



categories: Obtaining Evidence Showing Evidence
 (/resources#filter=.obtaining- (/resources#filter=.showing-
 evidence) evidence)

Automated Sectarianism and Pro-Saudi Propaganda on Twitter

Marc Owen Jones (<http://marcowenjones.wordpress.com/>) is a Research Fellow at the Institute for Arab and Islamic Studies at Exeter University, and a researcher and director at Bahrain Watch, an NGO doing investigative transparency work on the Gulf region. In this article Marc writes about his discovery of thousands of fake twitter accounts in the

Persian Gulf and his investigation into the distorting effect they have on public debate, how he identified them and the potentially humongous scale of this operation in the region.

By Marc Owen Jones

Possibilities of technological liberation through social media use has been marred by increasingly innovative use of counter-revolutionary or informational control tactics by national authorities, global corporations, intelligence agencies (<https://theintercept.com/2014/02/24/jtrig-manipulation/>) and Western PR firms. The Arab Uprisings have brought this into sharp relief, and far from being the stuff of conspiratorial fantasy, the dangers of social media are real. From using social media to deliver malicious links that identify activists, to using torture to extract login details for activists' social media accounts, technology has presented multiple opportunities (<http://www.aljazeera.com/programmes/peopleandpower/2015/06/syria-electronic-armies-150617151503360.html>) for regime surveillance and control. Furthermore, technologies like Prism, which track online sources for signs of information that can increase social tension, also reflect how personal sharing can facilitate information interventions by state agencies (<http://www.nytimes.com/2015/06/07/magazine/the-agency.html>).

Yet as well as being an important source of personal information, social media is still a crucial avenue for receiving real-time information, especially in countries where there is significant state censorship. As such, tactics to prevent the spread of uncensored information by regimes have varied. Turkey (<http://www.theatlantic.com/technology/archive/2016/07/social-media-goes-dark-during-chaos-in-turkey/491609/>), for example, has at times resorted to crude nation-wide blackouts of popular services such as Facebook, YouTube and Twitter, while Bahrain (<http://bahrainwatch.org/blog/2016/08/03/bahrain-internet-curfew/>) has been able to use internet curfews in specific 'troublespots'. Other tactics focus more on spreading propaganda rather than blocking, such as the pro-Putin troll armies/web brigades (<http://theatlantic.com/international/archive/2014/08/the-kremlins-troll-army/375932/>) used to spread disinformation on social media. Elsewhere

in China (<http://washingtonpost.com/news/monkey-cage/wp/2016/05/19/the-chinese-government-fakes-nearly-450-million-social-media-comments-a-year-this-is-why/>), a Harvard based academic estimates that up to 450 million social media accounts in China are operated by pro-government trolls, often referred to as the 50-cent army on the allegation that they receive 50 cents per post. In Bahrain too, there have been well-documented accounts (<http://www.westminsterpapers.org/articles/abstract/10.16997/wpsc.167/>) of state-sanctioned cyber vigilantism, trolling and intimidation online.

Industrial-scale propaganda in the Persian Gulf

Yet it recently came to light that the extent of social media disinformation in the Persian Gulf was a lot deeper than originally known. Back in June 2016, Twitter informed me that they had suspended 1,800 accounts for showing spam-like activity after I submitted an investigation to them. The investigation I had undertaken revealed that thousands of 'fake' Twitter accounts were polluting hashtags around the Persian Gulf with anti-Shia and anti-Iranian propaganda. Many were promoting discourses that mirrored that of the Islamic State, such as branding the Shia as 'rejectionist of the true Islamic faith' (Rawafid).

These thousands of accounts would, at certain times of the day, generate hundreds to thousands of tweets per hour, quickly flushing out legitimate tweets on various hashtags, including #Bahrain, #Yemen, #Saudi and others. The data suggests at least 10,000 of tweets per day were emanating from these suspicious accounts. While thousands of the tweets contained sectarian rhetoric, the majority of the tweets from these suspicious accounts focused on hashtags on #Saudi regions, from #Riyadh to #AlQatif, with the content generally lionising the Saudi government or Saudi foreign policy. However, a significant proportion of those tweets were also used to attack the #Bahrain and #Yemen hashtag.



A screenshot exemplifying individual tweets that were contaminating the #Bahrain hashtag

The tweeting of Isa Qassim, Yemen and Bahrain

The suspicious accounts first came to my attention when Isa Qassim, a prominent Shia Cleric in Bahrain, had his nationality removed by the Bahrain authorities in June 2016. This tactic of making perceived dissidents stateless is not uncommon in Bahrain. In Qasim's case, they accused him of creating an extremist sectarian atmosphere in Bahrain. What happened to Qassim is indicative of the authorities attempts to demonise the opposition in Bahrain as a sectarian and terrorist Iranian fifth column intending to install a Shia theocracy. Soon protests broke out in the village of Diraz, the place of Qassim's birth. A study by the NGO Bahrain Watch revealed that (<https://bahrainwatch.org/blog/2016/08/03/bahrain-internet-curfew/>) the authorities engaged in deliberate internet curfews on the village, a tactic designed to prevent people tweeting or dissenting on the internet.

At the same time, searching on Twitter for any results related to Isa Qassim yielded hundreds of identical tweets from accounts ostensibly run by Arab looking men. These weren't retweets, but tweets that looked like they had been copied and pasted. Written in both English and Arabic, they were all identical. These tweets, all clearly in quick succession, suggested deliberate spamming to convince those interested in searching for information on Isa Qassim that he was indeed, a Shia terrorist. Yet after conducting preliminary searches, it was clear that the suspicious activity was not localised to hashtags related to Isa Qassim. Instead, there was similar activity on other hashtags, such as #Bahrain and #Saudi. Soon, activists following #Yemen in depth alerted me to suspicious activity on that hashtag too. The operation was larger than initially assessed.

Pattern analysis to identify the suspicious accounts

To ascertain the scale of the automated or spam-like tweeting operation, I have analysed hundreds of thousands of tweets from Twitter on different hashtags, on different days. For the purpose of this article, I am drawing from a selection of these cases. Using a Google Add On (<https://chrome.google.com/webstore/detail/twitter-archiver/pkanpfekacaojdnfcgjbjadedbggbbphi>) to extract tweets from the Twitter API streams, tweets were put into a Google Sheet. The data pulled

included, among other information; an account's creation date, its number of followers, its location, and its biography. Tweet data also revealed important information, such as what platform the tweet was launched from. By performing data sorts and other useful functions allowed me to query the data to facilitate the identification of patterns that would suggest automation or spamming. For example, if you sort by the account formation date, you can identify how many of the accounts tweeting the same material were created on the same day. These patterns were true of all the hashtags mentioned above. It is important to bear in mind that the tweets pulled would be those visible to the public; i.e. if you're a Twitter user and use the search feature to look for a hashtag, and click on the 'live' stream, these are the tweets produced in real time.

The data

On June 21 2016, I queried Twitter's API for the phrase 'Isa Qasim'. I received 628 tweets in return. Of these 219 were identical; that is, tweets with the text 'Isa Qasim, the #Shiite #terrorist, telling followers to annihilate #Bahrain's Security Forces'. On June 22, I requested tweets from the Twitter API under the #Bahrain hashtag. This returned 10887 tweets from a 12 hour period. Between 10 and 12 July, I queried the Twitter API on the 'Yemen hashtag. 11,541 tweets were pulled from the Twitter API over an approximately 48 hour period. Analysing the tweets revealed patterns that indicated, beyond reasonable doubt, that the accounts were linked to some institution, individual, or organisation for the purpose of promoting certain ideas. Some of the patterns suggest that the accounts may be deliberately automated, yet it is also feasible that they are operated by a group of people working in the same organisation/institution.

1. There are only a limited number of unique tweets, each of which is cycled repeatedly by the Twitter accounts on loop.
2. Every one of the tweets from each suspicious account, with the exception of their first tweet, were launched from TweetDeck - a programme favoured by marketeers that allows one to manage multiple accounts from a single machine.

3. Examining new accounts highlighted an interesting pattern. The very first tweet from one of these accounts contained an unusual idiom, saying, or phrase in Arabic. This idiomatic phrase was always launched from 'Twitter Web Client', while the rest of the tweets were launched from TweetDeck. I tested this on about 10 of the accounts registered on 23 June 2016. One of the examples was 'اليس يحب الخناقاة', the idiomatic translation of which I am told is "People love their oppressors lit: cats love their stranglers/cats love to fight".
4. All the accounts have a similar, low number of followers and people they follow. This tends to range between 30 and 60 followers. Most of them also follow specific, mostly Saudi-based news sites, that engage in similar propaganda.
5. All the accounts were created in batches on consecutive days within certain months. For example, in the sample from the #Bahrain hashtag, approximately 101 accounts were created on either the 26, 27, 28 or 29 August 2013. Around 200 were created on 6, 8, 9, 10 and 11 June 2014. Again these figures suggest that these accounts were created in batches and in a co-ordinated fashion.
6. Accounts created on the same day tend to have a similar number of tweets. E.g. those accounts created on 2 February 2014 all have about 400 tweets to their name, all those created in March 2014 have about 3,500, and all those created in June 2014 have about 7000. Obviously this figure will change, yet the correlation between the date of account creation and number of tweets is interesting, especially given that the older accounts actually have less tweets.
7. The creation date of each count defines its biographical information. This would suggest that no one is updating the old biographies. For example the accounts created in 2016, unlike the other accounts, have a biography and a header image. Accounts created after June 2016 also have birthplace information, and user-inputted locational information. This location is always the name of a town or city in Saudi Arabia. Certainly the filling out of this biographical data is designed to make the accounts look more credible.
8. When you copy one of the Arabic tweets and paste it into Twitter's search facility, you get the Top Tweet, which in this case is the person who usually first tweets it. What almost always happens is that the

first person to tweet the information was also one of the suspicious accounts. In particular, a lot of the recent tweets seem to originate with those suspicious accounts that were set up in 2016.

9. The biographical information that does occur is fairly generic. By doing a crude corpus-based aggregation, and after removing a lot of commonly occurring prepositions, the most common word is 'Allah', and various other religious idioms or pleasantries. This is perhaps to convey a sense that the accounts are pious or inoffensive. The profile pictures and images tend to be rather generic, usually consisting of slightly pixelated traditionally dressed Gulf men or children (not unusual in the Middle East for a display picture). Former Guardian journalist Brian Whitaker looked into (<http://al-bab.com/blog/2016/06/how-robots-have-over-run-bahrain-hashtag-twitter>) these in more detail, noting that 'Another account, quoting the Prophet in its bio, apparently belongs to a religious gentleman called Mazher al-Jafari but the profile picture comes from the Facebook page of Omar Borkan al-Gala, an Iraqi-born male model who lives in Vancouver'.



A word-cloud offers a crude corpus based analysis of the Twitter bios; the most common word being 'Allah'

2	6/2/2016 14:27:49	@AbaharAlQasbi	ساعة التوسري	Isa Qasim, the #SHIK	745178962377574493	TweetDeck	6	40	0	0	No	04/02/2014
3	6/2/2016 14:51:08	@AbdulsalamAlmusa	عد لعدوم السيف	Isa Qasim, the #SHIK	745184729166496883	TweetDeck	8	39	0	0	No	04/02/2014
4	6/2/2016 14:58:37	@AbdulwahabAla	عد الفلمت	Isa Qasim, the #SHIK	745186651107421185	TweetDeck	6	40	0	0	No	04/02/2014
5	6/2/2016 14:51:08	@AbdulwahabAla	عد الفلمت	Isa Qasim, the #SHIK	745184729159389888	TweetDeck	3	39	0	0	No	04/02/2014
6	6/2/2016 14:51:08	@AbdulwahabAla	عد الفلمت	Isa Qasim, the #SHIK	745184729087721600	TweetDeck	6	40	0	0	No	04/02/2014
7	6/2/2016 14:51:08	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745184729097280116	TweetDeck	3	40	0	0	No	04/02/2014
8	6/2/2016 14:27:49	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	745178962398541824	TweetDeck	2	40	0	0	No	04/02/2014
9	6/2/2016 14:27:49	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745178962413324256	TweetDeck	6	40	0	0	No	05/02/2014
10	6/2/2016 14:27:49	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	745178962496337408	TweetDeck	8	40	0	0	No	05/02/2014
11	6/2/2016 14:58:37	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	7451866511178424	TweetDeck	5	40	0	0	No	05/02/2014
12	6/2/2016 14:27:49	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745178962496337408	TweetDeck	3	40	0	0	No	05/02/2014
13	6/2/2016 14:58:37	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	745186651306230416	TweetDeck	6	40	0	0	No	05/02/2014
14	6/2/2016 14:58:37	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	74518665138775297	TweetDeck	6	40	0	0	No	05/02/2014
15	6/2/2016 14:58:37	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745186651307300128	TweetDeck	4	40	0	0	No	05/02/2014
16	6/2/2016 14:51:08	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	745184729062857472	TweetDeck	4	40	0	0	No	05/02/2014
17	6/2/2016 14:58:37	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	7451789624153633626	TweetDeck	3	40	0	0	No	05/02/2014
18	6/2/2016 14:58:37	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	74518665153633626	TweetDeck	9	40	0	0	No	05/02/2014
19	6/2/2016 15:02:38	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745187642673110272	TweetDeck	3	35	0	0	No	02/03/2014
20	6/2/2016 15:39:13	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745196849378978104	TweetDeck	5	34	0	0	No	02/03/2014
21	6/2/2016 14:47:52	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745183947596739504	TweetDeck	8	32	0	0	No	02/03/2014
22	6/2/2016 15:31:51	@AbdulwahabAla	مبارك الفلمت	Isa Qasim, the #SHIK	745195017749461824	TweetDeck	4	31	0	0	No	02/03/2014
23	6/2/2016 14:29:25	@AbbasAlmusa	مبارك الفلمت	Isa Qasim, the #SHIK	745178304183629473	TweetDeck	4	34	0	0	No	02/03/2014

This table shows a typical snapshot of the type of data pulled; note the identical tweets, the similar account creation dates (far right), and the fact they were all launched from TweetDeck

What do they tweet about?

The contents of the tweets varies, depending on the hashtag. The majority are actually on the #Saudi hashtag, and contain links to videos from the satellite broadcaster Saudi 24. The text in the tweets usually lionizes the Saudi government, or praises Saudi's efforts and intervention in Yemen, pointing to pro-Saudi propaganda. Ongoing research by myself and Bahrain Watch is further exploring this link with Saudi 24. However, the sectarian tweets common on the Isa Qasim hashtags are of particular interest. If you look at the table below, you will see that 51% (5556) of tweets on the Bahrain hashtag during the 12 hour period sampled on 22 June 2016 were most likely produced by bots or spammers with a sectarian, hate-inciting agenda. Of these 5556 tweets, there were only four unique ones, each tweeted hundreds or thousands of times by multiple different accounts.

No. of tweets	Translation	Original Tweet
2742	Persians and the Majus (Zoroastrians) hate the Arabs	الفرس والمجوس حقد على العرب
946	The Iranian regime are the leaders in war and the manufacturing of terrorism and crises #withdrawalofnationalityfromIsaQasim #Bahrain #Fitna	النظام الإيراني رأس حربة في صناعة الإرهاب و الأزمات https://t.co/ZcQaOUvn1T ... # سحب جنسية عيسى قاسم # البحرين # الفقيه # القائد # Bahrain
95	Iran's Mullah's politicise the Hajj (pilgrimage) with slogans outside Islam and the Sunna of the Prophet #withdrawalofnationalityfromIsaQasim #Bahrain #Fitna	دولة ملالي # إيران تسييس الحج بشعارات خارجة عن الإسلام و السنة النبوي # البحرين # https://t.co/OHCOKhuJf6 ... آية الله الشيخ عيسى قاسم # Bahrain # Fitna
1773	Sowsan Sha'ir*: Bahrain will not listen to the threats of Qasim Suleimani**, or others like him. *Sowsan Sha'ir is a pro-status quo Bahriani Columnist **QS is an Iranian general who recently said that the move against Isa Qasim would cause a fire in the region	سوسن الشاعر: البحرين لا تعبأ بتهديدات قاسم سليمانى ولا غيره https://t.co/zkHjKY0u9Q ... سحب جنسية # البحرين # الفقيه # Bahrain # عيسى قاسم # البحرين # القائد #

Table showing breakdown of tweets from suspicious accounts on 22 June 2016

As you can see from the above table, 5,556 of the 10,866 tweets examined on 22 June 2016 were from these suspicious accounts. That is 51% of the total number of tweets. They condemn 'terrorist' acts in Saudi's mostly Shia Eastern province, and acts by the 'Shia' opposition in Bahrain. I use the term Shia here because it is mentioned frequently, as are derogatory terms, such as Rawafid (a term used to mean the Shia are 'rejectionists' of the true Islamic faith). This tweet, for example, 'رد قوي من شاعر سعودي على', translates as 'strong response from Saudi poet against the Rawafid' (a video (<https://www.youtube.com/watch?v=TTpccboRNoAc>) was attached to the poem in question). Iran is frequently criticised. Look, for example, at the tweet, 'Iran's Mullah's politicise the Hajj (pilgrimage) with slogans outside Islam and the Sunna of the Prophet #withdrawalofnationalityfromIsaQasim #Bahrain #Fitna'

The most common tweets on the Yemen hashtag were;

العوين:مجئ الخوميني هو بداية الشر و اشعال الحروب في المنطقة
<https://t.co/wkYWbGT3eG> ... #ايران #Iran #الله #yemen
#Iraq

(translation: The appearance of Khomeini was the beginning of evil and the spread of wars in the region #Shaykh_Isa_Qasim, #Hezbollah, #Yemen, #Iraq, #Iran

توجيهات خادم الحرمين بايصال رسالة المملكة للعالم كله ومنها برنامج الإمامة في الخارج
<https://t.co/QZif2VH3PP> ... #ايران #Iran #الله #yemen (The custodian of the holy mosques sends a message to the whole world....)

Again, we see a similar theme emerging, with Khomeini being criticised as the cause of wars in the region.

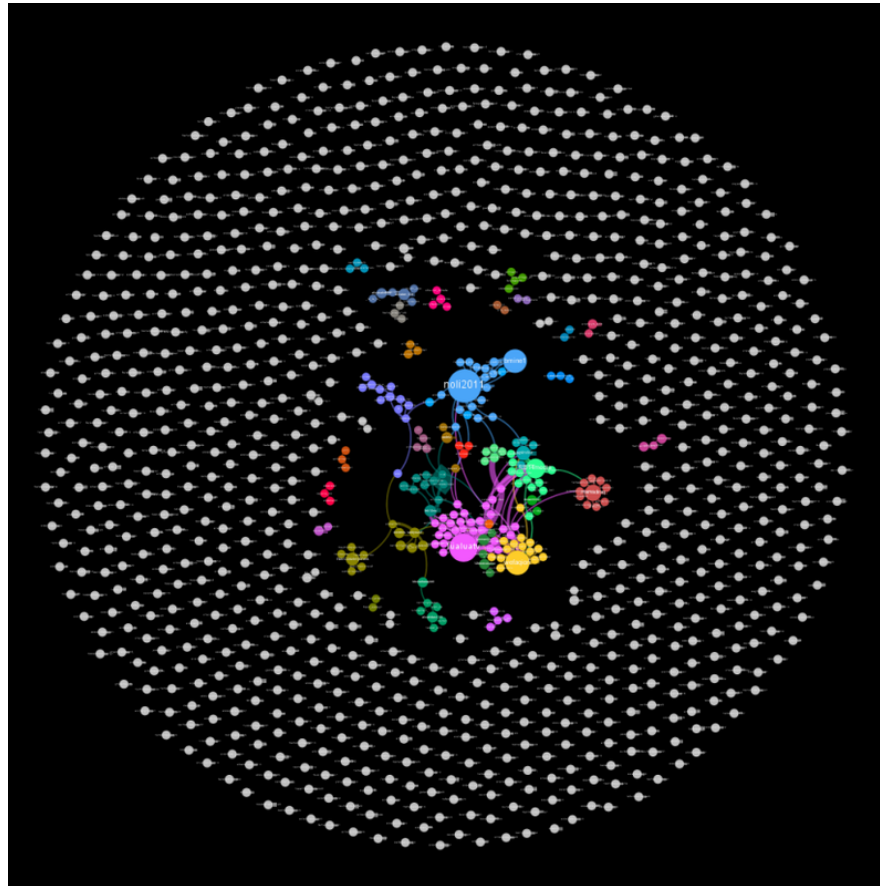
Anti-Social Media

The relevant thing is that hundreds of what seems to be automated or spam Twitter accounts are repeating propaganda that conflates acts of violence, terrorism and unrest, with both Arab Shia and Iran. This strongly suggests that institutions, people or agencies with significant resources are deliberately creating divisive, anti-Shia sectarian propaganda and disseminating it in a robotic and voluminous fashion. The problems here are numerous, yet such accounts can not only contribute to sectarianism (hard to infer causal relations from this), but create the impression that policies, such as the denationalisation of Isa Qasim, have widespread popular support. In addition, a vast majority of the accounts discovered actually tweeted nothing.

While the notion of bot accounts is probably not news to anyone, the evidence here hopefully highlights that much online sectarian discourse is perhaps inflated by those groups or individuals with specific ideological agendas who have the resources to do so. We are aware that PR and reputation management companies offer such services, yet their work is often done secretly and behind close doors.

In addition to spreading sectarian discourse and pro-Saudi government propaganda, the sheer volume of the tweets drowns out legitimate tweeting at certain times of the day. The below visualisation (using Gephi (<https://gephi.org/>) and NodeXL (<https://nodexl.codeplex.com/>)) of the #Bahrain hashtag on 22 June 2016 highlights this insulating effect. The colourful centre of the image represents mostly legitimate Twitter accounts tweeting on the Bahrain hashtag. The lines between them, and their proximity, represents that they are interacting with one another. The larger nodes represent those accounts with more 'influence', that is to

say, accounts connected to other highly connected accounts. Nodes of the same colour represent communities of tweeters, who have interacted regularly recently (in the sample of tweets). All the grey nodes are small, and disparate from the network, not connected to each other or anyone else. This means that they tweet, but do not interact with anyone.



Visualisation made on Gephi (<https://gephi.org/>) by Marc Owen Jones of the #Bahrain hashtag on 22 June 2016

Essentially it would suggest they are individuals tweeting but not interacting (a bot). The fact that the visualisation shows them on the periphery of the community is deliberate, as it highlights how they are not actually engaged with the community. A small number of these grey dots are legitimate people who just tweeted on the Bahrain hashtag, but the majority are bots. However, the visual strongly shows a grey shield around a core, attempting to make real debate impenetrable. It blocks out prying eyes! It also gives you a sense of how many bots there actually are.

The scale of the operation though is enormous, and also includes dormant Twitter accounts used as 'fake followers' to support the follower stats of other accounts. There are potentially up to a million of these accounts, many of which are, according to their profile, based in Saudi Arabia. While removing them all would put a modest dent in Twitter's overall world-wide users, it would certainly affect regional figures, which tout Saudi as the most connected Twitter nation in the MENA region. In 2014 it was believed that 40% of 2.4 million MENA Twitter users were based in Saudi. If up to a million of these accounts were fake, this would almost half Saudi's Twitter subscription stats.

However, even though Twitter announced that they had suspended 235,000 accounts since February 2016 because of their links to the promotion of terrorism, it is unclear whether these include any of the accounts studied in this investigation. Interestingly, one of the suspicious accounts we analysed whose followers mostly appear to be fake, saw a drop in followers of around 100,000 in the past few weeks.

Whether they are bots or people paid to use TweetDeck to promote a certain agenda is not yet clear. While there is evidence to suggest that the accounts are automated, it would not be uncommon in the gulf for an operation such as this to use cheap human labour. Despite what Twitter says, it is a difficult task to stop; an analysis of the Saudi hashtag on 30 August 2016 revealed that at least 1100 of these spam accounts have been created since July 2016. It would, however, be logical to deduce from this that an organisation in Saudi, with or without government approval, is operating an anti-Iranian, anti-Shia, but pro Saudi government propaganda campaign on a massive scale.

Conclusion

While the operation in Bahrain and Saudi has been going on for a while, it is not clear to what extent this exact tactic has been adopted elsewhere, or whether it is a solely regional variant of social media disinformation. Regardless, it fits in with the multiple variants of online troll armies, the impact of which is to insert disinformation or spread propaganda. In the case of what seems like a potentially automated operation, the high volume of tweets also leads to a flushing out of useful information. This means that those seeking legitimate information on the Gulf region are often being inundated with propaganda to such an extent, that it makes

finding real tweets far more difficult when looking in real-time. This is especially true when these accounts are turned to hijack certain hashtags.

It still remains to be seen whether this is a state-sponsored operation, the work of a PR company, or a wealthy individual's unilateral project, but the scale is such that it certainly impacts upon the credibility of Twitter as a tool that allows people to share and receive ideas without barriers. It also sheds doubt on Twitter's ability or commitment to tackling online hate speech. Yet even with the greatest intention by social media companies, it remains a sad truism that the profound influence of certain agencies or states distorts the online public sphere by disproportionately allowing those with wealth and power to shape the nature of discourse available to other netizens.

Marc Owen Jones is a Research Fellow at the Institute for Arab and Islamic Studies at Exeter University, and a researcher and director at Bahrain Watch, an NGO doing investigative transparency work on the Gulf region. He is also co-editor of Bahrain's Uprising: Resistance and Repression in the Gulf, published by Zed Books. Read more of his work here (<https://marcowenjones.wordpress.com/>) and find him on twitter here (https://twitter.com/marcowenjones?ref_src=twsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor).