Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)

A Large-Scale Study of ISIS Social Media Strategy: Community Size, Collective Influence, and Behavioral Impact

Majid Alfifi, Parisa Kaghazgaran, James Caverlee
Department of Computer Science & Engineering
Texas A&M University
College Station, TX
{alfifi, kaghazgaran, caverlee}@tamu.edu

Fred Morstatter
Information Sciences Institute
University of Southern California
Marina Del Rey, CA
fredmors@isi.edu

Abstract

The Islamic State of Iraq and Syria (ISIS) has received a tremendous amount of media coverage in the past few years for their successful use of social media to spread their message and to recruit new members. In this work, we leverage access to the full Twitter Firehose to perform a large-scale observational study of one year of ISIS social activity. We quantify the size of ISIS presence on Twitter, the potential amount of support it received, and its collective influence over time. We find that ISIS was able to gain a relatively limited portion from the total influence mass on Twitter and that this influence diminished over time. In addition, ISIS showed a tendency towards attracting interactions from other similar pro-ISIS accounts, while inviting only a limited anti-ISIS sentiment. We find that 75% of the interactions ISIS received on Twitter in 2015 actually came from eventually suspended accounts and that only about 8% of the interactions they received were anti-ISIS. In addition, we have created a unique dataset of 17 million ISIS-related tweets posted in 2015 which we make available for research purposes upon request.

1 Introduction

The past few years have seen social media as an effective tool for facilitating uprisings and enticing dissent in the Middle East (Lotan et al. 2011; Starbird and Palen 2012; Berger and Morgan 2015). The embrace of social media in the region has made it a battleground for ISIS (and similar groups) versus existing regimes, all spreading propaganda, recruiting sympathizers, and undermining rivals (Berger and Morgan 2015). Social media has given terrorists the ability to directly come into contact with their target audience and either spread terror or recruit. In fact, ISIS has been described by the FBI as the most adept terrorist group at using Internet and social media propaganda to recruit new members (Government-Publishing-Office 2016). Although several studies have examined how ISIS operates on platforms like Twitter (Magdy, Darwish, and Weber 2016; Benigni and Carley 2016; Ferrara et al. 2016), there is a research gap in the large-scale analysis of ISIS reach and impact. While many factors could have contributed to their success – such as gaining territories on the ground and appealing to the communities where they operate (Magdy, Darwish, and Weber 2016) – quantifying the impact of their social strategy is still crucial for a better understanding of their operations. In addition, although Twitter suspends accounts that violate the terms of service, ISIS accounts, in particular, manage to stay longer on the service posting more tweets than other malicious accounts (Figure 1) – making studying their impact even more important.

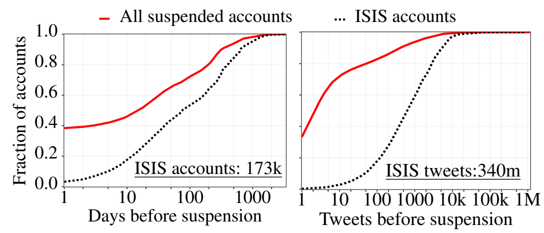


Figure 1: ISIS accounts stay longer on Twitter (left) spreading more content (right) before getting suspended compared with other eventually suspended accounts.

In this paper, we perform an observational study of the behavior of ISIS accounts and their interactions with the Twitter community. We utilize (i) their behavioral signals such as the number of URLs being used and the number of hashtags, (ii) their linguistic content, and (iii) their interactions via a retweet graph of 19 million nodes and 1.06 billion edges capturing the flow of influence in the whole Arabic Twitter-sphere in 2015. In addition, we propose a method that uses PageRank (Page et al. 1999) to quantify the Collective Influence (CI) of a group of users and use this method to show that although ISIS accounts were more active than average making more interactions with the community, their collective influence is relatively small and that it gradually diminished over the year 2015. We additionally find that while ISIS was heavily engaged in interactions with other users (4 times the average), they tended to attract similar pro-ISIS participants, with only about 8% of interactions coming from anti-ISIS accounts.

As we discuss in more detail in Section 7, several ISIS-related datasets have been made available to the research community. Our work complements these efforts by contributing a more comprehensive dataset spanning the year 2015. Concretely, we share with the research community a dataset of 24k ISIS users and their 17 million tweets.

2 Dataset

We have access to a large dataset of 9.3 billion tweets representing all tweets generated in the Arabic language in 2015 through full private access to the Twitter Firehose. We exploit a crowd-sourcing initiative by the Anonymous hacking group that invited Arabic speakers to report Twitter accounts that they think were associated with ISIS. This effort initially identified more than 25,000 ISIS sympathizers through crowdsourced reporting.[1] We extract an initial seed of ISIS accounts and expand it to a bigger dataset of accounts that are highly likely to be related to ISIS. We then extract accounts that interacted with ISIS along with all their tweets.

Validating ISIS seed accounts. We only use accounts that have actually been suspended by Twitter, indicating that (i) multiple users have reported those accounts as requested by the Anonymous hacking group, and (ii) the Twitter spam control team found those accounts in violation of the terms of service. We ended up with about 24k accounts. We then look for their tweets and their interactions with the Twitter community in our larger dataset. As a third level of validation that those accounts are related to ISIS, we built a minimal Twitter clone with a classification function and randomly sampled 1000 accounts along with a random sample of five tweets from their timeline and asked Arabic speaking volunteers to label those users as either pro-ISIS, anti-ISIS, or unrelated. 992 accounts were reported as ISIS supporters, two accounts as anti-ISIS, and six as unrelated accounts. We looked at the two accounts that were reported as anti-ISIS and found them to use sarcasm mocking those who were criticizing ISIS making it seem as though they were anti-ISIS for the labelers quickly skimming through these tweets.

A previous work (Ferrara et al. 2016) used this same ISIS users dataset but, due to Twitter limitations, they were limited to 10% of the total tweets through the Truthy project at Indiana University (Davis et al. 2016). In contrast, in this work, we were able to recover all (except for two disconnections that resulted in few hours of data loss) of the content these accounts generated in 2015 with our full private access to the Twitter Firehose. We select two groups of tweets – one that represents ISIS-related tweets and another randomly sampled group that we use to compare ISIS to other regular users. Concretely, we have the following sets of data (summarized in Table 1):

- **ISIS-Tweets:** Tweets posted by the reported ISIS-related accounts. There are 23,880 accounts that generated 17,424,323 tweets.
- **ISIS-Retweets:** All retweets of ISIS tweets including those from themselves. There are 10,436,603 retweets, posted by 551,869 users.
- **ISIS-Mentions:** All tweets that mention any of the ISIS accounts. There are 19,570,380 such tweets generated by 745,721 accounts.

[1] The website hosting these accounts has been taken offline, but accounts can be recovered from http://archive.is/A6f3L

Dataset	Accounts	Tweets
ISIS-Tweets	23,880	17,434,323
ISIS-Retweets	551,869	10,436,603
ISIS-Mentions	745,721	19,570,380
Big-ISIS-Tweets	173,340	341,365,270
Big-ISIS-Retweets	2,077,828	71,576,995
Legit-Tweets	23,880	17,454,068
Legit-Retweets	1,753,195	12,175,619
Legit-Mentions	2,161,106	17,479,990

Table 1: ISIS-Tweets are tweets posted by a known seed of ISIS-related accounts. Legit-Tweets is a randomly sampled set of users and their tweets. Retweets and mentions of these two sets (ISIS and Legit) by the overall Twitter community are also extracted. Big-ISIS group represents tweets associated with the suspended retweeters of the seed ISIS accounts. The population size of Legit tweets is 7,325,086,100 created by 20,536,162 accounts.

To better understand the behavior and interactions of ISIS accounts, we randomly sampled an equal-size set of legitimate accounts as a comparison set. That is, we sampled accounts that were still alive on Twitter by the end of 2016 – one year after our dataset was created – by querying each account page on Twitter. Concretely we have the following:

- **Legit-Tweets:** A random sample of 23,880 legitimate accounts. These accounts posted 17,454,068 tweets.
- **Legit-Retweets:** All retweets of the above legit tweets. We found 12,175,619 retweets generated by 1,753,195 users.
- **Legit-Mentions:** All tweets mentioning the above legit tweets. We found 17,479,990 tweets generated by 2,161,106 users.

Big ISIS. In addition to our initial seed of 24k ISIS-related accounts and inspired by the finding that retweets are usually an indicator of endorsement (Weber, Garimella, and Batayneh 2013) we want to utilize a much bigger dataset of 173k retweeters of our seed ISIS accounts shown in Figure 2 that have also been suspended most likely due to extremist behavior. We validate this bigger set is indeed ISIS-related by two methods:

- Language similarity: we utilize results recently reported by researchers in (Magdy, Darwish, and Weber 2016) and (Bodine-Baron et al. 2016) that could easily classify a user as a supporter or detractor based on keywords used that they found to be highly discriminative. For example, they found that the phrase “Islamic State” is an indicator of support to ISIS with 93% accuracy. Indeed, we find that ISIS retweeters who were eventually suspended made heavy use of pro-ISIS keywords (Table 2) with “Islamic

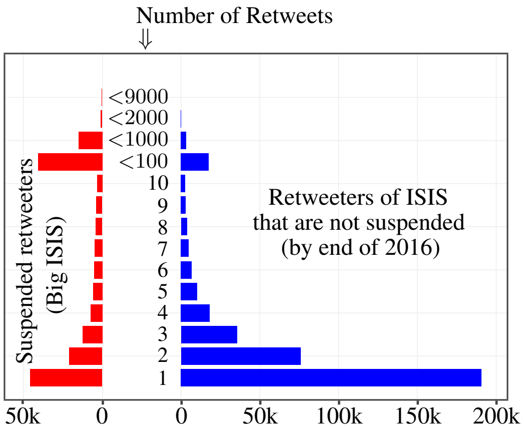


Figure 2: Retweeters of ISIS seed accounts that were either banned (red) or are still active (blue). Heavy retweeters are more likely to get banned while an occasional retweet from a regular user is less likely to warrant banning from the service. Examining a 1,000 random sample from the suspended retweeters show that 97% were pro-ISIS.

State" being the top hashtag used. (other similar keywords with a positive sentiment to ISIS include Caliphate State and Caliphate News) which hints that those retweeters are also pro-ISIS.

- Classification: we uploaded to our Twitter clone a random sample of 1,000 accounts from those suspended ISIS retweeters with five randomly sampled tweets from their tweets and asked Arabic speakers to label those accounts as pro-ISIS, anti-ISIS, or unrelated. Similar to the 24k seed accounts, the majority of ISIS retweeters who were suspended were classified as pro-ISIS (97%) and the rest as unrelated. No single account was classified as anti-ISIS.

Concretely, we have the following for the bigger ISIS set:

- **Big-ISIS-Tweets:** ISIS seed accounts and their suspended retweeters, there are 173,340 such accounts who posted 341,365,270 tweets. Note that this set of suspended retweeters includes about 21k accounts from the 24k seed accounts.
- **Big-ISIS-Retweets:** All retweets of the above Big-ISIS accounts. We found 71,576,995 retweets generated by 2,077,828 users.

Note that this dataset contains only tweets in the Arabic language, but since ISIS is mainly active in the Middle East and the majority of its members are Arabic speakers, we believe this dataset provides a clear window into ISIS social media strategy.

To study the impact of ISIS on the overall Twitter community, we focus here on their interactions with the community. We consider *retweeting* as a signal that a user has seen the tweet and, as we explain later, in most cases this means an interest in or endorsement of the tweet. We also consider *reply* or *mention* of a user as an interaction event that often involves discussion, agreement or disagreement. We measure

ISIS	**ISIS Retweeters (Suspended)**	**ISIS Retweeters (Not suspended)**
Islamic_State	*Islamic_State*	Saudi
Caliphate_News	Saudi	Decisive_Storm
Caliphate_State	Decisive_Storm	Al-Hilal_FC
Daesh	*Caliphate_State*	Quran
Saladin_Region	AlHilal_FC	Yemen
Decisive_Storm	Daesh	Syria
Ramadi	*Caliphate_News*	Hadith
Anbar_Area	AlNasr FC	AlNasr FC
Takrit	Riyadh	Peaceful_Tweeter
Aamaq_Agency	Egypt	Egypt

Table 2: Top Hashtags (translated) for ISIS and their retweeters. The top hashtag for suspended retweeters (#Islamic_State) is known to be highly associated with supporting ISIS (Magdy, Darwish, and Weber 2016). Similarly, the hashtags in *italic often carry positive sentiment towards ISIS.*

the size of the interaction which is useful in understanding ISIS penetration in the Twitter community.

We use both the smaller seed accounts and the bigger ISIS groups in the following analysis; we indicate which set is at study when ambiguous.

Location of ISIS users. Figure 3 shows a world map of accounts in our dataset where at least one of their tweets was posted with geolocation enabled. Unsurprisingly, we see that most ISIS accounts were based in the Middle East but with some activities happening in other parts of the world. Interestingly, however, we see that the seed 24k ISIS accounts concentrate in areas controlled by ISIS in Syria and Iraq while their retweeters came from neighboring countries, most notably Saudi Arabia. This may hint that the 24k seed accounts are for people officially associated with ISIS on the ground while their retweeters might have been sympathizers with ISIS but not members. We don't investigate this distinction in this paper.

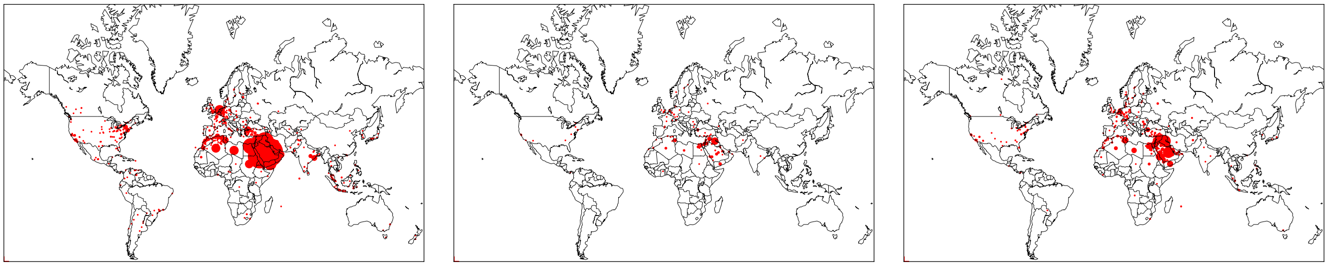
3 Research Questions and Setup

Having access to most ISIS-related content on Twitter in 2015, we seek to answer the following research questions:

1. How many supporters of ISIS were there?
2. How influential were ISIS collectively on the overall Twitter community?
3. What behavioral traces did ISIS leave behind that could have been used to identify them early in their lifespan?

Below we describe our methodology and experimental setup to address these questions, and in subsequent sections, we present our findings and their analysis.

"Retweet $\neq$ Endorsement" or is it? Several researchers have concluded that retweeting is a sign of agreement with either the sender or the message or both. For example (Boyd, Golder, and Lotan 2010) and (Cha et al. 2010) pointed out that retweeting is passing "along interesting pieces of information." (Welch et al. 2011) found that retweeting is a better



(a) All suspended accounts (b) 24k seed ISIS accounts (c) Suspended retweeters of ISIS (Big ISIS)

Figure 3: A world map of accounts that had geolocation enabled for at least one of their tweets. (b) Seed ISIS accounts are mainly based in areas in Syria and Iraq that were controlled by ISIS in 2015. (c) Retweeters of ISIS are concentrated in neighboring countries, most notably in Saudi Arabia. (a) all suspended accounts for reference.

indicator of interest than merely *following* a user. Directly relevant to our work, Weber et. al., found that "Retweeting signifies endorsement" while studying tweets in both Arabic and English about the Egyptian revolution (Weber, Garimella, and Batayneh 2013). Using retweet information, they were able to label users as either Islamist or secular. In addition, (Bruns and Burgess 2012) found that retweeting can be interpreted as an implicit endorsement for the message and the sender unless additional commentary is added. For this reason, we use retweeting as a proxy for measuring the influence of users on each other and discard other types of interactions such as replying to a tweet or quoting a tweet because it can be difficult to evaluate the added commentary.

Retweet Graph. We define the retweet graph G as a weighted directed graph $G = (V, E)$ where each node $u \in V$ represents a user in Twitter, and each edge $(p, q) \in E$ represents a retweet in the network, where the user p has retweeted a tweet posted by user q. The weight w of an edge (p, q) is the number of times user p has retweeted user q in the dataset. The number of all users in the graph is $N = |V|$. Notice that influence flows in the opposite direction of retweets, and thus we work on the graph with inverted edges. For example if a user p retweets another user q then the influence flows from q to p. In addition, for experiments involving PageRank and Personalized PageRank, we filter out edges that have weight ($w < 3$) for two reasons: to reduce complexity and to increase our confidence in the existence of an influence link between the two users. Table 3 shows the details of the whole and trimmed graph.

Collective Influence. PageRank (Page et al. 1999) is a well-known algorithm that uses link information to assign global importance to nodes in a given graph. PageRank or variations of it has been applied to many fields including Chemistry, Biology, Neuroscience, Sports, Social networks and others. In previous work, Personalized PageRank (Haveliwala 2002) has been used to discover spam nodes in a graph (Gyöngyi, Garcia-Molina, and Pedersen 2004; Krishnan and Raj 2006) and to find topic-sensitive influential users on Twitter (Weng et al. 2010) – all inspired by the general method of Random Walk with Restart (Tong, Faloutsos, and Pan 2006). In the following, we utilize PageRank to calculate Collective Influence of groups on Twitter and we use Personalized PageRank to approximate the size of ISIS community.

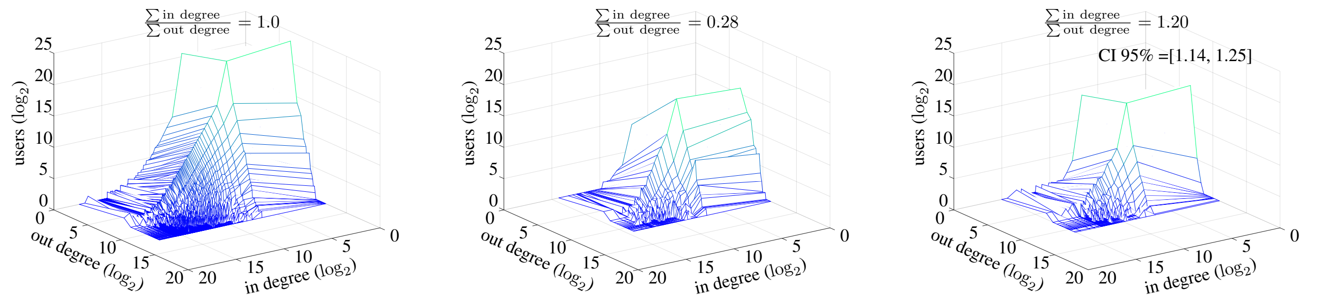
All Arabic Retweet Graph	**Nodes**	19,008,183
	Edges	1,063,949,461
Edge Weight ≥ 3	**Nodes**	8,976,644
	Edges	219,024,066
Seed ISIS	**Nodes**	23,051
	Edges (inward)	4,295,420
	Edges (outward)	4,705,535
Edge Weight ≥ 3	**Nodes**	18,633
	Edges (inward)	782,923
	Edges (outward)	833,542
Big ISIS	**Nodes**	172,511
	Edges (inward)	21,131,756
	Edges (outward)	67,473,685
Edge Weight ≥ 3	**Nodes**	157,522
	Edges (inward)	3,927,716
	Edges (outward)	14,272,022

Table 3: Arabic Retweet Graph in 2015. Nodes are users; edges are retweets. Big ISIS is our seed of 24k accounts and the retweeters who were suspended and spread pro-ISIS content.

4 ISIS Community Size

A first step towards understanding the influence of ISIS is to assess how many accounts are owned by the group or its sympathizers. We found that about 30% of retweeters of the 24k seed ISIS accounts were eventually suspended and were posting pro-ISIS content. However, that leaves 70% of the remaining retweeters for whom we don't know whether they are supporters of ISIS – making estimating the size of the ISIS community a crucial task in our study.

Measuring Community Size. Hence, we would like to approximate the size of the ISIS sympathizer community based on the interaction graph. We use a slightly modified version of Personalized PageRank as follows: we choose a subset of known ISIS accounts and run Personalized PageRank biased to these known accounts. We then find the top-k nodes relevant to this group and check how many of them are already



(a) All Arabic users (b) ISIS (Big) (c) Legit sample

Figure 4: Distribution of in-degrees and out-degrees for the Retweet graph. (b) ISIS tries to reach out by retweeting many other accounts while receiving less in return. A random sample of legitimate accounts by comparison (c) receives slightly more retweets than they give. Color intensity increases with the rarity of users (z axis).

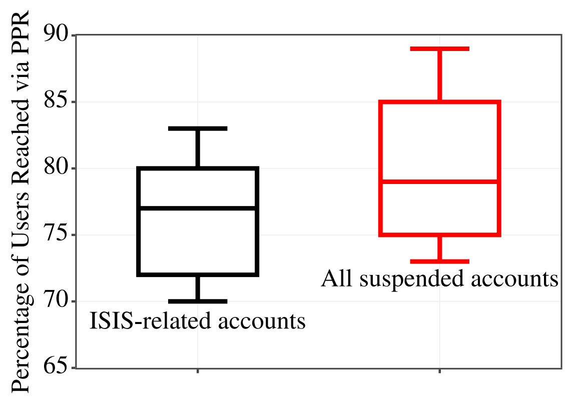


Figure 5: Results of running Personalized PageRank for 100 randomly selected ISIS accounts and checking the top-100 reached accounts, repeated ten times. About 75% of users reached are already in our Big-ISIS set, and 80% were eventually suspended. This gives us evidence that our Big-ISIS dataset represents about 75% of all potential ISIS accounts on Twitter in 2015.

in our ISIS group. We can then approximate the potential size of the group of ISIS-related accounts using the fraction of users that are already known to be ISIS-related. For example, if 50% of the accounts found by the Personalized PageRank are already known to be ISIS, then we can roughly say that the real size of the group of ISIS-related accounts is twice what we already know. We set k to 100, but it's possible to experiment with different values of k; if k is too small then the chance of having this k in our known ISIS dataset is high, and likewise if k is too big then it may reach unrelated users to ISIS and thus introduce errors in the approximated size. We leave experimenting with different values of k for future work.

How many ISIS supporters are there? We start with our seed of 24k ISIS-related accounts and the extended group of 170k retweeters that were suspended later by Twitter (Big ISIS). We find that in fact, 21k of those retweeters are already in the original 24k seed accounts showing that ISIS tend to inflate itself by self-retweeting. We then randomly sample 100 of these accounts and run the Personalized PageRank for this random set and record the top-100 reached users. We then find how many of these reached users are already in our extended ISIS dataset. We repeat this ten times with different seed accounts each time. Figure 5 shows that about 75% of accounts reached are already known to be ISIS-related and about 80% of the accounts reached have already been suspended by Twitter. This means the extended ISIS group we use in this paper is a good representative of ISIS accounts covering a majority of them. In addition, since we have about 173k ISIS-related accounts, the expected number of total ISIS accounts in 2015 is roughly 217k accounts (i.e., 25% more than 173k).

Twitter announced in February 2016 (Twitter 2016) that it had closed 125,000 ISIS accounts starting from mid-2015. Our estimate of 217k accounts suggests that even more ISIS accounts were alive. In August 2016, Twitter announced that they had closed an additional 235,000 accounts, but this includes accounts created after the period of our analysis.

5 ISIS Collective Influence

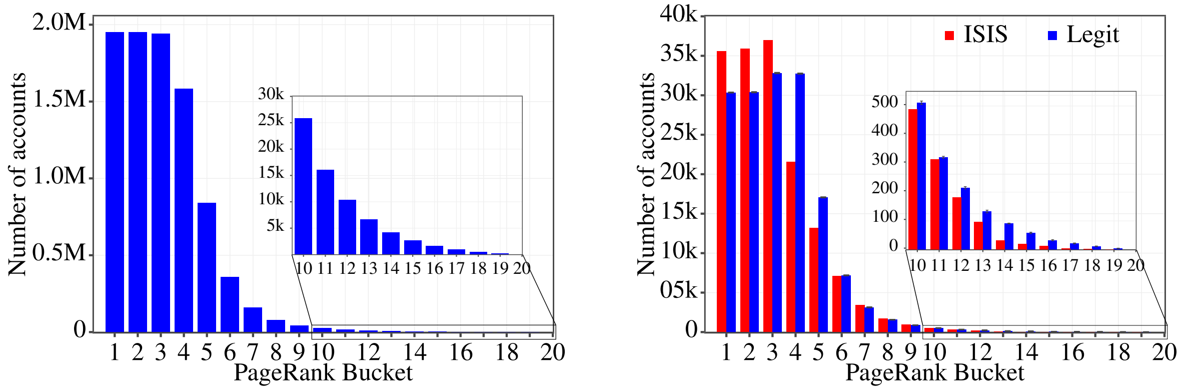
Now that we have a rough estimate of the size of ISIS community on Twitter, we would like to know how influential they were. As Figure 4 shows, ISIS tended to generate more out-links than an average community – perhaps in an effort to reach out to the community – but did they succeed in exerting influence on the community?

Measuring Collective Influence. In the context of Twitter and retweet graph introduced above, the PageRank can be iteratively defined as:

$$\mathrm{PR}(t) = (1-d) + d \times \sum_{v \in R_t} \frac{\mathrm{PR}(v)}{N_v}$$

where t is a user in the retweet graph and R_t are the set of users who retweeted t.

We run the PageRank algorithm on the retweet graph (with weight $w \geq 3$, and a decay factor d of 0.85 and 20 iterations) that assigns a value in the range $[0, 1]$ to each node (user) in the graph. We then calculate the collective influence CI of a user group U as the sum of the PageRank values for



(a) All users on Arabic Twitter in 2015

(b) ISIS (Big) and random sample of legit accounts

Figure 6: PageRank Distribution. Users in the first bucket are the least influential while bucket 20 contains the most influential users. (a) The distribution of all accounts on Arabic Twitter for comparison. (b) Distribution of accounts from ISIS and legit over buckets. ISIS tends to be skewed to the left occupying more of the lower level influence buckets. Error bars on the legitimate accounts indicate the standard error of the mean.

its members:

$$\mathrm{CI}(U) = \sum_{u \in U} \mathrm{PR_u}$$

For example, a CI of 0.02 for a group means that in the bigger Twitter community this smaller community attained 2% of the attention over the time these interactions happened.

Because we want to measure the influence of all users in a group (ISIS in this case) over the overall Twitter community, we use the original PageRank algorithm which assigns importance to nodes based on mutual reinforcement between all the nodes. Personalized PageRank (Haveliwala 2002), TrustRank (Gyöngyi, Garcia-Molina, and Pedersen 2004) and other algorithms that introduce bias are not suitable for measuring influence as we define it here but are useful in measuring the potential size of a given community (See Section 4).

How Influential are ISIS accounts? We first start by inspecting the retweet graph. Figure 4 shows the distribution of in-degrees and out-degrees over nodes (Twitter users). We see that ISIS accounts retweet about four times more than they themselves get retweeted. By contrast, a random sample of a similar number of legitimate users does receive slightly more retweets than they give. Moreover, we see that calculating the Collective Influence for ISIS also shows that the influence gained by ISIS from the total Influence mass is less than that of a random group: (Table 4). In other words, the accounts used by ISIS were not (relatively) influential accounts and did not manage to generate high-quality links that would have increased their CI.

To better understand the distribution of influence over users, we divide the PageRank mass into 20 buckets. The first bucket contains users in the lowest 5% of the PageRank scores (about two million users belong to this category) and accounts in the twentieth bucket are among the most influential accounts (a few hundred users). For example, @KingSalman account belongs to bucket 20 with the highest PageRank in 2015. Figure 6 shows that ISIS accounts tend to occupy more of the lower level buckets while a random sample of legitimate accounts tends to have relatively less mass on the lower level buckets and more mass on the higher level buckets. The conclusion here is that, although ISIS accounts may be more active – i.e., generating more "links" – their overall influence on the Twittersphere is limited.

Group	Collective Influence	95% Confidence Interval
ISIS (Big)	0.01434	N/A
Sample Legit	0.01964	[0.01886, 0.02041]

Table 4: Collective Influence of ISIS (Big) and a random sample of similar size set of legitimate users.

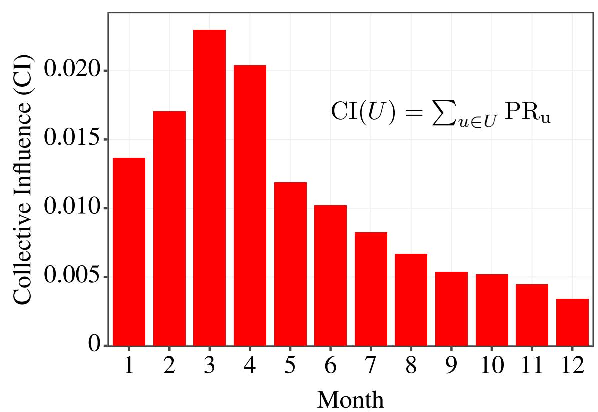


Figure 7: ISIS (Big) Collective Influence over months in 2015. We note a diminishing influence over time

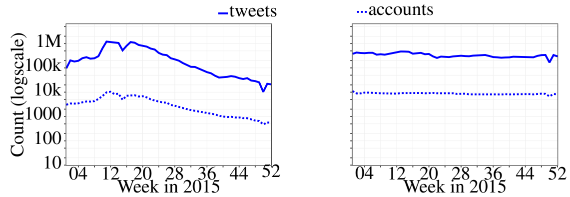
How does ISIS influence evolve? We would like to know if ISIS gained or lost influence over time in 2015. To answer this question, we recalculate the Collective Influence for each month and compare the portion gained by ISIS every month. Figure 7 shows that ISIS influence was improving in the first quarter of 2015 but it then sharply decreased until the end of 2015. This may be because Twitter cracked down on those accounts and closed many of them. ISIS kept creating new accounts (many accounts have "replacement account" in their description) after their old accounts were closed but they struggled to regain their lost influence due to the limited time those accounts were on the service. While some researchers have questioned the effectiveness of censorship for tackling the problem of extremism (Bartlett and Krasodomski-Jones 2015), this finding here shows that closing ISIS accounts is a practical method of reducing their impact and limiting their ability to spread their content.

6 Behavioral Observations

ISIS has gained a reputation for effectively using social media to their benefit. However, as we have seen, their influence on social media might have been overrated. In this section we complement our study so far by turning to an observational study of the content ISIS sends, their evolution over time, and how different they are from other common spam accounts on Twitter. Note that for analyses in this section we used ISIS seed accounts.

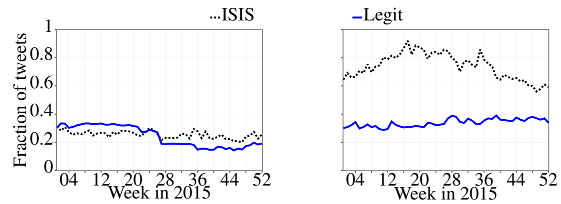
Are ISIS accounts Spam? We start our analysis here by looking into the overall activity of ISIS accounts as compared to other regular users. Do they exhibit a spammy behavior such as posting many tweets or many URLs linking to their own websites? Figure 8 shows that over time ISIS accounts maintain a similar level of activity when compared to other legitimate users although we see that their activity was negatively impacted through the year 2015, due to Twitter closing down their accounts after the crowd started reporting those accounts. In addition, we also see that ISIS accounts manage to post more than other suspended accounts (Figure 1) hinting that they might not have exhibited *banning* signals until later in their lifespan. Moreover, we don't notice an excessive use of URLs by ISIS (Figure 9), clearly distinguishing their accounts from other prevalent spam accounts that post many URLs (Grier et al. 2010). Another method by which ISIS may try to reach out to the community is by posting tweets in trending or unrelated hashtags. In fact, we find that ISIS heavily uses hashtags (Figure 9). The reach achieved through this *hashtag hijacking* is challenging to quantify because Twitter users may see their tweets while browsing the hashtag but will not take any action such as retweeting or replying that we can collect. For this reason, although important we don't consider reach obtained through *hashtag hijacking*.

Who retweets/mentions ISIS? Next, we move to studying the first-level users who interacted with ISIS accounts. We find that about 76% of the interactions ISIS received actually came from eventually suspended users (Figure 10a). Alfifi et. al., (Alfifi and Caverlee 2017) found that 23% of active Arabic Twitter users in 2015 were eventually suspended (contributing 21% to the Arabic Twitter volume that year). This means that roughly speaking, by mere chance a community of retweeters will be about 20% suspended and 80% not suspended. However, we see that retweets of ISIS con-



(a) ISIS tweets (b) Legit tweets

Figure 8: ISIS didn't post excessively in 2015 (a) compared with a random set of legit users (b). Additionally, ISIS lost accounts over time in 2015 due to Twitter suspending them.



(a) URLs (b) Hashtags

Figure 9: ISIS resembles normal users in their use of URLs (a) clearly different from spam users spreading lots of URLs (Bakshy et al. 2011). However, ISIS uses more hashtags (b), potentially for *hashtag hijacking*

tent are more than three times likely to be from suspended accounts, strongly suggesting that ISIS retweeters are also involved in malicious activities. The legit group, by comparison, has only 17% of its retweets coming from eventually suspended accounts which is in line with the above 80/20 rule.

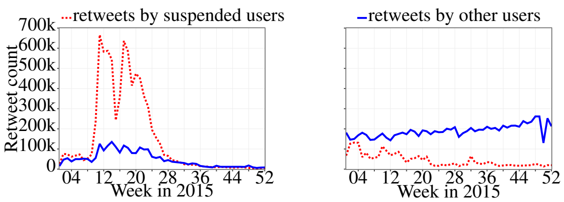
Similar to retweets, if we filter out mentions that were generated by eventually suspended users, we find that ISIS gets more negatively affected compared to a random sample of legit accounts (Figure 11) – another signal that ISIS accounts are more likely to attract other malicious users to interact with them. Thus, we next investigate if such accounts could actually be ISIS-related as well.

Are ISIS retweeters pro-ISIS? To answer this question, we collect all tweets generated by retweeters of ISIS accounts (i.e., we collect all their tweets whether ISIS-related or not). We first note that those accounts generated 1.4 billion tweets (15% of all Arabic content in 2015!). We then divide this big set of tweets into two groups: (i) tweets from suspended accounts and (ii) tweets from accounts still alive by the end of 2016 (Table 5). We then look into the content created by these two groups.

We have analyzed the group of suspended retweeters (Big ISIS) in Section 2 and showed that they are mostly pro-ISIS accounts but what about the big group of unsuspended retweeters? We again sampled 500 accounts from those unsuspended retweeters of ISIS and asked labelers to label them as either pro-ISIS, anti-ISIS, or unrelated. The majority ($\approx 90\%$) of those unsuspended accounts were labeled as unrelated (they may have made an occasional retweet of some ISIS account but didn't engage further) while 8% (40 ac-

Group	Users	Tweets
ISIS Retweeters (Suspended)	170,016	389,358,515
ISIS Retweeters (Not Suspended)	381,853	964,828,227
Total	551,869	1,354,186,742

Table 5: ISIS retweeters



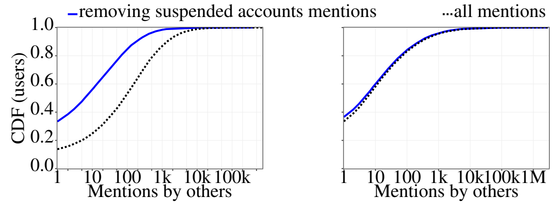
(a) ISIS accounts (b) Legit accounts

Figure 10: (a) Most (76%) of the retweets of ISIS came from eventually suspended accounts (31% of all retweeters) hinting that a large number of those interacting with ISIS are likely involved in malicious activities as well. (b) By comparison, only 17% of the retweets of legit users were from eventually suspended users (15% of all accounts).

counts) were labeled as pro-ISIS. As we discussed in Section 4, we think we have captured about 75% of ISIS accounts so it's not surprising to find that there are a small percentage of pro-ISIS accounts that were not suspended. Furthermore, only 2.4% (12 accounts) were labeled as anti-ISIS showing that overall there was little interest in arguing with ISIS.

Did ISIS receive negative sentiment? To answer this question, we consider replies to ISIS content as this can better capture sentiment than just retweeting where there is no added commentary. To that end, we grouped repliers to ISIS into two groups: suspended repliers and unsuspended repliers. We then sampled 500 users from each group with a randomly selected five of their tweets in which they replied to an ISIS tweet. We then used our Twitter clone showing labelers both the reply and the replied to tweet and asking them if a given user is supporting or undermining ISIS. We find that 97% of suspended repliers are labeled as pro-ISIS, and surprisingly even the unsuspended repliers are found to be 68% pro-ISIS. This shows that ISIS might have been using Twitter as a communication channel for their own members and that the overall Twitter community largely avoided interacting with them.

Are ISIS accounts recruited or born that way? As we see in Figure 1, ISIS accounts manage to post more content before getting suspended. Were those accounts *normal* accounts that later turned into bad ones (e.g., as a result of a recruitment campaign by ISIS) and hence had more time to spread content? Or were those accounts created to support ISIS from the beginning? While this problem deserves an independent investigation, we make a first observation that can be extended in future works. A straightforward way to check for the existence of recruitment would be to check the textual content of the accounts from their birth to their death.



(a) ISIS accounts (b) Legit accounts

Figure 11: When removing interactions initiated by eventually suspended accounts, ISIS accounts tend to lose a significant part of the community interactions (a) while a random sample of legit accounts is negligibly affected (b) – another indicator that a percentage of users interacting with ISIS are also involved in malicious activities.

We focus here on accounts that were born in January 2015 and were suspended before December 2015. There are 5,057 such accounts that generated 4,970,042 tweets. We check their top hashtags used at three points in their lifespan: at birth, mid-life, and at death (right before suspension). Table 6 shows that although a majority of accounts have been pro-ISIS from the beginning (because #Islamic_State is the top hashtag all the time), we still see a strengthening of ISIS support over time evident in the increasingly more ISIS-related hashtags appearing over users lifespan. For example, we see some innocuous hashtags (e.g., supplications and forgiveness) in the *at birth* column in Table 6 that we don't see at later stages. We acknowledge our dataset may not be optimal for studying recruitment as it spans only a year in the middle of ISIS existence; a more focused dataset spanning several years could uncover more results. However, within the span of one year, we see that most accounts were created to support ISIS from the beginning. Twitter was heavily cracking down on ISIS accounts forcing them to keep recreating new accounts to continue their operations (e.g., we see a lot of "replacement account" in the description of many accounts).

7 Related Work

In this section, we first discuss datasets that are relevant to measuring the social media communications of extremists and then we introduce other approaches that have investigated the online communication of extremists.

Berger and Morgan identified 46,000 ISIS supporter accounts on Twitter (Berger and Morgan 2015). They began with 454 accounts known as ISIS supporters as initial seeds. They then collected all accounts following those seeds. This approach continues to two further steps on the followed-by network. They introduced a classification task to determine if a user is an ISIS supporter.

Magdy et. al., collected 3 million Arabic tweets referring to ISIS (Magdy, Darwish, and Weber 2016). They classified users based on how they call ISIS. The full name of "Islamic State" is the indicator of support, while abbreviations like "ISIS" or "Daesh" indicate opposition. Zaman (Zaman 2016) introduced an ISIS dataset consisting of about 17,000 tweets from about 100 ISIS supporter accounts. Keywords,

At birth	Mid-life	At death
Islamic_State	*Islamic_State*	*Islamic_State*
Forgiveness	*Caliphate_News*	Syria
Supplication	*Caliphate_State*	*Caliphate_State*
Daesh	Daesh	Nusrah_Front
King_Abdullah_Death	Decisive_Storm	Aleppo
Charlie_Hebdo	Takrit	Daesh
Saudi	Saladin_Area	Fatah_Army
Supplications	Ramadi	Iraq
Forgiveness	Nusrah_Front	Aljazeera
Nusrah_Front	Anbar_Area	Aamaq_Agency

Table 6: Top Hashtags (translated) for ISIS accounts at three intervals in their life: at birth, mid-life, and at death. While the majority of accounts are pro-ISIS from day 1 (#Islamic_State being the top hashtag since birth), we see a strengthening message of ISIS overtime, Innocuous hashtags (supplications, forgiveness) disappear at mid-life onward. (Hashtags in *italic are highly associated with positive sentiment)*

Images, and network-based features were used to classify a user as pro-ISIS or not.

In another effort, Bodine-Baron et al., collected 23 million tweets that referenced the Arabic versions of either "Islamic State" or "Daesh" (Bodine-Baron et al. 2016). This corpus consists of both ISIS supporters and opponents whom they were separated by which phrase they use to describe ISIS: ISIS followers refer to ISIS as the "Islamic State" while detractors often use the term "Daesh". They lexically analyzed the tweets and found that users saying "Daesh" use other terms such as "Terrorist", "Kharijites", "militants", "dogs of fire", and "dogs of Baghdadi" that prove they are highly critical of ISIS. On the other hand, tweets with "Islamic State" contain glowing terms such as "monotheists Mujahideen", "Soldiers of the Caliphate", and "lions of the Islamic State".

Stanton et al., focus on 2,200 military incidents caused by ISIS and forces opposing ISIS (Stanton et al. 2015). Data was collected from resources other than Twitter including reports published by the Institute for the Study of War (ISW), MapAction, Google Maps, and Humanitarian Response.

In a recent effort, Badawy et al., investigated how ISIS supporters take advantage of social media to spread their propaganda and recruit militants by studying 1.9 million tweets posted by 25,000 accounts recognized as pro-ISIS and suspended by Twitter (Badawy and Ferrara 2017).

There exists a small number of datasets on violent acts. The Global Terrorism Database (GTD) consists of information about 170,000 terrorist attacks [2]. The Worldwide Incidents Tracking System (WITS) dataset on terrorist events is manually extracted from news websites and social media [3]. The International Terrorism: Attributes of Terrorist Events (ITERATE) dataset describes characteristics of terrorist groups, their activities that carry international impact and the environment in which they operate [4]. Minorities at Risk Organizational Behavior (MAROB) investigates the reasons why ethnic minorities are vulnerable in becoming radicalized and consequently forms extremist organizations mainly in the Middle East and North Africa[5]. The RAND Database of Worldwide Terrorism Incidents (RDWTI) contains information about terrorist incidents since 1968 [6].

The Islamic State heavily relies on social media for recruitment. Recently there has been work that focuses on analyzing the data made by this group. For example, Farwell (Farwell 2014) studied the media strategies of ISIS. They described two conflicting strategies followed by ISIS: they try to protect the identity and location of their leaders by minimizing Internet communications while they take advantage of social media for recruitment. Gates and Potter (Gates and Podder 2015) analyzed who creates the content of the organization's recruitment materials and found that they use a globally distributed network of volunteers who create content to fit the aesthetic of their particular region. Both of these works focus on the recruitment and creation of ISIS messaging. In this work, we investigate its effectiveness by measuring the reactions by others on social media.

8 Conclusion

This investigation has shed light on the extent to which extremist groups such as ISIS interact with the online social communities they operate in. Starting from a 24k seed accounts known to be related to ISIS, we were able to uncover a much larger group of potentially ISIS-related accounts (173k users). We saw how ISIS is fragile against suspended users with the shape of their interactions severely affected when we remove suspended accounts. We used PageRank to show that the Collective Influence of ISIS is limited and we applied Personalized PageRank to quantify the potential size of ISIS community. Although our proposed method showed that ISIS collective influence on Twitter is relatively less than the average, we would like to extend this work by evaluating and designing other Collective Influence measures that are more immune to the reciprocal interaction links that ISIS tend to engage in. In addition, we aspire to study other collective efforts on social media by groups other than extremists that seem to try to shape mass opinions. How do organic communities differ from organized campaigns? Do organized campaigns tend to have lower Collective Influence measure? How can our approach be immune to governments and organizations employing social media users that are already influential to spread a certain message?

References

Alfifi, M., and Caverlee, J. 2017. Badly evolved? exploring long-surviving suspicious users on twitter. In *International Conference on Social Informatics*, 218–233. Springer.

Badawy, A., and Ferrara, E. 2017. The rise of jihadist propaganda on social networks.

[2]https://www.start.umd.edu/gtd/

[3]https://ds-drupal.haverford.edu/aqsi/

[4]https://library.duke.edu/data/collections/iterate

[5]http://www.mar.umd.edu/mar_data.asp

[6]https://www.rand.org/labor/aging/dataprod/hrs-data.html

Bakshy, E.; Hofman, J. M.; Mason, W. A.; and Watts, D. J. 2011. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 65–74. ACM.

Bartlett, J., and Krasodomski-Jones, A. 2015. Counter-speech examining content that challenges extremism online. *DEMOS, October*.

Benigni, M., and Carley, K. M. 2016. From tweets to intelligence: Understanding the islamic jihad supporting community on twitter. In *Social, Cultural, and Behavioral Modeling: 9th International Conference, SBP-BRiMS 2016, Washington, DC, USA, June 28-July 1, 2016, Proceedings 9*, 346–355. Springer.

Berger, J. M., and Morgan, J. 2015. The isis twitter census: Defining and describing the population of isis supporters on twitter. *The Brookings Project on US Relations with the Islamic World* 3(20):4–1.

Bodine-Baron, E.; Helmus, T. C.; Magnuson, M.; and Winkelman, Z. 2016. Examining isis support and opposition networks on twitter. Technical report, RAND Corporation Santa Monica United States.

Boyd, D.; Golder, S.; and Lotan, G. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, 1–10. IEEE.

Bruns, A., and Burgess, J. 2012. Researching news discussion on twitter: New methodologies. *Journalism Studies* 13(5-6):801–814.

Cha, M.; Haddadi, H.; Benevenuto, F.; and Gummadi, P. K. 2010. Measuring user influence in twitter: The million follower fallacy. *Icwsm* 10(10-17):30.

Davis, C. A.; Varol, O.; Ferrara, E.; Flammini, A.; and Menczer, F. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 273–274. International World Wide Web Conferences Steering Committee.

Farwell, J. P. 2014. The media strategy of isis. *Survival* 56(6):49–55.

Ferrara, E.; Wang, W.-Q.; Varol, O.; Flammini, A.; and Galstyan, A. 2016. Predicting online extremism, content adopters, and interaction reciprocity. In *International Conference on Social Informatics*, 22–39. Springer.

Gates, S., and Podder, S. 2015. Social media, recruitment, allegiance and the islamic state. *Perspectives on Terrorism* 9(4).

Government-Publishing-Office. 2016. Isis online: Countering terrorist radicalization and recruitment on the internet and social media.

Grier, C.; Thomas, K.; Paxson, V.; and Zhang, M. 2010. @ spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, 27–37. ACM.

Gyöngyi, Z.; Garcia-Molina, H.; and Pedersen, J. 2004. Combating web spam with trustrank. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, 576–587. VLDB Endowment.

Haveliwala, T. H. 2002. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, 517–526. ACM.

Krishnan, V., and Raj, R. 2006. Web spam detection with anti-trust rank. In *AIRWeb*, volume 6, 37–40.

Lotan, G.; Graeff, E.; Ananny, M.; Gaffney, D.; Pearce, I.; et al. 2011. The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International journal of communication* 5:31.

Magdy, W.; Darwish, K.; and Weber, I. 2016. #failedrevolutions: Using twitter to study the antecedents of isis support. *First Monday* 21(2).

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Stanton, A.; Thart, A.; Jain, A.; Vyas, P.; Chatterjee, A.; and Shakarian, P. 2015. Mining for causal relationships: A data-driven study of the islamic state. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2137–2146. ACM.

Starbird, K., and Palen, L. 2012. (how) will the revolution be retweeted?: information diffusion and the 2011 egyptian uprising. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, 7–16. ACM.

Tong, H.; Faloutsos, C.; and Pan, J.-Y. 2006. Fast random walk with restart and its applications. In *Sixth International Conference on Data Mining (ICDM'06)*, 613–622. IEEE.

Twitter. 2016. An update on our efforts to combat violent extremism, https://t.co/d7pefwotf8, last access: 12/09/2018.

Weber, I.; Garimella, V. R. K.; and Batayneh, A. 2013. Secular vs. islamist polarization in egypt on twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 290–297. ACM.

Welch, M. J.; Schonfeld, U.; He, D.; and Cho, J. 2011. Topical semantics of twitter links. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 327–336. ACM.

Weng, J.; Lim, E.-P.; Jiang, J.; and He, Q. 2010. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*, 261–270. ACM.

Zaman, K. 2016. Dataset spotlight: How isis uses twitter, https://www.kaggle.com/datasets, last access: 24/11/2017.