2018, pp. 655–659). Even under the repressive regime of Saddam Hussein in Iraq, thousands of citizens volunteered for various paramilitary organizations despite dangerous conditions, even after the 1991 uprising against Saddam's rule (Blaydes, 2018). In the case of Egypt, for example, many prominent liberal Egyptian activists applauded

[1] Note that there are some ways around this, as seen in allegations that Facebook has deleted some pro-Kurdish pages after coming under pressure from the Turkish government (Spary, 2016).

[2] The distinction is from Margaret Roberts, with censorship or "friction" as a third strategy of authoritarian censorship (2018).

the coup and proceeded to whitewash the army's brutal crackdown on Islamists in the ensuing months (Fahmy & Faruqi, 2017).

It follows that expressions of support for authoritarian regimes on social media, now as then, may genuinely reflect popular sentiment. Furthermore, it is important to recall that social media platforms, and the notability or notoriety that accompany their most (in)famous users, are perfectly capable of incentivizing malicious online behavior all on their own. Andrew Marantz (2019), in a deep dive into the world of alt-right social media personalities in the United States, highlights the ways in which numerous online "edgelords" are motivated by fame, profits, or deep grievances against aspects of U.S. society (including outright racism) to spread fake news and latch onto yet-more-infamous Internet personalities such as current U.S. President Donald Trump. Tech writer Venkatesh Rao (2020) goes so far as to describe the clash of ideas in digital spaces as an "Internet of Beefs" (IoB), in which influential "knights" of Twitter flame wars rally their followers into digital battle largely for the joy of online "combat" and the fame that accompanies it.

Second, although authoritarianism certainly warps online speech, it does not necessarily require sophisticated techniques to do so. Most governments on either side of the Persian Gulf either censor social media platforms outright or have passed stringent cybercrimes laws that criminalize spreading online narratives that contain misinformation or that threaten national security—categories easily defined to suit the occasion in terms of forbidding "undesirable" free speech (Duffy, 2014). As discussed below, the highly skewed distribution of influence on Twitter implies a rarified class of "influencers" that increasingly appears to be the target of regime intimidation or co-option (Abrahams & van der Weide, 2020). Although Jennifer Pan and Alexandra Siegel (2020) find that offline repression fails to deter online criticism in Saudi Arabia, this may have resulted from a previous perception that Twitter dissent alone was insufficient to land a Saudi citizen in trouble. Since 2017, however, even minor criticisms have been met by harsh state repression (Khashoggi, 2017). Government crackdowns on a particular form of critical discourse can, in turn, empower other individuals to "ride the wave" of the moment by mobilizing in support of the state, even if they turn on the regime at some future point.[3]

Together, these two possible factors constitute our null hypothesis, namely, that proauthoritarian narratives on Middle East Twitter are basically "organic" (authentic, voluntary)—if not in the broad sense (citizens genuinely support their authorities), then at least in the more narrow sense (citizens feel pressured by repressive offline laws and policies to curb criticism and effuse proregime sentiment). The alternative hypothesis is to largely distrust online narratives under authoritarianism as substantially overrun with centrally commanded accounts (henceforth, "bots").[4]

[3] Eman Alhussein (2019) describes how this dynamic applies to Saudi religious conservatives in the 1980s and Saudi nationalists today.

[4] For the sake of brevity, we use "bots" to refer both to accounts centrally commanded by a computer program or "sock puppets" (operated by human beings; Gorwa & Guilbeault, 2018).

Process Tracing of Saudi-Related Hashtags

To adjudicate between our hypotheses, we begin by undertaking a case study of online mobilization related to a single country: Saudi Arabia. Although a study based largely on a single case study might appear problematic, doing so is more justifiable when we seek to disconfirm a broad hypothesis (Gerring, 2007): that large-scale Twitter activity under authoritarianism is driven by coordinated, inauthentic "bot" activity.

First, Saudi Arabia has emerged as an important site for theory generation and data collection about digital authoritarianism, given that media accounts, tech company analysis, and academic investigations have all suggested a significant Saudi capacity for social media manipulation and an intention to use that capacity (Benner, Mazetti, Hubbard, & Isaac, 2018 M. O. Jones, 2019a; M. O. Jones, 2019b; Leber & Abrahams, 2019; Nimmo, 2019). Furthermore, unlike China, Saudi Arabia is more representative of state-led efforts at top-down social media manipulation as the Kingdom does not design and operate its own social media platforms and hence cannot directly censor content (Pan & Siegel, 2020). Second, analysis of Saudi social media activity has emphasized the role of bot activity to the point that manipulation threatens to become the default explanation for any online content that accords with government narratives; a failure to reject the null here would encourage us to rethink how we distinguish between the role of direct (bot) manipulation and indirect (red-lining and repression) manipulation of online content, and how we design research projects around this distinction. Third, events of the past few years have generated considerable digital content from ostensibly Saudi accounts, following efforts by Saudi Arabia's current ruler, Muhammad bin Salman, to quell internal dissent, dominate foreign rivals, and embark on ambitious policies of social and political transformation.

Within Saudi Arabia, we focus on the use of Twitter for three reasons: data is readily available (compared with Snapchat or private WhatsApp groups), the open architecture of the platform allows the formation of shared meanings about the Saudi government, and Twitter is widely used by a broadly representative portion of the Saudi population (Northwestern University in Qatar, 2019).

Building the Data Set: Twitter's REST API

All Twitter data used in this article were acquired by the authors themselves by querying Twitter's free-tier REST API.[5] As tweets are tweeted around the world, Twitter ingests them into their in-house database. Any researcher with a Twitter developer account can query this database up to roughly 10 days in the past for all tweets mentioning a specific word or combination of words.[6] By keeping an eye on events transpiring in the region, we were able to query Twitter's database for trending hashtags around the time they trended, ensuring that we followed each hashtag from start to finish.

[5] For a general overview of how Twitter API data are acquired and applied to social science questions, see Zachary Steinert-Threlkeld's overview (2018).

[6] Twitter imposes some restrictions on what tweets may be retrieved from its database. For more information on this, see (Steinert-Threlkeld, 2018).

Bot Detection

To detect bots, we rely on the "anomaly detection" approach pioneered by Marc Jones (Abrahams & Jones, 2018; M. O. Jones, 2019b). This method is perhaps best understood as an innovation over methods that infer a Twitter account's (in)authenticity based on its own characteristics. The DFRLab (2017), for example, describes 12 "telltale" signs that a Twitter account is a bot, including that the account tweets "frequently" (more than 50 tweets per day); primarily engages in retweeting or quote-tweeting, rather than producing original content; and so on. In the Middle East, however, citizens may fear reprisal for posting to social media about political issues. Consequently, even a real user might attempt to conceal his or her identity by using stock imagery, or might find it safer to retweet influential accounts rather than using his or her own words. As for tweeting frequency, it makes sense that citizens frustrated by being unable to talk about politics in the workplace or coffeehouse might tweet quite vociferously on social media—especially during the course of dramatic political events. Altogether, then, classification of accounts based on their own characteristics is dangerous for our context and may lead to many false positives.

The birth anomaly detection method, by contrast, is a group-based detection method. Group-based detection methods look for multiple accounts that move together in a manner that defies probability. Whereas detection based on own characteristics may be problematic on Middle East Twitter for the reasons listed above, group-based detection makes good theoretical sense. In particular, bot-herders hoping to impact political discourse on a given topic (hashtag) need to deploy many accounts to tweet in tandem, both to win more attention and to fabricate the sense that a particular point of view is widely held. These "electronic armies" then become recognizable to the researcher by their high degree of correlation with each other. Thus, while it may be totally unremarkable that an account was born on June 10, 2017, it may be statistically improbable that, say, 100 accounts tweeting on a particularly partisan political hashtag trending in the Arabian Gulf were all born on June 10, 2017 (Abrahams & Jones, 2018).

To operationalize this intuition, we identify for each hashtag all the unique accounts that tweeted messages containing that hashtag. Among the metadata for each of these accounts is the date of account creation ("birth"). We proceed to plot the number of account births per day for all days since the earliest-born account. We then calculate a global threshold for anomalies at two standard deviations about the global mean number of births per day. We classify accounts born on days on which the number of births exceed this global threshold as globally statistically anomalous. The reader will note, however, that Twitter's popularity may trend over time or vary seasonally in a manner that renders whole months or years above the global anomaly threshold. To account for this, we define a rolling window of approximately one month (previous 30 days) and calculate a rolling threshold of two rolling standard deviations above the rolling mean. Figure 1 offers a visualization of this for one of our hashtags of interest.

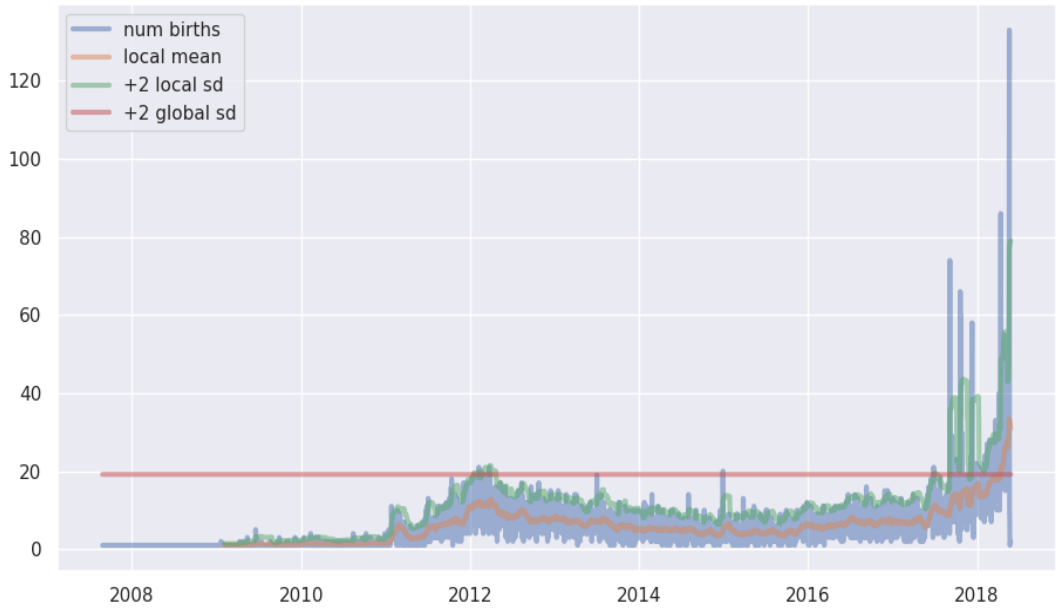

Figure 1. Local birthdate anomalies on #أين_الناشطين_الحقوقيين* [#where_are_the_rights_activists], *May 11–31, 2018.

In this Figure 1, locally anomalous birthdates are precisely those days where the blue curve (number of births) exceeds the green curve (rolling anomaly threshold).

Having identified anomalously born accounts, we then investigate these accounts for further anomalous similarities, following Jones's approach. For example, users in the Gulf tend to tweet from their mobile phones, so their tweeting platform is typically "Twitter for Android" or "Twitter for iPhone" (Northwestern University Qatar, 2019). If we see accounts born on the same day tweeting from "Twitter Web Client," this is again anomalous. Indeed, in our sample of tweets for #أين_الناشطين_الحقوقيين, we find that 55% of all tweets were posted using Twitter's iPhone client, while 30% were posted using the Android client. Only 6.4% were posted from Twitter Web Client. The tallest blue spike in the birth chart in Figure 1 corresponds to the anomalous birth of 133 accounts on May 20, 2018, right as the hashtag itself was trending. Just five of these accounts tweeted with the iPhone client, and only one of them tweeted with the Android client. The remaining 127 of 133 used Twitter Web Client. Moreover these 127 did not post original content, but instead retweeted just two tweets, both of them by @MohammedKAlsadi. We flag these 127 accounts as bots.

As a second take on bot detection, we run checkups on all accounts to see if they are still alive. In recent years, responding to growing international pressure, Twitter has moved to suspend or delete accounts

that are deemed abusive or in violation of its terms of service. We essentially piggyback on Twitter's abuse detection method, checking to see which accounts that participated in our hashtags of interest were subsequently suspended or deleted.[7]

Influencer Detection

For each hashtag, we classify certain users as "influencers." Although the notion of social media influencers has entered popular parlance, we use the term in a specific, technical way here. Specifically, by influencer we simply mean any user who belongs to the shortlist of users who, collectively, garnered 80% of retweets on the hashtag in question. To see the logic behind this definition, think of each hashtag as a conversation about a topic (the Suleimani assassination, Oscars 2020, and so on). Users who tweet original content and mention this hashtag in their tweets are essentially voicing their opinions or thoughts about this topic. Whether or not they see it in these terms, they are effectively in competition with each other for the attention and affirmation of others following the hashtag. Although the number of "impressions" (views) of each tweet is not available through Twitter API, we can see the number of retweets garnered per tweet and can thereby obtain the number of retweets garnered per user. From this we can calculate the share of total retweets garnered by each user and place these users in descending order of retweet share.

A common finding across a broad swath of hashtags from both inside and outside of MENA is that there is profound inequality across users in terms of retweet shares garnered, with most users obtaining a negligible share and a handful obtaining almost everything (Abrahams & van der Weide, 2020).

Figure 2 depicts profound inequality of influence on #جمال_خاشقجي [#jamal_khashoggi], one of our Saudi-related hashtags of interest that we investigate closely below. The "Lorenz" curves for followers, likes, and retweets, all start off flat and then turn sharply upward near the end, indicating that the lion's share of followers, likes, and retweets are monopolized by a narrow clique of influencers, while the majority of users are ignored.

[7] We perform these status checkups in December 2019 for all accounts in our Saudi-related hashtags of interest.

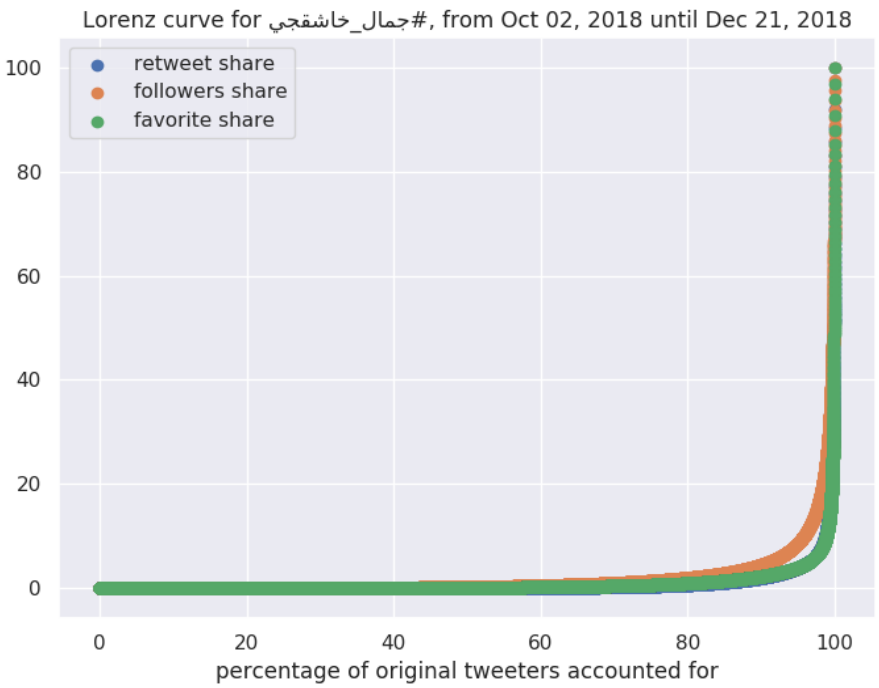

Figure 2. "Lorenz" curve depicting inequality on #جمال_خاشقجي [#jamal_khashoggi], October 2–November 26, 2018.

As we have argued in previous work, this stark inequality also constitutes an opportunity for reducing the data and making it tractable for qualitative investigation (Abrahams & Leber, 2020). Tweets from the shortlist of users who garnered 80% of all retweets including the hashtag in question constitute a first-order approximation of the prevailing narratives on this topic on Twitter. In this way, even though a hashtag may contain millions of tweets, we can still undertake a tractable, in-depth qualitative analysis by identifying the tens or hundreds of users who collectively drive the conversation (in the sense that they garnered most of the retweets).

Community Detection

We further reduce the workload of qualitative investigation by recognizing that these influential voices are by no means independent of each other. As Figure 3 shows, influencers often retweet each other—an act that, we have previously found, tends to accurately predict ideological affinity (Leber & Abrahams, 2019). Using a standard community detection algorithm ubiquitous to the social media literature, we can partition the influencers into different color-coded "clusters," and infer the ideology of each cluster by reviewing the tweets of a sample of its members (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). Figure 3 depicts the partitioned retweet network for influencers on #جمال_خاشقجي [#jamal_khashoggi].

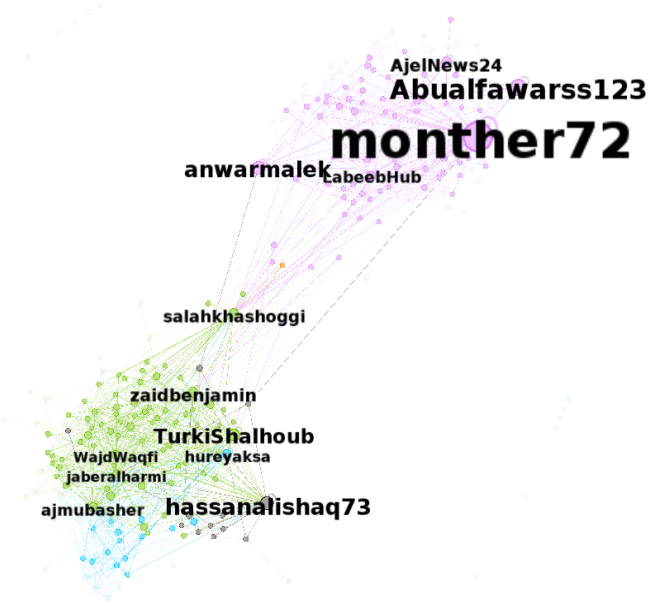

Figure 3. Influencer retweet network for #جمال_خاشقجي [#jamal_khashoggi], October 2–November 26, 2018.

To better understand the political world that Saudi- and Saudi-adjacent Twitter users inhabit, we refine our understanding of these interaction-based "clusters" by examining the Twitter content of key influencers both within our sample and in real time. This in turn allows us to sort influencers along major ideological cleavages that would be recognizable to the Twitter users under study. This fine-grained understanding of the meaning of Tweeting to users themselves is necessary to understand the ways that various hashtags are used in discourse—whether to promote it, criticize it, drown out those promoting the hashtag, or simply as a neutral observation of a popular hashtag (Simmons & Smith, 2017).

Nationalist Narratives

First, we examine tweets relating to two Saudi hashtag campaigns where we would plausibly expect to see considerable bot activity—efforts to portray several Saudi women's rights activists as enemies of the state. In May of 2018, activists Loujain al-Hathloul, Eman al-Nafjan, and Aziza Yousef were arrested along with several other individuals who had championed Saudi women's right to drive and were subsequently demonized in the Saudi press (Alsahi, 2018). Abroad, however, the arrests prompted criticism of the Saudi government and defenses of the detainees' peaceful activism (Anonymous, 2018).

These two frames for the activists—as peaceful agents of change, or as threats to the security of the nation—each attracted their own hashtags on Twitter. #أين_الناشطين_الحقوقيين ["Where are the activists?"] began first, highlighting the detention of several of the activists and calling for their release. The second, #عملاء_السفارات ["Agents of the Embassies"] spread tweets variously castigating and mocking the activists, suggesting that

they were spying on Saudi Arabia or otherwise conspiring against the Kingdom. These tweets present an ideal case to test whether online narratives within the Kingdom are inevitably manufactured by bots and hashtag "flooding" or might instead reflect offline messaging and organic online action.

Hashtag Hijacking: Where are the Activists?

Our Twitter ethnography for "Where are the Activists?" revealed two main camps of influencers: those defending the activists, and those who "hijacked" the hashtag to flood the hashtag's content with antiactivist views. Figure 4 shows the hourly totals of retweets ascribed to these factions. Word spread among offline activist communications even before the official announcement of the arrests (Trew, 2019). By contrast, the hijacking started only the day after the news was announced, first by user @SaudiFalcon9 and later by other nationalist-leaning accounts—albeit few who featured prominently in other nationalist hashtag campaigns. Antiactivist tweets and retweets quickly outnumbered proactivist tweets and retweets within the hashtag.

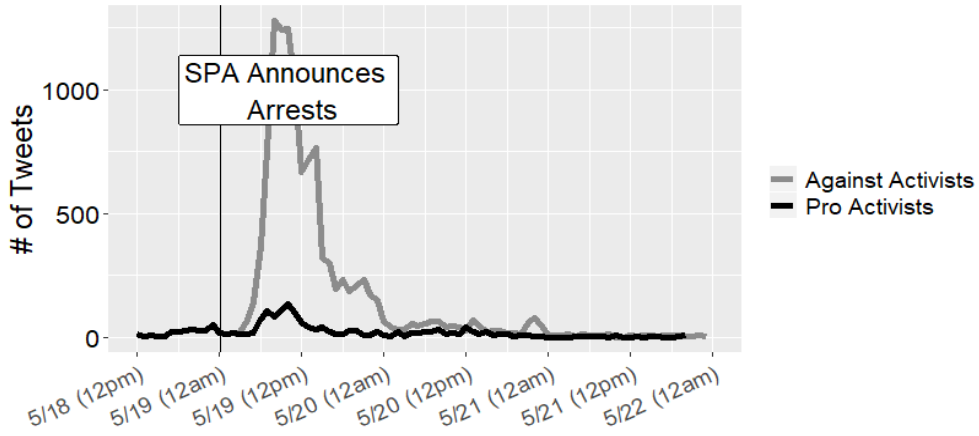

Figure 4. Dueling camps on #where_are_the_rights_activists; hourly totals of retweets for influencers in each camp.

Based on a comparison between these two camps, we fail to reject the null hypothesis that there is a similar presence of suspicious accounts in each. In other words, although there might be a "background" bot presence on all Twitter activity, there is little evidence of a concerted bot intervention against the activists. Proactivist and antiactivist accounts received a statistically indistinguishable portion of retweets from deleted (6.5% vs 4.5%) and suspended (10.1% vs. 10.7%) accounts. Even here, two antiactivist accounts —@SaheelKSA (an online news site) and @YL511 (a self-styled influencer)—accounted for a large portion of suspended-account tweets (36% and 51% of their respective retweets), suggesting account-specific manipulation rather than a concerted, hashtag-wide effort (Confessore, Dance, Harris, & Hansen, 2018). Removing these two accounts means that there are fewer suspicious retweets as a percentage of

the antiactivist than of the proactivist faction. The same holds in looking at suspicious accounts among retweets—if anything, it is proactivist tweets that are likelier to be retweeted by suspicious accounts.[8]

Still, 18% of tweets and retweets within the hashtag came from suspended accounts, raising the possibility that these accounts promoted the content of antiactivist influencers early on to "signal boost" the hijacking. Among proactivist influencers, only one (subsequently deleted) account featured a major tweet where we could safely reject the null of no signal boosting.[9] For antiactivist influencers, we reject the null hypothesis in favor of signal boosting for a few accounts, including @YL511, the account that featured many suspended account retweets—strengthening our confidence in this metric. The strongest piece of evidence in favor of manipulation of this hashtag is that we also reject the null hypothesis of "no boosting" for the initial tweet by @SaudiFalcon9 that "hijacked" the hashtag—although, unlike @YL511, other tweets by this account do not exhibit a suspicious clustering of suspended accounts.[10]

We next test for evidence of greater coordination among antiactivist retweets. When the timings for a collection of tweets are more skewed, it suggests a higher level of Twitter activity in the initial minutes or hours of the collection—whether a set of tweets from a hashtag or retweets of a single tweet—relative to the timing of all tweets' retweets covered by the collection (Leber & Abrahams, 2019). We think of the duration between an initial tweet and the last recorded retweet within our sample as the "life cycle" of the tweet. Skewness in retweets across this life cycle can therefore suggest tighter coordination to promote a particular message, as several accounts tweet out the same message or promote the same tweet. This might result from offline coordination—such as account owners being asked to tweet at the same time—or organically, through engaged users seeking to promote content by monitoring favored "influencers" and retweeting new statements as soon as they are made.

We find some evidence of greater coordination among the antiactivist community despite a relative dearth of bot accounts. First, antiactivist tweets and retweets display a higher overall skew of 2.5 versus 0.9 to 1.2 for the proactivist tweets (depending on whether proactivist skewness is calculated from the start of the hashtag or from the first antiactivist tweet). For a more robust comparison of the two camps, we examine the skew of retweets accruing to original tweets from each camp (34 tweets from proactivist influencers, 54 from antiactivist influencers).[11] Our main independent variable is whether the influencer in question is proactivist; if there is greater coordination among the antiactivist campaign, tweets from influencers in this faction should display higher skew on average (i.e., tweets should receive retweets more

[8] See appendix for the full regression (Table 4).

[9] Based on Wilcox tests of whether retweets from suspended accounts appeared later than nonsuspended accounts on average. Rejecting the null provides evidence that suspended-account retweets appeared early in the life cycle of the tweet, suggesting coordinated activity to signal boost a particular perspective.

[10] However, the SaudiFalcon9 account does not appear fully aligned with the Saudi government, hence this is likely an indicator of individual efforts at self-promotion. This account was later critical of the government for hosting the Russian circus that prompted considerable criticism. See appendix for comparative tests of select antiactivist accounts (Figure 10).

[11] A robustness check utilizing the kurtosis, or "peakedness," of retweets produces similar results (Table 5 in the data appendix).

rapidly). We find that antiactivist tweets do, indeed, exhibit a higher skew among their retweets on average (Table 1).

Table 1. Ordinary Least Squares Regression of Retweet Skew on Influencer Camp as well as Follower-Specific and Time Variables (Time Trend not Shown). Standard Errors Clustered by Influencer.

	All	Matching	>= 20 RT	>= 50 RT
Proactivist	−0.65	−1.37***	−1.27***	−1.76***
	(0.42)	(0.48)	(0.49)	(0.66)
Retweets of tweet (log)	0.33***	0.85***	−0.15	−0.49
	(0.11)	(0.30)	(0.16)	(0.32)
Followers (log)	0.38***	0.60***	0.31***	0.23***
	(0.09)	(0.20)	(0.09)	(0.08)
8:00 am to 4:00 pm	−0.66	−0.60	0.90	−1.18
	(0.46)	(2.15)	(1.04)	(1.93)
4:00 pm to midnight	0.13	0.64	2.55	0.87
	(0.59)	(2.98)	(1.64)	(1.83)
Constant	−1.99**	−4.92	−1.35	3.79*
	(0.82)	(3.31)	(1.81)	(1.95)
Observations	87	42	64	46
R^2	0.43	0.47	0.41	0.55

*p < 0.1; ** p <0.05; *** p <0.01

As Figure 5 shows, this difference is an average effect rather than a stark dichotomy—even some proactivist tweets display high skew, while some antiactivist accounts exhibit relatively low skewness.

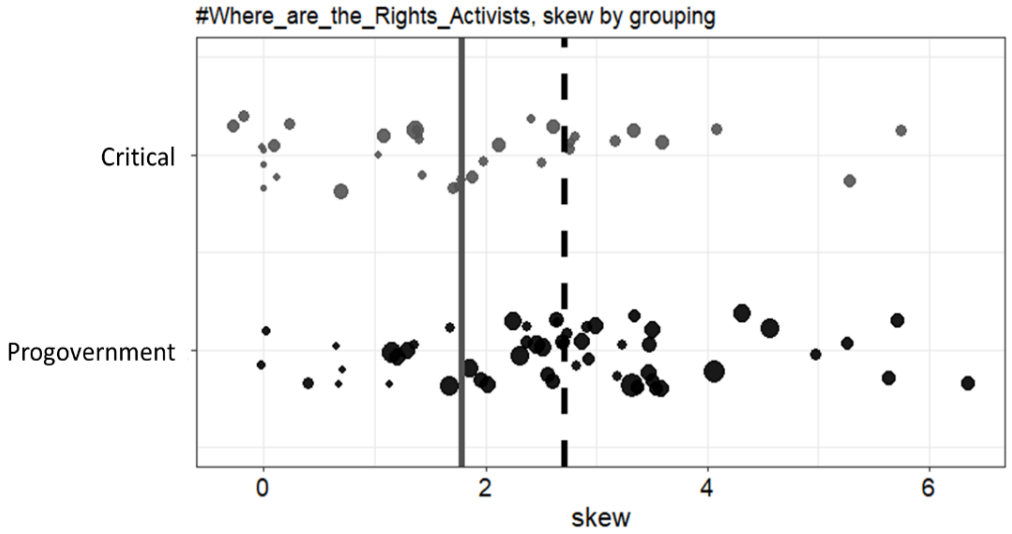

Figure 5. Skew of retweets by original tweet, pro- and antiactivist influencers.

To sum up, we find some (but not overwhelming) evidence of bots and suspicious accounts being used to promote particular (but far from all) progovernment accounts, including some that were highly influential. We do not, however, find enough evidence to rule out this "hijacking" being driven by opportunistic Twitter influencers and their engaged followers rather than a centrally coordinated campaign carried out by loyal agents. We instead find some evidence of an engaged Saudi nationalist Twitter community that is able to coordinate quickly around key tweets from copartisans to boost their preferred (and typically progovernment) narrative online.

Pushing Propaganda? Agents of the Embassies

The "Agents of the Embassies" hashtag sought to paint the women's rights activists as serving foreign powers and betraying their country. This amplified an offline narrative that was implied by the state-run Saudi Press Agency and relayed through other, heavily regulated news outlets. Online news aggregator account @SaudiNews50 was the first to use the "Agents of the Embassies" hashtag in reporting the arrest of "7 individuals" (SaudiNews50, 2018) shortly after midnight on May 19, 2018, Riyadh time. This account has been at least tangentially associated with Saudi influence operations on Twitter, raising the likelihood of observing at least some manipulation of this hashtag (Paul, 2019).

Among the influencers on this hashtag, we note several different categories: ordinary Saudis, "professional" influencers, media figures, news sites, and accounts for government officials and institutions. Of note: news sites and media figures accounted for the greatest portion of retweets at the outset of the hashtag, even though professional and progovernment influencers accounted for a greater portion of

retweets after the initial burst of activity (Figure 6). This would suggest that "disinformation" about the activists stems more from the overall controlled media environment within Saudi Arabia rather than specific efforts to manipulate online narratives—we examine this possibility in greater detail below.

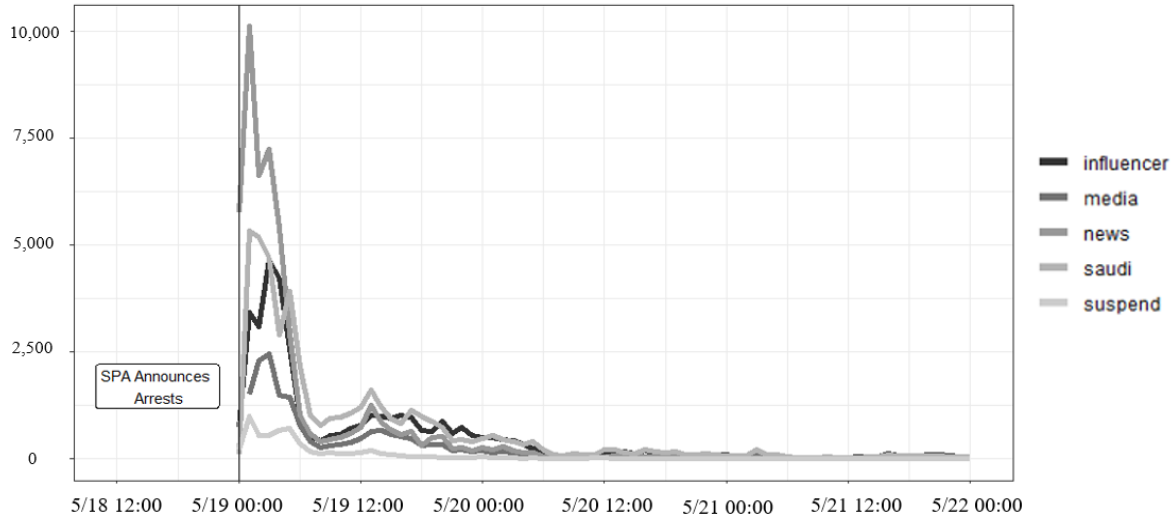

Figure 6. Retweets per hour on influencers' tweets from various classes of influencer.

Compared with the activists hashtag, a lower portion of agents tweets and retweets came from suspended (only 6% vs. 18%), deleted (2% vs. 5%), and likely bot accounts (2% vs. 4%). Additionally, a closer look at agents influencers' individual tweets provides little systematic evidence of boosting by suspicious accounts. To be sure, a few "professional influencers" appear to have had most or all of their tweets boosted by suspended accounts, along with (perhaps unsurprisingly) a Saudi nationalist account that claims to be a "cyber fighter" in its bio.[12] Yet while one key influencer, @Monther72, has been so prominent in online discourses that Twitter rumors have suggested the account is simply a cover for Saudi cybersecurity official Saud al-Qahtani (Mujtahidd, 2018), statistical tests cannot establish evidence of signal boosting. The same is true for @SaudiNews50.

We also compare the agents hashtag to a June 2018 hashtag entitled #روسيات_عاريات_في_الرياض ["Naked Russian Women in Riyadh"]—backlash to a Russian circus that performed in allegedly revealing outfits (Batrawy, 2018). Given that this hashtag led Saudi rulers to fire the head of the country's general entertainment authority, we can reasonably expect this to be an "organic" hashtag campaign (i.e., it is unlikely that the Saudi state would orchestrate a campaign against one of its own officials on the basis of an event it had already approved). We would therefore expect "agents" to display a higher percentage of suspicious-account activity (though perhaps attracting some efforts to flood the hashtag and diffuse popular anger). However, we again find a higher portion of suspended account tweets (17% of all tweets, more

[12] @RM7KSa, Twitter account at https://twitter.com/RM7KSa.

similar to activists rather than agents) and deleted account tweets (5%). Suspended account tweets in "Russian Women" exhibited greater skew (joined the hashtag faster) than nonsuspended accounts—again, the opposite of what we find for the agents hashtag.

We further compare the skewness of various retweets according to the type of influencer account. We find (Table 2) that media and official accounts do exhibit greater right-skewness (on average) among their tweets, relative to baseline news account tweets, suggesting an ability to convert offline reputation into online clout.

Table 2. Comparison of Different Groups within Agents of the Embassies Hashtag Campaign.

	Skewness		% Suspended	# Tweets (log)	
	(1)	(2)	(3)	(4)	(5)
Professional	1.32*	1.11	−0.05***	2.19***	1.59**
	(0.79)	(0.78)	(0.02)	(0.67)	(0.74)
Media and Official	0.89**	0.75*	−0.04*	0.89**	0.52
	(0.39)	(0.40)	(0.02)	(0.44)	(0.53)
Other Progovernment	0.21	0.09	−0.03	0.78*	0.45
	(0.30)	(0.30)	(0.02)	(0.44)	(0.50)
Suspended	0.47	0.40	−0.02	1.03**	0.80
	(0.38)	(0.38)	(0.02)	(0.41)	(0.49)
Followers (log)	0.30***	0.26***	−0.01***	0.35***	0.23**
	(0.05)	(0.05)	(0.003)	(0.09)	(0.10)
Tweets (log)	0.06	0.01	−0.01***		
	(0.08)	(0.09)	(0.003)		
% suspended		−3.96***			−8.16**
		(1.53)			(3.56)
Constant	−0.33	0.78	0.28***	0.90	3.11**
	(0.76)	(0.96)	(0.05)	(1.17)	(1.55)
Elapsed time control	Y	Y	Y	Y	Y
Time of day control	Y	Y	Y	Y	Y
Observations	279	279	279	279	279
R^2	0.38	0.39	0.34	0.25	0.32

*p < 0.1; **p < 0.05; ***p < 0.01

Media figure and official accounts are also less likely to exhibit high portions of suspended account retweets. We find further support for the idea that official accounts draw on their offline reputations because their tweets behave similarly to those of professional influencers—individuals who have invested considerable time and resources into cultivating online followings. To be sure, every class of influencer

garnered more retweets on average than news sites, controlling for each influencers' number of followers. Yet while controlling for suspended accounts tends to reduce the magnitude of the difference with news accounts, the sign on percent suspended is the opposite from what we would expect. More suspicious activity is associated with less retweet acceleration (on average) and fewer retweets (on average)—not what we would expect from a signal boosting operation.

Ultimately, the balance of evidence suggests that the campaign against the activists resulted from official and media attention on the issue rather than automated efforts to boost the hashtag. We further compare the "agents" hashtag to Twitter discussion of a rocket strike that hit Riyadh earlier the previous year, #صوت_انفجار_في_الرياض ["Sound of Explosion in Riyadh"]. The strike occurred roughly six months before the activist hashtags (November 4, 2017) and resulted from a Houthi rocket either being shot down over Riyadh or crashing in its outskirts (Almosawa & Barnard, 2017). We find that the "Explosion in Riyadh" hashtag attracted tweets and retweets at a faster rate than agents—exhibiting greater skewness. This suggests, at a minimum, that the surge of activity on the "agents" hashtag is within the realm of possibility for a major news story within Saudi Arabia.[13] Specific accounts (almost all news outlets) that featured as influencers in both hashtags likewise garnered attention more rapidly, if anything, in using the "explosion" hashtag. While progovernment influencers appear to play an important role in promoting antiactivist content, we cannot rule out the possibility that this reflects the cultivation of organic online followings rather than automated efforts.

Big Data Analysis of MENA Hashtags

We complement this in-depth, mixed-methods analysis by conducting a broad sweep for bots across 279 MENA-related hashtags collected by the authors during the period of October 2019 to January 2020. While relying on more superficial metrics to assess these hashtags, they offer an updated snapshot of bot prevalence on Middle Eastern Twitter. Additionally, broadening our analysis to include the region at large allows us to speak at greater scale and with greater perspective.

The MENA-related hashtags comprising our large dataset correspond to a politically charged period in the region. A wave of protests rippled across the region west to east, from Egypt to Lebanon to Iraq to Iran. Meanwhile, Turkey invaded northeastern Syria, while the United States assassinated a high-ranking Iranian general. A team of regional experts affiliated with the Citizen Lab monitored Middle East Twitter throughout this period, identifying hashtags relevant to events unfolding on the ground.[14] The resulting data set, though obviously nonrandom, constitutes the most comprehensive data set freely accessible to the authors for studying Twitter activity from this period.

To estimate bot prevalence, we rely on the same two methods discussed above: birth anomaly detection and account attrition. As not all accounts born during anomalous birth spikes are bots, the number of accounts born during anomalous spikes constitutes an upper bound (an overestimate) of the number of

[13] A Wilcox difference-in-median's test further confirms this. Overall skewness of "explosion" is 2.5 versus 1.6 for "agents" for the common duration period.

[14] The authors are especially grateful to Noura Aljizawi for flagging most of these hashtags for download.

bots. We exploit this fact to overestimate the percentage of users that are potentially bots on every hashtag, simply by counting the number of accounts born during anomalous spikes. The resulting estimates, therefore, are an exaggeration. Since we are arguing, however, that bot prevalence is relatively low, our approach is defensible in the sense that we are erring in a manner that works against our argument.

We find across 279 MENA hashtags that anomalously born accounts constitute merely 6.1% (s.d. 3.1%) of users per hashtag. Indeed, the median percentage of anomalously born accounts is merely 5.5%. If we weigh hashtags by the number of unique users posting to them, the percentage of anomalously born accounts falls even further, to 3.6%. Figure 7 plots hashtag size (number of unique tweets and retweets) against the percentage of anomalously born accounts. The best-fit line superimposed on the data suggests a downward trend, with smaller, fringier hashtags averaging a higher incidence of anomalous births than large hashtags with broad participation.

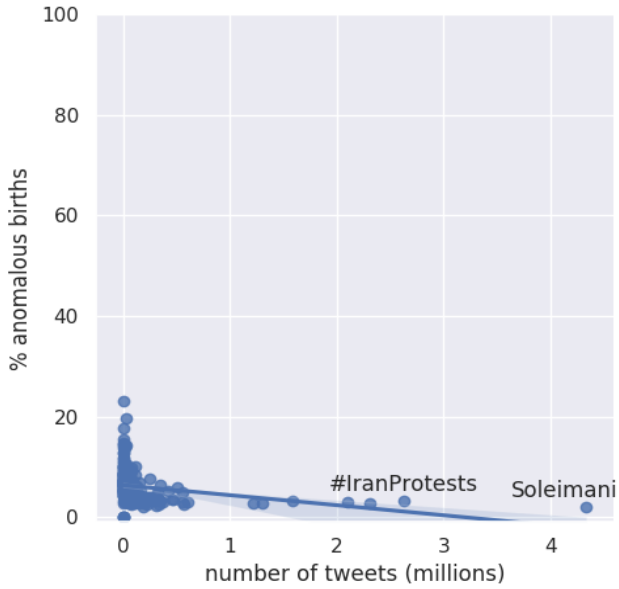

Figure 7. Share of anomalously born accounts declines with hashtag size.

For a second take on this finding, we randomly sample up to 1,000 users on each hashtag and query the Twitter API to check for account attrition (deletions and suspensions). The idea behind doing this is simply to piggyback on Twitter's own in-house algorithms for flagging abusive behavior. Figure 8 plots hashtag size against the percentage of randomly sampled accounts that proved to be deleted or suspended.

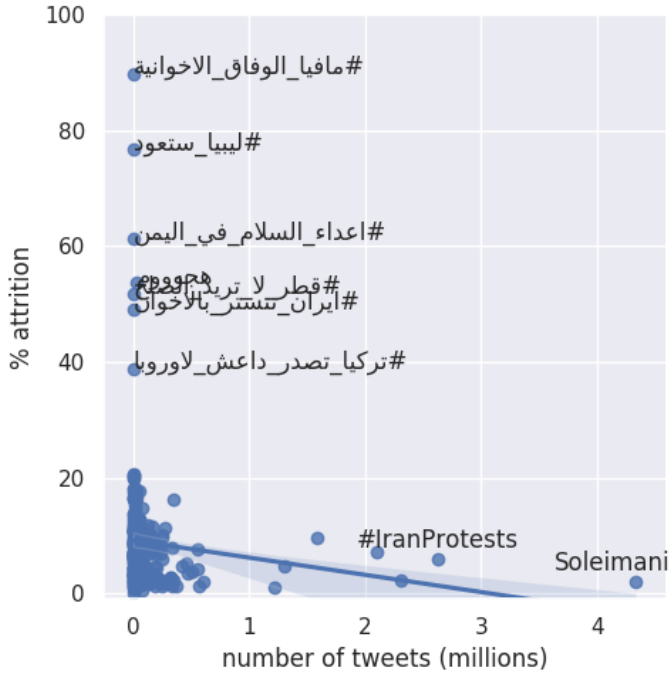

Figure 8. Share of suspended or deleted accounts declines with hashtag size.

As with the rate of anomalous birth, we find a downward-sloping relationship with fringier hashtags exhibiting higher rates of attrition relative to their more voluminous counterparts. On average, 8.8% (s.d. 9.8%) of accounts are deleted or suspended, but this higher mean is driven by a handful of highly suspicious accounts labeled in the figure. The median attrition rate, for example, is just 7%, and if we weigh hashtags by the number of unique users participating the attrition rate falls to only 4.4%.

Discussion

These low rates of anomalous births and attrition imply a low rate of bot prevalence (as detectable by these standard methods). If bot prevalence is truly so low on average, how have bots managed to seize the attention of scholars and journalists in recent years? Though answering this question lies beyond the scope of this article, a possible clue lies in the inverse relationship between hashtag size and (potential) bot incidence rates. Large hashtags, as we ourselves discovered, present technical challenges for researchers and journalists; they require more hardware and software infrastructure to download, store, and analyze. These technical difficulties, we suspect, push investigators to focus on smaller hashtags—those that tend to exhibit higher rates of attrition and anomalous births, both predictive of bot incidence. Moreover, researchers and journalists we have spoken to seem to intuit that smaller, fringier hashtags are likelier to yield a fruitful harvest of bots and conspiracy theories. This approach, however, may have led to oversampling of conversation threads that are almost by definition on the periphery of the social discourse.

Conclusion

In this article we have applied a mixture of methods to new Twitter data on the Middle East to evaluate the degree to which politically salient hashtags are affected by coordinated, inauthentic "bot" accounts. Across several prominent hashtags relating to Saudi Arabia, an in-depth investigation reveals relatively low rates of bot prevalence, weak evidence that bots were primarily proregime, and weak evidence that they meaningfully moved the discourse.

Complementing this deep investigation with a broad, big-data sweep of 279 hashtags from across the Middle East, we find a similarly low rate of bot prevalence. We do find, however, that bot incidence is higher on smaller hashtags.

Our findings suggest several major takeaways that future research can engage with. Firstly, the perception that MENA Twitter is uniquely dominated by bots ought to be revisited. While bots appeared across a broad swath of politically salient hashtags trending in recent months, they generally constituted only a small percentage of users. Bots did seem to enjoy a higher rate of incidence on smaller, fringier hashtags, but almost by definition such hashtags were followed by smaller audiences.

Secondly, however, this deemphasis on bot "soldiers" should be complemented by a reemphasis on cyber "knights." Though far fewer in number, these individuals enjoy wide followings and wield considerable clout. Their power to manipulate opinion has not gone unnoticed by authoritarian regimes, who in dozens of known cases have attempted to co-opt or intimidate influencers to toe the party line. Further research on Twitter influencers would be very welcome, though this will inevitably lead researchers off Twitter and require more bespoke investigative methods.

Finally, although the authors share with other scholars and journalists a fascination for social media technology, we hope that this article, insofar as it asserts the plausibility of proauthoritarian discourses on Middle East Twitter being organic and "authentic," will encourage readers to see social media not only as a vector for top-down propaganda but also as a potential outlet for the expression of bottom-up, proauthoritarian sentiment. Citizens of authoritarian states, after all, are molded by the conformist content of state-controlled education systems (Darden & Grzymala-Busse, 2006, pp. 101–107) and fed a steady diet of statist thought through mass media (Stockmann & Gallagher, 2011). Acknowledging the considerable potential for individuals to build up vocal, proregime social media followings largely on their own helps us to make sense of developments such as Saudi state-run television criticizing pro-Saudi accounts that have "given [themselves] the appearance of being supported by the state" (Chopra, 2020, para. 8) in accusing other Saudi users of insufficient patriotism—especially when those criticized fire back that their critics are in league with foreign enemies (Al Saud, 2020).

Social media, therefore, constitutes not only a new vector for authorities to reach their citizens, but a venue for citizens who have absorbed this indoctrination to express the loyalty previously instilled in them—even if redlines and naked repression have cleared critical voices from online spaces. Social media's novelty as a tool of control would therefore appear to lie not so much in how it helps authorities spam their

intended audiences, but, on the contrary, in the way it facilitates the laundering of authoritarian thought through the voluntary expressions of ordinary supporters.

References

Abrahams, A. (2019). *Regional authoritarians target the Twittersphere*. https://merip.org/2019/12/regional-authoritarians-target-the-twittersphere/

Abrahams, A., & Jones, M. O. (2018). *Bladerunning the GCC: A hashtag-based anomaly detection approach to propaganda bots in the Arabian Gulf*. Unpublished manuscript.

Abrahams, A., & Leber, A. (2020). Framing a murder: Twitter influencers and the Jamal Khashoggi incident. *Mediterranean Politics*. Advance online publication. doi:10.1080/13629395.2019.1697089

Abrahams, A., & van der Weide, R. (2020). *Ten thousand whispering: Measuring inequality of voice on Twitter*. Unpublished manuscript.

Alhussein, E. (2019). *Saudi first: How hypernationalism is transforming Saudi Arabia* (Policy Brief, 19). https://www.ecfr.eu/publications/summary/saudi_first_how_hyper_nationalism_is_transforming_saudi_arabia

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. doi:10.1257/jep.31.2.211

Almosawa, S., & Barnard, A. (2017, November 4). Saudis intercept missile fired from Yemen that came close to Riyadh. *The New York Times*. https://www.nytimes.com/2017/11/04/world/middleeast/missile-saudi-arabia-riyadh.html

Alsahi, H. (2018, June 20). Saudi women can drive starting Sunday. Why are feminists there still labeled traitors? *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2018/06/20/saudi-women-can-drive-sunday-why-are-feminists-there-still-labeled-traitors/

Al Saud, A. [sattam_al_saud]. (2020, June 14). البعض يهاجم المغرد السعودي ويصفه بأقبح وصف. [Some attack the Saudi Tweeter and describe him in the ugliest way]. [Tweet]. https://twitter.com/sattam_al_saud/status/1272238634549284865

Anonymous. (2018, October 16). *Saudi Arabia: No country for bold women*. https://pomed.org/saudi-arabia-no-country-for-bold-women/

Batrawy, A. (2018, June 19). Saudi entertainment chief sacked after conservative backlash. *Associated Press*. https://apnews.com/577489a1c3ae48e58a45d3754d52bb6c/Saudi-entertainment-chief-sacked-after-conservative-backlash

Benner, K., Mazetti, M., Hubbard, B., & Isaac, M. (2018, October 20). Saudis' image makers: A troll army and a Twitter insider. *The New York Times*. https://www.nytimes.com/2018/10/20/us/politics/saudi-image-campaign-twitter.html

Blaydes, L. (2018). *State of repression: Iraq under Saddam Hussein*. Princeton, NJ: Princeton University Press.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, R. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 10*(P10008). doi:10.1088/1742-5468/2008/10/P10008

Bradshaw, S., & Howard, P. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation* (Working Paper, 2019.2). Project on Computational Propaganda, Oxford University. https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf

Chopra, A. (2020, August 7). Saudi Arabia seeks to tame powerful cyber armies. *Agence-France Presse (AFP*). http://u.afp.com/3B9Q

Confessore, N., Dance, G., Harris, R., & Hansen, M. (2018, January 27). The follower factory. *The New York Times*. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

Darden, K., & Grzymala-Busse, A. (2006). The great divide: Literacy, nationalism, and the Communist collapse. *World Politics*, *59*(1), 83–115. doi:10.1353/wp.2007.0015

Davies, H. (2015, December 11). Ted Cruz using firm that harvested data on millions of unwitting Facebook users. *The Guardian*. https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data

DFRLab. (2017, August 27). #BotSpot: Twelve ways to spot a bot. *Medium*. Retrieved June 16, 2020 from https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c

Diamond, L., & Plattner, M. F. (Eds.). (2012). *Liberation technology: Social media and the struggle for democracy*. Baltimore, MD: Johns Hopkins University Press.

DiResta, R. (2016, September 15). *Crowds and technology*. https://www.ribbonfarm.com/2016/09/15/crowds-and-technology/

Duffy, M. J. (2014). Arab media regulations: Identifying restraints on freedom of the press in the laws of six Arabian peninsula countries. *Berkeley Journal of Middle Eastern & Islamic Law*, *6*(1), 1–31. doi:10.15779/Z384S3C

Fahmy, D., & Faruqi, D. (Eds.). (2017). *Egypt and the contradictions of liberalism: Illiberal intelligentsia and the future of Egyptian democracy*. New York, NY: Oneworld Publications.

Forelle, M., Howard, P., Monroy-Hernández, A., & Savage, S. (2015). *Political bots and the manipulation of public opinion in Venezuela*. Retrieved from https://arxiv.org/ftp/arxiv/papers/1507/1507.07109.pdf

Gerring, J. (2007). Is there a (viable) crucial-case method? *Comparative Political Studies, 40*(3), 231–253. doi:10.1177/0010414006290784

Gerschewski, J. (2018). Legitimacy in autocracies: Oxymoron or essential feature? *Perspectives on Politics*, *16*(3), 655–659. doi:10.1017/S1537592717002183

Gleicher, N. (2019, August 1). *Removing coordinated inauthentic behavior in UAE, Egypt and Saudi Arabia* [Facebook blog]. Retrieved from https://about.fb.com/news/2019/08/cib-uae-egypt-saudi-arabia/

Gorwa, R., & Guilbeault, D. (2018). Unpacking the social media bot: A typology to guide research and policy. *Policy & Internet, 12*(2), 225–248. doi:10.1002/poi3.184

Greenberg, N. (2019, October 4). *Russia opens digital interference front in Libya*. Retrieved from https://merip.org/2019/10/russia-opens-digital-interference-front-in-libya/

Howard, P., & Bradshaw, S. (2018). "The global organization of social media disinformation campaigns." *Journal of International Affairs, 71*(1.5). Retrieved from https://jia.sipa.columbia.edu/global-organization-social-media-disinformation-campaigns

Howard, P. N., Woolley, S., & Calo, R. (2018). Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics, 15*(2), 81–93. doi:10.1080/19331681.2018.1448735

Jones, M. O. (2013). Social media, surveillance and social control in the Bahrain uprising. *Westminster Papers in Communication and Culture, 9*(2), 69–92. doi:10.16997/wpcc.167

Jones, M. O. (2016). *Automated sectarianism and pro-Saudi propaganda on Twitter*. Retrieved from https://exposingtheinvisible.org/resources/automated-sectarianis

Jones, M. O. (2019a, September 25). Saudi-UAE Twitter takedowns won't curb rampant disinformation on Arab Twitter. *The Washington Post*. https://www.washingtonpost.com/politics/2019/09/25/saudi-uae-twitter-takedowns-wont-curb-rampant-disinformation-arab-twitter/

Jones, M. O. (2019b). The gulf information war: propaganda, fake news, and fake trends: The weaponization of twitter bots in the gulf crisis. *International Journal of Communication, 13*, 1389–1415. https://ijoc.org/index.php/ijoc/article/view/8994/2604

Jones, M. O., & Abrahams, A. (2018, June 5). "A plague of Twitter bots is roiling the Middle East." *The Washington Post*. https://www.washingtonpost.com/news/monkey-cage/wp/2018/06/05/fighting-the-weaponization-of-social-media-in-the-middle-east/

Jones, R. (2019, November 7). In Saudi Arabia, Twitter has become a tool to crack down on dissent. *The Wall Street Journal*. https://www.wsj.com/articles/in-saudi-arabia-twitter-has-become-a-tool-to-crack-down-on-dissent-11573126932

Khashoggi, J. (2017, September 18). Saudi Arabia wasn't always this repressive. Now it's unbearable. *The Washington Post*. https://www.washingtonpost.com/news/global-opinions/wp/2017/09/18/saudi-arabia-wasnt-always-this-repressive-now-its-unbearable/

Kuran, T. (1997). *Private truths, public lies: The social consequences of preference falsification*. Cambridge, MA: Harvard University Press.

Leber, A., & Abrahams, A. (2019). A storm of tweets: Social media manipulation during the Gulf Crisis. *Review of Middle East Studies*, *53*(2), 241–258. doi:10.1017/rms.2019.45

Lohmann, S. (1994). The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91. *World Politics, 47*(1), 42–101. doi:10.2307/2950679

Marantz, A. (2019). *Antisocial: Online extremists, techno-utopians, and the hijacking of the American conversation*. New York, NY: Viking.

Marczak, B., Scott-Railton, J., McKune, S., Razzak, B. A., & Deibert, R. (2018, September 18). Hide and seek: Tracking NSO Group's Pegasus Spyware to operations in 45 countries (Research Report No. 113). *The Citizen Lab.* University of Toronto. https://citizenlab.ca/2018/09/hide-and-seek-tracking-nso-groups-pegasus-spyware-to-operations-in-45-countries/

Michael, K. (2017). Bots trending now: Disinformation and calculated manipulation of the masses. *IEEE Technology and Society Magazine*, *36*(2), 6–11. doi:10.1109/MTS.2017.2697067

Mujtahidd. [mujtahidd]. (2018, December 14). سعود القحطاني. [Sa'ud al-Qahtani]. [Tweet]. https://twitter.com/mujtahidd/status/1073537970060312576

Nadler, A., Crain, M., & Donovan, J. (2018, October 17). Weaponizing the digital influence machine: The political perils of online ad tech. *Data & Society Research Institute*. https://datasociety.net/wp-content/uploads/2018/10/DS_Digital_Influence_Machine.pdf

Nakashima, E., & Bensinger, G. (2019, November 7). Former Twitter employees charged with spying for Saudi Arabia by digging into the accounts of kingdom critics. *The Washington Post*. Retrieved from https://www.washingtonpost.com/national-security/former-twitter-employees-charged-with-spying-for-saudi-arabia-by-digging-into-the-accounts-of-kingdom-critics/2019/11/06/2e9593da-00a0-11ea-8bab-0fc209e065a8_story.html

Nimmo, B. (2019). Measuring traffic manipulation on Twitter (Working Paper, 2019.1). *Computational Propaganda Research Project, Oxford Internet Institute.* https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Manipulating-Twitter-Traffic.pdf

Northwestern University Qatar. (2019). *Media use in the Middle East 2019*. http://www.mideastmedia.org/survey/2019/

Pan, J., & Siegel, A. (2020). How Saudi crackdowns fail to silence online dissent. *American Political Science Review*, *114*(1), 109–125. doi:10.1017/S0003055419000650

Patin, N. (2019, June 26). *Lord of the flies: An open-source investigation into Saud al-Qahtani*. https://www.bellingcat.com/news/mena/2019/06/26/lord-of-the-flies-an-open-source-investigation-into-saud-al-qahtani/

Paul, K. (2019, December 20). Twitter suspends accounts linked to Saudi spying case. *Reuters*. https://www.reuters.com/article/us-twitter-saudi/twitter-suspends-accounts-linked-to-saudi-spying-case-idUSKBN1YO1JT

Rao, V. (2020, January 16). *The Internet of beefs*. https://www.ribbonfarm.com/2020/01/16/the-internet-of-beefs/

Ritzen, Y. (2019, July 15). How armies of fake accounts "ruined" Twitter in the Middle East. *Al Jazeera*. https://www.aljazeera.com/news/2019/07/armies-fake-accounts-ruined-twitter-middle-east-190715165620214.html

Roberts, M. (2018). *Censored: Distraction and diversion inside China's great firewall*. Princeton, NJ: Princeton University Press.

SaudiNews50. [SaudiNews50]. (2018, May 19). عاجل ..أمن الدولة يقبض على 7 أشخاص [Urgent . . . state security arrests 7 individuals]. [Tweet]. https://twitter.com/SaudiNews50/status/997585703948124165

Simmons, E. S., & Smith, N. R. (2017). Comparison with an ethnographic sensibility. *PS: Political Science & Politics, 50*(1), 126–130. doi:10.1017/S1049096516002286

Spary, S. (2016, April 8). *Facebook is embroiled in a row with activists over censorship*. https://www.buzzfeed.com/saraspary/facebook-in-dispute-with-pro-kurdish-activists-over-deleted

Steinert-Threlkeld, Z. (2018). *Twitter as data*. Cambridge, UK: Cambridge University Press. doi:10.1017/9781108529327

Stockmann, D., & Gallagher, M. (2011). Remote control: How the media sustain authoritarian rule in China. *Comparative Political Studies, 44*(4), 436–467. doi:10.1177/0010414010394773

Stubbs, J., & Bing, C. (2018, November 30). Special report: How Iran spreads disinformation around the world. *Reuters*. https://www.reuters.com/article/us-cyber-iran-specialreport/special-report-how-iran-spreads-disinformation-around-the-world-idUSKCN1NZ1FT

Stukal, D., Sanovich, S., Bonneau, R., & Tucker, J. A. (2017). Detecting bots on Russian political Twitter. *Big Data, 5*(4), 310–324. doi:10.1089/big.2017.0038

Trew, B. (2019, October 1). "Even in death, Jamal is protecting us": Saudi activists mark anniversary of Khashoggi's killing which rocked the world. *The Independent*. https://www.independent.co.uk/news/world/middle-east/jamal-khashoggi-killing-anniversary-murder-saudi-arabia-death-a9127551.html

Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From liberation to turmoil: Social media and democracy. *Journal of Democracy, 28*(4), 46–59. https://www.journalofdemocracy.org/articles/from-liberation-to-turmoil-social-media-and-democracy/

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication*, *62*(2), 363–379. doi:10.1111/j.1460-2466.2012.01629

Twitter. (2019a, September 20). Disclosing new data to our archive of information operations. *Twitter Safety blog*. https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019.html

Twitter. (2019b, December 20). New disclosures to our archive of state-backed information operations. *Twitter Safety blog*. https://blog.twitter.com/en_us/topics/company/2019/new-disclosures-to-our-archive-of-state-backed-information-operations.html

Waltzman, R. (2017). The weaponization of information: The need for cognitive security. *RAND Corporation.* https://www.rand.org/content/dam/rand/pubs/testimonies/CT400/CT473/RAND_CT473.pdf

Wedeen, L. (1999). *Ambiguities of domination: Politics, rhetoric, and symbols in contemporary Syria*. Chicago, IL: University of Chicago Press.

Weidmann, N. B., & Rød, E. G. (2019). *The Internet and political protest in autocracies*. Oxford, UK: Oxford University Press.