Studies in Conflict & Terrorism ›

Volume 42, 2019 - Issue 1-2: Islamic States Online Activity and Responses

Open access

13,578 Views | 34 CrossRef citations to date | 34 Altmetric

Listen

Original Articles

Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts

Maura Conway, Moign Khawaja, Suraj Lakhani, Jeremy Reffin, Andrew Robertson & David Weir

Pages 141-160 | Published online: 29 Oct 2018

Cite this article https://doi.org/10.1080/1057610X.2018.1513984 Check for updates

Abstract

This article contributes to public and policy debates on the value of social media disruption activity with respect to terrorist material. In particular, it explores aggressive account and content takedown, with the aim of accurately measuring this activity and its impacts. The major emphasis of the analysis is the so-called Islamic State (IS) and disruption of their online activity, but a catchall “Other Jihadi” category is also utilized for comparison purposes. Our findings challenge the notion that Twitter remains a conducive space for pro-IS accounts and communities to flourish. However, not all jihadists on Twitter are subject to the same high levels of disruption as IS, and we show that there is differential disruption taking place. IS’s and other jihadists’ online activity was never solely restricted to Twitter; it is just one node in a wider jihadist social media ecology. This is described and some preliminary analysis of disruption trends in this area supplied too.

In the aftermath of the London Bridge attack in June 2017, the British prime minister, Theresa May, warned social media companies, including Twitter and Facebook, that they must eradicate extremist "safe spaces."[1] She reiterated this in her speech to the World Economic Forum at Davos in January 2018, stating "technology companies still need to do more in stepping up to their responsibilities for dealing with harmful and illegal online activity. Companies simply cannot stand by while their platforms are used to facilitate … the spreading of terrorist and extremist content."[2] Prime Minister May's concerns about the use of the Internet, particularly social media, by violent extremists, terrorists, and their supporters are shared by an assortment of others, including academics, policymakers, and publics. Much of this is due to apparent connections between the consumption of, and networking around, violent extremist and terrorist online content and the internalization[3] of extremist ideology (i.e., "(violent) online radicalization"); recruitment into violent extremist or terrorist groups or movements; and/or attack planning and preparation. Apparently easy access to large volumes of potentially influencing violent extremist and terrorist content on prominent and heavily trafficked social media platforms is a cause of particular anxiety. The micro-blogging platform, Twitter, has been subject to particular scrutiny, especially regarding their response (or alleged lack of same) to use of their platform by the so-called Islamic State (IS), also known as Daesh or Da'ish.

Internet companies have responded both individually and collectively. On 26 June 2017, Facebook, Microsoft, Twitter, and YouTube jointly announced, via an agreed text posted on each of their company's official blogs, the establishment of the Global Internet Forum to Counter Terrorism (GIFCT).[4] They described the purpose of the GIFCT as "help[ing] us continue to make our hosted consumer services hostile to terrorists and violent extremists" and went on to state: "We believe that by working together, sharing the best technological and operational elements of our individual efforts, we can have a greater impact on the threat of terrorist content online."[5] In terms of individual companies' responses, in November 2017 Facebook published a blog post in their "Hard Questions" series addressing the question "Are We Winning the War on Terrorism Online?" Facebook announced in the post that it is able to remove 99 percent of IS and Al Qaeda material prior to it being flagged by users "primarily" due to advances in artificial intelligence techniques. Once Facebook becomes aware of a piece of terrorist

their being uploaded, the company said.[6] Missing from the update however were figures on how much terrorist content (e.g., posts, images, videos) is removed from Facebook on a daily, weekly, or monthly basis. Twitter is much less reticent on this point.

According to the section “Combating Violent Extremism” in Twitter’s twelfth Transparency Report, published in September 2017, in the period 1 January to 30 June 2017:

> … a total of 299,649 accounts were suspended for violations related to promotion of terrorism, which is down 20% from the volume shared in the previous reporting period. Of those suspensions, 95% consisted of accounts flagged by internal, proprietary spam-fighting tools, while 75% of those accounts were suspended before their first tweet.[7]

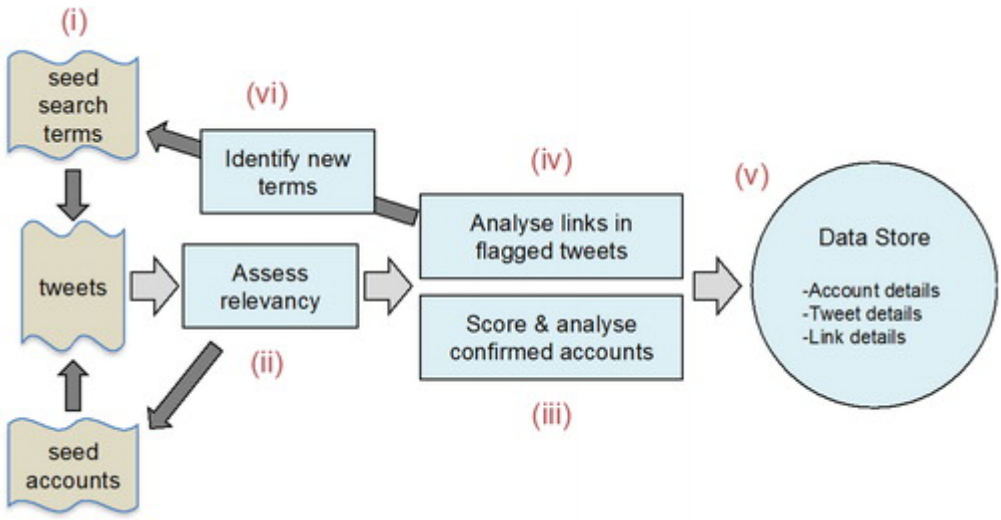
All told, Twitter claim to have suspended a total of 1,210,357 accounts for “violations related to the promotion of terrorism” in the period from 1 August 2015 to 31 December 2017.[8]

A disparity therefore exists between the assertions of policymakers, on the one hand, and major social media companies, on the other, as regards the levels and significance of their disruption activity. Although Twitter claims severe disruption of IS is occurring on their platform, detailed description and analysis of the precise nature of this disruption activity and, importantly, its effects are sparse,[9] particularly within the academic literature. This article aims to contribute to public and policy debates on the value of disruption activity, particularly aggressive account and content takedown, by seeking to accurately measure this activity and its impacts. The research findings challenge the notion that Twitter remains a conducive space for IS accounts and communities to flourish, although IS continue, to some lesser extent, to distribute propaganda through the channel. Not all jihadists on Twitter are subject to the same high levels of disruption as IS; however, this research demonstrates that a level of differential disruption is taking place. Additionally, and critically, the online presence of IS and other jihadists is not restricted to Twitter. The platform is merely one node in a wider jihadist online ecology. The article describes and discusses this, and supplies some preliminary analysis of disruption trends in this area too.

Methodology

To undertake the research, a semi-automated methodology for identifying pro-jihadist accounts on Twitter was developed (see Figure 1) and implemented using the social media analysis platform known as Method 52.[10] In the first instance, a number of candidate accounts of interest were identified. The approach was grounded in finding tweets that contained specific terms of interest (i.e., “seed search terms”), and/or the identification of accounts that were, in some way, related to other accounts known to be of interest (i.e., “seed accounts”) (see step (i) in Figure 1).

Figure 1. Detailed flow diagram for semi-automated social media analysis.



Display full size

If a tweet matched these search criteria, it was automatically analyzed to determine if it was relevant, using a machine-learning classifier trained to mimic the classification decisions of a human analyst.[11] A key task of the relevancy classifier was to separate target Twitter accounts from other Twitter accounts using similar language, such as those held by journalists or researchers, for example. If the tweet was deemed as relevant, further historic tweets were automatically extracted for the candidate account and assessed for relevancy (see step (ii) in Figure 1), providing the system with an aggregate view of the tweet history of the account. This overview of the tweet history was combined with other account metadata that could be extracted. These pieces of information were scored automatically and candidate Twitter accounts that exceeded

Figure 1). As portrayed in step (iv) in Figure 1, if one of the research analysts on the project confirmed that the account was pro-jihadist, then the out-links contained in all of the account's tweets were automatically analyzed, and details of the account, its tweets, and its links were stored in the database (see step (v) in Figure 1).

Information from new confirmed accounts was used by the system in a feedback loop to continually improve its efficiency, thereby identifying new seed search terms (see step (vi) in Figure 1) and providing additional seed accounts (see step (ii) in Figure 1).

Caveats

There were, however, a number of caveats attached to the data-collection that deserve mention. First, the bulk of the data was gathered over two months in early 2017 (February to April). The system to implement the semi-automated methodology was created, tested, and evolved throughout this period. The online accounts returned by the system were integrated with those found via traditional, manual search for accounts of interest. The overall approach was, therefore, a combination of automated and manual, and snowball and purposive sampling methods.

Second, not all available data were captured. There were some periods of downtime for the semi-automated system throughout this period as the methodology was developed and modified. In addition, certain accounts found via automated means were unable to be included due to them being taken down before the human analyst could assess and confirm their affiliation,[12] providing an early indication of the high levels of disruption taking place. By the end of the research, when the system was working optimally, 100 percent of these accounts were identified by the software as pro-IS, again reflecting the high level of disruption of IS-related accounts (discussed in further detail below).

Third, the semi-automated system primarily focused on pro-IS accounts operating in English and Arabic (or some combination of these languages). There is, then, the possibility that accounts tweeting primarily in, for example, Bahasa,[13] Russian, or Turkish were overlooked. This is worth noting, but probably negligible as the system's effectiveness improved as the research team learned more about pro-IS users' Twitter activity and refined the methodology accordingly. By early April 2017, for example, the software was able to detect accounts directly distributing IS propaganda with very high precision, regardless of what language was used. In addition, it is believed that the system also identified the majority of accounts linking to that propaganda.

Data

The research dataset comprised 722 pro-IS accounts (labeled Pro-IS hereafter) and 451 other jihadist accounts (labeled “Other Jihadist” hereafter), with at least one follower[14] active on Twitter at any point between 1 February and 7 April 2017 (see Table 1). Accounts were determined to be Pro-IS if their avatar or carousel images contained explicitly pro-IS imagery and/or text, and/or at least one recent tweet by the user (i.e., not a retweet) contained explicitly pro-IS images and/or text, such as referring to IS as “Dawlah” or their fighters as “lions.” Accounts maintained by journalists, academics, researchers, and others who tweeted, for example, Amaq News Agency content for informational purposes, were manually excluded. The Other Jihadist category included, among others, those supportive of Hay'at Tahrir al-Sham (HTS), Ahrar al-Sham, the Taliban, and al-Shabaab. Similar parameters were employed to categorize these accounts.

Table 1. Description of final dataset.

Download CSV Display Table

Accounts in the research database were located and identified in three different ways (see Table 2). The first set of accounts was manually identified by the research team, principally by examining known jihadi accounts (or those known to be of interest to jihadi supporters) and inspecting accounts within their networks (i.e., those following or being followed by them). A second group of accounts was identified “semi-automatically”—that is, automatically by the above-described social media monitoring system and then manually inspected by a human analyst who confirmed: (a) whether or not they were jihadist accounts; and (b), if they were, of what type. Several approaches were used to identify or generate seed accounts. This included analyzing vocabulary used in known jihadi accounts that were active during the time period studied or had recently been active. This enabled the team to determine which terms were being used much more often than would be expected statistically, and searching for tweets that contained these terms. These candidates were then winnowed based on the relevancy of their tweets in general (see above) and other metadata. Finally, a third set of accounts was identified automatically by the social media monitoring system,

[URLs] linking to official IS content hosted on some other platform or in some database on the Internet). These links were first identified through other tracking procedures, including (but not limited to) being spotted in confirmed IS tweets.

Table 2. Location and identification of Twitter accounts.

Download CSV Display Table

It is important to underline here that the Pro-IS account dataset is as close as possible—taking into account the caveats already made—to a full dataset of explicitly IS-supportive accounts with at least one follower for the period studied. On the other hand, the Other Jihadist dataset is a convenience sample of non-IS jihadist Twitter accounts collected for comparison purposes and in no way reflects the actual number of these accounts present on Twitter.

Measuring Disruption and Its Effects

Twitter was one of the most preferred online spaces for IS and their "fans,"[15] even prior to the establishment of their so-called caliphate in June 2014. It was estimated that there were between 46,000 and 90,000 pro-IS Twitter accounts active in the period September to December 2014.[16] However, their activity was subject to disruption by Twitter from mid-2014 and, although initially low level and sporadic, significantly increasing levels of disruption were instituted throughout 2015 and 2016. From mid-2015 through January 2016, for example, Twitter claimed to have suspended in the region of 15,000 to 18,000 IS-supportive accounts per month. From mid-February to mid-July 2016, this increased to an average of 40,000 IS-related account suspensions per month,[17] according to the company.[18] Despite the growing costs attached to remaining on Twitter (such as greater effort to maintain a public presence while relaying diffused messages and deflated morale), during this period IS supporters routinely penned online missives exhorting "Come Back to Twitter."[19] The question raised here is whether, in 2017, it was any longer worthwhile for pro-IS users to continue to seek to retain a presence on the platform?

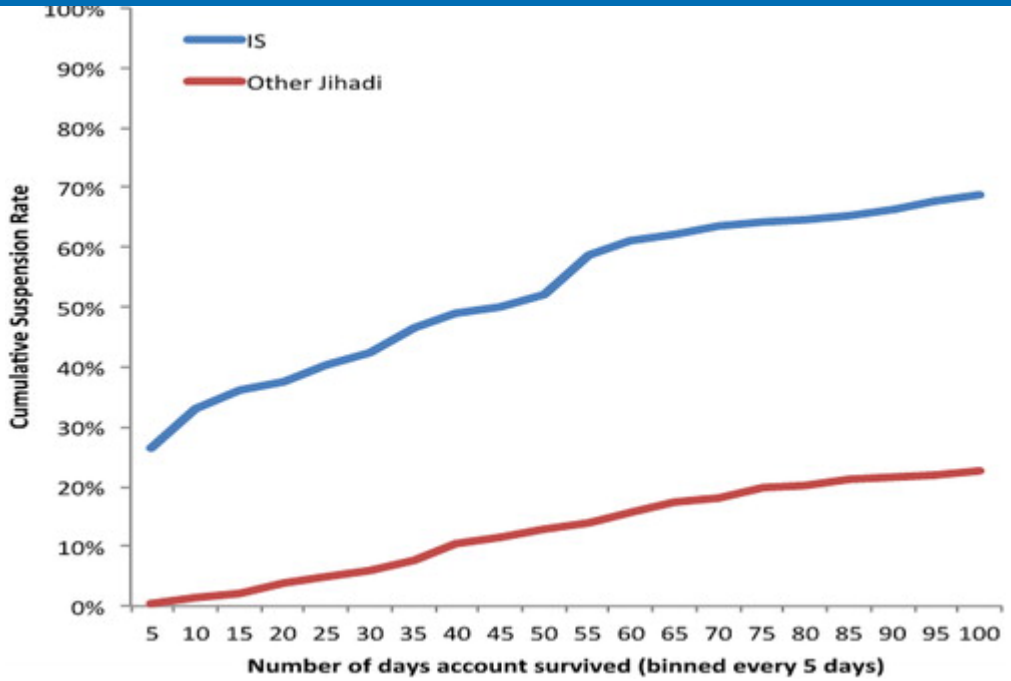
Until now, the small amount of publicly available research on the online disruption of IS has focused on the impact of Twitter's suspension activities on follower numbers for reestablished accounts.[20] As well as updating these data, this research also examined the longevity or survival time of accounts, and compared Pro-IS to Other Jihadist accounts on both measures (i.e., follower numbers and longevity). The overall finding was that IS-supportive accounts were being significantly disrupted, which in turn has effectively eliminated IS's once vibrant Twitter community. Differential disruption is taking place, however, meaning Other Jihadist accounts were subject to much less pressure.

Account Longevity

This section addresses the survival time of accounts in the research database. All were active at the point they were identified and classified as Pro-IS or Other Jihadist. Once an account was added to the database, its status was monitored and the system recorded when it was suspended, if this subsequently occurred. This enabled the research team to measure the age of each account (i.e., the time elapsed since the account's creation) at the date of suspension. Worth underlining here is that the below-described survival rates of Pro-IS accounts would likely have been considerably shorter if the analysis included those accounts suspended—often within minutes of creation—before they could be captured by the research team for inclusion in the dataset.

Figure 2 depicts the estimated cumulative suspension rate for all Twitter accounts in the dataset, outlining the probability of an account being suspended against its age (represented in days) for the 722 Pro-IS accounts and 451 Other Jihadist accounts. The majority—around 65 percent—of Pro-IS accounts were suspended before they reached 70 days since inception. At the same time point, less than 20 percent of Other Jihadist accounts had been suspended. In fact, in terms of differential disruption, more than 25 percent of Pro-IS accounts were suspended within five days of inception; a negligible number (less than 1 percent) of Other Jihadist accounts were subject to the same swift response.

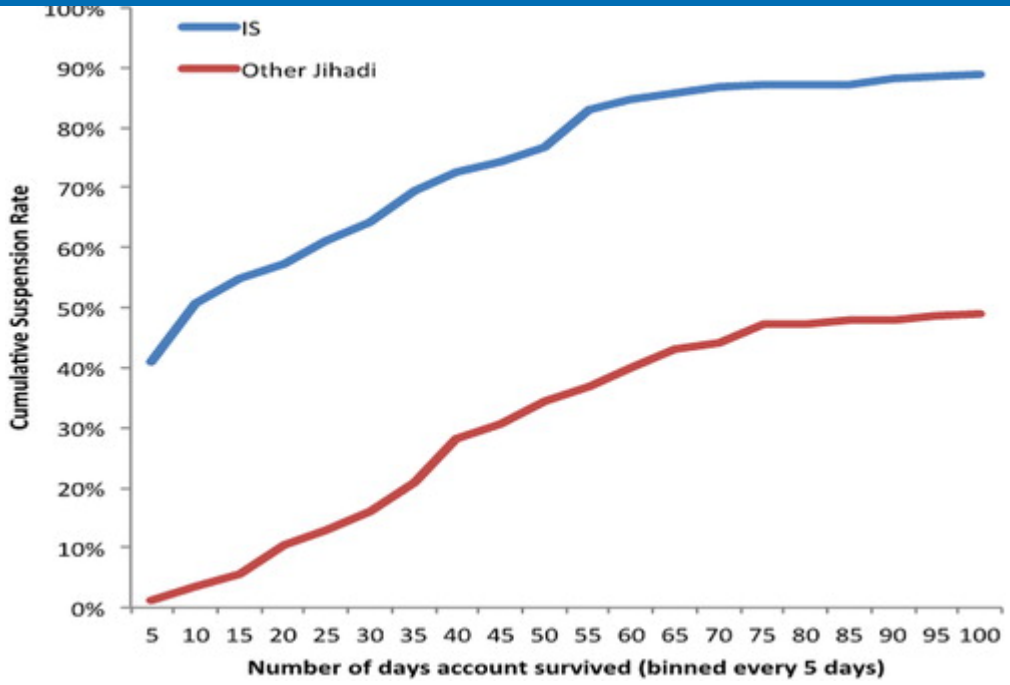
Figure 2. Cumulative suspension rate for all accounts in database.



Display full size

The categorization of these accounts as jihadist in orientation was necessarily subjective. It is possible that others may disagree with our decisions. To address this possibility, Figure 3 focuses on those accounts in the dataset that were eventually suspended: 455 Pro-IS accounts and 163 Other Jihadist accounts. The rationale is that these accounts were independently judged to have breached Twitter's terms of service. Again, regarding differential disruption, the data illustrates that 85 percent of Pro-IS accounts were suspended within the first 60 days of their life, compared to 40 percent of accounts falling into the Other Jihadist category. Further, more than 30 percent of Pro-IS accounts were suspended within two days of their creation; less than 1 percent of Other Jihadist accounts met the same fate.

Figure 3. Cumulative suspension rate for accounts eventually suspended.



Display full size

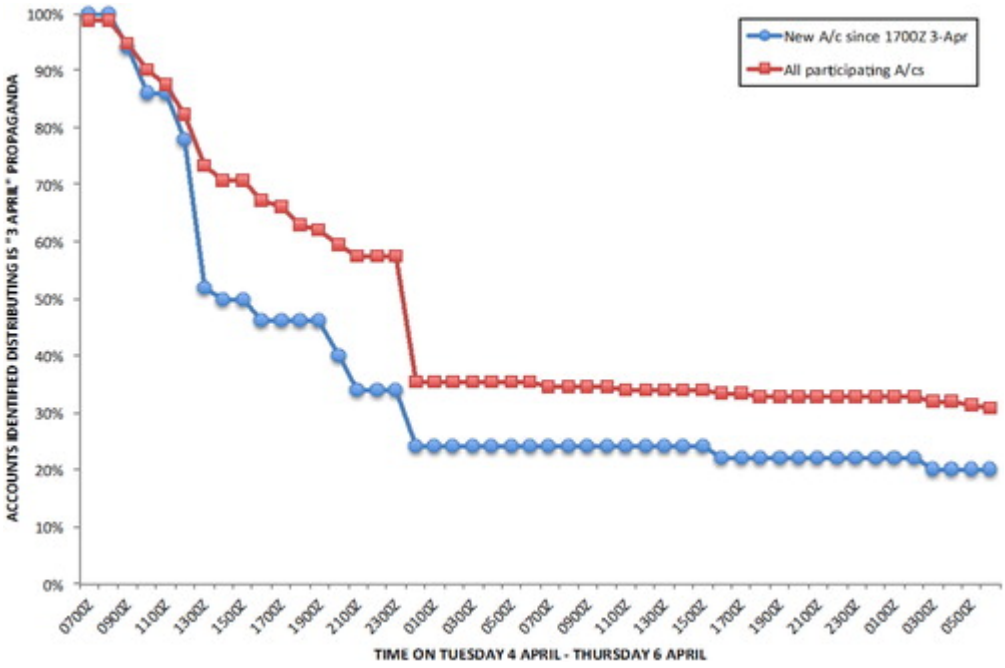
Further analysis of suspended accounts revealed that the three subsets of Pro-IS accounts (i.e., those identified manually, semi-automatically based on general tweet content, and advanced semi-automatically as a result of linking to official IS propaganda) also displayed different survival and activity patterns. From the 722 Pro-IS accounts in the dataset, the manually identified accounts (27 percent) survived disruption for longer periods and were primarily tweeting about general IS and non-IS related news. The accounts identified through semi-automated means (30 percent) had a somewhat shorter lifespan and were tweeting content generically related to IS involvement in the Syria conflict, including daily battle updates from several of what were then IS frontlines, such as Mosul, Al-Bab, Deir Ez-Zor, and eastern Aleppo. Accounts located via advanced semi-automated means (43 percent) experienced the shortest lifespans. They were initially identified as a result of sending at least one tweet specifically disseminating “official” IS propaganda (e.g., from the Amaq News Agency). Many were thus found to be exclusively tweeting links to official IS propaganda.

Mini-Case Study: Intervention Effectiveness

Throughout the period of data collection, IS operated a 24-hour “news cycle,” disseminating a new batch of propaganda on a daily basis via Twitter and other online platforms, using links to content hosted elsewhere on the Internet. These were probably “ghazwa” or social media “raids” orchestrated using an alternative online platform, such as Telegram.[21] The rapid takedown of Twitter accounts sending tweets containing

the effectiveness of intervention over a single 24-hour period. Figure 4 depicts survival curves for those Twitter accounts in the research database that disseminated links to one or more pieces of official IS propaganda produced on Monday, 3 April 2017[22] (based on data collected on Monday, 3 April and Tuesday, 4 April 2017).[23]

Figure 4. Case study of intervention effectiveness.



Display full size

On Monday, 3 April 2017, IS uploaded its daily propaganda content to a variety of social media and online content-hosting platforms. This content generally included videos (in daily news format and other propaganda videos), “picture stories” (a photo montage that tells a story), brief pronouncements similar to short press releases, radio podcasts, and other documents (e.g., magazines). Over the course of Monday afternoon and evening, 153 unique Twitter accounts were identified that sent a total of 842 tweets with links to external (non-Twitter) Web pages, each loaded with an item or items of IS propaganda. It was found that only 10 of those Twitter accounts (7 percent) were independent, third-party “mainstream” accounts. The balance of accounts was identified as pro-IS. Fifty of these appeared to be throwaway accounts (i.e., accounts with no followers set up solely to distribute propaganda and sending only IS propaganda tweets until suspended) created on Monday evening.

Method 52 was used to track all accounts disseminating this propaganda—those

GMT on the morning of Tuesday, 4 April 2017. Figure 4 shows the survival curves for all 153 Twitter accounts tweeting IS propaganda from Monday, 3 April and for the subset of 50 throwaway accounts specifically created on the Monday evening. The data shows that at 07.00 GMT on Tuesday, 4 April 2017, 100 percent of these accounts were active. However, by 13.00 GMT, this figure had reduced to just 73 percent, falling to 58 percent by 23.00 GMT. This then dropped sharply to 35 percent surviving un-suspended by midnight on Tuesday. Very few of these surviving accounts were suspended over the subsequent 48 hours tracked. The fifty throwaway accounts created on Monday evening specifically to disseminate propaganda were suspended or deleted even faster: by 13.00 GMT only 52 percent were still active, falling to 34 percent by 23.00 GMT and 24 percent by midnight on Tuesday.

This demonstrates that the response to official IS propaganda being distributed via Twitter was reasonably effective in terms of identifying and taking down disseminator accounts in the first 24 hours after they linked to official IS content. Pro-IS accounts disseminating this official IS propaganda were taken down at a higher rate, compared to those Pro-IS accounts not disseminating it. However, it must be borne in mind that some Pro-IS accounts were operating on a 24-hour "news cycle" and a large number of accounts were created daily to disseminate this propaganda. As these accounts were being taken down during Tuesday, a similar number of fresh accounts were being created and used to distribute the next day's official IS content. Therefore, it could be argued that, while efforts to remove permanent traces of IS propaganda links from Twitter were relatively successful, pro-IS users were still able to broadcast links to its daily propaganda using Twitter in 24-hour bursts during the research period.

Community Breakdown

What are the effects of this disruption on IS-supportive users and accounts? The truncated survival rates for Pro-IS accounts meant that their relationship networks were much sparser than for the Other Jihadist accounts in the dataset and compared to previously mapped IS-supporter networks on Twitter. Taking a qualitative perspective, this means that the pro-IS Twitter community was virtually non-existent during the research period.

To demonstrate this, Table 3 compares the median number of tweets, followers, and friends[24] of Pro-IS accounts versus those of Other Jihadists. The short lifespan of Pro-IS

and follow other accounts. This meant Other Jihadist accounts had the opportunity to: send six times as many tweets; follow or "friend" four times as many accounts; and importantly, gain 13 times as many followers as the Pro-IS accounts. An even more stark comparison is between median figures for Pro-IS accounts in 2017 versus those recorded for similar accounts in 2014. The median number of followers for pro-IS accounts in 2017 was 14 versus 177 in 2014,[25] a decrease of 92 percent. The median number of accounts followed by IS supporters in 2014 was 257, whereas this research found a median of thirty-three "friends" per pro-IS account—a decrease of 87 percent.[26] In an analysis of 20,000 IS supporter accounts in a five-month period between September 2014 and January 2015, Berger and Morgan observed suspension of just 678 accounts,[27] a total loss of 3.4 percent. In the research dataset outlined in this article, the total loss of Pro-IS accounts in just four months (between January and April 2017) was conservatively 63 percent.

Table 3. Median number of tweets, followers, and friends for accounts not yet suspended.

Download CSV Display Table

Throughout what may be referred to as the IS Twitter "Golden Age" in 2013 and 2014, a variety of official IS "fighter" and an assortment of other IS "fan" accounts were accessible with relative ease. For the uninitiated user, once one IS-related account was located, the automated Twitter recommendations on "who to follow" accurately supplied others.[28] For those "in the know," pro-IS users were easily and quickly identifiable through their choice of carousel and avatar images, along with their user handles and screen names. Thus, if one wished, it was quick and easy to become connected to a large number of like-minded other Twitter users. If sufficient time and effort was invested, it was also relatively straightforward to become a trusted—even prominent—member of the IS "Twittersphere."[29] Not only was there a vibrant overarching pro-IS Twitter community in existence at this time, but also a whole series of strong and supportive language (e.g., Arabic, English, French, Russian, Turkish) and/or ethnicity-based (e.g., Chechens or "al-Shishanis") and other special interest (e.g., females or "sisters"[30]) pro-IS Twitter sub-communities. Most of these special interest groups were a mix of: a small number of users on the ground in the

stated preference to do so); and an even larger number of "jihobbyists."[31] The latter had no formal affiliation to any jihadist group, but spent their time lauding fighters, celebrating suicide attackers and other "martyrs," and networking around and disseminating IS content.

In 2014, pro-IS accounts were already experiencing some pressure from Twitter; for example, official IS accounts were some of the first to be suspended that summer. Twitter's disruption activity increased significantly over time, forcing pro-IS users to develop and institute a host of tactics to allow them to maintain their Twitter presences, remain active, and preserve their communities of support on the platform.[32] For example, the group employed particular hashtags, such as #baqiyyafamily ("baqiyya" means "remain" in Arabic), to announce the return of suspended users to the platform, in an attempt to regroup after their suspension. Twitter eventually responded by including these hashtags in their disruption strategies. Interestingly, this increased disruption only strengthened some IS supporters' resolve and they became more determined to reestablish their accounts, even after repeated suspensions. During this time, suspension was, for some, considered to be a "badge of honor." Thus, although disruption may have resulted in decreased numbers of pro-IS users, it may also have contributed to the generation of more close-knit and unified communities, as those who remained needed a high level of commitment and virtual community support to do so.[33]

Eventually, however, the costs of remaining on Twitter began to outweigh the benefits. Research from 2016 shows that "the depressive effects of suspension often continued even after an account returned and was not immediately re-suspended. Returning accounts rarely reached their previous heights,"[34] in terms of numbers of followers and friends. This was probably due to the eventual discouragement of many IS supporters subjected to rapid and repeated suspension. Even those who persisted were forced to take countermeasures such as locking their accounts so they were no longer publicly accessible, or diluting the content of their tweets so their commitment to IS was no longer so readily apparent. By April 2017, these measures had taken such hold that the vast majority of Pro-IS account avatar images were default "eggs" or other innocuous images, and many of the account user handles and screen names were meaningless combinations of letters and numbers (see Table 4).[35] A conscious, supportive, and influential virtual community is almost impossible to maintain in the face of loss of access to such group or ideological symbols and the resultant breakdown in

activity elsewhere, primarily to Telegram, which is no longer merely a back-up for Twitter.[36]

Table 4. Changes in account name types due to disruption activity.*

Download CSV Display Table

From a quantitative perspective, the data discussed in this section demonstrate three key findings. First, IS and their supporters were being significantly disrupted by Twitter, where the rate of disruption depended on the content of tweets and out-links. Second, although all accounts experienced some type of suspension over a period of time, Pro-IS accounts experienced this at a much higher rate compared to the Other Jihadist accounts in the dataset. Third, this severely affected IS's ability to develop and maintain robust and influential communities on Twitter. As a result, pro-IS Twitter activity has largely been reduced to tactical use of throwaway accounts for distributing links to pro-IS content on other platforms, rather than as a space for public IS support and influencing activity.

Beyond Twitter: The Wider Jihadi Online Ecology

Research on the intersections of violent extremism and terrorism and the Internet have, for some time, been largely concerned with social media. Studies have often had a singular focus on Twitter due to its particular affordances (e.g., ease of data collection due to its publicness, the nature of its application programming interface), which is problematic.[37] For example, Europol's Internet Referral Unit reported that, as far back as mid-2016, they had identified "70 platforms used by terrorist groups to spread their propaganda materials."[38] This section of the article is therefore concerned with the wider online ecology where IS supporters and other non-IS jihadist users operate, with a particular focus on out-links from Twitter.

Owing partly to its character limit,[39] Twitter can function as a "gateway"[40] platform to other social networking sites and a diversity of other online spaces. In 2014, it was estimated that one in every 2.5 pro-IS tweets contained a Uniform Resource Locator

but this was not undertaken due to complications around Twitter's URL-shortening practices.[41] The roll-out of auto-expanding link previews by Twitter in July 2015 remedied this difficulty. In terms of link activity in the data collected for this research, most links were not out-links, but rather in-links (i.e., within Twitter): 8,086 or 14 percent for Pro-IS and 4,650 or 7.5 percent for Other Jihadist tweets. Of the Pro-IS and Other Jihadist Twitter accounts identified, one in eight (around 13 percent) contained non-Twitter URLs or out-links. This is a considerable reduction from the 40 percent of tweets reportedly containing URLs in 2014. Analysis of Twitter out-links nonetheless provides an interesting snapshot of the Top 10 platforms linked to by Pro-IS and Other Jihadist accounts during our data-collection period (see Table 5).

Table 5. Top 10 other platforms (based on out-links from Twitter).

Download CSV Display Table

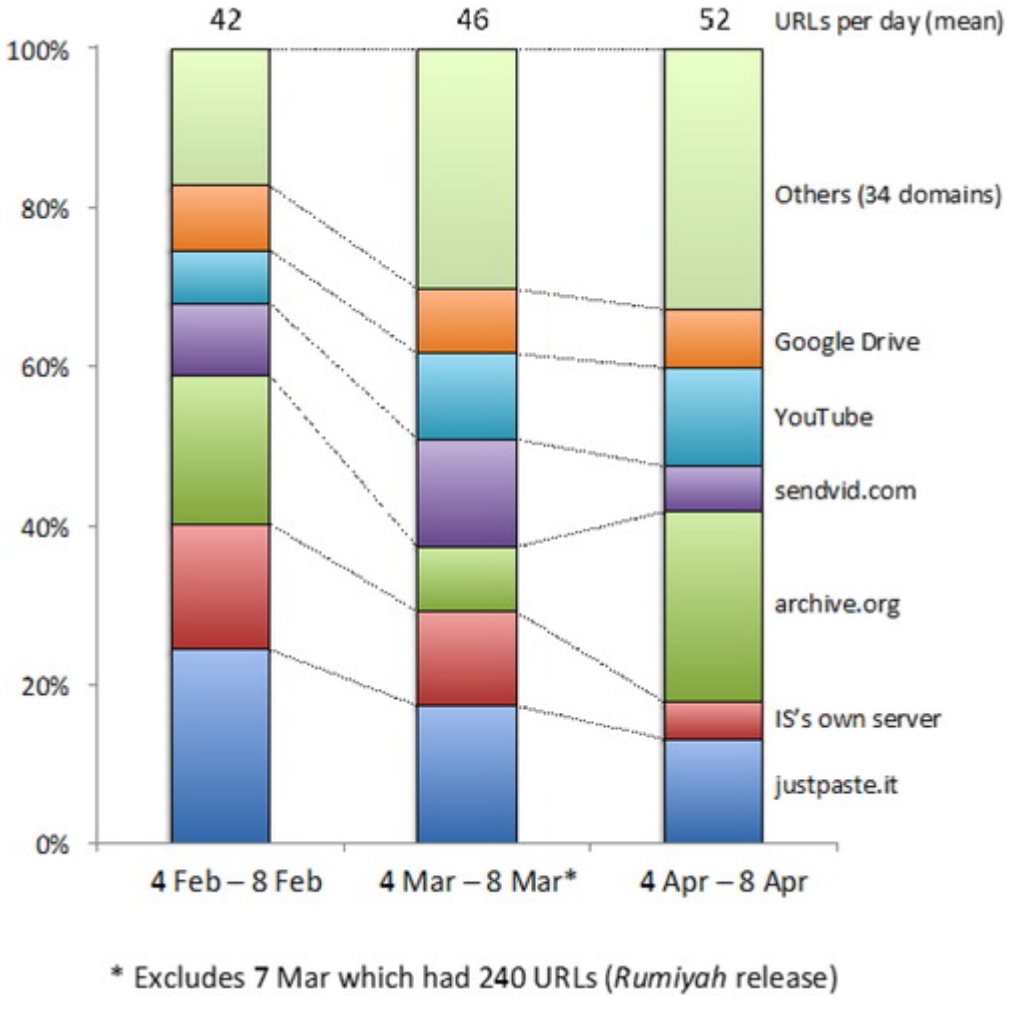
YouTube was the top linked-to platform for both Pro-IS and Other Jihadist accounts, pointing to the overall importance of the site—and of video generally in Web 2.0—in the jihadist online scene. Facebook does not appear in the Top 10 out-links for Pro-IS accounts, albeit a recent report claims that IS content and IS-supportive users are still easily locatable on Facebook.[42] What our findings indicate is that, like Twitter, Facebook is engaged in differential disruption as it is the second most preferred platform for out-linking by Other Jihadists. The somewhat obscure justpaste.it content upload site has been known for some time as a core node in the "jihadisphere,"[43] and its high-ranking status for both Pro-IS and Other Jihadist accounts is thus relatively unsurprising.

Other content upload destinations preferred by Pro-IS users, including Google Drive, Sendvid, Google Photos, and the Web Archive, do not appear in the Other Jihadist Top 10. One particular reason for this is probably the focus of Other Jihadists on linking to traditional proprietary websites, such as the Taliban's suite of sites. It is worth mentioning that, while Telegram slips into the Top 10 for Other Jihadists, only twenty (0.04 percent) of all tweets from Pro-IS accounts contained a telegram.me link. The paucity of such links caused us to explore further; we were surprised to find that just two of 722 Pro-IS users' biographies and two of 451 Other Jihadist users' biographies contained Telegram links. Neither group of accounts were using Twitter to advertise

Case Study: Destinations of Official IS Propaganda

As mentioned, during the research period, IS was operating a 24-hour “news cycle,” disseminating a daily batch of new official propaganda via social media channels, including Twitter. Links to this propaganda were circulated through tweets and other means. These links pointed to a wide variety of other social media and content hosts that contained newly uploaded propaganda daily. A sample of these propaganda destinations were analyzed at three time points: 4–8 February, 4–8 March (excluding 7 March, see below), and 4–8 April 2017. The research team obtained the full daily roster of IS propaganda and the sites where it appeared for each of these time periods. This allowed the identification of the most frequently linked-to platforms, along with how many pieces of propaganda were posted by host domains, and what proportion of these URLs were subsequently taken down (see Figure 5).

Figure 5. Destinations of official IS propaganda: Number of URLs and URL destinations February to April 2017.



Overall, over these three time periods, Pro-IS users linked to thirty-nine different third-party platforms or sites, as well as IS running its own server[44] to host its propaganda material. It is important to note that the former were exclusively, it is believed, “leaf” destinations. That is, they contained content but no links to other sites, so did not have a networking or community-building aspect. Someone visiting such a page would not be able to discover more about the network of other sites. Important exceptions to this were YouTube and a small number of other sites that algorithmically “recommend” similar content in their inventory, which may have resulted in their pointing to other available IS propaganda.[45] During the period of the research, the average number of URLs populated rose from forty-two per day in February to fifty-two per day in April. This hints at increasing fragmentation and dispersal, possibly in response to takedown activity by a variety of platforms and sites. However, there was a large inter-day variation (twenty to sixty-five) and one outlier day on 7 March was excluded, as it was the publication date of issue 7 of IS’s Rumiyah online magazine. On this day, IS pushed 240 separate URLs, a quarter of which contained direct reference to Rumiyah in the link, and many more that probably linked to the new issue of the magazine.[46]

Of the forty domains used (thirty-nine external, one internal server), a consistent “big 6” became apparent across the three time periods: justpaste.it; IS’s own server; archive.org; sendvid.com; YouTube; and Google Drive. These six domains accounted for 83 percent, 70 percent, and 67 percent of the URLs in the February, March, and April sampling periods, respectively. However, there was a noticeable declining trend in the use of justpaste.it and IS’s own servers. Between them, this accounted for 40 percent of URLs in February declining to only 18 percent by April. At around that time the Amaq News Agency website had come under repeated attack, which may have been responsible for its relative downgrading.[47] Use of sendvid.com and archive.org varied across the time periods, while Google Drive and YouTube were consistently heavily used. In fact, YouTube use showed an increasing trend (7 percent, 11 percent, and 12 percent, respectively). The remaining URLs (17 percent in February rising to 33 percent of URLs by April) were spread across a wide variety of mainly, although not exclusively, content upload sites (thirty-four in total).

The proportion of IS propaganda successfully taken down was also analyzed. The takedown rate (as of 12 April, 2017) was 72 percent, 66 percent, and 72 percent for the

live on 12 April. This suggests that takedown activity was relatively rapid (occurring over a matter of days after propaganda was posted) and widespread (across a multiplicity of sites and platforms).

Conclusion

The costs for most pro-IS users of engaging on Twitter (in terms of deflated morale, diffused messages and persistent effort needed to maintain a public presence) now largely outweigh the benefits. This means that the IS Twitter community is now almost non-existent. In turn, this means that radicalization, recruitment, and attack planning opportunities on this platform have probably also decreased. However, a hard core of users remain persistent. In particular, a subset of established throwaway disseminator accounts pushed out “official” IS content in a daily cycle during our data-collection period and continue to do so. These accounts were generally suspended within 24 hours, but not before they promoted links to content hosted on other platforms.

This article was mainly concerned with pro-IS Twitter accounts and their disruption. However, IS are not the only jihadists active on Twitter, and a host of other violent jihadists were shown to be subject to much lower levels of disruption by Twitter. Also, IS and other jihadist groups remain active on a wide range of other social media platforms, content hosting sites and other cyberspaces, including blogs, forums, and dedicated websites. While it appears that official IS content is being disrupted in many of these online spaces, the extent is yet to be fully determined.

The Telegram messaging application was mentioned a number of times in this article and is worth treating here in slightly more depth as it is IS supporters’ currently most preferred platform. Telegram is as yet a lower profile platform than Twitter—and obviously also Facebook—with a smaller user base and higher barriers to entry (e.g., provision of a mobile phone number to create an account, time-limited invitations to join channels[48]). These are probably positive attributes from the perspective of cutting down on the numbers of users exposed to IS’s online content and thereby in a position to be violently radicalized by it. On the negative side, this may mean that Telegram’s pro-IS community is more committed than its Twitter variant. Also, while IS’s reach via Telegram is less than it was via Twitter, the echo chamber effect may be greater as the “owners” of Telegram channels and groups have much greater control over who joins

also attractive to pro-IS users is its in-platform content upload and cloud storage function(s). While Telegram restricts users from uploading files larger than 1.5GB—roughly a two-hour movie—it provides seemingly unlimited amounts of storage.

In terms of proactive steps taken by Telegram with respect to IS and their supporters' use of their service, in December 2016, Telegram established a dedicated "ISIS Watch" channel, which provides a running tally of numbers of "ISIS bots and channels banned" by them. On 11 March, 2017 a message on the channel stated "Our abuse team actively bans ISIS content on Telegram. Following your reports, an average of 70 ISIS channels are terminated each day before they reach any traction." Between January and May 2018, the average number of terminations per days had jumped sixfold to 422. All told, Telegram claims to have banned 106,573 "ISIS bots and channels" in the period December 2016 to 31 May 2018, with May 2018 (9,810) having the highest number of bans yet recorded.[49] While it is clear therefore that Telegram routinely bans pro-IS users, channels, and bots, interpreting the numbers that Telegram has supplied is difficult absent knowing the overall numbers of users, channels, and bots actually active on the platform at any given time. Also worth pointing out is that in addition to exploiting the channels feature, IS began taking advantage of Telegram's groups function around summer 2017. So-called Supergroups allow for intra-group communication among up to 30,000 members[50] and like all other group chats on Telegram are private among participants; Telegram does not, in other words, process requests, including termination requests, related to them.[51]

Recommendations

The recommendations arising from this analysis are threefold. First, modern social media monitoring systems have the ability to dramatically increase the speed and effectiveness of data gathering, analysis, and (potentially) intervention. To work effectively, however, they must deploy a combination of suitable technology solutions, including analytical systems, with trained human analysts who are versed in the relevant domain(s) and preferably also the appropriate languages. This is particularly the case where an adversary is actively trying to evade tracking efforts. Technology such as Method52 assists by allowing the analyst to rapidly develop new analytical pipelines that take into account day-to-day changes in modes of operation. However, technology cannot detect such changes; these can generally only be spotted by a human well-versed in the particular domain of interest.

Second, some IS supporters remain active on Twitter. Content disseminators using throwaway accounts could probably be degraded further—although this may have both pros (e.g., detrimental impact on last remaining significant IS supporter Twitter activity) and cons (e.g., further degradation of Twitter as a source of data or open source intelligence on IS). Like all disruption activity, whether this is viewed positively or negatively depends on one's perspective and institutional interests. For example, law enforcement tends to favor this approach, whereas free-speech advocates warn against corporate policing of political speech, even if that speech is deeply objectionable. Some intelligence professionals, on the other hand, advocate for greater attention to social media intelligence.[52]

Third, the focus of this article has not just been Twitter, but the importance of the wider jihadist online ecology was also pointed to. The analysis was also not restricted to IS users and content; the presence and often uninterrupted online activity of non-IS jihadists was underlined too. In recent years, many counterterrorism professionals tasked with examining the role of the Internet in violent extremism and terrorism have narrowed their focus to IS. Scholarly researchers have acted similarly, many narrowing their focus further to IS Twitter activity. Continued analytical contraction of this sort should be guarded against. Maintenance of a wide-angle view of online activity by diverse other jihadists across a variety of social media and other online platforms is recommended. This is particularly important due to the shifting fortunes of IS and HTS on the ground in Iraq and Syria. In the face of increasing loss of physical territory, the continued—and potentially increasing— importance of online "territory" should not be underestimated. It is not that a focus on IS should be dispensed with, but the significantly less-impeded online activity of HTS is surely an important asset for them and worth monitoring. Because data collection and analysis of other terrorist groups and their online platforms has been neglected, very few historical metrics are available for comparative analyses; this should be guarded against in future too.

Future research

Finally, some comments as regards future research. Our Other Jihadist category was a convenience sample of non-IS jihadist accounts. It is therefore proposed to replicate the present research, but with a larger and more equal sample of HTS, Ahrar al-Sham, and Taliban accounts. This would allow for a more systematic and comparative analysis of the disruption levels for a range of non-IS jihadists, including those with a significant

regional terrorism profile (i.e., Taliban), and a party to the Syria conflict (i.e., Ahrar al-Sham).[53] Such an analysis could help to ascertain the vibrancy of their contemporary Twitter communities and Twitter out-linking practices, and allow their preferred other online platforms to be identified.

Additional research is also clearly warranted into the wider violent jihadist online ecology. Wider and more in-depth research into the following is therefore recommended:

1. patterns of use, including community-building and influencing activity;
2. levels of disruption on other platforms besides Twitter, including other major platforms such as YouTube, but also other smaller or more obscure platforms, such as justpaste.it and others.

Analysis of pro-IS and other violent jihadist activity on Telegram and comparing this with our present findings is suggested too. It would also be worthwhile analyzing out-linking trends on Telegram to determine how the functionalities of different platforms have an impact on linking practices.

Acknowledgment

This article is a slightly revised and updated version of the VOX-Pol report with the same title, available at http://www.voxpol.eu/download/vox-pol_publication/DCUJ5528-Disrupting-DAESH-1706-WEB-v2.pdf.

Notes

Additional information

Funding

This research was supported by funding from the UK Home Office and the European Union's Framework Programme 7 [Grant number 312827: VOX-Pol Network of Excellence].

Notes

1 George Parker, “Theresa May Warns Tech Companies: ‘No Safe Space’ for Extremists,” Financial Times, 4 June 2017, https://www.ft.com/content/0ae646c6-4911-11e7-a3f4-c742b9791d43?mhq5j=e3 (accessed 1 Oct. 2018).

2 Theresa May, “PM’s Speech at Davos 2018,” Prime Minister’s Office, 25 January 2018, https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january (accessed 1 Oct. 2018).

3 For a detailed discussion on “internalization” see Herbert C. Kelman, “Compliance, Identification and Internalization: Three Processes of Attitude Change,” The Journal of Conflict Resolution 2, no. 1 (1958), 51–60; in relation to radicalization and violent extremism, see Suraj Lakhani, Radicalisation as a Moral Career: A Qualitative Study of How People Become Terrorists in The United Kingdom, Ph.D. thesis (Universities Police Science Institute: Cardiff University, 2014).

4 Facebook, “Facebook, Microsoft, Twitter and YouTube Announce Formation of the Global Internet Forum to Counter Terrorism,” Facebook Newsroom, 26 June 2017, https://newsroom.fb.com/news/2017/06/global-internet-forum-to-counter-terrorism/ (accessed 1 Oct. 2018).

5 Ibid.

6 Monika Bickert, “Hard Questions: Are We Winning the War on Terrorism Online?” Facebook Newsroom, 28 November 2017, https://newsroom.fb.com/news/2017/11/hard-questions-are-we-winning-the-war-on-terrorism-online/ (accessed 1 Oct. 2018).

7 Twitter, “Government TOS Reports: January to June 2017,” Twitter Transparency, n.d., https://transparency.twitter.com/en/gov-tos-reports.html (accessed 1 Oct. 2018).

8 Twitter Public Policy, “Expanding and Building #TwitterTransparency,” Twitter Blog, 5 April 2018, https://blog.twitter.com/official/en_us/topics/company/2018/twitter-transparency-report-12.html (accessed 1 Oct. 2018).

9 J. M. Berger and Jonathan Morgan, “The ISIS Twitter Census: Defining and Describing the Population of ISIS Supporters on Twitter,” The Brookings Project on U.S. Relations with the Islamic World, Analysis Paper #20 (2015), https://www.brookings.edu/wp-

content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf (accessed 1 Oct. 2018).

10 Method52 was developed by the TAG Laboratory at the University of Sussex. For more information, see www.taglaboratory.org (accessed 1 Oct. 2018).

11 Classifiers were trained using supervised machine-learning approaches. Method52 provides components that enable this to be done swiftly and in a manner that is bespoke to a project.

12 Berger and Morgan, "The ISIS Twitter Census," 41 and 44.

13 J. M. Berger and Heather Perez, "The Islamic State's Diminishing Returns on Twitter: How Suspensions are Limiting the Social Networks of English-speaking ISIS Supporters," Washington DC: George Washington University Program on Extremism, 2016, 6, https://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/Berger_Occasional%20Paper.pdf (accessed 1 Oct. 2018).

14 The data from the latter stages of this project suggest that fifty or more throwaway IS accounts were produced daily. These accounts appeared to be set up solely to distribute propaganda, typically had no followers, and sent only IS propaganda tweets until they were suspended. If the throwaway accounts from across the whole research period had been included in the dataset, the total number could have reached as many as 2,000–3,000.

15 Walid Magdy, Kareem Darwish, and Ingmar Weber, "#FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support," Doha: Qatar Foundation, 2015, 3, https://arxiv.org/abs/1503.02401 (accessed 1 Oct. 2018).

16 Berger and Morgan, "The ISIS Twitter Census," 9.

17 Twitter, "Combating Violent Extremism," Twitter Blog, 5 February 2016, https://blog.twitter.com/2016/combating-violent-extremism (accessed 1 Oct. 2018).

18 Twitter, "An Update on our Efforts to Combat Violent Extremism," Twitter Blog, 18 August 2016, https://blog.twitter.com/2016/an-update-on-our-efforts-to-combat-violent-extremism (accessed 1 Oct. 2018).

19 Cole Bunzel, '"Come Back to Twitter': A Jihadi Warning Against Telegram," Jihadica,

20 Berger and Perez, “The Islamic State’s Diminishing Returns on Twitter.”

21 Nico Prucha, “IS and the Jihadist Information Highway: Projecting Influence and Religious Identity via Telegram,” Perspectives on Terrorism 10, no. 6 (2016), 51–52.

22 By early April 2017, the research had reached the stage where there was complete access to IS’s main Twitter propaganda apparatus. This enabled the semi-automated system to determine what IS and supporter tweets would be linking to before those tweets were sent. It is thought that this occurred several hours before Twitter themselves become aware of these accounts and their tweets. Much of this may have been due to the research team being able to access data and intelligence across multiple sites, allowing early prediction of tweet material, where Twitter’s disruption team were likely restricted to monitoring their own platform only. The system was thus able to immediately identify when an account disseminated one of these propaganda links on Twitter. It was then possible to capture the rate and speed of suspension.

23 It should be noted that this date was chosen at random. Propaganda represented in this graph had no relation to the chemical attack on the town of Khan Shaykhun, also on 4 April, as content was produced by IS on Monday 3 April 2017.

24 The term “friends” refers to accounts the Pro-IS accounts were following.

25 Berger and Morgan, “The ISIS Twitter Census,” 30.

26 BerIbid.

27 Ibid.

28 Online recommender systems present content to users of specific platforms that they might not otherwise locate based on, for example, prior search or viewing history on that platform; see Derek O’Callaghan, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham, “Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems,” Social Science Computer Review 33, no. 4 (2015), 459–478.

29 See, for example, the extensive media coverage of Twitter user @ShamiWitness who, in December 2014, was revealed to be Mehdi Biswas, a 24-year-old Bangalore-based business executive, who prior to his arrest was one of the most prominent IS supporters on social media. Interestingly, his Twitter account was only suspended in early 2017, despite being dormant since his arrest. At the time of writing Biswas is

30 Pearson, “Wilayat Twitter and the Battle Against Islamic State’s Twitter Jihad.”

31 Jarret M. Brachman, Global Jihadism: Theory and Practice (Abingdon: Routledge, 2009), 19.

32 For examples, see Berger and Perez, “The Islamic State’s Diminishing Returns on Twitter,” 15–18.

33 Pearson, “Wilayat Twitter and the Battle Against Islamic State’s Twitter Jihad.” See also Elizabeth Pearson, “Online as the New Frontline: Affect, Gender, and ISIS-takedown on Social Media,” Studies in Conflict & Terrorism 5 (September 2017).

34 Berger and Perez, “The Islamic State’s Diminishing Returns on Twitter,” 9.

35 Such meaningless combinations of letters and numbers are also characteristic of “bots” (i.e., automated social media accounts that pose as real users). See Ben Nimmo, Digital Forensics Research (DFR) Lab, “#BotSpot: Twelve Ways to Spot a Bot: Some Tricks to Identify Fake Twitter Accounts,” Medium, 28 August 2017, https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c (accessed 1 Oct. 2018).

36 Berger and Perez, “The Islamic State’s Diminishing Returns on Twitter,” 15.

37 Maura Conway, “Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research,” Studies in Conflict and Terrorism 40, no. 1 (2017), 9 and 12.

38 EUROPOL, “EU Internet Referral Unit: Year One Report,” EUROPOL (2016), 11, www.europol.europa.eu/content/eu-internet-referral-unit-year-one-report-highlights (accessed 1 Oct. 2018).

39 Twitter’s character limit per tweet was 140 when this research was undertaken. It was increased to 280 characters per tweet platform-wide on 7 November, 2017.

40 Derek O’Callaghan, Derek Greene, Maura Conway, Joe Carthy, and Pádraig Cunningham, “Uncovering the Wider Structure of Extreme Right Communities Spanning Popular Online Networks,” in WebSci 13: Proceedings of the 5th Annual ACM Web Science Conference (New York: ACM Digital Library, 2013), 276–285.

42 Gregory Waters and Robert Postings, “Spiders of the Caliphate: Mapping the Islamic State’s Global Support Network on Facebook,” New York and London: Counter Extremism Project, 2018, https://www.counterextremism.com/sites/default/files/Spiders%20of%20the%20Caliphate%20%28May%202018%29.pdf (accessed 1 Oct. 2018).

43 The term “Twittersphere” is used to refer to Twitter users as a collectivity. The term “jihadisphere” has been used to refer to online jihadis as a collectivity; see Benjamin Ducol, “Uncovering the French-speaking Jihadisphere: An Exploratory Analysis,” Media, War & Conflict 5, no. 1 (2012).

44 Due to the domain names experiencing rapid removal, this server had five names over the three research periods studied.

45 O’Callaghan et al., “Down the (White) Rabbit Hole,” 37.

46 For a detailed accounting of Twitter activity around the release of an issue of Rumiyah’s precursor publication, Dabiq, see Daniel Grinnell, Stuart Macdonald, and David Mair, “The Response Of, and On, Twitter to the Release of Dabiq Issue 15,” paper presented at the 1st European Counter Terrorism Centre (ECTC) conference on online terrorist propaganda, 10–11 April 2017, Europol Headquarters, The Hague, Netherlands, https://www.europol.europa.eu/publications-documents/response-of-and-twitter-to-release-of-dabiq-issue-15 (accessed 1 Oct. 2018).

47 Lizzie Dearden, “ISIS Losing Ground in Online War Against Hackers After Westminster Attack Turns Focus on Internet Propaganda,” The Independent, 1 April 2017, http://www.independent.co.uk/news/world/europe/isis-islamic-state-propaganda-online-hackers-westminster-whatsapp-amaq-cyber-attacks-paranoia-a7662171.html (accessed 1 Oct. 2018).

48 Mia Bloom, Hicham Tiflati, and John Horgan, “Navigating ISIS’s Preferred Platform: Telegram,” Terrorism and Political Violence (July) Vol. 29, pp.’s 1–13.

49 The previous highest number of bot and channel terminations took place in October 2017 and amounted to 9,270.

50 Telegram, “Q: What's the Difference Between Groups, Supergroups, and Channels?” Telegram FAQ, n.d., https://telegram.org/faq#groups-supergroups-and-channels

51 Telegram, “Q: There's Illegal Content on Telegram. How do I Take it Down?” Telegram FAQ, n.d., https://telegram.org/faq#q-what-are-your-thoughts-on-internet-privacy. See also Maura Conway, with Michael Courtney, “Violent Extremism and Terrorism Online in 2017: The Year in Review,” Dublin, VOX-Pol, 2018, 4, http://www.voxpol.eu/download/vox-pol_publication/YiR-2017_Web-Version.pdf (accessed 1 Oct. 2018).

52 David Omand, Jamie Bartlett, and Carl Miller, “Introducing Social Media Intelligence (SOCMINT),” Intelligence and National Security 27, no. 6 (2012).

53 Nationally, Syria, Russia, Iran, Egypt, and the UAE have designated Ahrar al-Sham as a terrorist organization. Internationally, the United States, Britain, France, and Ukraine blocked a May 2016 Russian proposal to the United Nations to take a similar step.

Download PDF

Related research

People also read | Recommended articles | Cited by 34

Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research >

Maura Conway
Studies in Conflict & Terrorism
Published online: 12 Apr 2016

Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists >

Guri Nordtorp Mølmen et al.
Behavioral Sciences of Terrorism and Political Aggression
Published online: 30 Oct 2021

Online Radicalisation: Moving Beyond a Simple Dichotomy >

Chamin Herath et al.

Article contents | Related research

View more

Information for

Authors
R&D professionals
Editors
Librarians
Societies

Open access

Overview
Open journals
Open Select
Dove Medical Press
F1000Research

Opportunities

Reprints and e-prints
Advertising solutions
Accelerated publication
Corporate access solutions

Help and information

Help and contact
Newsroom
All journals
Books

Keep up to date

Register to receive personalised research and resources by email

Sign me up

Copyright © 2026 Informa UK Limited Privacy policy

Cookies Terms & conditions Accessibility

Registered in England & Wales No. 01072954
5 Howick Place | London | SW1P 1WG