From Isolation to Radicalization: Anti-Muslim Hostility and Support for ISIS in the West

Online Appendix

Contents

List of Tables

List of Figures

S1 Identifying ISIS activist and follower accounts on Twitter

In this project, I track lists published publicly by several anti-ISIS hacking groups to identify ISIS supporters' accounts on Twitter. Using the Twitter APIs,[1] I designed an algorithm that continually monitored and recorded ISIS accounts identified by the hacktivist group @CtrlSec.[2] Immediately upon observing a new account in the @CtrlSec list, I downloaded the complete "timeline" of tweets for the account, as well as its user profile, which includes various user-level fields, and list of the account's friends and followers. The full list of user profile fields is given in Table S1. The database contains "snapshots" of each user's profile at various points in time. In particular, between December 2015 and May 2016, user profile snapshots were saved when the user was encountered on the @CtrlSec list or included as part of 5,000 randomly selected follower accounts for content sampling every 24 hours. Between May 2016 and January 2017, new snapshots are obtained for all non-suspended user accounts every 1-2 days, on average. The full list of data fields for each tweet is given in Table S2.

Downloading Twitter timelines

The dimensionality of the friends and followers is particularly challenging for historical timeline data collection. While I have identified approximately 15,000 activists from the @CtrlSec postings, this has led to over 1.6 million followers and about 450,000 friends of these followers. Due to rate limits, it is impossible using the publicly available Twitter API to obtain full content timelines for all of these accounts. Thus, I began by downloading the full historical tweet timelines of all @CtrlSec-identified "ISIS activist" accounts ($N = 15,088$), as well as of all the friends of a sub-sample of the activists who were first observed in the database as a follower or friend, and subsequently 'flipped' and became flagged as activists ($N = 193,973$). After completing an initial round of location prediction, I downloaded the complete historical tweet timelines of additional accounts of ISIS followers and friends predicted to be located in Europe and North America.

There are two additional sources of tweet timeline content in the dataset. The first is a so-called "random sample with holes." Since the Twitter Streaming API imposes rate limits on usage, I was only able to stream content for 5,000 users in a 24-hour period. The streaming began on December 19, 2015, and with the exception of occasional technical glitches, has been collecting data on the content posted by a random sample of 5,000 followers each day. Moreover, as noted previously, user profile information was downloaded at the same time. This ensures that user-level information (such as profile picture, number of friends, etc.), as well as account suspension status, were updated daily for this random sample.

The second source of tweet timeline data is a daily "total refresh" that began in May 2016. The Twitter API permits obtaining a current profile snapshot for a user, which contains their most

[1] `https://dev.twitter.com/overview/documentation`

[2] Lists are available in these handles: @ctrlsec, @ctrlsec0, @ctrlsec1, @ctrlsec2, @ctrlsec9.

recently posted tweet, at a much faster rate limit than a full historical content download. Thus, I began to cycle through the entire database of over 1.6 million accounts on a daily basis, requesting latest profile and tweet, which led to a complete refresh of user profiles and the latest tweet for each user in the system, as well as their suspension status, every 1-2 days on average. The total number of tweets scraped with this method was over 100 million as of January 2017.

Table S1: List of data fields at the user level

Field Name	**Description**
user_id	The integer representation of the unique identifier for this User.
date_added	The datetime the user profile snapshot was added to the database.
name	The name of the user, as they've defined it. Not necessarily a person's name.
screen_name	The screen name, handle, or alias that this user identifies themselves with.
location	The user-defined location for this account's profile. Not necessarily a location nor parseable.
description	The user-defined UTF-8 string describing their account.
url	A URL provided by the user in association with their profile.
protected	When true, indicates that this user has chosen to protect their Tweets.
followers_count	The number of followers this account currently has.
friends_count	The number of users this account is following (AKA their "followings").
listed_count	The number of public lists that this user is a member of.
created_at	The UTC datetime that the user account was created on Twitter.
favourites_count	The number of tweets this user has favorited in the account's lifetime.
utc_offset	The offset from GMT/UTC in seconds.
time_zone	A string describing the Time Zone this user declares themselves within.
geo_enabled	When true, indicates that the user has enabled the possibility of geotagging their Tweets.
verified	When true, indicates that the user has a verified account.
statuses_count	The number of tweets (including retweets) issued by the user.
lang	The BCP 47 code for the user's self-declared user interface language. May or may not have anything to do with the content of their Tweets.
profile_background_image_url	A HTTP-based URL pointing to the background image the user has uploaded for their profile.
profile_image_url	A HTTP-based URL pointing to the user's avatar image.
profile_image_file	A local copy of the user's profile image.
profile_banner_url	The HTTPS-based URL pointing to the standard web representation of the user's uploaded profile banner.
profile_banner_file	A local copy of the user's profile banner.
followers	The list of the user's followers, as of the date of this "snapshot." (Only obtained for certain users such as ISIS activists.)
friends	The list of the user's followers, as of the date of this "snapshot." (Only obtained for certain users such as ISIS activists.)
suspended	A flag for whether the account was suspended.

Table S2: List of data fields at the tweet level

Field Name	**Description**
id	The integer representation of the unique identifier for this Tweet.
user_id	The integer representation of the unique identifier for the author of the Tweet.
date_added	The datetime that the Tweet was added to the database.
created_at	The datetime that the user account was created on Twitter.
text	The actual UTF-8 text of the status update.
source	Utility used to post the Tweet, as an HTML-formatted string. Tweets from the Twitter website have a source value of web.
truncated	Indicates whether the value of the text parameter was truncated, for example, as a result of a retweet exceeding the 140 character Tweet length. Truncated text will end in ellipsis, like this ...
in_reply_to_status_id	If the represented Tweet is a reply, this field will contain the integer representation of the original Tweet's ID.
in_reply_to_user_id	If the represented Tweet is a reply, this field will contain the integer representation of the original Tweet's author ID.
in_reply_to_screen_name	If the represented Tweet is a reply, this field will contain the screen name of the original Tweet's author.
retweet_count	Number of times this Tweet has been retweeted.
favorite_count	Indicates approximately how many times this Tweet has been "liked" by Twitter users.
lang	When present, indicates a BCP 47 language identifier corresponding to the machine-detected language of the Tweet text, or "und" if no language could be detected.
possibly_sensitive	This field is an indicator that the URL contained in the tweet may contain content or media identified as sensitive content.
coordinates	Represents the geographic location of this Tweet as reported by the user or client application.
withheld_in_countries	When present, indicates a list of uppercase two-letter country codes this content is withheld from.
quoted_status	This field only surfaces when the Tweet is a quote Tweet. This attribute contains the Tweet object of the original Tweet that was quoted.
retweeted_status	This attribute contains a representation of the original Tweet that was retweeted.

Note: Descriptions are copied verbatim from the Twitter REST API at https://dev.twitter.com/overview/api

Figure S1: Scraping ISIS accounts

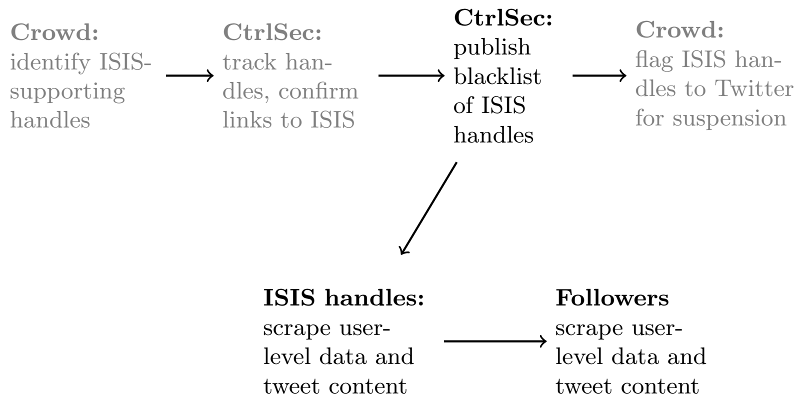


Table S3: Number of tweets posted by all users in database, by year

Year	# tweets
2007	849
2008	4,740
2009	42,667
2010	113,625
2011	376,627
2012	1,299,006
2013	3,285,090
2014	6,552,219
2015	17,887,290
2016	69,900,477
2017	4,903,609

Note: The number of tweets is accurate to 1/30/2017.

S2 Predicting geographic location of ISIS activists and followers

S2.1 Spatial Label Propagation algorithm

The spatial label propagation (SLP) algorithm used to predict the geographic locations of Twitter users in this paper implements the method developed by Jurgens (2013). The algorithm works as follows. First, define U to be a set of Twitter users in a social network, and for each user, let N be a mapping from the user to her friends (i.e., users to whom the user is directly connected), such that $u \rightarrow [n_i, ..., n_m]$. Also, let L be a mapping of users to their known geographic locations: $u \rightarrow (latitude, longitude)$, and E the current mapping from users to locations. E is being updated with each iteration of the algorithm.

The algorithm works as follows. First, it initializes E, the current mapping from users to locations, with L, the ground truth data. Then, for each user who does not have location data and has friends with location data, the algorithm creates a vector, M, which stores a list of the friends' locations. Using this list of latitude and longitude coordinates, the algorithm predicts the user's location by calculating the geometric median of the locations in M. The new predicted locations from the first round are added to E, the new mapping from users to locations. The algorithm repeats itself by predicting additional users' locations in the second round, using the ground truth and predicted location data from the previous round. The algorithm stops when the stopping criterion is met (in this paper, three rounds of prediction).

```
Data: U, L, and N
Let E be the current mapping from user to location;
Initialize E with L;
while Convergence criteria are not met do
    Let E′ be the next mapping from user to (predicted) location;
    for u ∈ (U − domain(L)) (i.e., users who do not currently have location information) do
        Let M be a list of locations;
        for n ∈ N(u) (i.e., friends of user u) do
            if E(n) ≠ ∅ (i.e., if the friend n has location information) then
                add E(n) to M;
            end
        end
        if M ≠ ∅ (i.e., user u's friends have location information) then
            E′(u) = arg min_{x∈L} Σ_{y∈L} distance(x, y) (the predicted location of user u is the
            geometric median of her friends' locations)
        end
    end
    E = E′
end
Result: Estimated user locations, E
```

Algorithm 1: Spatial Label Propagation (Jurgens, 2013)

Figure S2 illustrates the way in which spatial label propagation algorithms work. First, location data from users who have them are used as "ground truth" to predict the locations of users to whom they are directly connected. If a user has more than one friend with ground truth data, the geometric median is calculated to predict his or her location. The geometric median is preferred over the geometric mean, as it represent the actual location of users in the network and not a meaningless average of coordinates. In addition, it is less sensitive to outliers, which might happen when users post geo-located tweets while traveling. To give a concrete example, in Panel (a) the location of user a is predicted as the geometric median of users b, d, and e.

In the second stage, after the first round of prediction is completed and new users have predicted location information, the algorithm carries out a second round of location predictions, which uses richer location data that is distributed across the network, incorporating both ground truth and predicted location data points. Panel (b) shows that in the second round, it is possible to predict the location for user c using data on the location of users a, b, and e. In the same round, the location of user a is re-estimated, using a new data point from the predicted location of user f, in addition to the location information used in the first round, from users b, d, and e. This process is repeated a fixed number of times or until a minimum proportion of users have predicted location data.[3]

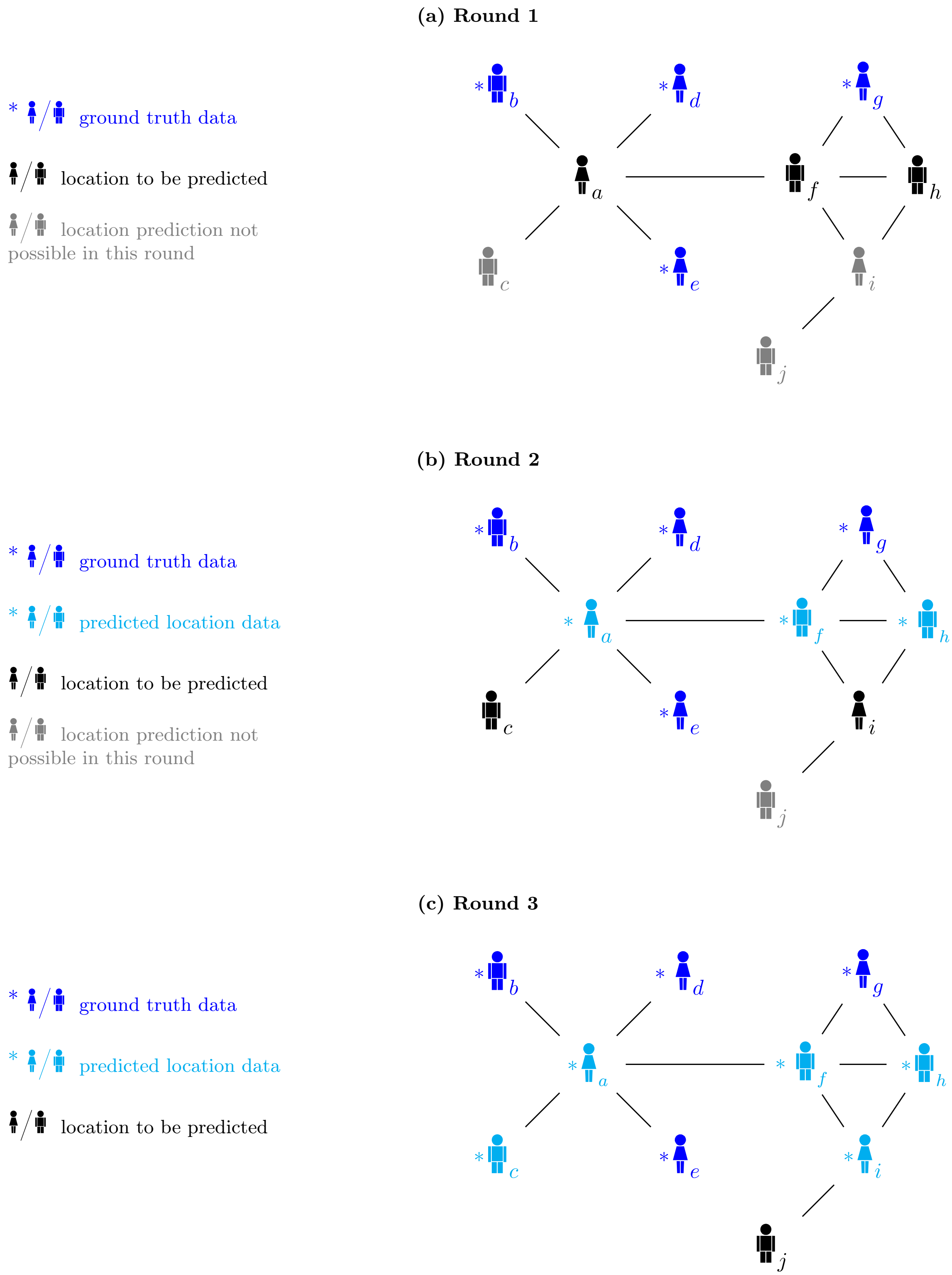
I implement a slight deviation from the procedure described in Jurgens (2013). The original algorithm is designed to operate on a random sample of tweets, and not on a deep network of users who have timeline data and full lists of friends and followers. Thus, it identifies connections between individuals on the basis of "bidirectional mentions," i.e., user A mentions user B in a tweet and vice-versa. Bidirectional mentions are used in the original algorithm as a proxy for friends on social media, as it is impractical to obtain lists of friends and followers from a random sample of tweets. However, in my database, I have actual lists of friends and followers of accounts flagged as ISIS activists. As such, while I adopt the Jurgens (2013) algorithm as-is and allow connections between individuals to be identified on the basis of bidirectional mentions, I also generate "artificial" tweets containing bidirectional mentions between activists and their followers and friends. This ensures that the network structure contained in my database will be faithfully reproduced in the spatial label propagation algorithm.

The SLP algorithm requires so-called "ground truth" data, i.e., users with a known location, to base the prediction of the location for users without a known location. I obtained ground truth data as follows. For users with at least one geolocated tweet, I used the coordinates from an arbitrarily selected geolocated tweet. For users without any geolocated tweets but with a location field in their user profile, I looked up the location using the Google Maps and/or Bing Maps APIs (the specific API is selected arbitrarily).[4] If there was a match, I used the coordinates corresponding to this location as the user's ground truth location. To be sure, both of these methods are measured with error, but there is no reason to believe that these errors are systematically biased in any specific direction. Thus, by the law of large numbers, across the total universe of accounts with ground

[3] I employed three iterations, which predicted locations for 1,676,419 users in the database.

[4] Google Maps API: `https://developers.google.com/places/web-service/details`; Bing Maps API: `https://msdn.microsoft.com/en-us/library/ff701711.aspx`.

Figure S2: Spatial Label Propagation Algorithm



truth data ($N = 287,482$), these errors should be inconsequential.

S2.2 Stability of location predictions

I verify the accuracy of the location prediction algorithm in the following way. The network structure in my database is relatively deep, centered around ISIS activists for whom I have full lists of followers, as well as friends of a subset of the followers. Thus, individuals distributed across the network with ground truth data are connected to each other mainly through the ISIS activists' accounts. This is different from flat networks studied in other SLP applications using data from random samples of tweets (Jurgens et al., 2015). As a result, cross validation using only data from accounts with ground truth information is not useful for estimating the performance of the model.

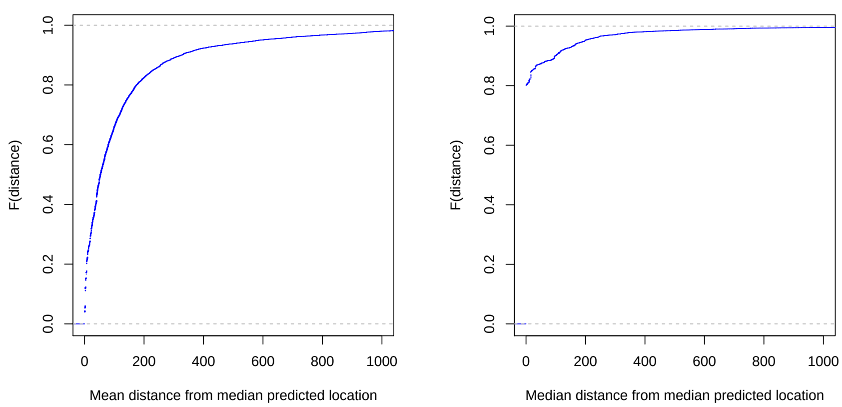
In non-network data, cross validation on the training set is useful because observations do not depend on each other. Thus, $\hat{y}_i$, the prediction for observation i, is simply some function of the covariates for unit i and some parameters: $\hat{y}_i = f(x_i, \theta)$. Taking observations out in cross validation to test the model's prediction works well, because of the limited dependency between observations. In network data, cross validation is more problematic, because observations are dependent: $\hat{y}_i = f(\sum_j y_j, \theta)$. Therefore, taking observations out in cross validation does not only change θ, the parameters of the model, but also $\sum_j y_j$, the data used to predict $\hat{y}_i$. As a result, the estimations in the cross validation are likely to be biased, with greater bias for deeper networks in which the dependency between observations is higher.

To overcome this challenge and estimate the algorithm's performance, I designed a 10-fold out-of-sample stability test. I divided the training set into ten folds, and in each fold I randomly excluded 1/10 of the ground truth data when estimating the model. The algorithm therefore ran ten times, each time using only 90% of the training data to predict the locations of all users in the dataset ($N = 1,676,419$). I assume that the out-of-sample stability of the location prediction for each user i across ten folds can proxy the algorithm's location prediction accuracy. The logic behind this assumption is that highly unstable (stable) predictions across ten different prediction exercises likely means that the prediction is not very accurate (accurate). If a given user's friends are distributed geographically in a manner that renders the prediction highly unstable when excluding a random portion of the friends, then it means that the geometric median of the friends' locations is probably not a good proxy for the user's true location. On the other hand, if leaving out friends with location data does not affect the stability of the user's predicted location, then it means that many of the user's friends are located in the same area, making prediction stable, and likely more accurate.

After obtaining ten different location predictions for each user in the dataset, I calculated, for each user i, the mean and median distance from the median location predicted for user i. Figure S3 shows the performance for the ISIS activists' accounts. Figure S4 shows the performance for the ISIS followers' accounts. The figures plot the cumulative distribution function of the location predictions' stability across ten prediction estimations. In Panel (a), the stability is calculated as the mean of the predicted locations' deviations from the median predicted location for each user across the ten folds. In Panel (b), the stability is calculated as the median of the predicted locations' deviations from the median prediction. When using the mean stability measure, the majority of users' predicted locations are stable around a radius of about 50 kilometers or less for activists, and 70 kilometers or less for followers. When using the median stability measure, for over 80% of the users locations are predicted with a median stability of 10 kilometers or less.

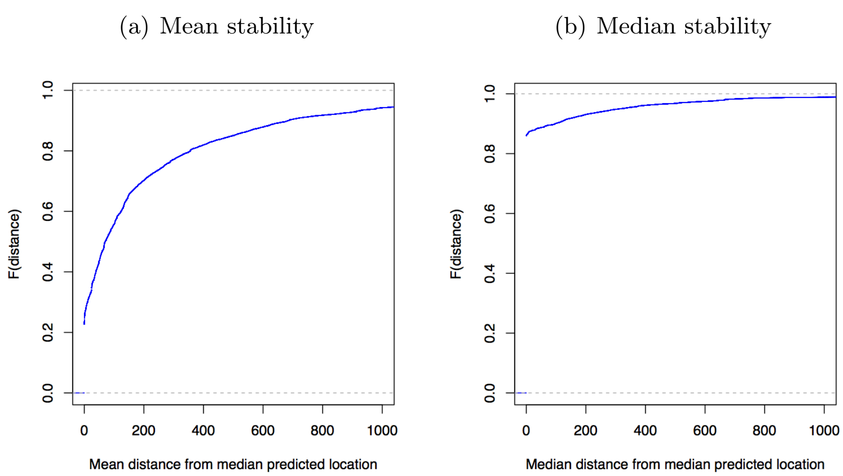
Figure S3: 10-Fold out-of-sample stability test (ISIS activists' accounts)

(a) Mean stability (b) Median stability



Note: The figure plots the cumulative distribution function of the stability of location predictions of ISIS activists across ten prediction estimations when leaving out one-tenth of the training data each time. In Panel (a), the stability is calculated as the mean of the predicted locations' deviations from the median predicted location for each user across the ten folds. The x axis shows the mean distance from the median predicted location for each user. The y axis shows the probability that mean deviation is x distance or less from the user's median predicted location. In Panel (b), the stability is calculated as the median of the predicted locations' deviations from the median prediction. When using the mean stability measure, the majority of users' predicted locations are stable around a radius of about 50 kilometers or less. When using the median stability measure, for over 80% of the users locations are predicted with a median stability of 10 kilometers or less.

Figure S4: 10-Fold out-of-sample stability test (ISIS followers' accounts)



Note: The figure plots the cumulative distribution function of the stability of location predictions of ISIS followers across ten prediction estimations when leaving out one-tenth of the training data each time. In Panel (a), the stability is calculated as the mean of the predicted locations' deviations from the median predicted location for each user across the ten folds. The x axis shows the mean distance from the median predicted location for each user. The y axis shows the probability that mean deviation is x distance or less from the user's median predicted location. In Panel (b), the stability is calculated as the median of the predicted locations' deviations from the median prediction. When using the mean stability measure, the majority of users' predicted locations are stable around a radius of about 70 kilometers or less. When using the median stability measure, for over 80% of the users locations are predicted with a median stability of 10 kilometers or less.

S2.3 Comparing the ISIS sample with a random Twitter sample

One might worry that predicting locations with the algorithm described above may not be suited for ISIS networks, as individuals in these networks are likely to be very different from ordinary citizens. While this concern is valid, and is probably true for ISIS activists that disseminate the organization's propaganda, this should not be the case for followers (who comprise over 99% of the sample). The followers are users who follow one or more ISIS activist accounts, and include a range of users, from individuals who actively support the organization, through accounts of interested citizens, to accounts that seek to counter ISIS. This means that ISIS followers are likely to be more similar to ordinary citizens than not.

To test this proposition, I obtained a random sample of Twitter users from the Twitter Streaming API, and compared it to follower and activist accounts. I used various user-level fields to examine the similarity between the samples, including the length of screen names and profile descriptions, the amount of time the accounts have been active on Twitter, whether the accounts are geo-enabled, the number of friends, followers, and twitter posts, as well as the language used by the users.

Table S4 compares the ISIS followers sample to the random Twitter sample. In most fields, ISIS followers do not significantly differ from random Twitter users: both groups have similar length of screen names, similar network sizes, and are likely to geo-enable their accounts at a similar rate. There are four fields where the samples differ: ISIS followers are more likely to have a shorter profile description, shorter statuses, are more likely to have protected accounts, and more of them have accounts set to Arabic. Overall, however, ISIS followers are not notably different from a random Twitter sample, especially in the most important field – the size of their networks.

Table S4: Balance table: ISIS followers versus a random sample

	Random sample		**ISIS followers sample**			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff.	P-value
Screen name (# characters)	10.38	2.54	10.53	2.78	-0.15	0.57
Description (# characters)	69.65	46.95	39.56	50.14	30.09	0.00***
Geo-enabled	0.34	0.48	0.26	0.44	0.08	0.12
Statuses count	38412.97	84915.98	5785.84	16758.87	32627.13	0.00***
Followers count	3677.96	12579.99	76482.71	1911304.68	-72804.75	0.23
Friends count	1769.17	7254.44	2936.38	21076.87	-1167.21	0.24
Protected	0.00	0.00	0.07	0.26	-0.07	0.00***
Account set to English	0.42	0.50	0.34	0.47	0.08	0.10
Account set to Arabic	0.11	0.31	0.44	0.50	-0.33	0.00***
Account set to French	0.07	0.26	0.07	0.26	-0.00	0.97

Table S5: Balance table: ISIS activists versus a random sample

	Random sample		**ISIS activists**			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff.	P-value
Screen name (# characters)	10.38	2.54	10.21	2.69	0.17	0.52
Description (# characters)	69.65	46.95	49.15	52.10	20.50	0.00***
Geo-enabled	0.34	0.48	0.41	0.49	-0.07	0.15
Statuses count	38412.97	84915.98	10882.06	28366.96	27530.91	0.00***
Followers count	3677.96	12579.99	11847.67	71547.36	-8169.71	0.00***
Friends count	1769.17	7254.44	3694.59	17415.86	-1925.41	0.04**
Protected	0.00	0.00	0.09	0.29	-0.09	0.00***
Account set to English	0.42	0.50	0.37	0.48	0.05	0.34
Account set to Arabic	0.11	0.31	0.42	0.49	-0.31	0.00 ***
Account set to French	0.07	0.26	0.08	0.27	-0.01	0.77

S2.4 Location prediction for individuals whose friends traveled to Syria

Another concern that may arise with the location prediction approach described above is that predictions will be biased for individuals whose friends have traveled to Syria. As the algorithm relies on the network of friends and their locations to predict geo-location, a person who has many friends that traveled to Syria is likely to be predicted to be in Syria. In the analysis in this paper, such an individual would be excluded from the sample, as this study only analyzes users whose locations are predicted to be in France, Germany, Belgium, and the UK.

It is important to note that the algorithm predicts locations by calculating the *geometric median* of the coordinates of a user's friends. Using the geometric median is crucial, since it predicts locations using the distribution of friends' actual locations. If I were to use the geometric mean — which I am not doing in this project — a user's location would be predicted to lie in places where they have no friends, or even in meaningless locations like the middle of the ocean. In addition, in the context of this study, using the geometric mean could bias the results by pulling out individuals located in cities to more rural areas where far-right parties might be more popular. This problem does not occur with the geometric median, where predicted locations are never pulled out of cities into rural areas if there are no friends in rural areas.

To visually show how this works, consider Figures S5 and S6, which display results from simulations using the geometric mean and the geometric median to predict a hypothetical user's location. In the simulation, user i is located in Paris, France, and has 100 friends. In each iteration, the distribution of user i's friends' locations changes, such that in the beginning most friends are located in Paris, and as the simulation progresses more and more move to Syria, Turkey, or Iraq. The simulation parameters are set such that out of the friends that travel abroad (whose number increases in each iteration), 60% are located in Raqqa, Syria, 30% in Mosul, Iraq, and 10% in Gaziantep, Turkey. The simulation shows what happens to the predicted location of user i as more of his or her friends travel to the Syrian civil war.

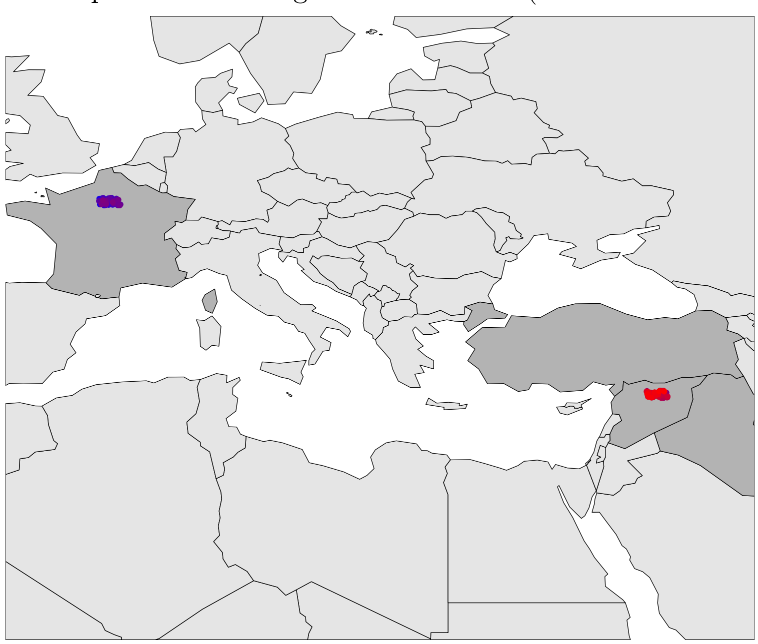
Figure S5 shows the results from the simulation using the geometric mean. Each point represents the predicted location of user i in each of the 100 simulation iterations. The color of the points changes from blue to red with each iteration, as more friends move out of Paris to Syria, Turkey or Iraq. The figure shows that the geometric mean introduces a lot of bias. In the early phases of the simulation, user i is still predicted to be in France, but having friends who traveled abroad pulls the user's predicted location out of Paris into more rural areas in France. Furthermore, as the proportion of user i's friends who travel abroad increases, user i's location is predicted to be outside of France, sometimes in arbitrary places like in the waters of the Mediterranean Sea. This example illustrates how serious the bias can be when using the geometric mean to predict a users' geo-locations.

However, we find a very different result when using the geometric median. Figure S6 shows that as more friends move out of Paris, user i's location shifts from Paris to Raqqa in Syria, but is never predicted to be in arbitrary locations outside of these two points. Specifically, with the parameters set in this simulation, user i's predicted location moves from Paris to Raqqa in the 52^{nd}

iteration, when 48 of the user's friends are in Paris, 31 in Raqqa (Syria), 15 in Mosul (Iraq), and 5 in Gaziantep (Turkey). Figure S6 shows that the blue points (which mark the earlier phases of the simulation) are located in Paris, and the red points (which mark the later phases of the simulation) are located in Raqqa in Syria. For easier visualization, in both figures I jittered the coordinates of user i's predicted location.

Figure S5: Location prediction with geometric mean (NOT used in this paper)

Figure S6: Location prediction with geometric median (the method used in this paper)



S2.5 Location prediction error and far-right vote share

Finally, one might worry that location prediction errors might have the effect of spreading out users in cities to rural areas outside of cities, where voters may be more inclined to vote for far-right parties. If measurement errors diffuse users in this way, then the results might be biased by erroneously having more users in areas with greater far-right support.

I examine this possibility in two ways. First, in Table S6 I test if there is in-sample correlation between location prediction errors and far-right vote share. Using the mean and median location prediction stability measures described in section S2.2 as the dependent variable, I estimate regressions on far-right vote share. The results show that there is no systematic relationship between far-right support and in-sample location prediction errors.

Second, I test if there is evidence of out-of-sample prediction error by examining the spatial distribution of predicted locations. As the Spatial Label Propagation algorithm predicts locations on the basis of the number of friends in each area, we would intuitively expect prediction errors to be biased in favor of cities (rather than rural areas), simply due to population size. Indeed, this can be seen in Columns (1) and (3) in Table S7, which regresses the number of users in each locality on the population and far-right support. These columns show that, as expected, more users are predicted to be located in areas with larger populations.

Nonetheless, if the effect of the prediction error is to spread out users outside of cities to rural areas where far-right parties are more popular, we would expect to observe a link between far-right vote share and the number of accounts in each location, even after adjusting for population size. There are two possible patterns that might emerge in the data. One possibility is that location prediction errors place *more users* in certain localities with high levels of far-right support. If this were the case, then we would observe a positive relationship between far-right vote share and the number of users in each locality. Another possibility is that measurement errors place users in a *greater number of areas* with high levels of far-right support. In the second scenario, we will observe a negative relationship between far-right vote share and the number of accounts, as diffusion will result in fewer users in each location.

As can be seen in Columns (2) and (3) in Table S7, the relationship between the number of users in each locality and far-right vote share is statistically insignificant. This finding holds whether population size is accounted for or not, as well as when adjusting for country fixed effects. Overall, these findings suggest that prediction errors are not spreading out users into areas with greater voting for far-right parties. In addition, it is worth recalling that the dependent variable in this study does not examine the number of users in each locality, but the correlation between the content that they produce and far-right vote share. A greater number of users in high far-right areas does not necessarily imply any such correlation.

Table S6: Location prediction errors and far-right vote share

	Mean error (km)	Median error (km)
Far-right vote share (%)	3.25	1.68
	(3.60)	(1.28)
Constant	248.22***	53.69***
	(32.68)	(12.24)
Country fixed effects	✓	✓
R^2	0.021	0.004
Number of clusters	3,136	3,136
Number of observations	116,465	116,465

Robust standard errors in parentheses, clustered at the locality level.
Base category is Belgium.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S7: Predicted locations and far-right vote share

	Dependent variable: Number of accounts		
	(1)	(2)	(3)
Population	0.01***		0.00***
	(0.00)		(0.00)
Far-right vote share (%)		-1.89	-1.43
		(1.66)	(1.76)
Constant	3.44	21.95	19.95
	(71.96)	(54.79)	(58.59)
Country fixed effects	✓	✓	✓
R^2	0.191	0.001	0.007
Number of observations	2987	3140	2803

Standard errors in parentheses. Base country is Belgium.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

S3 Classifying Twitter content

To generate the textual content outcomes in this study, I used supervised machine learning to classify tweets into several categories. The categories classified by the model included (1) Anti-West, (2) Sympathy with ISIS, (3) Life in ISIS territories, (4) Travel to Syria or foreign fighters, and (5) Syrian war. When developing my training set, I coded content into additional categories, including references to Islam (expressions of faith, Islamic quotes, and prayers and/or requests for prayers), as well as Islamophobia (content describing discrimination against Muslims). These two categories are not utilized in my analysis, which focuses on radicalization. For each of the four languages: English, Arabic, French and German, I obtained a random sample of tweets posted by ISIS activists (i.e., the accounts that have been flagged by @CtrlSec). These tweets served as a training set for a classification model. The sizes of the training sets varied by language: English ($N = 9,926$), Arabic ($N = 10,631$), French ($N = 6,158$), and German ($N = 3,011$). Each tweet was assigned one or more of the categories by three distinct Amazon Mechanical Turk and/or Crowdflower workers, and label(s) were retained for a given tweet if and only if there was "majority agreement," i.e., at least two out of the three workers assigned the same label(s) to the tweet. See Figure S11 for an example of instructions for the classification task in the Crowdflower platform.

After obtaining the training set labels, I pre-processed the tweet text as follows. For tweets in the English, French and German languages, I removed punctuation, numbers, stop words, and applied standard word stemming algorithms for each language. For tweets in the Arabic language, I similarly removed punctuation and numbers. To pre-process Arabic tweets, I used the R package arabicStemR to stem Arabic text (Nielsen, 2017). See `https://CRAN.R-project.org/package=arabicStemR` for more details.

With the pre-processed text, I generated a document-term matrix composed of unigrams and bigram tokens. That is, I obtained the frequency of individual words and two-word phrases that appeared in these tweets. I combined unigrams and bigrams in order to provide more textual structure and increase the predictive accuracy of the models. Any term included in the document-term matrix must have had appeared in at least two tweets in order to be included in the classification model. Then, I applied a term-frequency / inverse-document-frequency (tf-df) transformation to down-weight the frequency of very common phrases across the whole corpus, as is standard in automated content analysis (Ramos, 2003).

Since Twitter textual data are very noisy, and radical pro-ISIS content is rare, many tweets in the database were coded as unrelated to any of the above categories. Class proportions for each language in the training set are shown in Tables S8 – S11. To facilitate statistical prediction, I followed King and Zeng (2001), randomly over-sampling pro-ISIS tweets and randomly under-sampling unrelated tweets to obtain a class proportion of 0.5 for each of the categories, for each topic, for each language.

I trained separate logit models using the labeled rebalanced training sets for each category in each language. For all specifications, I used the the elastic-net generalized linear model (Friedman, Hastie and Tibshirani, 2010), selecting the regularization parameter λ by cross-validation to maximize the area under the ROC curve. Figures S7 – S10 show the cross-validation curves for each language and topic. The classification models for each topic and language were then employed on the full set of tweets in the database to classify each unlabeled tweet as belonging to one or more of these categories.

S3.1 Model performace

Model performance statistics from 10-fold cross validation for each topic and language are shown in Tables S12 – S11. It can be seen that the models were able to predict the content categories with high levels of in-sample accuracy. For example, for the Sympathy with ISIS topic in English, the accuracy rate is over 99.3%. This means that the misclassification rate is less than 1% for this topic and language. For the same topic in Arabic, the in-sample accuracy is 99.4%, for French it is 99.4% and for German it is 96.2%. As can be seen below, we find similar metrics for other topics and languages.

These high accuracy rates are driven by the fact that tweets labeled as these pro-ISIS topics are extremely different from tweets on other topics. The difference in content is related to the rare frequency of these categories: in the entire population of tweets, there may very well be content that has similar words and phrases to these pro-ISIS topics, but occurs so infrequently that it was not included in my training set. Those population tweets may be incorrectly classified as belonging to one of these pro-ISIS categories as a result. It is thus reasonable to suppose that my sample may contain more false positives than false negatives.

However, it is unlikely that my sample contains many such false positives because the proportion of tweets containing these topics in the sample is extremely small (see Tables S8 - S11 for an illustration of the distribution of these topics in the training set). Further, if there are a small number of false positives, there is little reason to think they would be concentrated in far-right areas. The consistency of my text-based results with non-text measures like being flagged as an ISIS activist, suspension, and the number of activist accounts followed suggests that false positives in the textual variables are not biasing my estimates.

Table S8: Class proportions by topic (English)

	0	1
Anti-West	0.984577	0.015423
Sympathy with ISIS	0.982727	0.017273
Life in ISIS territories	0.963603	0.036397
Travel to Syria or foreign fighters	0.996607	0.003393
Syrian war	0.924532	0.075468

Table S9: Class proportions by topic (Arabic)

	0	1
Anti-West	0.998104	0.001896
Sympathy with ISIS	0.996777	0.003223
Life in ISIS territories	0.996777	0.003223
Travel to Syria or foreign fighters	0.999526	0.000474
Syrian war	0.981043	0.018957

Table S10: Class proportions by topic (French)

	0	1
Anti-West	0.971370	0.028630
Sympathy with ISIS	0.965607	0.034393
Life in ISIS territories	0.965607	0.034393
Travel to Syria or foreign fighters	0.982711	0.017289
Syrian war	0.947388	0.052612

Table S11: Class proportions by topic (German)

	0	1
Anti-West	0.959585	0.040415
Sympathy with ISIS	0.932124	0.067876
Life in ISIS territories	0.915026	0.084974
Travel to Syria or foreign fighters	0.947668	0.052332
Syrian war	0.915026	0.084974

Table S12: Model performance (English)

	anti-west	is-sympathy	is-life	syria-travel-ff	syrian-war
Accuracy	0.9899	0.9868	0.9784	0.9960	0.9802
Sensitivity	0.9855	0.9781	0.9628	0.9921	0.9699
Specificity	0.9941	0.9955	0.9943	1.0000	0.9907
Pos Pred Value	0.9939	0.9954	0.9940	1.0000	0.9906
Neg Pred Value	0.9862	0.9787	0.9635	0.9920	0.9702
Precision	0.9939	0.9954	0.9940	1.0000	0.9906
Recall	0.9855	0.9781	0.9628	0.9921	0.9699
F1	0.9897	0.9867	0.9781	0.9960	0.9801
Prevalence	0.4936	0.4962	0.5019	0.5020	0.5019
Detection Rate	0.4865	0.4853	0.4831	0.4979	0.4867
Detection Prevalence	0.4895	0.4876	0.4860	0.4979	0.4914
Balanced Accuracy	0.9898	0.9868	0.9785	0.9960	0.9803

Table S13: Model performance (Arabic)

	anti-west	is-sympathy	is-life	syria-travel-ff	syrian-war
Accuracy	0.9866	0.9828	0.9928	0.9948	0.9816
Sensitivity	0.9843	0.9825	0.9855	0.9965	0.9635
Specificity	0.9889	0.9831	1.0000	0.9931	1.0000
Pos Pred Value	0.9887	0.9828	1.0000	0.9929	1.0000
Neg Pred Value	0.9846	0.9830	0.9858	0.9967	0.9643
Precision	0.9887	0.9828	1.0000	0.9929	1.0000
Recall	0.9843	0.9825	0.9855	0.9965	0.9635
F1	0.9865	0.9826	0.9927	0.9947	0.9814
Prevalence	0.4972	0.4942	0.4984	0.4925	0.5029
Detection Rate	0.4894	0.4856	0.4912	0.4908	0.4845
Detection Prevalence	0.4950	0.4941	0.4912	0.4943	0.4845
Balanced Accuracy	0.9866	0.9828	0.9928	0.9948	0.9818

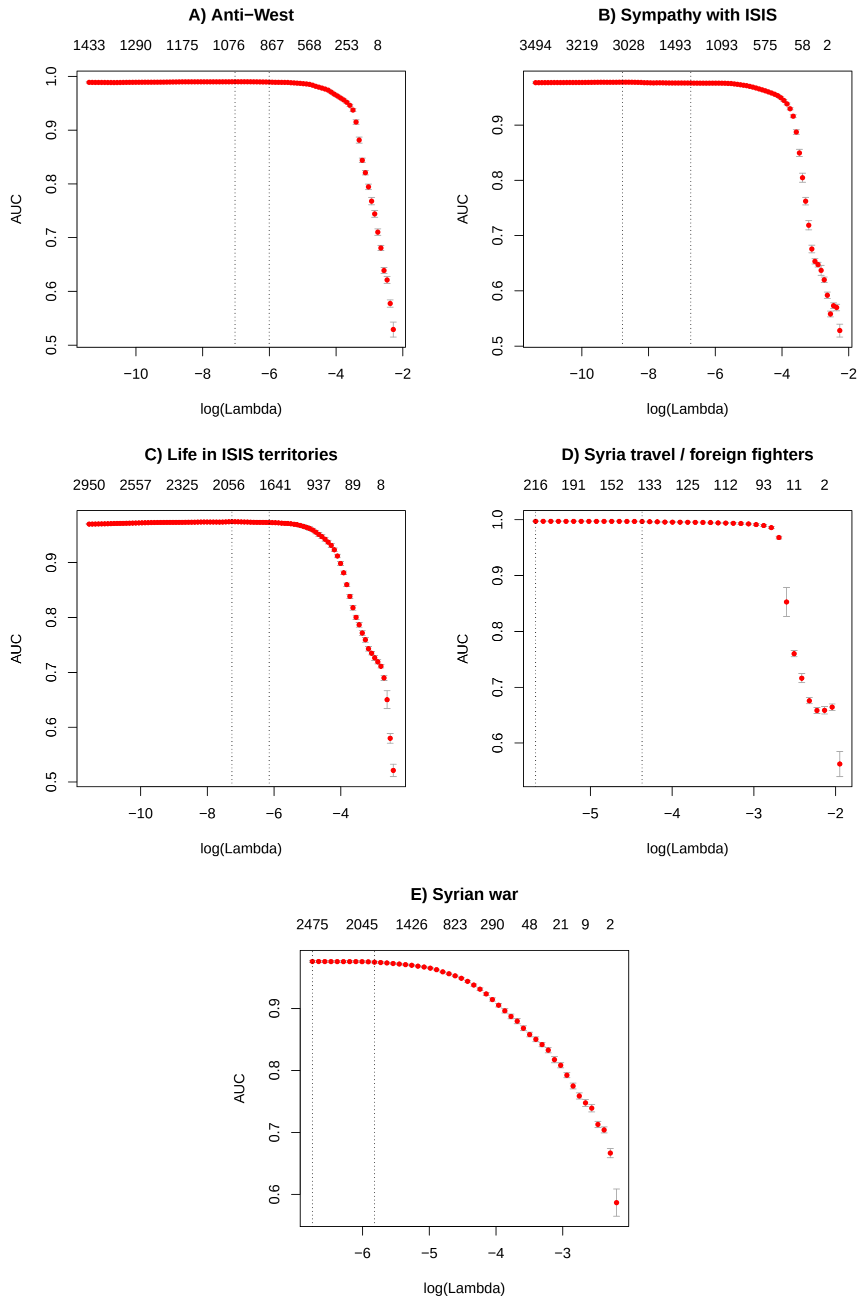
Table S14: Model performance (French)

	anti-west	is-sympathy	is-life	syria-travel-ff	syrian-war
Accuracy	0.9955	0.9948	0.9927	0.9968	0.9940
Sensitivity	0.9909	0.9933	0.9887	0.9938	0.9885
Specificity	1.0000	0.9963	0.9969	1.0000	0.9992
Pos Pred Value	1.0000	0.9963	0.9971	1.0000	0.9993
Neg Pred Value	0.9913	0.9933	0.9884	0.9936	0.9892
Precision	1.0000	0.9963	0.9971	1.0000	0.9993
Recall	0.9909	0.9933	0.9887	0.9938	0.9885
F1	0.9954	0.9948	0.9928	0.9969	0.9938
Prevalence	0.4998	0.5054	0.4993	0.5065	0.4911
Detection Rate	0.4953	0.5020	0.4935	0.5034	0.4855
Detection Prevalence	0.4953	0.5039	0.4950	0.5034	0.4858
Balanced Accuracy	0.9954	0.9948	0.9928	0.9969	0.9939

Table S15: Model performance (German)

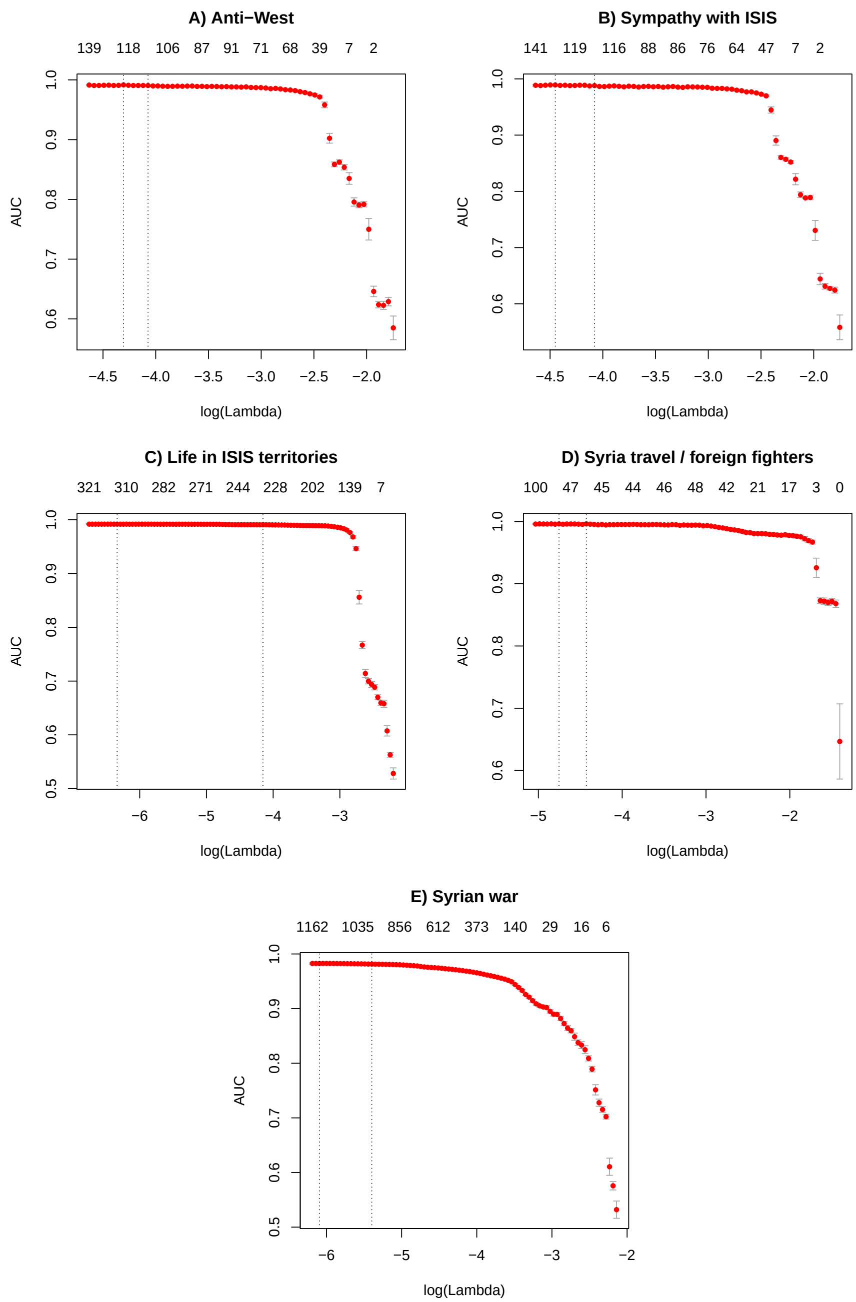
	anti-west	is-sympathy	is-life	syria-travel-ff	syrian-war
Accuracy	0.9793	0.9648	0.9710	0.9772	0.9777
Sensitivity	0.9696	0.9564	0.9693	0.9879	0.9772
Specificity	0.9896	0.9717	0.9727	0.9662	0.9775
Pos Pred Value	0.9894	0.9693	0.9711	0.9679	0.9793
Neg Pred Value	0.9688	0.9609	0.9705	0.9869	0.9775
Precision	0.9894	0.9693	0.9711	0.9679	0.9793
Recall	0.9696	0.9564	0.9693	0.9879	0.9772
F1	0.9793	0.9627	0.9701	0.9778	0.9780
Prevalence	0.5057	0.4756	0.4896	0.5150	0.4974
Detection Rate	0.4902	0.4549	0.4746	0.5088	0.4860
Detection Prevalence	0.4953	0.4694	0.4886	0.5254	0.4969
Balanced Accuracy	0.9796	0.9641	0.9710	0.9771	0.9774

Figure S7: Cross validation for model choice (English tweets)



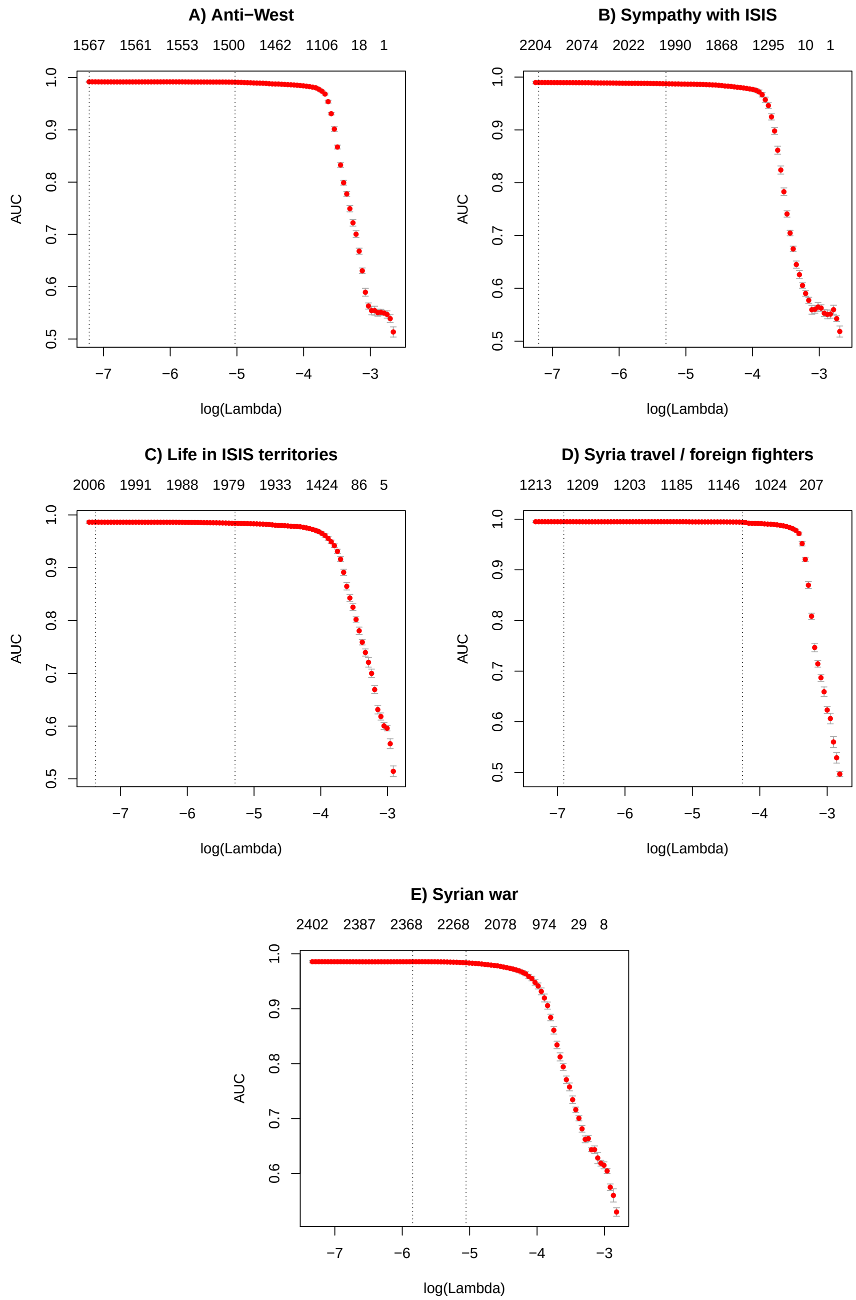
Note: The figure shows cross-validation curves for model choice in text classification of English language tweets for six topics. The cross-validation estimates for each model are shown in red dots, surrounded by error bars, plotted against the λ sequence. The y axis marks the Area Under the ROC Curve (AUC). Two selected λs are marked by vertical dotted lines. The numbers at the top of the figures represent the number of tokens (unigrams and bigrams) used in each model.

Figure S8: Cross validation for model choice (Arabic tweets)



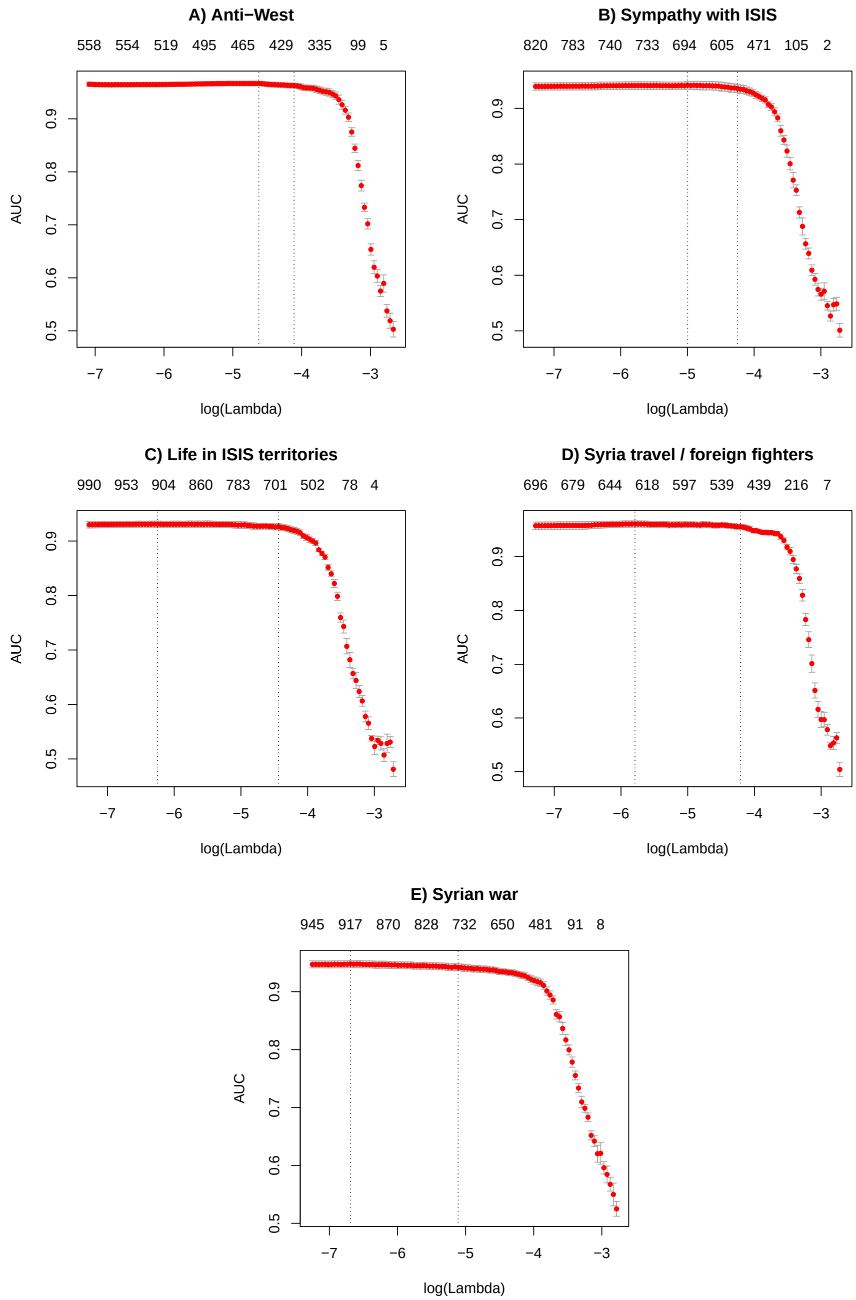
Note: The figure shows cross-validation curves for model choice in text classification of Arabic language tweets for six topics. The cross-validation estimates are shown in red dots, surrounded by error bars, plotted against the λ sequence. The y axis marks the Area Under the ROC Curve (AUC). Two selected λs are marked by vertical dotted lines. The numbers at the top of the figures represent the number of tokens (unigrams and bigrams) used in each model.

Figure S9: Cross validation for model choice (French tweets)



Note: The figure shows cross-validation curves for model choice in text classification of French language tweets for six topics. The cross-validation estimates are shown in red dots, surrounded by error bars, plotted against the λ sequence. The y axis marks the Area Under the ROC Curve (AUC). Two selected λs are marked by vertical dotted lines. The numbers at the top of the figures represent the number of tokens (unigrams and bigrams) used in each model.

Figure S10: Cross validation for model choice (German tweets)



Note: The figure shows cross-validation curves for model choice in text classification of German language tweets for six topics. The cross-validation estimates are shown in red dots, surrounded by error bars, plotted against the λ sequence. The y axis marks the Area Under the ROC Curve (AUC). Two selected λs are marked by vertical dotted lines. The numbers at the top of the figures represent the number of tokens (unigrams and bigrams) used in each model.

Figure S11: Tweet content classification task instructions for CrowdFlower workers

Classify Syrian Civil War Tweets (English)

Instructions ▴

Please label each tweet by checking all labels that correctly describe its content. If a tweet does not fit any of the labels, check "None of the Above".

Category	Description
Anti-West	Anti-West rhetoric, criticizing Western countries' foreign policy and military operations in the Middle East
Islamic faith	Expressions of faith in the Islamic religion, Islamic quotes, and prayers and/or requests for prayers
IS sympathy	Expressions of support or sympathy with the Islamic State, its ideology and its activities in territories under its control
Life in IS territories	Tweets from Islamic State activists describing their life in the territories controlled by the Islamic State; includes descriptions of daily activities under Islamic State rule, fighting; things that 'market' the life in Syria to potential foreign fighters
Travel to Syria / foreign fighters	Tweets describing interest or intent to travel to Syria, and/or discussion of foreign fighters
Syrian war	Tweets describing events in the Syrian civil war and/or discussion/analysis of those events
Islamophobia	Tweets describing unfair treatment of Muslims and/or discrimination against Muslims in non-Muslim majority countries

Islam is not a religion as Christianity/Judaism nor a political belief as Capitalism/Communism but rather it is a comple…

Classification:

- ☐ Anti-West
- ☐ Islamic faith
- ☐ IS sympathy
- ☐ Life in IS territories
- ☐ Travel to Syria / foreign fighters
- ☐ Syrian war
- ☐ Islamophobia
- ☐ None of the Above

UK extremist's sharia law photo used in free speech ad

Classification:

- ☐ Anti-West
- ☐ Islamic faith
- ☐ IS sympathy
- ☐ Life in IS territories
- ☐ Travel to Syria / foreign fighters
- ☐ Syrian war
- ☐ Islamophobia
- ☐ None of the Above

Note: This is an example of a CrowdFlower task to classify English language tweets on various dimensions. Classified tweets are included in a training set to predict the content of unclassified tweets. The classification was carried out in English, French, Arabic, and German.

Figure S12: Supervised machine learning

Random sample of tweets* ⟶ **Label tweets** ⟶ **Train models** ⟶ **Predict unlabeled tweets' content**

Label tweets ↓ 971 coders (Crowdflower, Mturk)**

Train models ↓ Using labeled tweets***

Note: * English: 9,926; Arabic: 10,631; French: 6,158; German: 3,011.
** Each tweet coded by 3 coders, label retained if there was majority agreement.
*** Over-sample pro-ISIS content, under-sample unrelated tweets.

S4 Collecting administrative data from European countries

To assign independent variables to each user in my database, I collected administrative data from France, Germany, Belgium and the United Kingdom on far-right vote share, percent unemployment, share of foreigners, population size, and additional variables described below. I matched each variable to its corresponding spatial polygon using shape files from official government databases. Then, I used Twitter users' predicted geo-location data and the shape files of local administrative areas to assign users to areas with local-level socioeconomic data. This process was done in `R`, and the code to replicate the point-to-polygon matching is available upon request.

S4.1 Far-right vote share

France I obtained data on voting results in the 2015 French Departmental Elections at the polling station level from France's open platform of public data.[5]. The data contain information on the votes for each party in each polling station, the total eligible votes, as well as the electoral canton in which each polling station is located, among other variables. I aggregated the votes for the Front National party to the electoral canton level, and then divided the raw vote total for the party by the total eligible votes in each electoral canton. I used the electoral canton level vote share because of the availability of shape files at that level.

Germany I obtained data on voting results in the 2013 Federal Elections in Germany at the constituency level from Germany's Federal Returning Officer's Office.[6] For each constituency, I calculated the percent vote share in the Second Vote for the National Democratic Party of Germany (NPD) and the Alternative for Germany (AfD) party.

United Kingdom I obtained information on the vote share of the United Kingdom Independence Party (UKIP), British Democrats, British National Party, Liberty GB party, and the National Front party in the United Kingdom's 2015 General Elections from the country's Electoral Commission website.[7] For each constituency, I calculated the percent vote share for these parties.

Belgium I downloaded voting results from the 2014 Belgian Federal Elections at the municipality level from the country's Election Board website.[8] I calculated the vote share for Vlaams Belang for each constituency.

[5] `https://www.data.gouv.fr/fr/datasets/elections-departementales-2015-resultats-par-bureaux-de-vote/`
[6] `https://www.bundeswahlleiter.de/en/bundestagswahlen/BTW_BUND_13/ergebnisse/wahlkreisergebnisse/index.html`
[7] `http://www.electoralcommission.org.uk/our-work/our-research/electoral-data`
[8] `http://www.elections.fgov.be/index.php?id=3265&L=1`

S4.2 Socioeconomic data

France I obtained data on unemployment, share of foreigners, number of asylum seeker centers, and population size from the National Institute of Statistic and Economic Studies (INSEE).

1. *Unemployment (2011).* Unemployment at the municipality level the 2011 census.[9]
2. *Share of foreigners (2011).* The share of non-nationals in each municipality from the 2011 census.[10]
3. *Asylum seekers (2014).* The number of asylum seeker centers in each municipality as of 2014.[11]
4. *Population (2011).* Population size in each municipality from the 2011 census.[12]

Germany I downloaded data on unemployment, immigration, asylum seeker benefit receivers, and population size at the municipality level from The Regional Database Germany.[13] In order to access the data, it is necessary to create an account. Thus, I provide the names of the tables that I downloaded from the database.

1. *Unemployment (2015).* Unemployed individuals by selected groups of persons (Arbeitslose nach ausgewählten Personengruppen)
2. *Share of foreigners (2014).* Immigration and emigration by gender and age groups, over municipal boundaries, yearly total (Zu- und Fortzüge nach Geschlecht und Altersgruppen, über Gemeindegrenzen, jahressumme)
3. *Asylum seeker benefits receivers (2014).* Recipients of asylum seekers standard benefits, by gender, type of service, and age groups (Empfänger von Asylbewerberregelleistungen, Geschlecht, Art der Leistung, Altersgruppen)
4. *Population size (2011).* Population size at the municipality level from the 2011 census.

United Kingdom I obtained data from the 2011 census on unemployment, immigration, population size, religion, and ethnicity at the level of the Mid-layer super output area (MSOA), which is roughly equal to the size of a neighborhood, from the United Kingdom's Office of National Statistics.[14] I provide the names and numbers of tables that I downloaded from the database.

1. *Unemployment (2011).* KS601UK – Economic activity
2. *Share of foreigners (2011).* QS803EW – length of residence in the UK
3. *Population (2011).* KS101EW – Usual resident population

[9] http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=td-population-13
[10] http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=td-nationalite-13
[11] http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=equip-serv-action-sociale
[12] http://www.insee.fr/fr/themes/detail.asp?reg_id=99&ref_id=td-population-13
[13] https://www.regionalstatistik.de/genesis/online/online;jsessionid=EE45147898822814978BE734145275C4?operation=sprachwechsel&option=en
[14] https://www.ons.gov.uk

4. *Religion (2011).* LC1202EW – Household composition by religion of Household Reference Person (HRP)

5. *Ethnic group (2011).* KS201EW – Ethnic group

Belgium I downloaded data on unemployment, immigration, and population at the statistical sector (sub-municipality) level from the 2011 Belgian census.[15] I provide the names of the tables that I downloaded from the database.

1. *Unemployment (2011).* Employed population by gender and age group - Total population - Statistical Sector (Werkende bevolking naar geslacht en leeftijdsklasse - Totale bevolking - Statistische sector)

2. *Share of foreigners, population (2011).* Population of Belgian and foreign nationality by gender – Statistical sector (Bevolking van Belgische en vreemde nationaliteit naar geslacht - Statistische sector)

S4.3 Stability of socioeconomic data over time

Since the Twitter data in this study covers content posted between 2014–2016, and the local administrative data captures socioeconomic conditions in earlier years (2011–2015), one might wonder how this gap might affect the results. As long as local-level socioeconomic data stay stable over time, the results should hold. To test the stability of these data, I collected information on every variable on which I could find over time information. Since yearly data in the relevant years is only available for France and Germany, I present results for these countries.[16] Table S16 presents the over-time correlations in unemployment and share of foreigners for each locality in France and Germany. In the analysis, I regressed each year's local data at time t on the data at time $t-1$. It can be seen that local-level socioeconomic data are highly stable over time.

Table S16: Stability of local-level socioeconomic data over time

	France		*Germany*	
	(1)	(2)	(3)	(4)
	Foreigners	Unemployment	Foreigners	Unemployment
	(2011-2014)	(2009-2014)	(2013-2015)	(2013-2015)
$t-1$	0.964***	0.607***	1.066***	0.143***
	(0.001)	(0.005)	(0.011)	(0.003)
Constant	0.002***	0.042***	0.002**	0.020***
	(0.00004)	(0.0004)	(0.001)	(0.0001)
R^2	0.958	0.291	0.326	0.102
Number of observations	106,170	35,900	20,503	20,415

Note: *p<0.1; **p<0.05; ***p<0.01

[15] http://census2011.fgov.be/download/statsect_nl.html

[16] The latest local-level socioeconomic data from the U.K. and Belgium comes from the 2011 census.

S4.4 Shape files

France I obtained shape files for the electoral cantons in France's 2015 Departmental Elections from the country's open platform of public data.[17]. For other administrative data, I obtained shape files of the contours of France's municipalities from France's open platform for public data.[18]

Germany I downloaded shape files of electoral constituencies in the 2013 German Federal Elections from Germany's Federal Returning Officer's Office.[19] For other socioeconomic variables, I used shape files from the contours of Germany's administrative boundaries.[20]

United Kingdom I obtained shape files for UK parliamentary constituencies from *MapIt*, a charity that provides data on contours of administrative areas in the United Kingdom.[21] I then matched the constituency-level vote share of far-right parties to the relevant polygon. For census data at the MSOA level, I used shape files from the Office of National Statistics.[22]

Belgium I downloaded the shape files of the contours of Belgium's statistical sectors (sub-municipality level) from Statistics Belgium, the official website of national statistics.[23]

[17] https://www.data.gouv.fr/fr/datasets/contours-des-cantons-electoraux-departementaux-2015/
[18] https://www.data.gouv.fr/fr/datasets/geofla-communes/
[19] https://www.bundeswahlleiter.de/en/bundestagswahlen/BTW_BUND_13/wahlkreiseinteilung/kartographische_darstellung.html
[20] https://www.zensus2011.de/DE/Infothek/Begleitmaterial_Ergebnisse/Begleitmaterial_node.html
[21] https://mapit.mysociety.org/areas/WMC.html
[22] http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/products/census/spatial/2011/index.html
[23] http://statbel.fgov.be/nl/statistieken/opendata/datasets/tools/geografisch/

S5 Social media usage by ISIS supporters in the United States

Table S17 provides details on the social media usage of over a hundred of individuals charged in the United States with providing material support for ISIS or plotting a violent attack on the organization's behalf. Data come from criminal complains filed against these individuals in United States courts, which describe in detail these individuals' pro-ISIS activities. I coded each case according to whether the individual used social media platforms such as Twitter or Facebook during their radicalization process. In addition, I documented whether the individual expressed publicly his or her support for the Islamic State and its ideology. Understanding whether radicalizing individual post *public* social media posts is important for this paper's data collection method, which assumes that it is possible to observe (at least part of) one's radicalization process by scraping information on his or her online behavior. The data show that the majority of these individuals used social media when radicalizing (about 62%). Among those who used social media, the vast majority (about 86%) posted publicly their support for ISIS.

Table S17: Social media usage by ISIS supporters in the United States

	Name	Location	Used social media	Posted public posts
1	Samy el-Goarany	New York	1	1
2	Ahmed Mohammed El Gammal	Arizona	1	1
3	Abdul Malik Abdul Kareem	Phoenix, AZ	0	0
4	Elton Francis Simpson	Phoenix, AZ	1	1
5	Nader Ehuzayel	Santa Ana, California	1	1
6	Muhanad Badawi	Santa Ana, California	1	1
7	Nicholas Michael Teausant	Acampo, CA	1	1
8	Adam Dandach	Orange County, CA	0	0
9	Enrique Marquez Jr.	Riverside, CA	0	0
10	Aws Mohammed Younis al-Jayab	Sacramento, CA	1	0
11	Mahamad Saeed Koadimati	San Diego, CA	1	0
12	Shannon Maureen Conley	Denver, CO	1	0
13	James Gonzalo Medina	Hollywood, FL	0	0
14	Harlem Suarez	Key West, FL	1	1
15	Gregory Hubbard	West Palm Beach, FL	1	0
16	Dayne Antani Christian	Lake Park, FL	0	0
17	Darren Arness Jackson	West Palm Beach, FL	0	0
18	Miguel Moran Diaz	Miami-Dade, FL	1	1
19	Robert B. Jackson	Pensacola, FL	1	1
20	Leon Nathan Davis	Augusta, GA	0	0
21	Hasan R. Edmonds	Aurora, IL	1	1
22	Jonas M. Edmonds	Aurora, IL	0	0
23	Mhammed Hamzah Khan	Bolingbrook, IL	0	0
24	Ramiz Zijad Hodzic	Saint Louis, MO	1	1
25	Sedina Unkic Hodzic	Saint Louis, MO	1	1
26	Nihad Rosic	Utica, NY	1	1
27	Mehida Medy Salkicevic	Schiller Park, IL	1	1
28	Armin Harcevic	Saint Louis, MO	1	1
29	Jasminka Ramic	Rockford, IL	1	1
30	Abdullah Ramo Pazara	Saint Louis, MO	1	0
31	Akrami I. Musleh	Brownsburg, IN	1	1
32	Alexander E. Blair	Topeka, KS	0	0
33	John T. Booker	Topeka, KS	1	1
34	Alexander Ciccolo	Adams, MA	1	1
35	David Wright	Everett, MA	0	0
36	Mohamed Elshinaway	Edgewood, MD	1	1
37	Khalil Abu Rayyan	Dearborn Heights, MI	1	1
38	Sebastian Gregerson	Detroit, MI	0	0
39	Al-Hamzah Mohammad Jawad	East Lansing, MI	0	0
40	Abdirizak Mohamed Warsame	Eagan, MN	0	0
41	Abdul Raheem Habil Ali-Skelton	Glencoe, MN	0	0
42	Mohamed Abdihamid Farah	Minneapolis, MN	0	0
43	Adnan Abdihamid Farah	Minneapolis, MN	1	1
44	Abdurahman Yasin Daud	Minneapolis, MN	0	0
45	Zacharia Yusuf Abdurahman	Minneapolis, MN	0	0
46	Hanad Mustafe Musse	Minneapolis, MN	0	0
47	Guled Ali Omar	Minneapolis, MN	0	0
48	Hamza Ahmed	Minneapolis, MN	1	1
49	"H.A.M"	Burnsville, MN	1	1
50	Abdullahi Yusuf	Inver Grove Heights, MN	1	1
51	Abdi Nur	Minneapolis, MN	1	1
52	Yusra Ismail	St. Paul, MN	0	0
53	Safya Roe Yassin	Bolivar, MO	1	1
54	Jaelyn Delshaun Young	Starkville, MS	1	1
55	Muhammad Oda Dakhlalla	Starkville, MS	0	0

Note: The table provides details on the social media usage d of individuals charged in the United States with providing material support for ISIS or plotting a violent attack on the organization's behalf. Data come from criminal complains filed against these individuals in United States courts, which describe in detail these individuals' pro-ISIS activities.

Social media usage by ISIS supporters in the United States

	Name	Location	Used social media	Posted public posts
56	Justin Nojan Sullivan	Burke County, NC	1	0
57	Erick Jamal Hendricks	Charlotte, NC	1	1
58	Avin Marsalis Brown	Raleigh, NC	0	0
59	Akba Johad Jordan	Raleigh, NC	0	0
60	Donald Ray Morgan	Rowan County, NC	1	1
61	Nader Saadeh	Rutherford, NJ	1	1
62	Alaa Saadeh	West New York, NJ	0	0
63	Samuel Rahamin Topaz	Fort Lee, NJ	1	1
64	Tairod Nathan Webster Pugh	Neptune, NJ	0	0
65	Sajmir Alimehmeti	Bronx, NY	1	0
66	Abdursasul Hasanovich Juraboev	Brooklyn, NY	1	1
67	Akhror Saidakhmetov	Brooklyn, NY	1	1
68	Arbor Habibov	Brooklyn, NY	0	0
69	Dilkhayot Kasimov	Brooklyn, NY	0	0
70	Almal Zakirov	Brooklyn, NY	0	0
71	Mohimanul Bhuiya	Brooklyn, NY	0	0
72	Noelle Velentzas	Queens, NY	0	0
73	Asia Siddiqui	Queens, NY	1	1
74	Arafat M. Nagi	Lackawanna, NY	1	1
75	Ali Saleh	Fort Wayne, IN	1	1
76	Munther Omar Saleh	Queens, NY	1	1
77	Emanuel L. Luchtman	Rochester, NY	1	0
78	Mufid A. Elfgeeh	Rochester, NY	1	1
79	Farred Mumuni	Staten Island, NY	0	0
80	Terrence Joseph Mcneil	Akron, OH	1	1
81	Christopher Lee Cornell	Cincinnati, OH	1	1
82	Amir Aid Abdul Rahman Al-Ghazi / Robert C. McCollum	Sheffield Lake, OH	1	1
83	Munir Abdulkader	West Chester, OH	1	1
84	Jalil Ibn Amer Aziz	Harrisburg, PA	1	1
85	Keonna Thomas	Philadelphia, PA	1	1
86	David Wright	Everett, MA	0	0
87	Nicholas Rovinski	Warwick, RI	1	1
88	Usama Rahim	Roslindale, MA	0	0
89	Michael Todd Wolfe	Houston, TX	0	0
90	Omar Faraj Saeed Al Hardan	Houston, TX	0	0
91	Asher Abid Khan	Spring, TX	1	0
92	Sixto Ramiro Garcia	Houston, TX	1	1
93	Bilal Abood	Mesquite, TZ	1	1
94	Mohamad Jamal Khweis	Alexandria, VA	1	1
95	Haris Qamar	Burke, VA	1	1
96	Nicholas Young	Fairfax, VA	0	0
97	Amine El Khalifi	Fairfax, VA	1	1
98	Yusuf Abdirizak Wehelie	Failfax, VA	0	0
99	Heather Elizabeth Coffman	Richmond, VA	1	1
100	Mohamed Bailor Jalloh	Sterling, VA	1	1
101	Ali Shukri Amin	Woodbridge, VA	1	1
102	Joseph Hassan Farrokh	Woodbridge, VA	0	0
103	Mhamoud Amin Mohamed Elhassan	Woodbridge, VA	0	0
104	Daniel Seth Franey	Montesano, WA	1	1
105	Joshua Van Haften	Madison, WI	1	1
Proportion using social media			0.62	
Proportion posting public posts (among those using social media)				0.86

Note: The table provides details on the social media usage d of individuals charged in the United States with providing material support for ISIS or plotting a violent attack on the organization's behalf. Data come from criminal complains filed against these individuals in United States courts, which describe in detail these individuals' pro-ISIS activities.

S6 Hate crimes and far-right vote share

This study proxies anti-Muslim hostility with local-level vote share for far-right parties in the United Kingdom, France, Germany, and Belgium. However, since support for the far-right is driven by various factors, such as unemployment and rising levels of immigration (Golder, 2016), it is important to examine whether locations with high far-right support have greater levels of hate towards Muslims. Empirically, this is a challenging task, since systematic local-level data on hate crimes is not publicly available in most countries. Nonetheless, local data on hate crimes are available in the United Kingdom. This section examines the relationship between far-right support, hate crimes motivated by religion, and support for ISIS in the U.K.

Using official data from the U.K. police, I matched Twitter accounts of ISIS activists and followers in the U.K. with information on hate crimes motivated by religion in each police force area,[24] as well as granular geo-spatial data on public order crimes.[25] Public order crimes include incidents that "cause fear, alarm or distress" and subsume most hate crimes in the U.K.[26] Since official police-force area data on hate crimes is reported at a very aggregate level that includes both areas with high and low support for far-right parties,[27] I use incident-level, geo-tagged data on public order crimes that are reported at more granular levels. A test of the correlation between public order crimes and religiously motivated hate crimes, at the Twitter user level, shows a very strong relationship: the correlation coefficient is 0.9 with a p-value < 0.01. This means that Twitter users in areas with higher levels of public order crimes are also located in police force areas with higher levels of hate crimes.

Tables S18 and S19 show the relationship between hate crimes, public order offneses, far-right support and and pro-ISIS discourse in the U.K. Both tables report the same specifications, but vary in the outcome variable. In Table S18, the dependent variable is a composite measure of all pro-ISIS topics: sympathy with ISIS, life in ISIS territories, foreign fighters or travel to Syria, and the Syrian war; in Table S19 the dependent variable includes only sympathy with ISIS.

Hate crimes motivated by religion. Columns (1), (3), and (4) in both tables show that users located in police force areas with greater levels of hate crimes motivated by religion significantly tweet more pro-ISIS content. This result holds even when controlling for a battery of other variables, including far-right support, unemployment, the share of foreigners, Muslims, and Arabs in each local area.

Public order offenses. Columns (2) and (3) show a very similar relationship when using public order incidents to proxy for hate crimes. Users located in local-areas (Mid-layer super output area

[24] Hate crime data in each police force area cover the years 2015-2017. See `https://www.gov.uk/government/statistics/hate-crime-england-and-wales-2015-to-2016` and `https://www.gov.uk/government/statistics/hate-crime-england-and-wales-2016-to-2017`

[25] See `https://data.police.uk/data/`

[26] See `https://www.police.uk/about-this-site/faqs/#what-do-the-crime-categories-mean`

[27] These data are reported at the police force area level; there are 45 police force areas in the UK. See `https://www.police.uk/forces/`

(MSOA), which is roughly equal to the size of a neighborhood) that have greater levels of public order offenses are also more likely to tweet more pro-ISIS content.

Far-right vote share. As found in the main paper, all models show that far-right vote share at the local level is strongly associated with posting greater pro-ISIS content. Column (4) interacts far-right vote share with the number of offenses in each local area to examine whether users located in areas with higher far-right support post greater pro-ISIS content if they are exposed to more public order crimes (which, as mentioned above, are a plausible proxy for hate crimes). Both tables show that this is the case. The interaction term *Number of offenses in local area* × *Far-right vote share* is positive and significant at the 10% level. This evidence suggests that exposure to hate crimes is a mechanism that might be driving ISIS support in areas with greater support for far-right parties. The data also show that exposure to hate crimes in and of itself has a strong relationship with pro-ISIS support, which provides further support for the hypothesis tested in this paper, that anti-Muslim hostility might be driving pro-ISIS radicalization in Europe.

Table S18: Hate crimes and pro-ISIS discourse in the UK

Dependent variable: *Number of tweets on pro-ISIS topics**	(1)	(2)	(3)	(4)
Number of hate crimes motivated by religion†	0.59***		0.60***	0.65***
	(0.18)		(0.18)	(0.19)
Far-right vote share (%)	0.49***	0.57***	0.60***	0.22
	(0.06)	(0.06)	(0.06)	(0.23)
Number of offenses in local area‡		1.30***	1.28***	0.49
		(0.25)	(0.25)	(0.52)
Number of offenses in local area x Far-right vote share				0.06*
				(0.03)
Muslims (%)	−0.05	−0.08	−0.07	−0.12**
	(0.05)	(0.05)	(0.06)	(0.06)
Arabs (%)	−0.01	0.23	0.02	0.13
	(0.26)	(0.26)	(0.27)	(0.27)
Unemployment (%)	0.29**	0.13	0.08	0.02
	(0.14)	(0.14)	(0.14)	(0.15)
Foreigners (%)	0.15*	−0.07	−0.15	−0.09
	(0.08)	(0.09)	(0.10)	(0.10)
Constant	11.05***	6.05***	3.34*	8.66**
	(1.34)	(1.83)	(2.01)	(3.66)
R^2	0.001	0.001	0.001	0.001
Observations	80,058	79,134	79,132	79,132

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
* Pro ISIS topics include tweets sympathizing with ISIS, discussing life in ISIS territories or foreign fighters, and describing the Syrian civil war.
† Hate crimes motivated by religion reflect the logged number of hate crimes reported in each police force area in the UK during 2014-15.
‡ Number of offenses in local area reflects the logged number of public order crimes, which subsume most hate crimes, in each local area (middle layer super output areas).

Table S19: Hate crimes and sympathy with ISIS in the UK

Dependent variable: *Number of tweets sympathizing with ISIS*	(1)	(2)	(3)	(4)
Number of hate crimes motivated by religion[†]	0.13^{***}		0.13^{***}	0.14^{***}
	(0.04)		(0.04)	(0.04)
Far-right vote share (%)	0.11^{***}	0.13^{***}	0.13^{***}	0.04
	(0.01)	(0.01)	(0.01)	(0.05)
Number of offenses in local area[‡]		0.28^{***}	0.28^{***}	0.10
		(0.05)	(0.05)	(0.12)
Number of offenses in local area x Far-right vote share				0.01^{*}
				(0.01)
Muslims (%)	-0.02	-0.03^{**}	-0.02^{*}	-0.03^{**}
	(0.01)	(0.01)	(0.01)	(0.01)
Arabs (%)	0.01	0.06	0.02	0.04
	(0.06)	(0.06)	(0.06)	(0.06)
Unemployment (%)	0.07^{**}	0.03	0.02	0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Foreigners (%)	0.03^{**}	-0.01	-0.03	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)
Constant (%)	2.34^{***}	1.25^{***}	0.68	1.89^{**}
	(0.30)	(0.40)	(0.44)	(0.81)
R^2	0.001	0.001	0.001	0.001
Observations	80,058	79,134	79,132	79,132

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
[†] Hate crimes motivated by religion reflect the logged number of hate crimes reported in each police force area in the UK during 2014-15.
[‡] Offenses in local area reflects the logged number of public order crimes, which subsume most hate crimes, in each local area (middle layer super output areas).

S7 Unemployment, far-right vote share, and support for ISIS on Twitter

One concern that may arise with the analysis presented in the paper is that far-right vote share and pro-ISIS rhetoric may both be driven by unemployment. While all specifications control for unemployment at the local level, this might not be enough to rule out the confounding effect of unemployment. To address this issue, I carry out several additional tests. First, as presented in the main paper, I conduct high frequency studies around events that may mobilize support for ISIS, and examine whether pro-ISIS rhetoric increases after these events more strongly in areas with higher levels of far-right vote share. The idea is that if far-right areas make people more likely to support ISIS, then we should also observe this pattern in the high frequency time dimension. The results show systematic evidence that across various events, including terrorist attacks, anti-Muslim marches, and ISIS propaganda releases, users express greater support for ISIS after these events in localities where far-right parties are more popular. In particular, when I examine heterogeneous changes following these events for both far-right vote share and unemployment (see Table S23), it is clear that these high-frequency changes are linked to the former and not the latter.

Second, I carry out a more comprehensive examination using a matching design. In the matching approach, I compare users located in areas with high and low far right vote share that are matched on levels of unemployment, the proportion of foreigners, population size, and the country in which they are located.[28] I created a binary variable for areas with high far-right support that is coded 1 when a location is at or above the median far-right vote share, and 0 otherwise. I then estimated a logistic regression of the high-far right variable on these covariates, choosing the single nearest neighbor as a control. I use propensity scores from this matching procedure as a weight in a regression comparing the difference in ISIS support between users located in areas with low and high levels of far-right vote share, as well as around events that mobilize support for ISIS.

Before discussing the results, I examine whether the matching method was able to achieve balance. Table S20 shows results from regressions of local-level far-right vote share (measured with the binary variable described above) on the covariates used in the matching. Columns (1) and (2) show that in the unbalanced model ("UB"), greater levels of unemployment are significantly correlated with high far-right vote share. This is expected, as the popularity of far-right parties in Europe is driven to a great extent by unemployment. However, this correlation disappears in the balanced model ("B") presented in Column (2). I find the same results when adding covariates to the model in Columns (3) – (6). Interestingly, in Column (7), which presents the unbalanced regression when adding country fixed effects, the relationship between unemployment and far-right vote share also goes to zero. This suggests that this model successfully accounts for the confounding effect of unemployment. As these are the covariates used in the paper's main specifications, it reduces the concern that unemployment drives the results.

[28] I use these covariates since they are available for all countries in the study.

Next, I examine the results from the matching design. Table S21 shows the relationship between far right vote share and pro-ISIS support when comparing users in high and low far-right areas. In Column (1) the variable is coded 1 for individuals who are at the top 1% of the distribution of posting pro-ISIS content, and 0 otherwise. Column (2) is measured similarly, but uses only sympathy with ISIS to measure radical content. In Columns (3) and (4) the dependent variable is a binary measure of being flagged as an ISIS activist and being suspended from Twitter, respectively. Column (5) uses the number of ISIS accounts that a user follows. Overall, the matching results show very similar findings to those found in the main paper. Moving from areas that are below the median far-right vote share to matched areas that are above the median significantly increases the probability that a user is at the top 1% posters of tweets sympathizing with ISIS, is flagged as an ISIS activist, suspended from Twitter, and follows a greater number of ISIS accounts.

Table S22 presents results from the event studies using matching. Panel A shows the impact of the events on pro-ISIS content in all areas, using data from three days before and after the events. Panel B examines whether this effect differs between areas with low and high support for far-right parties. I find that in most models, users in areas with greater far-right vote share post significantly more pro-ISIS content after terrorist attacks, the ISIS propaganda release, and the anti-Muslim marches. These results, together with the cross-sectional matched design described above, suggest that the link between far-right vote share and support for ISIS on Twitter are not driven by unemployment.

Table S20: Balance test

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High far-right vote share	UB	B	UB	B	UB	B	UB	B
Unemployment (%)	0.03^{***}	0.00	0.03^{***}	-0.00	0.03^{***}	-0.00	0.00	-0.00
	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)
Foreigners (%)			-0.01^{***}	0.00	-0.01^{***}	0.00	-0.01^{***}	0.01
			(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Population					-0.00^{***}	-0.00	0.00	-0.00
					(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.58^{***}	0.49^{***}	0.60^{***}	0.49^{***}	0.60^{***}	0.49^{***}	0.68^{***}	0.46^{***}
	(0.02)	(0.11)	(0.02)	(0.10)	(0.02)	(0.11)	(0.05)	(0.10)
Country fixed effects	✗	✗	✗	✗	✗	✗	✓	✓
R^2	0.025	0.000	0.064	0.001	0.071	0.003	0.312	0.003
Number of observations	2790	2,367	2,790	2,367	2,786	2,367	2,786	2,367

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S21: Far-right vote share and support for ISIS on Twitter (Matched design)

	(1) Top 1% radical content	(2) Top 1% sympathy with ISIS only	(3) Flagged as an ISIS activist	(4) Suspended from Twitter	(5) Number of ISIS accounts following
High far-right = 1	1.97	3.75***	2.16*	10.16***	1.86***
	(1.29)	(1.29)	(1.21)	(3.78)	(0.55)
Constant	8.69***	7.48***	2.90***	33.77***	4.02***
	(1.15)	(1.14)	(0.78)	(3.04)	(0.27)
R^2	0.000	0.000	0.000	0.001	0.002
Number of clusters	2,367	2,367	2,367	2,367	2,367
Number of observations	157,873	157,873	157,873	157,872	157,872

Standard errors in parentheses
Coefficients in columns 1-4 are $\times$ 1,000 to account for the skewed distribution of the dependent variables.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S22: Events and changes in pro-ISIS rhetoric (matched design)

	Paris attacks		**Brussels attacks**		**ISIS propaganda release**		**PEGIDA marches**		
	(1) Sympathy with ISIS	(2) ISIS topics	(3) Sympathy with ISIS	(4) ISIS topics	(5) Sympathy with ISIS	(6) ISIS topics	(7) Sympathy with ISIS	(8) ISIS topics	(9) ISIS topics + Anti-West
A. Changes in pro-ISIS content (standard deviation units)									
After event = 1	0.146***	0.118***	0.022*	0.024***	0.017	0.063**	0.008	-0.020	-0.011
	(0.030)	(0.034)	(0.013)	(0.009)	(0.025)	(0.031)	(0.010)	(0.014)	(0.011)
Constant	0.081	-0.082	0.166	0.002	-0.641***	-0.478***	-0.059	0.316***	0.283***
	(0.156)	(0.069)	(0.171)	(0.113)	(0.222)	(0.029)	(0.050)	(0.064)	(0.083)
R^2	0.011	0.008	0.002	0.003	0.001	0.005	0.004	0.001	0.002
Number of clusters	293	293	445	445	113	113	417	417	417
Number of observations	32,451	32,451	60,224	60,224	5,016	5,016	50,874	50,874	50,874
B. Changes in pro-ISIS content (standard deviation units), by far-right support									
After event = 1	0.016	0.070	0.041*	-0.041	-0.128*	-0.074	0.015	-0.030	-0.030
	(0.031)	(0.062)	(0.021)	(0.037)	(0.075)	(0.100)	(0.021)	(0.028)	(0.020)
Far-right vote share (%)	-0.003	-0.006*	0.005**	-0.002	-0.007	-0.009	0.003*	0.001	0.001
	(0.002)	(0.003)	(0.002)	(0.002)	(0.008)	(0.006)	(0.002)	(0.002)	(0.002)
After event = 1 × Far-right vote share (%)	0.004**	0.004	-0.001	0.004**	0.011**	0.004	-0.000	0.002	0.002*
	(0.002)	(0.003)	(0.001)	(0.002)	(0.004)	(0.005)	(0.002)	(0.001)	(0.001)
Constant	0.028	-0.172	0.141	0.452***	-0.400***	-0.312***	-0.371***	0.061	-0.029
	(0.202)	(0.379)	(0.216)	(0.101)	(0.108)	(0.110)	(0.052)	(0.194)	(0.193)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.007	0.011	0.002	0.002	0.008	0.008	0.004	0.001	0.003
Number of clusters	281	281	428	428	110	110	405	405	405
Number of observations	19,235	19,235	40,419	40,419	2,777	2,777	34,391	34,391	34,391

Standard errors in parentheses, clustered by location. Base country is Belgium.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

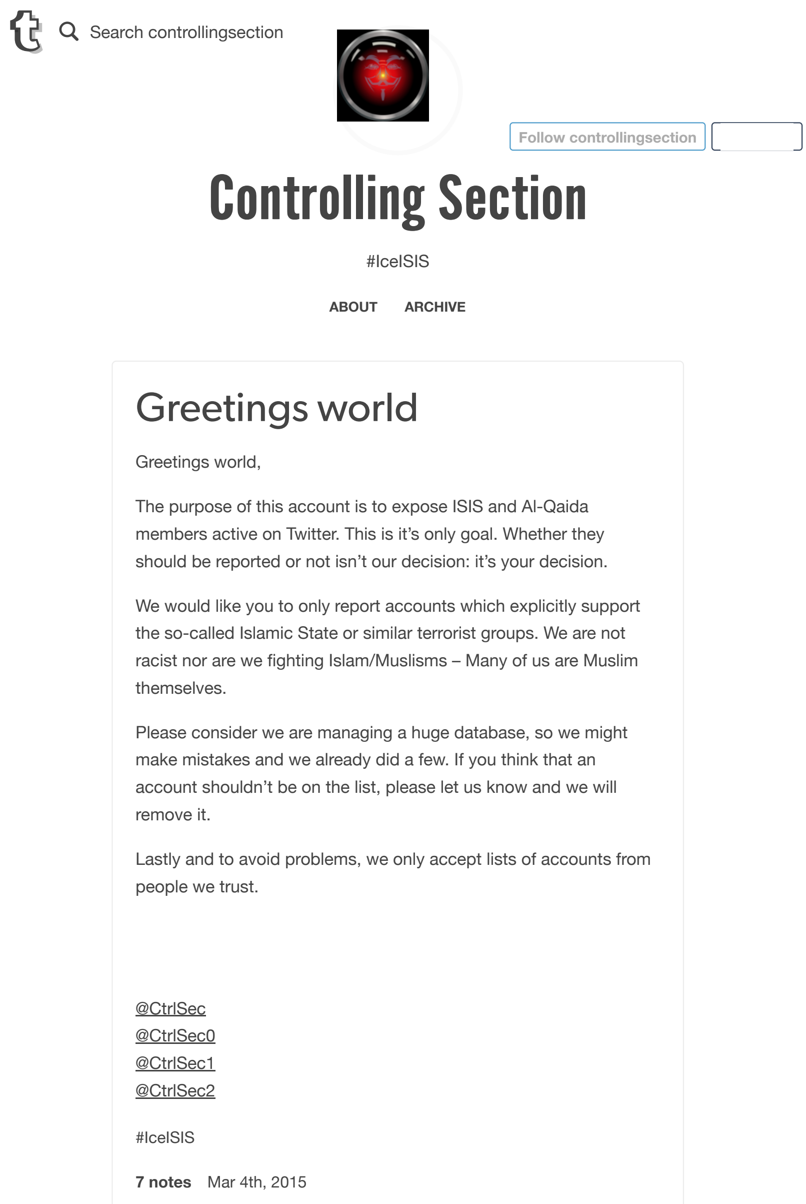
Table S23: Events and changes in pro-ISIS rhetoric, by far-right vote share and unemployment

	Paris attacks		Brussels attacks		ISIS propaganda release		PEGIDA marches		
	(1) Sympathy with ISIS	(2) ISIS topics	(3) Sympathy with ISIS	(4) ISIS topics	(5) Sympathy with ISIS	(6) ISIS topics	(7) Sympathy with ISIS	(8) ISIS topics	(9) ISIS topics + Anti-West
Changes in pro-ISIS content (standard deviation units), by far-right vote share and unemployment									
After event = 1	-0.006	-0.008	0.071***	-0.009	-0.069	-0.085	-0.019	-0.039	-0.053*
	(0.041)	(0.043)	(0.020)	(0.024)	(0.078)	(0.088)	(0.028)	(0.028)	(0.027)
Unemployment (%)	-0.018**	-0.010	0.006*	-0.003	0.012	-0.009	-0.003	-0.006	-0.004
	(0.008)	(0.009)	(0.004)	(0.004)	(0.017)	(0.015)	(0.005)	(0.004)	(0.005)
After event = 1 × Unemployment (%)	0.012	0.015*	-0.010***	-0.000	-0.008	0.005	0.005	0.001	0.003
	(0.009)	(0.008)	(0.004)	(0.005)	(0.019)	(0.011)	(0.006)	(0.004)	(0.004)
Far-right vote share (%)	-0.002	-0.004*	0.002	-0.000	-0.009	-0.012**	0.001	0.000	0.000
	(0.002)	(0.002)	(0.001)	(0.002)	(0.006)	(0.005)	(0.002)	(0.001)	(0.001)
After event = 1 × Far-right vote share (%)	0.002	0.003*	-0.000	0.003**	0.010**	0.004	0.000	0.002	0.002**
	(0.002)	(0.002)	(0.001)	(0.001)	(0.005)	(0.004)	(0.002)	(0.001)	(0.001)
Constant	0.279**	0.384***	-0.055	0.086	-0.354	0.048	0.106	0.138*	0.163*
	(0.132)	(0.133)	(0.072)	(0.074)	(0.223)	(0.200)	(0.099)	(0.078)	(0.085)
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.006	0.006	0.002	0.002	0.010	0.008	0.003	0.001	0.002
Number of clusters	362	362	529	529	140	140	508	508	508
Number of observations	21,459	21,459	46,460	46,460	3,216	3,216	38,527	38,527	38,527

Standard errors in parentheses, clustered by location. Base country is Belgium.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

S8 Additional figures

Figure S13: @CtrlSec request to expose ISIS members on Twitter



Search controllingsection

Follow controllingsection

Controlling Section

#IceISIS

ABOUT ARCHIVE

Greetings world

Greetings world,

The purpose of this account is to expose ISIS and Al-Qaida members active on Twitter. This is it's only goal. Whether they should be reported or not isn't our decision: it's your decision.

We would like you to only report accounts which explicitly support the so-called Islamic State or similar terrorist groups. We are not racist nor are we fighting Islam/Muslisms – Many of us are Muslim themselves.

Please consider we are managing a huge database, so we might make mistakes and we already did a few. If you think that an account shouldn't be on the list, please let us know and we will remove it.

Lastly and to avoid problems, we only accept lists of accounts from people we trust.

@CtrlSec
@CtrlSec0
@CtrlSec1
@CtrlSec2

#IceISIS

7 notes Mar 4th, 2015

Source: `http://controllingsection.tumblr.com/post/112703617620/greetings-world`

Figure S14: Example of @CtrlSec real-time flagging of ISIS acounts

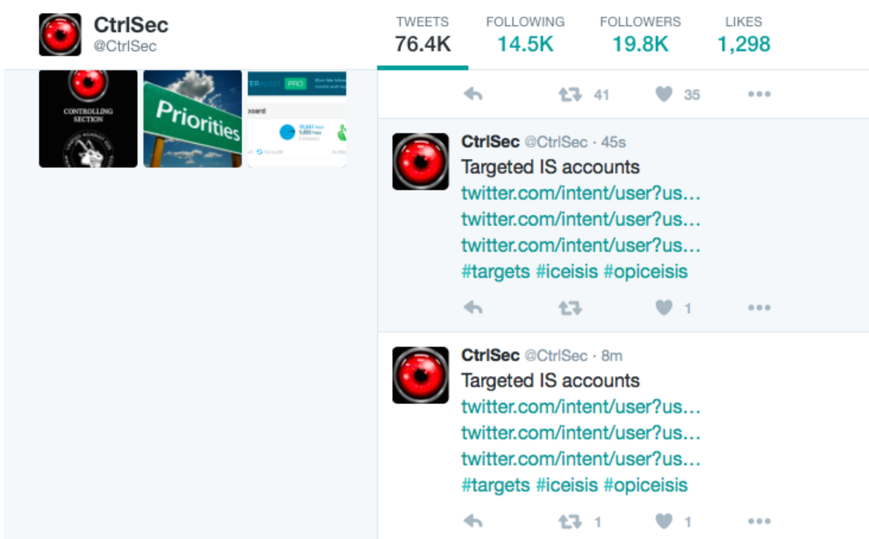
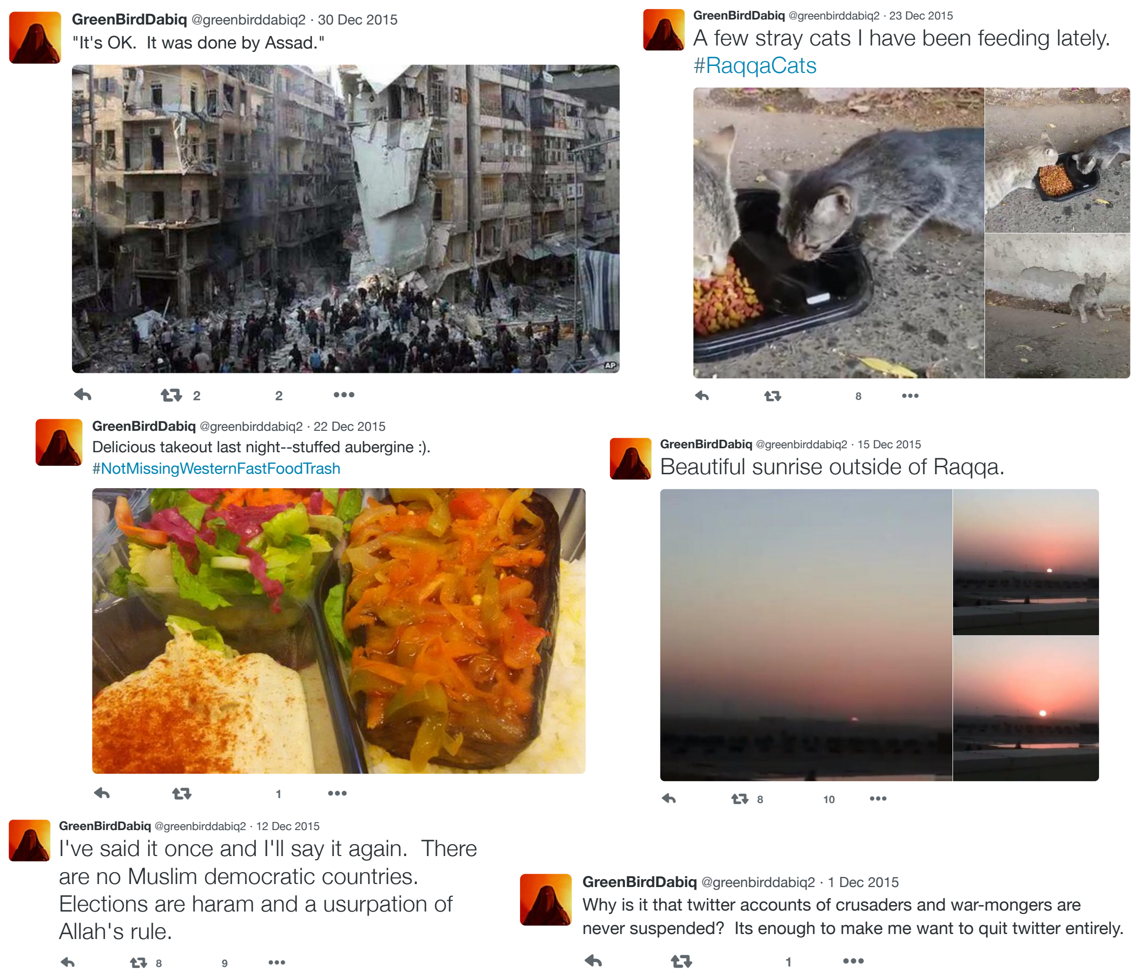


Figure S15: Example of ISIS accounts

Figure S16: Example of a Western fighter tweeting from Syria



Note: This account has already been suspended as of February 2016.

Figure S17: Example of a suspended account

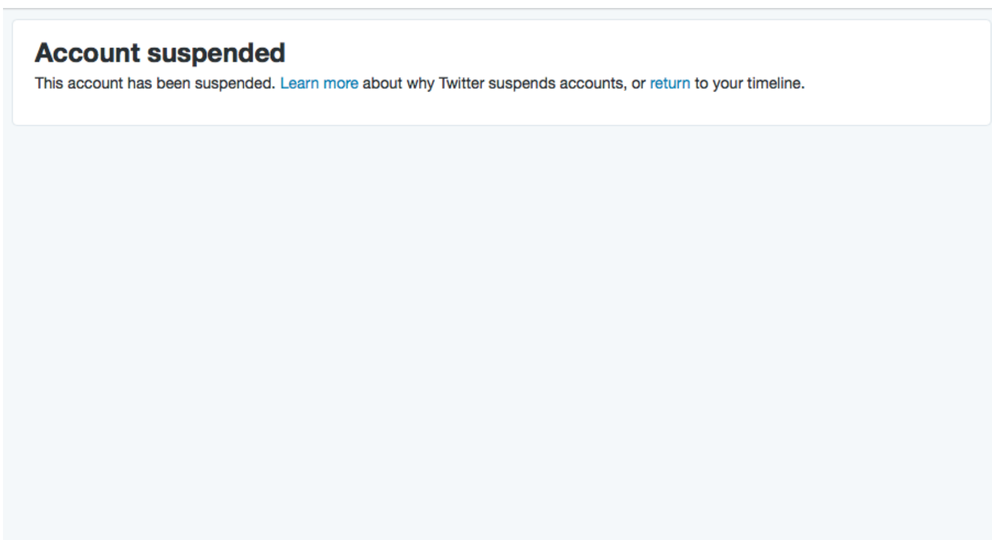
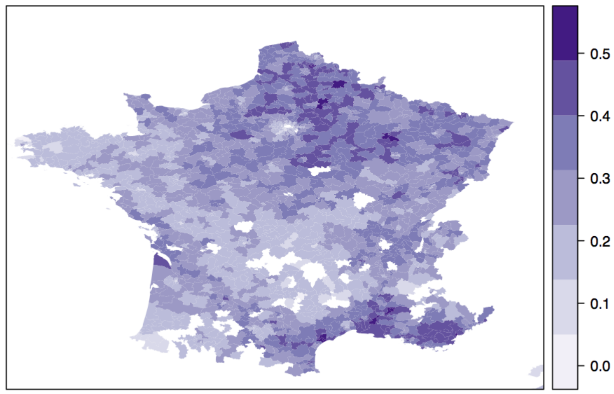
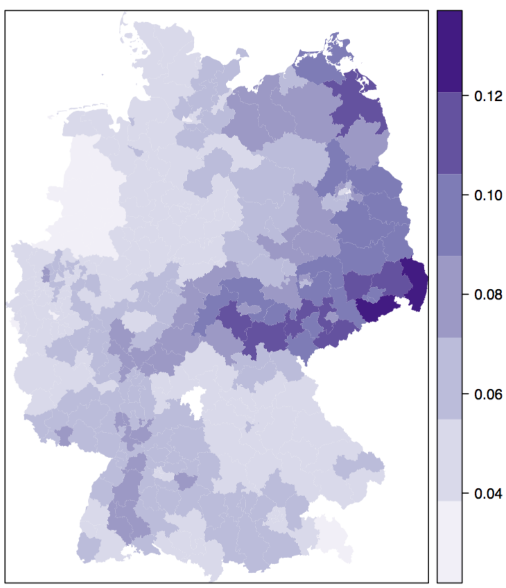


Figure S18: Vote share for far-right parties

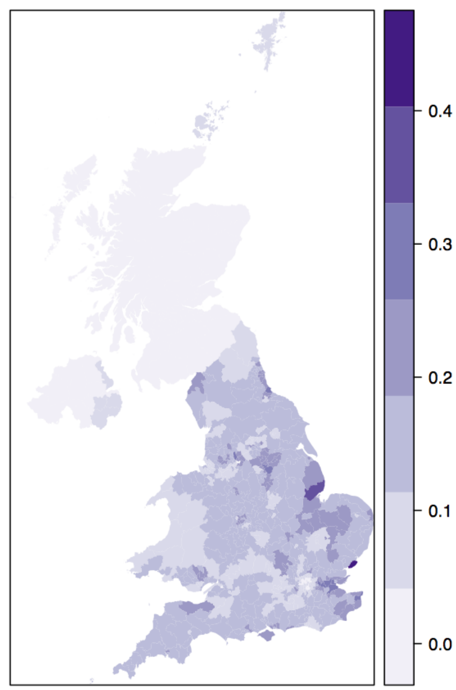
(a) France



(b) Germany

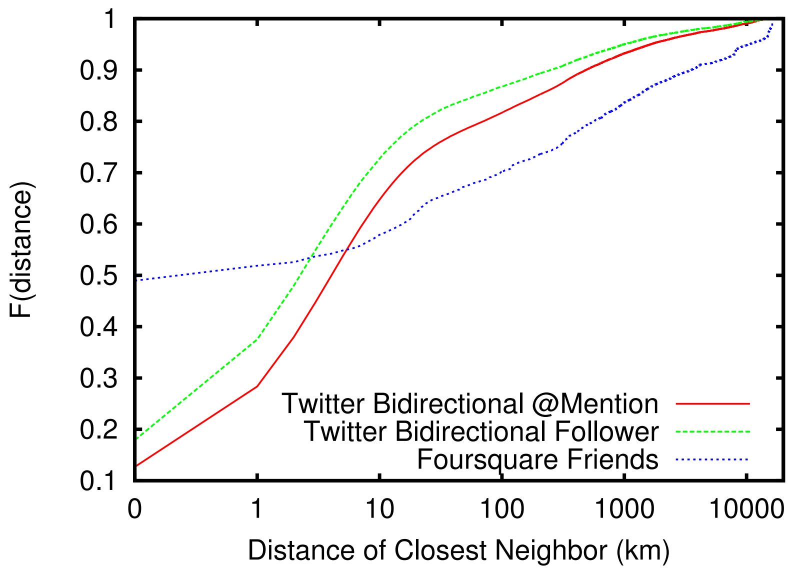


(c) United Kingdom



Note: For France, the map displays the vote share for the Front National party in the 2015 departmental elections at the electoral canton level. For Germany, the map displays the vote share for the Alternative for Germany (AfD) party and the National Democratic Party (NPD) in the 2013 federal elections. For the UK, the map represents the vote share for the British Democrats, British National Party, Liberty GB party, National Front party, and United Kingdom Independence Party in the 2015 UK parliamentary general elections.

Figure S19: The cumulative distribution functions for the distance to a user's geographically closest friend (Figure taken from Jurgens (2013))

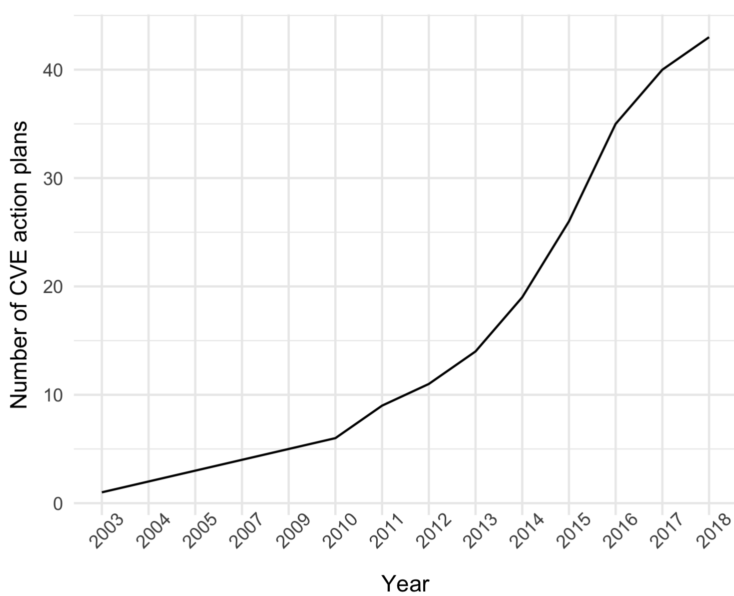


Note: The figure, taken from the study of Jurgens (2013), shows cumulative distribution functions (CDFs) of users' geographical distance to their closest neighbor in three social media networks. In the figure, the x axis shows distance in kilometers, and the y axis shows the probability that the closest neighbor for each user is located x distance or less from that user. It can be seen that more than half of the users in these three networks had neighbors that were located within 4 kilometers from them, thereby allowing location prediction within 4-kilometer bounds.

Figure S20: Anti-Muslim marches organized by PEGIDA across Europe

Note: Photos credit: *Radio Free Europe Radio Liberty* (2016) and Malm (2015)

Figure S21: National action plans to counter violent extremism



Note: The figure presents the number of official national action plans to counter violent extremism by year. National action plans to counter violent extremism are official policies adopted by countries, and are reflected in formal documents collected by the author. It can be seen that official strategies to counter extremism have dramatically increased in recent years.

S9 Additional results

Table S24: Western foreign fighters and online radicalization by country

	(1) Number of Twitter users flagged as ISIS activists	(2) Number of Twitter users posting highly radical content	(3) Number of ISIS accounts followed	(4) Number of Twitter users suspended from Twitter
Number of foreign fighters (official count)	0.135***	0.159	79.253***	0.305***
	(0.030)	(0.133)	(20.192)	(0.107)
Constant	11.916	3.446	4,734.094	42.960
	(15.882)	(71.194)	(10,773.890)	(56.946)
Population controls	✓	✓	✓	✓
R^2	0.396	0.336	0.434	0.360
Number of observations	46	46	46	46

Note: The table reports the correlation between online radicalization measures and foreign fighter counts in European countries, controlling for population size. It can be seen that all online radicalization variables are positively correlated with the number of foreign fighters in each country, with the number of users flagged as ISIS activists, number of ISIS accounts followed, and the number of users suspended from Twitter significant at the 5% level.
*p<0.1; **p<0.05; ***p<0.01

Table S25: Different cutoffs for classifying top posters of radical content

	(1) Top 5%	(2) Top 10%	(3) Top 15%	(4) Top 20%	(5) Top 25%
Far-right vote share (%)	0.81**	0.88*	1.00*	1.63*	1.99
	(0.35)	(0.51)	(0.60)	(0.88)	(1.25)
Unemployment (%)	1.19	3.15**	3.66**	5.38**	7.66**
	(0.75)	(1.27)	(1.60)	(2.54)	(3.49)
Foreigners (%)	0.40	0.48	-0.05	-0.68	-1.03
	(0.29)	(0.46)	(0.54)	(0.77)	(1.06)
Constant	45.45***	75.29***	149.63***	215.52***	281.94***
	(16.97)	(23.13)	(28.12)	(40.00)	(55.40)
Population controls	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓
R^2	0.001	0.002	0.001	0.002	0.003
Number of clusters	2,654	2,654	2,654	2,654	2,654
Number of observations	112,253	112,253	112,253	112,253	112,253

Robust standard errors in parentheses, clustered at the locality level.
Base country is Belgium.
All coefficients are × 1,000 to account for the skewed distribution of the DV.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S26: Correlates of activists

	(1) Top 1% radical content	(2) Top 1% sympathy with ISIS only	(3) Suspended from Twitter	(4) Number of ISIS accounts following
Flagged as an IS activist	0.15***	0.13***	0.45***	128.22***
	(0.01)	(0.02)	(0.04)	(15.25)
Constant	0.01**	0.01	0.04***	2.37
	(0.00)	(0.00)	(0.01)	(3.11)
Controls	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓
R^2	0.010	0.009	0.028	0.131
Number of clusters	2,654	2,654	2,653	2,653
Number of observations	112,253	112,253	112,249	112,249

Standard errors in parentheses, clustered at the locality level. Base country is Belgium.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents the relationship between various radicalization outcomes and being an ISIS activist on Twitter. The regressions control for local-level vote share for far-right parties, unemployment, the share of foreigners, and population size, and include country fixed effects. It can be seen that ISIS activists on Twitter are significantly more likely to show signs of radicalization, when compared to ISIS followers.

Table S27: Far-right vote share and support for ISIS on Twitter

	(1) Top 1% radical content	(2) Top 1% sympathy with ISIS only	(3) Flagged as an ISIS activist	(4) Suspended from Twitter	(5) Number of ISIS accounts following
Far-right vote share (%)	0.25***	0.20***	0.30***	0.09	0.09***
	(0.06)	(0.07)	(0.04)	(0.13)	(0.02)
Unemployment (%)	0.25	0.23	-0.20*	-1.24***	-0.11***
	(0.17)	(0.17)	(0.12)	(0.32)	(0.03)
Foreigners (%)	0.11*	0.14**	0.26***	-0.06	0.08***
	(0.06)	(0.06)	(0.04)	(0.12)	(0.02)
Constant	7.89**	4.35	-9.78***	35.10***	1.12
	(3.74)	(3.48)	(1.91)	(6.70)	(0.74)
Population controls	✓	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓	✓
R^2	0.000	0.000	0.006	0.002	0.006
Number of observations	112,253	112,253	112,253	112,249	112,249

Robust standard errors in parentheses. Base category is Belgium.

Coefficients in columns 1– 4 are × 1,000 to account for the skewed distribution of the dependent variables.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S28: Far-right vote share and posting pro-ISIS content on Twitter

	(1) Sympathy with ISIS	(2) ISIS life/ Foreign fighters	(3) Syrian war	(4) Anti-West
Far-right vote share (%)	0.05***	0.09***	0.07***	0.04***
	(0.01)	(0.02)	(0.01)	(0.01)
Unemployment (%)	0.12***	0.24***	0.15***	0.13***
	(0.02)	(0.04)	(0.03)	(0.02)
Foreigners (%)	0.02***	0.04**	0.03***	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)
Constant	3.54***	7.32***	5.79***	3.21***
	(0.43)	(0.82)	(0.61)	(0.42)
Population controls	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓
R^2	0.001	0.002	0.002	0.003
Number of observations	112,253	112,253	112,253	112,253

Robust standard errors in parentheses. Base country is Belgium.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table S29: Unemployed immigrants, asylum seekers and support for ISIS on Twitter

	(1) Top 1% radical content	(2) Flagged as an ISIS activist	(3) Suspended from Twitter	(4) Number of ISIS accounts following
Far-right vote share (%)	0.24**	0.52***	0.62***	0.23***
	(0.09)	(0.07)	(0.18)	(0.02)
Unemployed immigrants (%)	0.70*	0.39	0.09	0.36***
	(0.42)	(0.26)	(0.77)	(0.09)
Asylum seekers (%, sd units)	-0.40	-11.77***	-14.21***	-2.62***
	(0.93)	(1.15)	(1.87)	(0.23)
Constant	-4.27	-63.72***	-41.47***	-14.52***
	(5.81)	(6.86)	(12.03)	(2.93)
Population controls	✓	✓	✓	✓
Country fixed effects	✓	✓	✓	✓
R^2	0.001	0.012	0.003	0.005
Number of observations	30,373	30,373	30,372	30,372

Robust standard errors in parentheses. Data available only for France and Germany. Base category is Germany.

Coefficients in columns 1-3 are $\times$ 1,000 to account for the skewed distribution of the dependent variables.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Friedman, Jerome, Trevor Hastie and Robert Tibshirani. 2010. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software* 33(1):1–22.
URL: *http://www.jstatsoft.org/v33/i01/*

Golder, Matthew. 2016. “Far Right Parties in Europe.” *Annual Review of Political Science* 19(1).

Jurgens, David. 2013. “That’s What Friends Are For: Inferring Location in Online Social Media Platforms Based on Social Relationships.” *ICWSM* 13:273–282.

Jurgens, David, Tyler Finethy, James McCorriston, Yi Tian Xu and Derek Ruths. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. In *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

King, Gary and Langche Zeng. 2001. “Logistic regression in rare events data.” *Political analysis* 9(2):137–163.

Malm, Sara. 2015. “Muslim is stabbed to death while shopping in Dresden, where thousands have gathered for right wing rallies.” Daily Mail.
URL: *http://www.dailymail.co.uk/news/article-2909950/PEGIDAs-anti-Muslim-calls-shake-German-politics.html*

Nielsen, Rich. 2017. “arabicStemR: Arabic Stemmer for Text Analysis.”.
URL: *https://cran.r-project.org/web/packages/arabicStemR/index.html*

Radio Free Europe Radio Liberty. 2016. “Prague Refugee Center Fire-Bombed After Anti-Muslim Protests.”.
URL: *http://www.rferl.org/content/pegida-refugee-crisis-protests-prague-attack/27536604.html*

Ramos, Juan. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*.