Russian Influence Operations on Twitter

Summary

This short paper lays out an attempt to measure how much activity from Russian state-operated accounts released in the dataset made available by Twitter in October 2018 was targeted at the United Kingdom. Finding UK-related Tweets is not an easy task. By applying a combination of geographic inference, keyword analysis and classification by algorithm, we identified UK-related Tweets sent by these accounts and subjected them to further qualitative and quantitative analytic techniques.

We find:

- There were **three phases in Russian influence operations**: under-the-radar account building, minor Brexit vote visibility, and larger-scale visibility during the London terror attacks.
- Russian influence operations linked to the UK **were most visible when discussing Islam**. Tweets discussing Islam over the period of terror attacks between March and June 2017 were retweeted 25 times more often than their other messages.
- The most widely-followed and visible troll account, @TEN_GOP, shared 109 Tweets related to the UK. Of these, **60 percent were related to Islam**.
- The topology of tweet activity underlines the vulnerability of social media users to disinformation in the wake of a tragedy or outrage.
- Focus on the UK was **a minor part of wider influence operations in this data**. Of the nine million Tweets released by Twitter, 3.1 million were in English (34 percent). Of these 3.1 million, we estimate 83 thousand were in some way linked to the UK (2.7%). Those Tweets were shared 222 thousand times. It is plausible we are therefore seeing how **the UK was caught up in Russian operations against the US**.
- Influence operations captured in this data show attempts to **falsely amplify other news sources and to take part in conversations around Islam**, and rarely show attempts to spread 'fake news' or influence at an electoral level.

Background

On 17 October 2018, Twitter released data about 9 million tweets from 3,841 blocked accounts affiliated with the Internet Research Agency (IRA) – a Russian organisation founded in 2013 and based in St Petersburg, accused of using social media platforms to push pro-Kremlin propaganda and influence nation states beyond their borders, as well as being tasked with spreading pro-Kremlin messaging in Russia. It is one of the first major datasets linked to state-operated accounts engaging in influence operations released by a social media platform.

Caveats

The analysis presented here is based on a dataset released by Twitter in October 2018. Although large, we cannot say with confidence what proportion of Russian state-operated accounts that were active over the period the data represents. Given the number of users posting in English is around four thousand, we expect there to be significantly more accounts that have either not been detected or were not contained in the data released. Although this is a useful window into Russian influence operations, we cannot be sure it is a representative one. We are equally dependent on Twitter's determination that these are indeed Russian state-operated accounts.

One of the major questions that this analysis cannot answer is whether this data set reveals Russian operations against the UK directly, or a small part of a Russian operation against the USA which happened to include UK-related messaging. It is plausible that the data is limited to US-focused influence operations, and mentions of the UK are included as collateral and as a means of influencing public opinion in the US. We cannot therefore discount the possibility that we are examining US-focused data, and that unreleased data may shed further light on UK-focused operations.

As part of this research, we rely on probabilistic classification – both to locate those Twitter users mentioned by Russian state-operated accounts to the UK, and to classify the contents of the messages they sent. We believe the use of natural language processing (NLP) classification to be an improvement over keyword analytics, but despite our classifiers operating at high levels of accuracy – shown in Appendix 1 – they are not perfect. A full methodology is included.

Analysis

Estimate of UK Focus

Analysts looked to estimate the extent to which the accounts in the dataset targeted the United Kingdom. To do this, three layers of classification were applied.

1. **UK Mentions & Retweets: 75,787 Tweets**

All users who had been mentioned or retweeted by one of the state-operated accounts was passed through a geolocation algorithm, using available evidence contained in the data to determine their likely country. Of the three million or so English-language Tweets sent, 76 thousand mentioned or retweeted an account we estimate to be UK-based. This included ordinary Twitter users, journalists, MPs and media outlets.

2. **UK Replies: 3,106 Tweets**

Following the methodology for retweets and mentions above, the process was repeated for users to whom a Russian-linked state-operated account had replied. Just over three thousand Tweets were identified in this way.

3. **UK Keywords: 16,381 Tweets**

Tweets were also checked against a list of keywords tying them to the UK. This included uniquely British political events (Brexit, UK General Elections), other UK events (terrorist attacks, television programmes), and UK political figures (MPs, journalists). This process added 16 thousand Tweets.

Tweets could be identified as being linked the UK through overlapping methods. In total, 83,075 unique Tweets were classified as being connected to the UK.

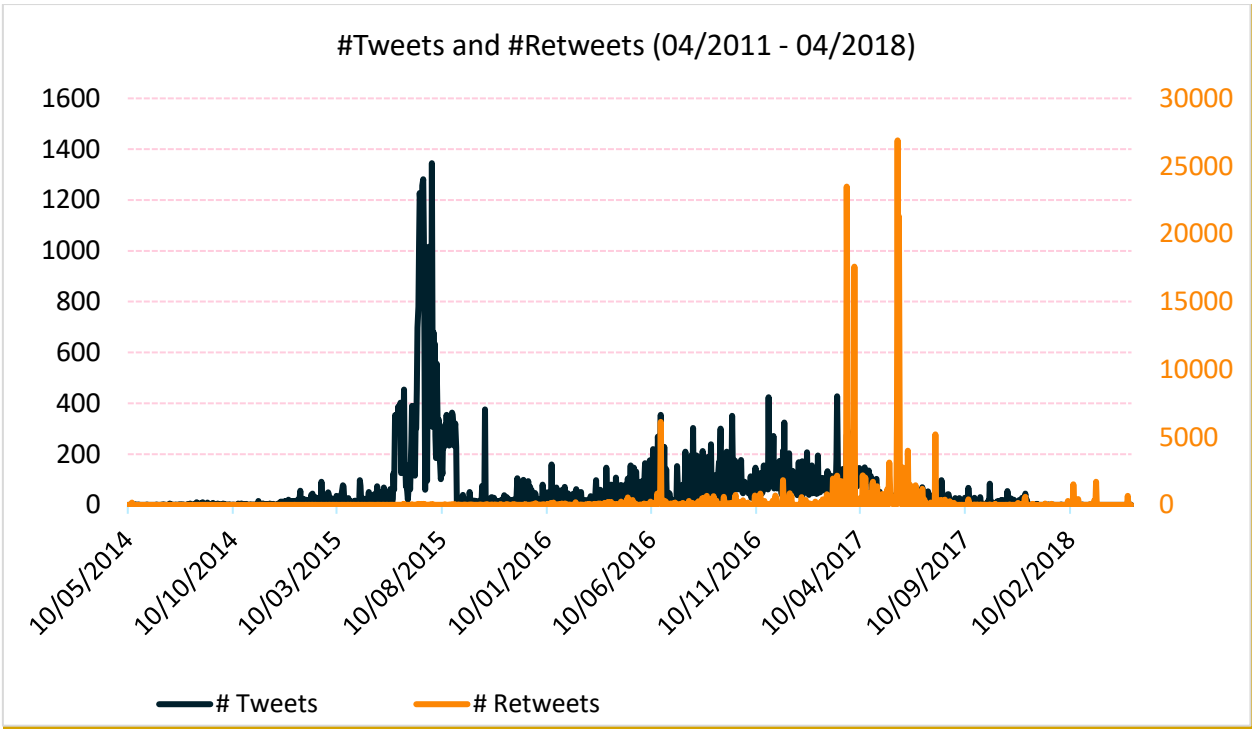
This number is low, given that recent estimates say Twitter users post nearly 500 million tweets per day in 2018, though we must keep in mind that this dataset likely represents a fraction of the accounts operated or controlled by Russian state-linked operatives. With two exceptions, detailed in 'Tweets over Time' below, it is likely that the majority of content produced by state-operated accounts was not widely read or interacted with.

Tweets over Time

Researchers investigated the activity and reception of Tweets sent by Russian state-operated accounts over time.

Examining activity levels over time shows how state-operated accounts operated and when their messaging was most likely to have reached the widest audience. The graph below shows the number of daily Tweets sent by these accounts (in black, left axis) and the number of shares those Tweets received (in orange, right axis).

Chart 1: Tweets and Retweets over Time



Between 1 January 2015 and 1 January 2016, the Russian state-operated accounts posted 43.5 thousand Tweets linked to the United Kingdom, peaking at 1,345 Tweets sent on 27th July. That year, the tweets were shared a total of three thousand times, or 0.07 times per Tweet on average.

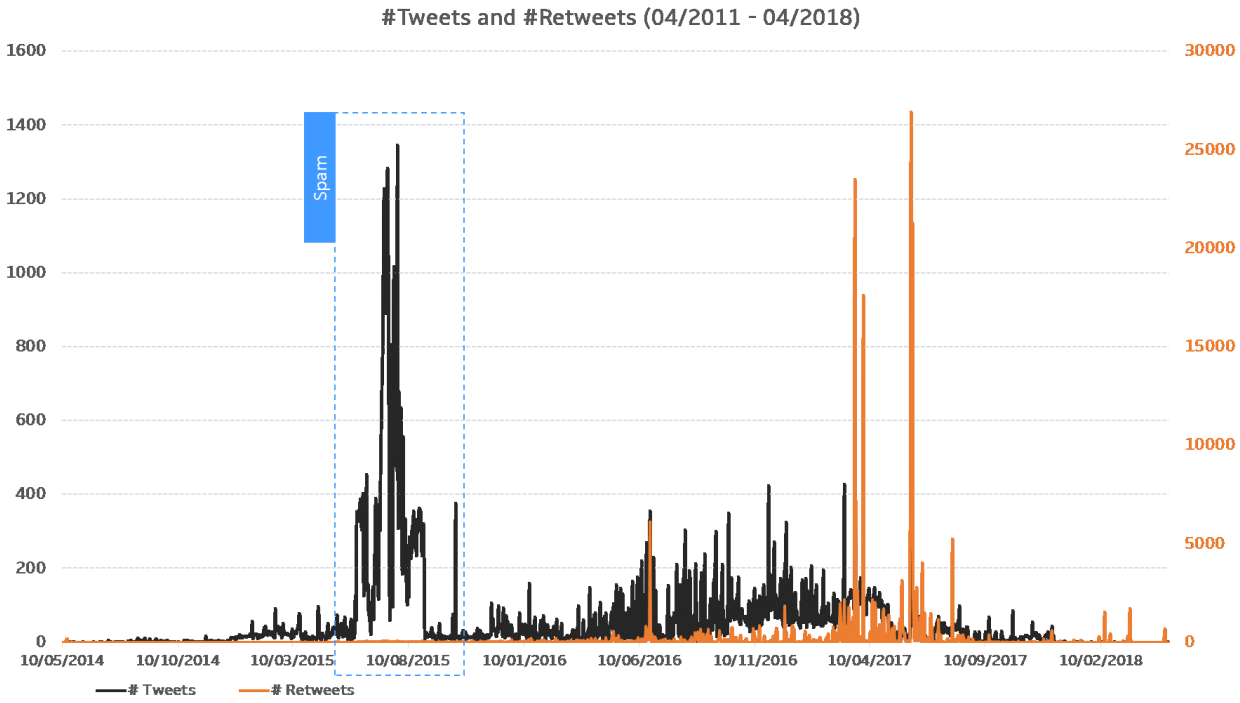
By contrast, between 1 January 2017 and 1 January 2018, the accounts posted 16.8 thousand Tweets. These tweets were shared 178 thousand time, or 10.6 times per Tweet on average, peaking around the London Bridge terror attacks in June 2017.

This shows a stark difference in the how far state-operated account messaging was likely to have carried into Twitter users' timelines. In 2015, these accounts went largely under the radar, sharing messaging that was not amplified and did not leave an impression on the platform. In 2017, the content was significantly more widely shared.

We understand the graph to break into three broad areas of interest.

Phase One: Spam and the Process of Building of Credible Accounts

Chart 2: Tweets and Retweets over Time



Between late May and late August 2015, English-language Tweets from accounts in the dataset increased significantly, averaging over 300 Tweets per day over the period, representing the single highest bust of activity across the timeframe. Engagement, as noted above, was extremely low. A manual coding of 100 of these Tweets is shown in the table below.

Table 1: Coding of Tweets over Phase One

Category	% Tweets
Fitness & Exercise	59
Chain Tweet/Spam	30
News Sharing	6
Other	5

The majority of Tweets were related to fitness and exercise. On examination, Tweets look like they were procedurally generated, either from a corpus of fitness related sentences or from other peoples' fitness and exercise Tweets. Examples are shown below:

> *I'm ready to eat healthy and workout. @xhibellamy @William_Stokes @guru_paul @ThomasAmor1 @jennyc08318 @richtweten http://t.co/TAZ9Co1QF9*

> *http://t.co/t1wInUwpjd Eat healthy b's exercise @jenannrodrigues @Embarrasthykids @brawlinbaby @x_Jems_x @RozaPayne4 @BlueEagle212*

Chain tweets also appear to be procedurally-generated strings of Twitter users linked together with the first name of the user, though unlike fitness and exercise Tweets they appear to be completely nonsensical. These tweets made up 30 percent of the sample. Examples are shown below.

> *.@pedrareyes148 pedra @Chloe0354 ASDFGchloeHJKLL? @pulmonxry Yeezus @Nick281051 Nick @puffylore163 lore http://t.co/ZLpIlrsV33*

> *.@__ilianaromero Iliana @faniiifordaze free @5hljp chris @AnjanettKay21 Anjanett @Theblessone_11 *IWillMakeIt11* http://t.co/CnEUewfDRT*

The remaining 11 percent were a mix of news sharing and other Tweets, usually short replies to a news agency or another Twitter user with no discernible political or social aim. Examples of both categories are shown below.

> *@BBCBreaking That's horrible!*

> *@TheEconomist life is alawys preferable*

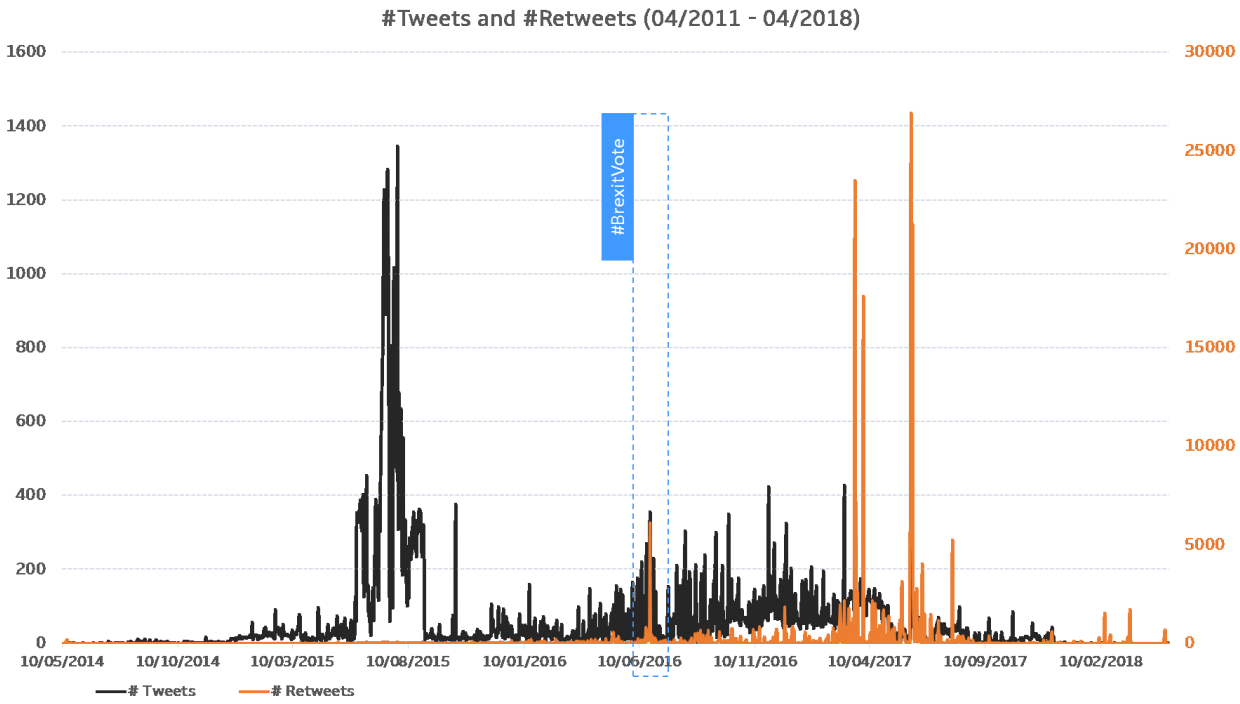
> *British TV personality Cilla Black dies aged 72 #life*

> *Swiss stocks - Factors to watch on June 29 - Reuters UK @swissbusiness http://t.co/O3veEmmqeY*

The likely purpose of these Tweets was to build the authenticity of an account by increasing the followership and visibility of that account, and by building metrics of Twitter activity. Sustained low-level engagement suggest those behind the accounts were not looking to push this content into the public domain. Rather, this was an attempt to camouflage fake accounts and begin to infiltrate the wider conversation.

Phase Two: #BrexitVote

Chart 3: Tweets and Retweets over Time



The 24th June 2016 saw the first Tweets that show a major spike in retweets in our dataset, on the day the results of the UK's Referendum on EU membership was announced. Russian state-operated accounts in the dataset sent 354 English-language Tweets on the day. On a manual review of 100 of these Tweets, 91 percent were related to Brexit, comprising a mix of news sharing, celebration and a small number of references to Islam and multiculturalism.

The table below shows the ten most frequently shared Tweets on the 24th June within the dataset.

Table 2: Top Tweets (by Retweets) on 24 June 2016

Tweet Text	User Screen Name	# Retweets
'We want our country back' first appeared in 1950's Britain it invoked this: #BrexitVote #Brexit https://t.co/iJlApW3UmS	Crystal1Johnson	764
Those who are still EU members can enjoy their political correctness and tolerance #BrexitVote https://t.co/VeMW7bagDQ	TheFoundingSon	735
This is the simplest explanation. Just like UK we too want to stop globalist liberals from ruining us! #BrexitVote https://t.co/XkNFpNof1c	redlanews	341
Oh the irony of #BrexitVote https://t.co/xZQMfV2cvz	User Screen Name Hashed	332
Algerian illegally in Britain attacked 8 women in ten days! Send the Muslim back to EU! #BrexitVote https://t.co/NEbSaJnQzV	TEN_GOP	326
Just found one more priceless picture! This time it's UK twitter trends. #BrexitVote https://t.co/5Vsx0L4rvP	User Screen Name Hashed	302
That is the sign that #BrexitVote was right choice! https://t.co/rgZcK2l7MA	SouthLoneStar	285
UK has no masters UK is free #BrexitVote https://t.co/0OQXv9ovvj	TheFoundingSon	241
Brits MADE UK GREAT AGAIN! It's time for us to MAKE AMERICA GREAT AGAIN! #BrexitVote #MAGA #IndependenceDay https://t.co/xbZYGt5tfC	TEN_GOP	232
I hope UK after #BrexitVote will start to clean their land from muslim invasion! https://t.co/C9JR9m9ewt	SouthLoneStar	201

We believe these tweets represent the first time that state-operated accounts were successful in reaching a wider audience, as measured by the spike in retweets. These popular tweets tend to be celebratory of Brexit.

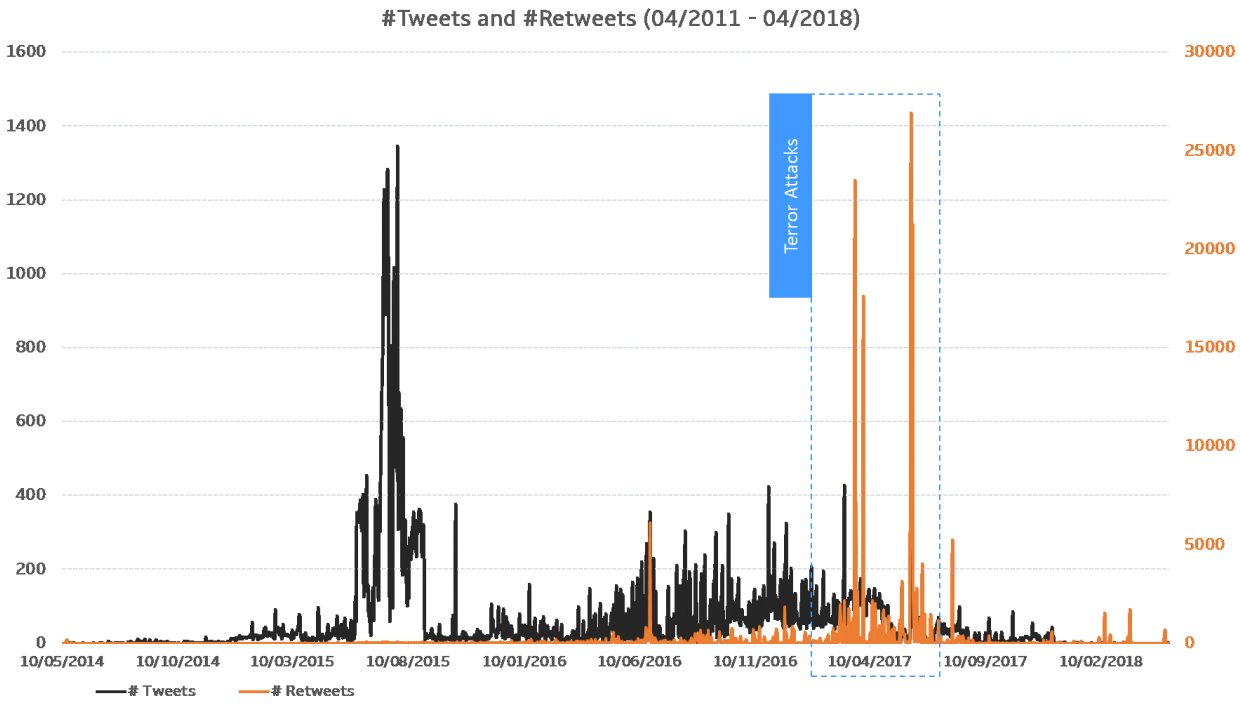
Interestingly, the accounts in the dataset were not highly vocal on Brexit in the six months prior to the vote, sharing 206 Tweets containing the keyword across that period, with the majority sharing news reportage, and with each Tweet shared 2.3 times on average. The most widely shared tweets in those periods related to Islam in the UK: over the six months prior to the Brexit vote, account sent 435 Tweets classified as related to Islam, with each Tweet shared ten times on average. This can be seen in the table below, showing the most widely shared Tweets in the six months prior to Brexit – six were directly related to Islam, and the use of #LondonHasFallen in a seventh was commonly used in connection to discussion of immigration and terrorism.

Table 3: Top Tweets (by Retweets, 24 Jan – 24 June 2016)

Tweet Text	User Screen Name	# Retweets	Topic
London: Muslims running a campaign stall for Sharia law! Must be sponsored by @MayorofLondon! #BanIslam https://t.co/9LAJrYfNrY	TEN_GOP	518	Islam
A boy just ran up to live the @FoxNews truck and said 'Can u report my smile?' This is Cameron.His smile's reported https://t.co/STVSmjVdrf	gloed_up	442	US/Other
Sharia NO-GO areas in BRITAIN. Citizens blocked from their own suburbs. Only #Trump can stop this here! https://t.co/IuQDe8rvPA	PamelaKealer13	319	Islam
That's what happened in Sweden, Germany and UK These countries invited #rapefugees and now they're paying the price https://t.co/g8WDF4eL7z	SouthLoneStar	242	Islam
Muslim Migrant Burns Five British Teenagers With Acid, One Victim May Be Permanently Blinded! https://t.co/qodCy8su4F	TEN_GOP	240	Islam
Welcome To The New Europe! Muslim migrants shouting in London “This is our country now, GET OUT!” #Rapefugees https://t.co/GCiFT96h76	PamelaKealer13	219	Islam
Britain wanted to ban Trump from entering the UK. Looks like British people don't agree with their government. https://t.co/Y7navHIFyg	SouthLoneStar	201	Trump
Churchill turning in his grave. Will be shocked to see Britain today. #LondonHasFallen https://t.co/LmdKDkFrpW	PamelaKealer13	195	Islam/Other
Queen Elizabeth II supported Brexit, thinks European courts “denigrate” Britain. God save the Queen! https://t.co/NNob65magN	USA_Gunslinger	184	Brexit
London is the first victim of Islamization! Will America be the next? #WorldRefugeeDay #NeverHillary https://t.co/UNqKI4ATjS	TEN_GOP	181	Islam

Phase Three: London Terror Attacks

Chart 4: Tweets and Retweets over Time



As shown in the chart above, there is a period of six weeks where UK-related Tweets sent by the Russian state-operated accounts were most widely shared. They coincide with the terror attacks carried out in London on the 22nd March and 3rd June 2017 and in Manchester on 22nd May 2017. During and after each attack, messages sent by Russian state-operated accounts were widely shared on the platform.

- London, Westminster (22nd March to 25th March) – **815 Tweets, 32,501 Retweets**
- Manchester – (22nd May – 25th May) - **156 Tweets, 5,674 Retweets**
- London, London Bridge (3rd June – 6th June) – **445 Tweets, 60,324 Retweets**

We cannot tell how many of these retweets were from other state-controlled or otherwise malicious accounts. Nevertheless, the spikes shown above represent the moments in this dataset when state-operated accounts were likely most visible to the average Twitter user.

Notably, the accounts in this dataset were significantly less active over the course of the Manchester terror attack. The tweets during this time period that received the largest number of shares are shown below. Without wider historical data against which to compare it is difficult to establish how widely this messaging would have been viewed, however, we believe that retweets above the tens of thousands are highly likely to be viewed by a significant number of UK twitter users.

Of the ten most retweeted messages, there appears to be an overwhelming focus on Islam in the UK.

Table 4: Top Tweets (by Retweets, 1 March – 30 June 2017)

Tweet Text	User Screen Name	# Total Shares for UK-Related Content	Related to Islam
Fahma Mohamed, 19, has made history by becoming one of the youngest people in the UK to receive a doctorate. https://t.co/Dhb2VZsDuc	Crystal1Johnson	17575	Other
Reminder: Mayor of London Sadiq Khan was a lawyer for a 9/11 terrorist Zacarias Moussaoui and has ties with Islamist movement worldwide. https://t.co/Fl56GdCrFk	TEN_GOP	13532	Islam
Mayor of London: "Terror attacks are part and parcel of living in a big city." Tokyo: biggest city in the world.. NO ISLAMIC TERRORISM. https://t.co/kQ01wSoX0d	TEN_GOP	6400	Islam
Hi @CNN, you were caught literally creating fake news,scripting a Muslim protest in London.Delete your account. Signed, the American People	TEN_GOP	6308	Islam
Just a gentle reminder that the Mayor of London Sadiq Khan called moderate Muslims "Uncle Toms". #LondonBridge https://t.co/wtnMIUhJLy	TEN_GOP	5021	Islam
Just a gentle reminder that the Mayor of London Sadiq Khan called moderate Muslims "Uncle Toms". #PrayForLondon #Westminster https://t.co/kCKMBS99Il	TEN_GOP	4910	Islam
Mayor of London calls for cancelling Trump's visit to the UK. But didn't say a word about banning terrorists from entering the country!	TEN_GOP	4801	Islam
#London terrorist Abu Izzadeen was a well known British Jihadist. This is him calling for Jihad 5 yrs ago. Why does Britain put up with it? https://t.co/bLvN3NEMi5	TEN_GOP	4198	Islam
Look how all of these "moderate Muslims" on Al Jazeera react to the London terrorist attack. #LondonBridge https://t.co/4VaKIVtCqn	TEN_GOP	3498	Islam
7 more dead in London because of climate change. Oh wait, nope, it's Islamic terrorism again. #LondonAttacks https://t.co/PUqc6wYf44	TEN_GOP	2851	Islam

This activity strongly suggests that attempts by Russian state-operated accounts to influence and increase the volume of Islam-related conversations in the wake of the terror attacks was successful.

Given the large numbers of interactions received by Islam-related messages in this period, analysts looked to isolate tweets from Russian-linked accounts that were about Islam and analyse them comparatively.

Islam-related Tweets

To further investigate the extent to which Islam was a primary area of focus by Russian-linked accounts during this period, a classifier was trained to recognise Islam-related messages and separate them from the remainder of the dataset. The classifier was trained on examples of tweets, and reported a 97 percent accuracy.

Tweets that were deemed relevant included messages about Islam and Muslims, about terror attacks known to be carried out by Islamist extremists, and about known Islamist terror organisations (such as ISIL and al Quaeda). Tweets categorised as 'other' included any message unrelated to Islam.

Between 1st March and 30th June, Russian-linked accounts sent 9,365 messages of which 1,159 were categorised as being related to Islam (12 percent).

The majority of Russian-account activity over this period was the amplification of other sources of news, discussed further below. 85 percent of Tweets were retweets of other Twitter users. Of the remaining 15 percent – original content from Russian-linked accounts – 29 percent were related to Islam.

However, not all accounts had similar followership, visibility or activity. The table below shows the activity over the period of the ten most visible accounts, measured by multiplying the number of Tweets they sent by the number of retweets those messages received on average.

Table 5: % Tweets related to Islam (10 most influential accounts)

User Screen Name	# Tweets x Average # Retweets per Tweet	# Tweets	# Tweets (Islam)	% Tweets (Islam)
TEN_GOP	94927	109	65	60%
Crystal1Johnson	23442	24	5	21%
Pamela_Moore13	19590	69	34	49%
SouthLoneStar	9792	27	15	56%
Jenn_Abrams	1229	76	16	21%
TheFoundingSon	1086	52	18	35%
USA_Gunslinger	658	23	15	65%
BlackNewsOutlet	456	2	0	0%
wokeluisa	213	4	1	25%
DailyLosAngeles	140	42	19	45%

This shows that the most visible accounts were even more focused on Islam as a subject of discussion. In our estimation, the 'big guns' focused on Islam more often than the smaller, less-well followed state-operated accounts. 43.9 percent of their 428 Tweets were related to Islam.

The comparison is starker when extended across the dataset as a whole. Across all 88,075 Tweets we judged to be related to the UK, just four percent were related to Islam. Yet for the ten most visible accounts, that average was four times higher – 17.6 percent. For the most visible account (TEN_GOP), the number was 43.5 percent.[1]

To estimate the relative visibility Russian messaging around Islam had by comparison with the remainder of their messaging, analysts compared the average number of retweets and likes the two categories of messaging received. The results for the period over the London terror attacks are shown below in the table below.

[1] Tennessee GOP (@TEN_GOP) was a Twitter account falsely claiming to be run by the Republicans in Tennessee, but was in fact operated by a Russian state operative. The account had at 136,000 followers at one point, and premiered on pro-Trump, anti-Liberal partisan content.

Table 6: # Tweets, average # Retweets and # Likes per Tweet, per category (1 March – 30 June)

Category	# Tweets	Average # Retweets per Tweet	Average # Likes per Tweet
Islam	1159	102.0	98.3
Other	8206	3.7	4.2

Tweets that were related to Islam were far more widely shared and interacted with than tweets from other categories. During the period 1 Match – 30 June, Tweets about Islam were retweeted 25 times more often than other tweets on average, and liked 23 times more often on average. Expanding the analysis period to the dataset as a whole shows a similar pattern.

Table 7: # Tweets, average # Retweets and # Likes per Tweet, per category (April 2011 – May 2018)

Category	# Tweets	Average # Retweets per Tweet	Average # Likes per Tweet
Islam	3296	40.0	37.1
Other	79779	0.9	1.2

Again, tweets related to Islam were shared around forty times more widely than tweets related to other categories on average. We conclude that Islam-related Tweets were the messages that Russian influence operations had greatest success with when discussing UK subjects. Whether aimed at a US audience or a UK one, the three 2017 terror attacks in London and Manchester were exploited by these accounts, and the high levels of interaction suggest that their audiences were particularly receptive or vulnerable to this category of operation.

‘Astroturfing’ and non-Islam-related Content

Although Islam-related messages appear to be the content circulated by Russian-linked accounts that was most widely shared and interacted with, they made up 12 percent of the accounts’ output. Their remaining activity was primarily sharing other content, most often news sources, artificially amplifying content. 88 percent of ‘Other’ activity were retweets of other Twitter users. The table below shows the result of a manual coding of 100 Tweets in this category.

Category	% Tweets
UK News, Politics & Brexit	25
Pop History, Culture & Music	24
Foreign Affairs/News	16
UK Television and Culture	7
Sport	6
Other	21

On this sample, the ‘other’ category was wide-ranging. A quarter focused on the reporting of UK news, with multiple mentions of the Grenfell tower disaster and of the election campaign. A further quarter was the sharing of Twitter accounts dealing in popular history, culture and music, with one account - @oldpicsarchive – featuring prominently. Tweets about foreign affairs made up 16 percent of this sample, with a focus on UK/US politics and UK interactions with Russia. The remaining messages focused on media, culture and sport, including tweets about television programs ‘The Voice’ and the film ‘The Hate U Give’, as well as a number of tweets about football.

Conclusion

This report outlines the ways in which accounts linked to the Russian Internet Research Agency (IRA) carried out influence operations on social media and the ways their operations intersected with the UK.

The UK plays a reasonably small part in the wider context of this data. We see two possible explanations: either influence operations were primarily targeted at the US, or this dataset is limited to US-focused operations in which British Twitter users were impacted as collateral – that is to say, events in the UK were highlighted in an attempt to impact US public, rather than a concerted effort against the UK. It is plausible that such efforts also existed but are not reflected in this dataset.

Nevertheless, the data offers a highly useful window into how Russian influence operations are carried out, as well as highlighting the moments when we might be most vulnerable to them.

Between 2011 and 2016, these state-operated accounts were camouflaged. Through manual and automated methods, they were able to quietly build up the trappings of an active and well-followed Twitter account before eventually pivoting into attempts to influence the wider Twitter ecosystem. Their methods included engaging in unrelated and innocuous topics of conversation, often through automated methods, and through sharing and engaging with other, more mainstream sources of news.

Although this data shows levels of electoral and party-political influence operations to be relatively low, the day of the Brexit referendum results showed how messaging originating from Russian state-controlled accounts might come to be visible – on June 24th 2016, we believe UK Twitter users discussing the Brexit Vote would have encountered messages originating from these accounts.

As early as 2014, however, influence operations began taking part in conversations around Islam, and these accounts came to the fore during the three months of terror attacks that took place between March and June 2017. In the immediate wake of these attacks, messages related to Islam and circulated by Russian state-operated Twitter accounts were widely shared, and would likely have been visible in the UK.

The dataset released by Twitter begins to answer some questions about attempts by a foreign state to interfere in British affairs online. It is notable that overt political or electoral interference is poorly represented in this dataset: rather, we see attempts at stirring societal division, particularly around Islam in the UK, as the messages that resonated the most over the period.

What is perhaps most interesting about this moment is its portrayal of when we as social media users are most vulnerable to the kinds of messages circulated by those looking to influence us. In the immediate aftermath of terror attacks, the data suggests, social media users were more receptive to this kind of messaging than at any other time.

It is clear that hostile states have identified the growth of online news and social media as a weak spot, and that significant effort has gone into attempting to exploit new media to influence its users. Understanding the ways in which these platforms have been used to spread division is an important first step to fighting it.

Nevertheless, it is clear that this dataset provides just one window into the ways in which foreign states have attempted to use online platforms as part of wider information warfare and influence campaigns. We hope that other platforms will follow Twitter's lead and release similar datasets and encourage their users to proactively tackle those who would abuse their platforms.

Methodology

Further details on the analytical tools used to categorise the dataset are presented below.

The data released by Twitter contained tweets in several languages including English, French, Arabic, Russian and Spanish and about a variety of topics not related to Britain such as the latest US and French elections. In order to isolate tweets likely to be targeted at the UK researchers employed a range of classification techniques.

Language Annotation

First, tweets in English were separated from tweets in other languages –approximately 2.8 million tweets were in English (31 percent of the total data released).

Geoannotation

Using the screennames of users retweeted, mentioned or replied to by state-operated accounts, we built a set of usernames that, provided they were still active on the platform, could be geolocated.

Across replies, mentions and retweets, 78.9 thousand tweets referenced a user located to the UK.

Keyword Annotation

Tweets were also compared to a list of UK-specific keywords as a test for UK relevance. Beginning with a review of language relevant to the UK within the dataset, an iterative list was created and expanded to include:

1. Names and Twitter usernames of British MPs & MEPs
2. Names and Twitter usernames of prominent British journalists and media outlets
3. British political keywords
4. British cultural keywords

In various instances, keywords selected both relevant and irrelevant tweets; for instance, the keyword “England” selected both tweets containing the word “England and “New England” (USA). To filter out the irrelevant tweets without excluding the relevant ones we compiled a list irrelevant keywords and excluded tweets containing them from the final dataset. In total we collected 16 thousand tweets highly likely to be related to the UK using this technique. A full list of keywords is contained in Appendix 2.

Appendix 1

Method52

Data drawn from social media are often too large to fully analyse manually, and also often not amenable to the conventional research methods of social science. The research team used a technology platform called Method52, developed by CASM technologists based at the Text Analytics Group at the University of Sussex. [2] It is designed to allow non-technical researchers to analyse very large datasets like Twitter.

Data Analysis

Method52 allows researchers to train algorithms to split apart (‘to classify’) Tweets into categories, according to the meaning of the Tweet, and on the basis of the text they contain. To do this, it uses a technology called natural language processing. Natural language processing is a branch of artificial intelligence research, and combines approaches developed in the fields of computer science, applied mathematics, and linguistics.

An analyst ‘marks up’ which category he or she considers a tweet to fall into, and this ‘teaches’ the algorithm to spot patterns in the language use associated with each category chosen. The algorithm looks for statistical correlations between the language used and the categories assigned to determine the extent to which words and bigrams are indicative of the pre-defined categories.

The Accuracy of Algorithms

To measure the accuracy of algorithms into the categories chosen by the analyst, we used a ‘gold standard’ approach. For each, around 100 user descriptions were randomly selected from the relevant dataset to form a gold standard test set for each classifier. These were manually coded into the categories defined above. These tweets were then removed from the main dataset and so were not used to train the classifier.

As the analyst trained the classifier, the software reported back on how accurate the classifier was at categorising the gold standard, as compared to the analyst’s decisions. On the basis of this comparison, classifier performance statistics – ‘recall’, ‘precision’, and ‘F-score’ are

[2] This group is led by Professor David Weir and Dr Jeremy Reffin. More information is available about their work at: http://users.sussex.ac.uk/~davidw/styled-3/

created and appraised by a human analyst. Each measures the ability of the classifier to make the same decisions as a human in a different way:

Overall accuracy:
This represents the percentage likelihood of any randomly selected description within the dataset being placed into the appropriate category by the algorithm. It is based on three other measures (below).

Recall:
The number of correct selections that the classifier makes as a proportion of the total correct selections it could have made. If there were 10 relevant descriptions in a dataset, and a relevancy classifier successfully picks 8 of them, it has a recall score of 80 per cent.

Precision:
This is the number of correct selections the classifiers makes as a proportion of all the selections it has made. If a relevancy classifier selects 10 descriptions as relevant, and 8 of them actually are indeed relevant, it has a precision score of 80 per cent.

F-Score:
All classifiers are a trade-off between recall and precision. Classifiers with a high recall score tend to be less precise, and vice versa. The 'overall' score reconciles precision and recall to create one, overall measurement of performance for each decision branch of the classifier.

The values for each algorithm (called a classifier) are presented within appendix methodology of this report. The values are expressed as value up to 1: a value of 0.76, for instance, indicates a 76% accuracy.

Caveats:
The research of large social media datasets is a reasonably new undertaking. It is important to set out a series of caveats related to the research methodology that the results must be understood in the light of:

- The algorithms used are very good, but not perfect: throughout the report, some of the data will be misclassified. The technology used to analyse tweets is inherently probabilistic, and none of the algorithms trained and used to produce the findings for this paper were 100% accurate. The accuracy of all algorithms used in the report are clearly set out in this report.
- Twitter, and especially political Twitter, is not a representative window into British society: Twitter is not evenly used by all parts of British society. It tends to be used by groups that are younger, more socio-economically privileged and more urban. Additionally, the poorest, most marginalised and most vulnerable groups of society are least represented on Twitter; an issue especially important when studying the prevalence of xenophobia, Islamophobia and the reporting of hate incidents.

The accuracies of the algorithm used to classify for content related to Islam has a precision of 0.80 and a recall of 0.89 for Islam-related messages and a precision of 0.99 and a recall of 0.98 on other tweets. The final F-score was 0.97.

Appendix 2

List of keywords used for UK-related Keyword message analysis

Link to keywords on Demos website.