Threat Network Detection and Tracking

Olga Simek, Danelle Shah, John Passarelli, Andrew Heier
MIT Lincoln Laboratory, Lexington, MA
PUB N1C01

Abstract

Identifying and profiling threat actors are high priority tasks for a number of governmental organizations. These threat actors may operate actively, using the Internet to promote propaganda, recruit new members, or exert command and control over their networks. Alternatively, threat actors may operate passively, demonstrating operational security awareness online while using their Internet presence to gather information they need to pose an offline physical threat. This paper presents a flexible new prototype that allows analysts to automatically detect, monitor and characterize threat actors and their networks using publicly available information. It fills a need in the intelligence community for a capability to automate manual construction and analysis of online threat networks.

We adapt graph sampling approaches from literature in order to perform targeted data collection of jihadi accounts and their networks on Twitter. We design workflow and analytics for tracking network changes over time. We develop and incorporate into the prototype new approaches for threat network role classification and radicalization detection using insights from social science studies of extremism. In addition, we integrate into the prototype several novel machine learning algorithms for threat actor discovery and characterization, namely classification of user posts into discourse categories, user post summaries and gender prediction.

1. Introduction

Many Department of Defense (DoD) organizations need to maintain situational awareness of hundreds if not thousands of cities, communities and networks. Often publicly available information is the primary source for this intelligence. Identifying and profiling threat actors are high priority tasks for a number of DoD organizations. These threat actors may operate actively, using the Internet to promote propaganda, recruit new members, or exert command and control over their networks. For instance, terrorist organizations actively engage in global radicalization and recruitment on social media, targeting people of diverse backgrounds. Alternatively, threat actors may operate passively, demonstrating operational security (OPSEC) awareness online while using their Internet presence to gather information they need to pose an offline physical threat to U.S. forces and interests.

To date, no capability exists to summarize online activity and to detect, follow, and profile these actors at scale. Instead, analysts are limited to manual online browsing, manual construction and analysis of threat networks, and exploitation tools designed to summarize user-volunteered profiles and network links rather than to uncover hidden behavior.

In our work, we address these deficiencies by developing analytics and incorporating them into a prototype, which enables the analysts to perform the following tasks:

- Automatically collect targeted threat network social media data and other publicly available information (PAI)
- Automatically monitor network changes over time
- Detect susceptibility to radicalization
- Detect and classify threat network roles and activities
- Identify users from difficult-to-obfuscate online signatures relating to temporal, linguistic, geographic, and social network patterns

This work is sponsored by the Department of Defense under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

- Discover additional related accounts on other data platforms
- Glean information about threat actors from OPSEC failures among individuals in their operational or social networks

The rest of the paper is organized as follows: Section 2 reviews related work, Section 3 describes our prototype for PAI exploitation, Section 4 explains our targeted data collection approach, our threat actor discovery analytics are presented in Section 5, Section 6 discusses automated dual account detection, Section 7 presents our network role detection algorithms and Section 8 concludes.

2. Related Work

There are multiple studies focused on ISIS users and their social networks. One of the initial studies characterizes the volume, behavioral characteristics, and organization of Twitter ISIS users [1]. A follow-on study also co-authored by Berger found that the reach of ISIS had been somewhat curtailed by the beginning of 2016 due to the efforts of Twitter to suspend accounts with radical content [2].

Our dual accounts detection approach is similar to [3; 4]. We design automated workflow meant to be integrated with the automatic data collection. The specific parameters (e.g., number of returned prioritized search results) for this workflow would need to be customized to particular end user's data access and computational resources. We also discuss advantages and disadvantages of different strategies for account acquisition.

Several papers have also studied approaches for identifying extremist content in radical extremist groups. In [5] the authors apply supervised learning techniques and natural language processing (NLP) to automatically identify forum posts used for recruitment of new violent extremist members. Similar work in [6] uses machine learning to identify extremism-promoting content.

In [7], the author identifies and discusses roles of women in ISIS and Al-Queda networks. Rowe and Saif in [8] identify specific terms indicative of support versus opposition to ISIS and use them to investigate how users develop to exhibit signals of pro-ISIS behavior. Chatfield [9] and others use community detection to help identify topics and functions of individual communities in Islamic state supporters Twitter networks. In contrast, we use supervised learning and rule-based approaches to develop algorithms to identify specific roles of interest.

Several publications on radicalization trajectories are related to our work. Silber and Bhatt [10] identify four stages of radicalization and Klausen et. al [11] applies their model to study convicted homegrown American terrorism offenders inspired by Al Qaedas ideology. In [12; 13], the authors study linguistic markers for various stages of radicalization trajectory. We use ideas in these works to develop our algorithm for detection of potential recruits.

We adapt algorithm in [14] for our automatic targeted data collection. The follow-back algorithm in [15] is implemented in the prototype to identify relevant users for counter-terrorism messaging.

3. Prototype Description

Our prototype for discovery and exploitation of publicly available information (PAI) is designed to allow rapid integration of new analytics in order to keep pace with technology advances and evolving adversary ConOps. To achieve rapid integration cycles and ease the transition and development for third parties, the prototype utilizes Elasticsearch (https://www.elastic.co/), Python and Django (https://djangoproject.com), which are all common Open Source Big Data architecture components. Elasticsearch is our chosen database because of its efficient search capabilities and its built-in analytic capabilities. Elasticsearch is very well suited for our application as it supports aggregation of structured metadata, parsing of unstructured text as well as storing of images and analytic outputs. Django a lightweight Python web framework which makes integration of Python code, our language of choice for rapid implementation of algorithms, very seamless. The lightweight aspect of Django ensures minimal time spent configuring, hosting, and managing the web servers, and Django's modularity provides a simple framework for plugging in new analytics. Overall, this web stack design limited our time spent on web developing so that we could focus on developing and implementing analytic capabilities. The system design is illustrated in Figure 1. The prototype is deployed on AWS GovCloud for evaluation and testing by selected organizations.

The prototype provides a user-friendly interface (Fig. 2). Figure 2 shows the available search panels; names, user names and ids have been obscured. To initiate PAI discovery, analysts can enter searches based on meta data, free text and images. Since micro-blogs often include slang, hashtags and spelling errors, the prototype includes an interface for expanding search terms to include related terms. The model for query

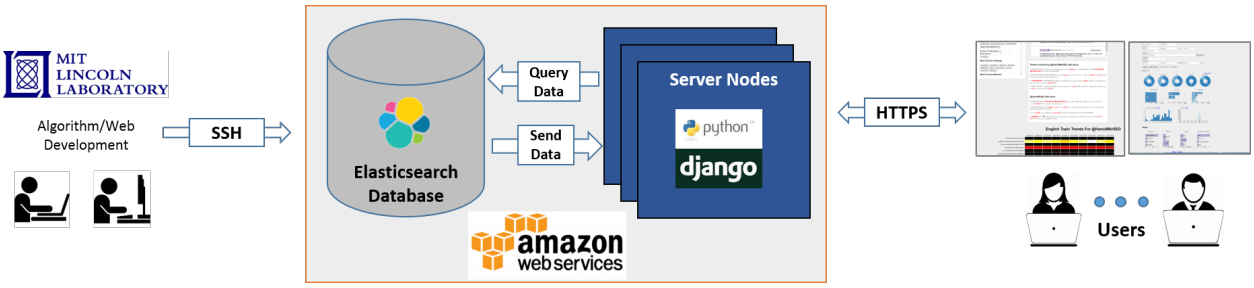

Figure 1: System design

expansion is trained using a deep learning technique against the dataset and can be re-trained against any other text corpus. The Section 5 describes additional analytics for threat actor discovery.

4. Targeted Data Collection

We collect Twitter data on Middle-eastern terrorist networks. We start with a subset of truthed jihadist accounts available online (e.g., suspected jihadist accounts posted by @CTRLSec Twitter account). At the next step, we scrape friends, followers and mentions of these seed accounts. We could continue to build n-hop network in this fashion but this snowball sampling approach would result in exponential growth of our graph that would include many irrelevant nodes. Instead, to find other relevant accounts in our network, we adapt targeted sampling approach described in [14]. The central assumption of Smith et al. sampling approach is homophily between vertices, specifically common content of interest to relevant vertices. The approach leverages a trained classifier to bias the sampling procedure toward only the most relevant portions of the graph based on a priori domain knowledge of relevance. We manually truthed a dataset of several hundred Twitter users and assigned these users to two classes based on their posted content: either relevant or irrelevant to the threat network. Several annotators were used to classify each user. We then trained supervised classifier using a variety of natural language processing (NLP) features such as bag of words, n-grams and semantic word embedding [16]. The F1 score for the classifier is above 0.7. We can tolerate a fairly high level of irrelevant nodes in our network, the goal here is to limit exponential growth of the threat network graph. We then use this classifier to expand nodes in our graph. Our dataset contains about 45 million tweets and about 60 thousand users.

5. Threat Actor Discovery Analytics

While social media sites like Twitter include a diverse set of metadata that can be used for determining a network of users, such as friends, followers, mentions and retweets, they do not reveal all the relationships of users in a social media network. In addition, OPSEC-aware actors purposefully obscure their networks to avoid detection. The prototype includes three methods for prioritizing explicit and hidden links in the network:

- Similarity of two-hop social networks
- Semantic similarity of all text content
- Accounts with similar usernames

Often analysts may also wish to better understand the network or history of a message. The prototype supports semantic similarity search at the single micro-blog level and fuzzy image search. For each account in the dataset, the tool predicts the gender of the user and characterizes users activity patterns which we group into five categories: conversation, address followers, self, opinion and information. The prototype also implements the follow-back algorithm in [15] which suggests an interaction policy to get a target user to follow agents deployed in the social network. The following subsections describe selected analytics in more detail.

5.1. Classification of User's posts

The prototype puts each user's posts into five groups: conversation, address followers, self, opinion and information (Figure 4). To train a Nave Bayes classifier, three annotators labeled around 500 tweets using the five categories. Tweets that had perfect inter-rater agreement were used as the truth set. This analytic is language-agnostic and features include: number of hashtags, number of mentions, number of question marks, number of URLs, number of pronouns, number

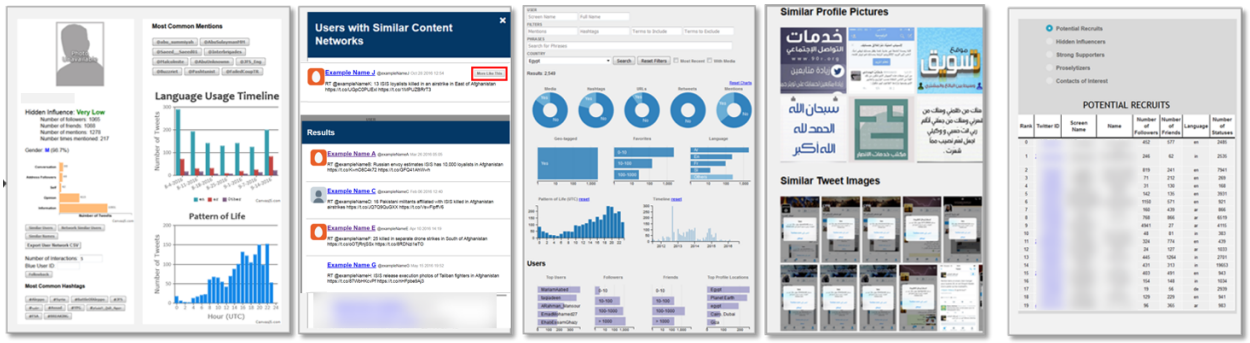

Figure 2: PAI prototype interface

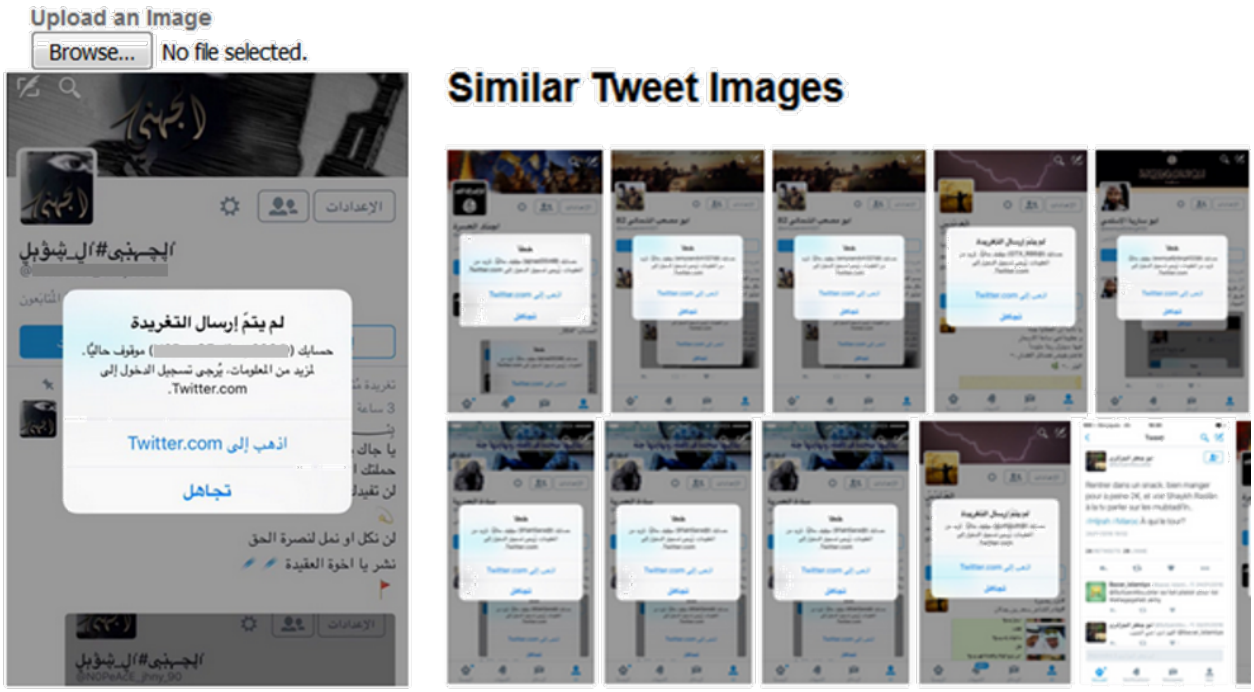

Figure 3: Fuzzy image search

of retweets, if it is a retweet, number of favorites, tweet user's number of statuses, tweet user's number of followers, tweet user's number of favorites.

5.2. User Tweet Summary

This analytic calculates top N indicative tweets for a given user (Figure 5). These tweets are meant to represent the content the user frequently tweets about. The algorithm consists of four steps:

- Tokenize words
- Calculate term-frequency and inverse document frequency (TF-IDF) [17] where the words are the terms and the document contains all of user's tweets
- Rank tweets using the formula: $\text{TweetScore} = \sum_{i}^{n_{tokens}} \frac{\text{TF-IDF}_i}{n_{tokens}}$
- Return top N tweets. To encourage diversity of summary tweets, we require each tweet to be 'different enough' from the other tweets. We determine how different tweets are by taking the tweets' corresponding rows in the TF-IDF matrix (the columns are the terms) and calculating the

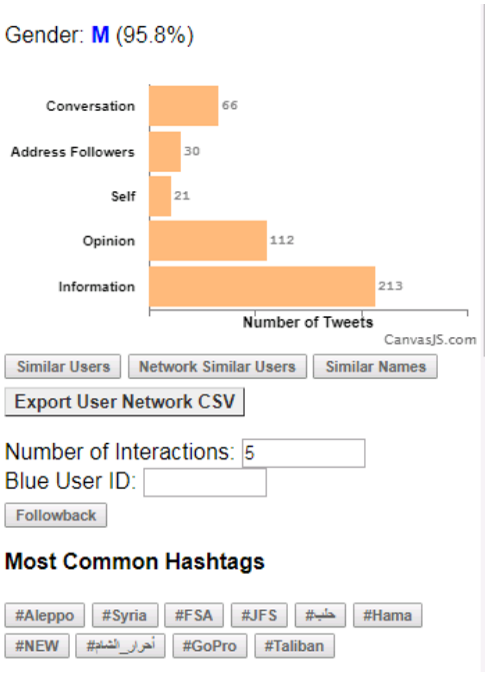

Figure 4: User tweet categories

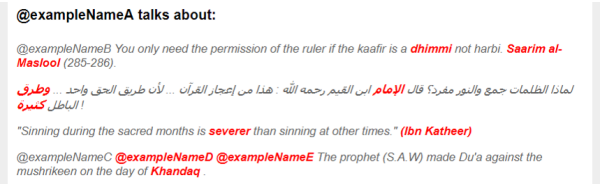

Figure 5: User tweet summary

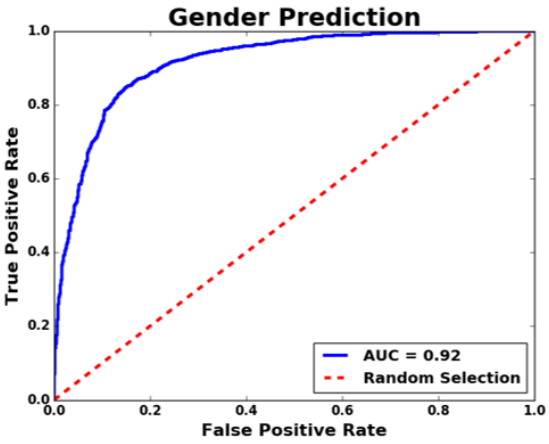

Figure 6: Performance for gender prediction algorithm

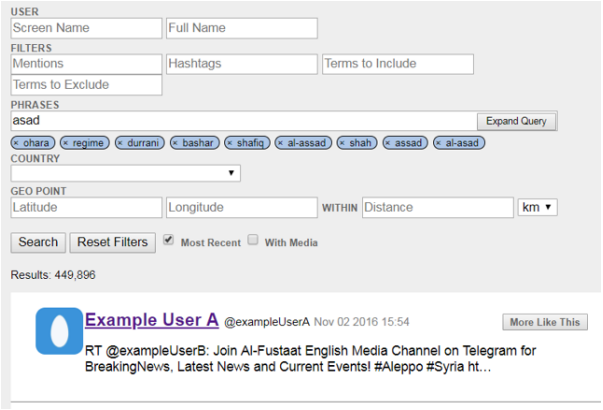

Figure 7: Query expansion interface

cosine similarity distance between the tweets. If the distance is greater than a threshold the tweets are considered 'different enough.'

5.3. Gender Prediction

We predict gender based on all of user's English tweets. For training and testing data, we gathered names that were typically just female names (e.g., Jessica) and typically just male names (e.g., David). These typically gender discrete names were used to label the training and testing data as male or female. The training set composed of 32,792 users and the test set composed of 3,792 users. The gender predictions were generated by a neural network using the high-level neural network library Keras and the Theano libray for back-end computational processing [18; 19]. The input for a single user is the Global Vectors for Word Representation (GloVe) representations [16] for each word in all of the user's tweets. The neural network consisted of one convolution layer and one hidden layer. Figure 6 demonstrates the performance.

5.4. Query Expansion

When a user submits a query searching for certain text or phrases, the query expander suggests terms that are semantically related (Figure 7). A GloVe model was trained on all the tweets of our dataset. Upon submission of a user query, each word's corresponding vector is used to calculate the Euclidean distance to all other words. The top semantically similar words, the words that have the shortest Euclidean distance, are returned.

5.5. "More Like This" Search for Similar Tweets

In this prototype a 'more like this' button is adjacent to all tweets and clicking on these buttons will display similar tweets (Figure 8). 'More like this' is a capability provided by Elasticsearch. Elasticsearch's 'more like

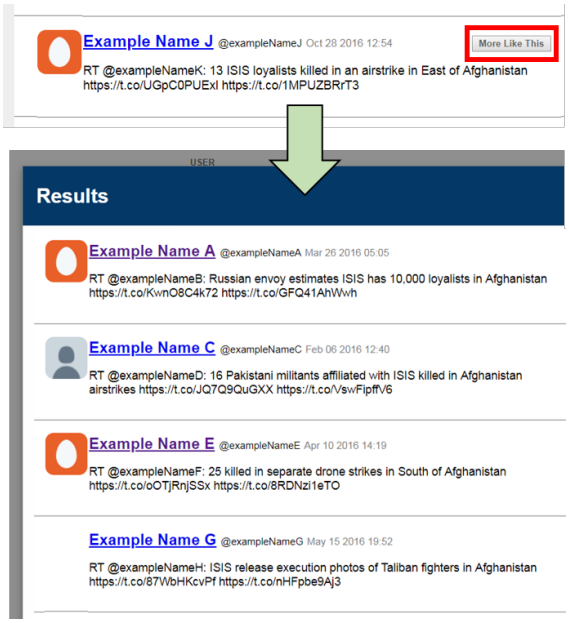

Figure 8: "More Like This" button displays similar tweets

this' query uses TF-IDF to select similar text. In this instance, the document is the tweet, and the terms are the words in the tweet.

5.6. Image Search

A user may upload an image or click on an image in the web application to find similar images from tweets or profile pictures A LIRE (Lucene Image REtrieval) plugin for Elasticsearch named elasticsearch-image enables image content-based retrieval (https://github.com/kzwang/elasticsearch-image). Our chosen image feature are PHOG features (Pyramid Histogram of Oriented Gradients). PHOG features emphasize edge lines, which are robust to image adaptations commonly associated with social media. Dataset of 5000 images was collected that included original images and the original images' 5 Instagram stylizations 9. PHOG features had the highest performance (Figure 10).

6. Automated Dual Account Detection

In order to correctly characterize threat networks, we need to ensure their accuracy as they change over time. In order to do so, once an account is suspected of belonging to an online extremist network, we need to continue monitoring its activity. This seemingly straightforward task is complicated by the fact that social media platforms have been aggressively suspending accounts for violating Terms of Use, particularly those that promote violence or share terrorist content. In April of 2018, Twitter reported that over 1.2 million accounts had been suspended for terrorist content since August 2015, many of which were suspended before their first Tweet [20].

While the removal of violative content is expedient, suspended users can quickly and effectively return to the network, creating an unfortunate game of "whack-a-mole" for authorities wishing to keep track of influential actors and to monitor their networks. The process by which suspended users acquire new accounts can vary, and each method generates unique challenges for dual account estimation.

Creating a new account is free and takes only a few minutes. While Twitter now requires a phone number or email address at sign-up, it is trivially easy to create these as well. Newly-created accounts can often reconnect with their networks rapidly by a combination of actions including using the same or similar screen name and profile image (see Fig. 11, authenticating their suspension by screenshot (see Fig. 12), tweeting at designated "follow-bots" that exist purely to facilitate reconnection, etc. Users that anticipate frequent suspension often create many new accounts at once, generating dense subnetworks for propagating content and maintaining connections (see Fig. 13). Particularly for accounts with large and active follower-networks, it is common for one account to reference another as a "reserve account" (see Fig. 14). These activities are intended to make the users unambiguous and easy to find.

Assuming an existing account can be accomplished via donation, purchase, or by hacking. Some suspended users reacquire their networks using the same methods as described above for new accounts, making them easy to identify. Other accounts, however, are amassed and connected to the network in advance, awaiting dormant until needed. These accounts need not have a familiar name or screen name, and may instead leverage only a recognizable profile picture. These accounts can be several years old and have pre-existing timelines that can make them appear benign (see Fig. 15).

The examples provided in this paper illustrate just a few of the many techniques employed by users to withstand repeated account suspensions. While some reports suggest that multiple suspensions limit the effectiveness of extremist networks online [2; 21; 22], these networks have demonstrated significant adaptability and nevertheless continue to operate on Twitter and other digital platforms [23].

Figure 18 plots the social networks of two users with distinct re-acquisition strategies. In this figure each

Figure 9: Sample training dataset images for image search

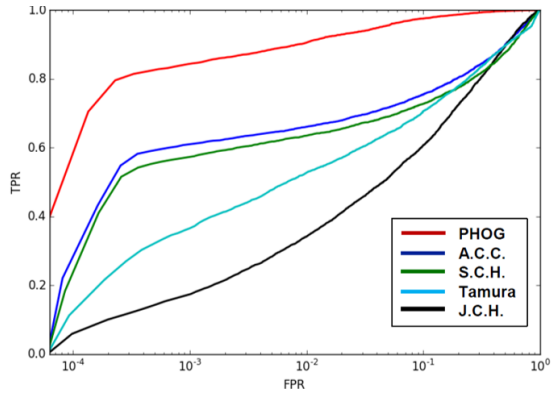

Figure 10: Image Search Feature Comparison (PHOG: Pyramid Histogram of Oriented Gradients; ACC: Auto Color Correlogram; SCH: Simple Color Histogram; Tamura: Tamura's Texture Features; JCH: JPEG Coefficient Histogram)

Profile Picture	Screen Name	Name	Date Created
	@l483516_ayla	Ayla	Feb 28, 2017
	@l880300	Ayla	Mar 3, 2017
	@Ayla49042246	Ayla	Mar 3, 2017
	@l1185543	Ayla	Mar 6, 2017
	@l335707	Ayla	Mar 7, 2017

Figure 11: Twitter user "Ayla" assumes several accounts over the span of one week. The employment of similar screen names and repetitive thematic imagery make these accounts recognizable to other users.

Figure 12: Twitter user @Islamicstate_57 advertises a new account by sharing a screenshot of the suspension notice for a previous account, @Islamicstats_56. Original tweet in Arabic, translated into English here.

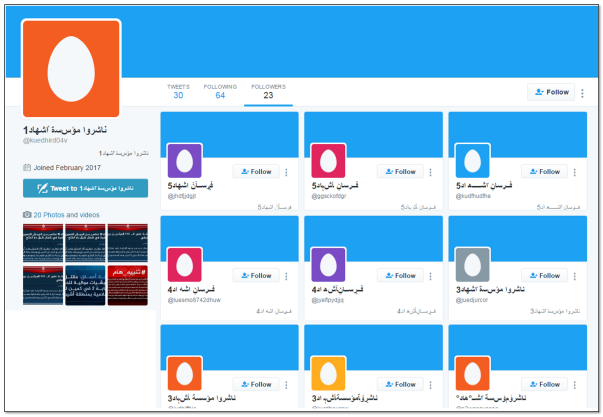

Figure 13: Newly-created ISIS propaganda Twitter account @kuedhird04v, likely a bot, is already connected to dozens of similarly disposable accounts.

colored node has been identified as a unique account belonging to a single user (determined manually by our research team). The gray nodes are all the friends and followers of those accounts.

For the user represented in Fig. 18 (left), new accounts are created as active accounts are suspended. This requires the user to reconnect with the network over and over again. This is not only inefficient, but the user's friends and followers are also constantly being

Figure 14: Twitter user @isisom59 declares that this is a "Reserve Account" and should be followed in anticipation of the primary account being deleted. Original tweet in Arabic, translated into English here.

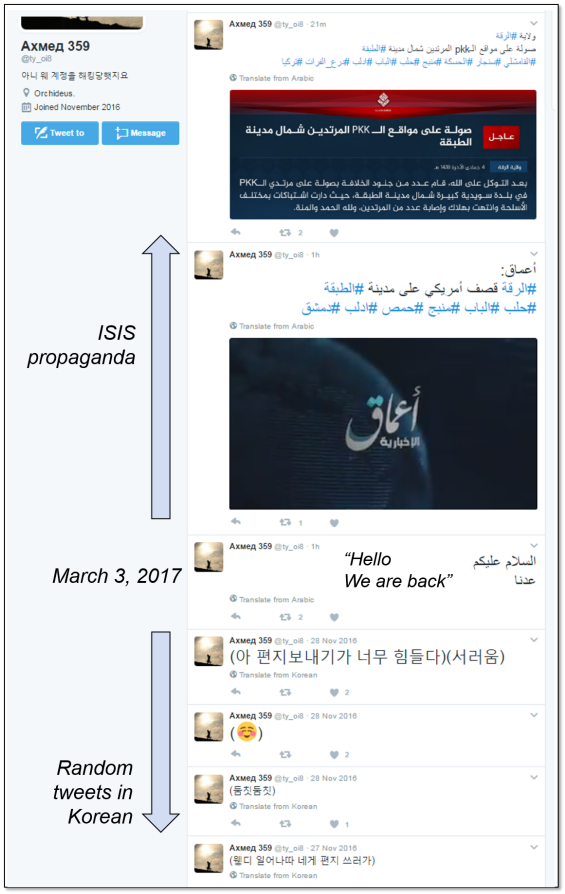

Figure 15: Twitter user @ty_oi8 appears to be an innocuous Korean-language account until March 3, 2017 when it resurfaces as an ISIS propaganda account.

suspended. The result is a very low network overlap between subsequent incarnations of this user, however the name, screen name, and profile picture are similar. This account is very easy to identify and is suspended every few hours. In the month of March 2017 we observed 53 unique accounts for this user.

When the user represented in Fig. 18 (right) is suspended, years-old dormant accounts are re-activated and re-purposed with networks already partially intact. In this case, the new account's name, screen name, and profile picture do not match that of the suspended account, however the network overlap between accounts is relatively high. In fact, many of the target's accounts are themselves connected. This account was able to stay active longer (sometimes several days) and gain more followers between suspensions. In March 2017, 22 accounts for this user were observed, and analysis of this user's network revealed a massive "bot farm" containing hundreds of inert and seemingly innocuous accounts waiting to be inherited.

Figure 16 illustrates the workflow developed for dynamically following high-interest users through account suspensions. A human-selected "watch list" of users is persistently monitored using Twitter's Streaming Application Programming Interface [24] for collecting tweets, profile changes, and friend/follower networks. When an account is observed to be suspended, prioritized queries are submitted to Twitter's REST API to search for candidate dual-accounts. If we have not observed previous suspension-reincarnation cycles for the user, we query for all accounts with similar names or screen names, users that have mentioned the account in their tweets, and the 2-hop friend-follower network of the suspended account. The 2-hop network may be extremely large, particularly if the suspended account followed many high-degree users, so we only collect the most recent N friends/followers where N can be set dynamically depending on resources and Twitter API limits. If the user has been suspended before, the search procedure is adapted according to the user's previously-observed patterns and techniques for account acquisition. Since the user's new account may not immediately exist, this search procedure may be conducted more than once.

After prioritized search is conducted, similarity metrics are calculated between the original (now-suspended) account and each candidate account. As described in in this section and similar to methods proposed by [3], we calculate similarity metrics from profile features (name, screen name, profile picture, etc.), timeline features (temporal activity, language use, links and media), and network features (overlap of friends, followers, and mentions). Figure 17 illustrates an example of how one user was automatically followed from a suspended account to a newly-created account. Candidate accounts with high similarity scores are presented to an analyst for confirmation, and subsequently added to the account watch list.

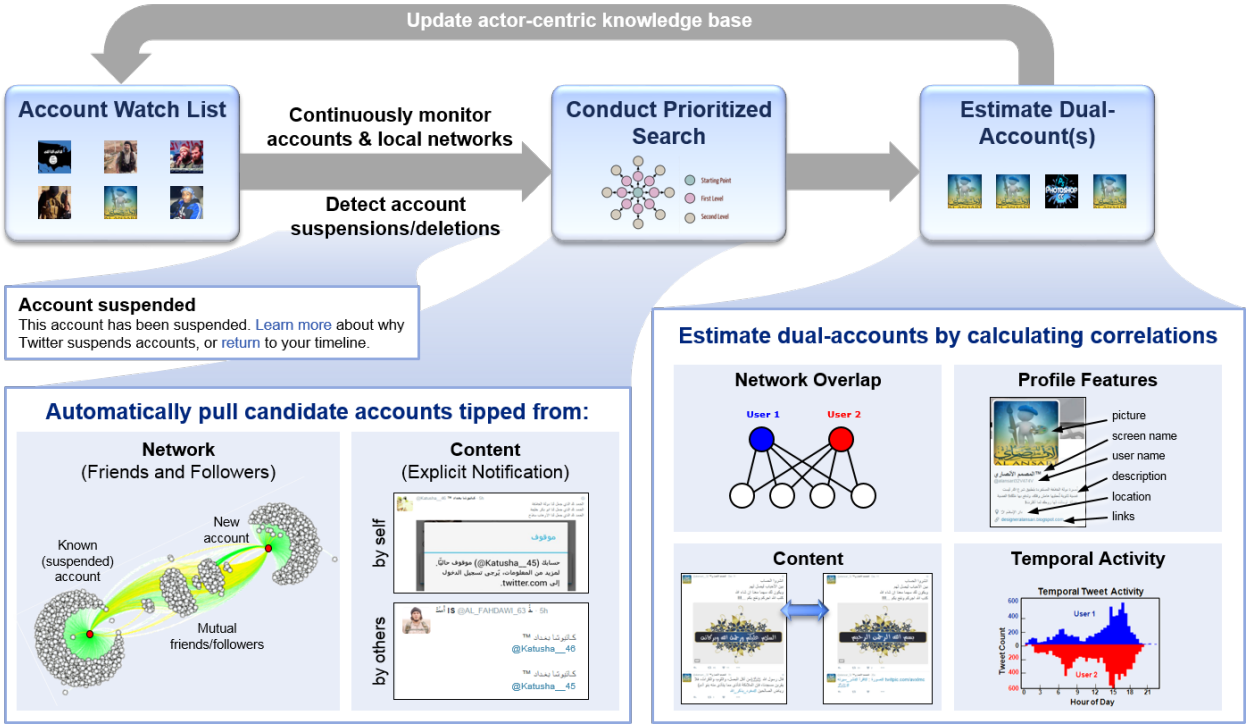

Figure 16: Semi-automated workflow for account tracking through suspensions.

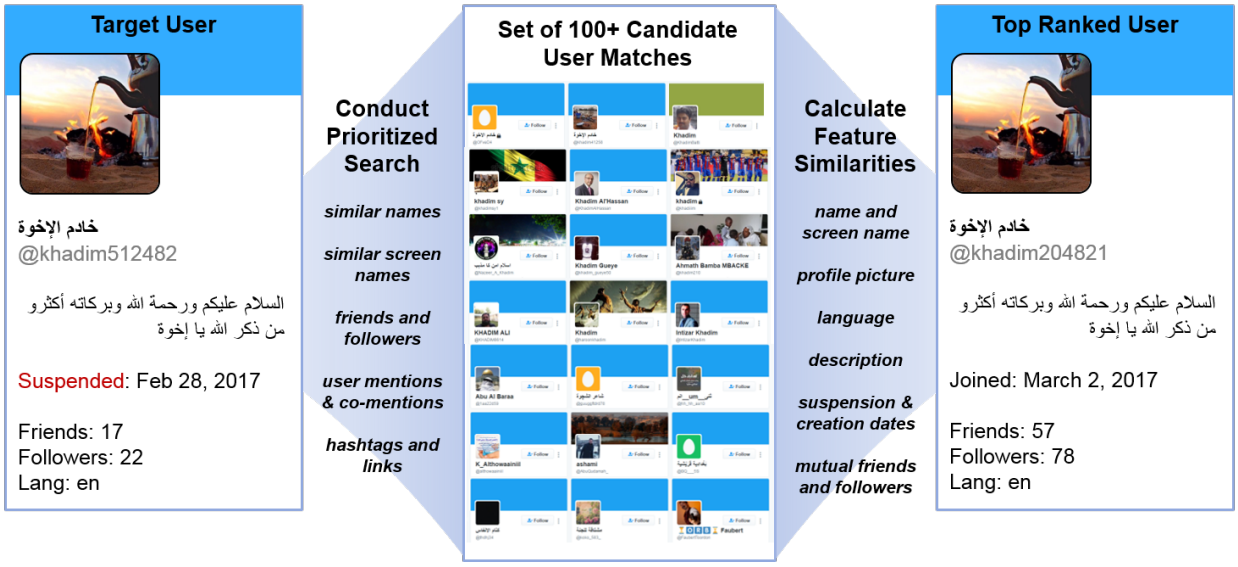

Figure 17: Following the suspension of user @khadim512482, a new account is automatically discovered as a likely match.

7. Network Role Detection

Many of the user accounts in online terrorist networks have specific roles [9; 7], e.g. propagandists, proselytizers, recruiters, sympathizers, fighters. We are interested in detecting these roles in order to characterize the threat networks, with main focus on detecting potential recruits.

Figure19 shows a small subset of the threat network and consists of friends and followers of 10 seed nodes (blue labels). The nodes labeled in gray are high-centrality nodes which are often propagandists or

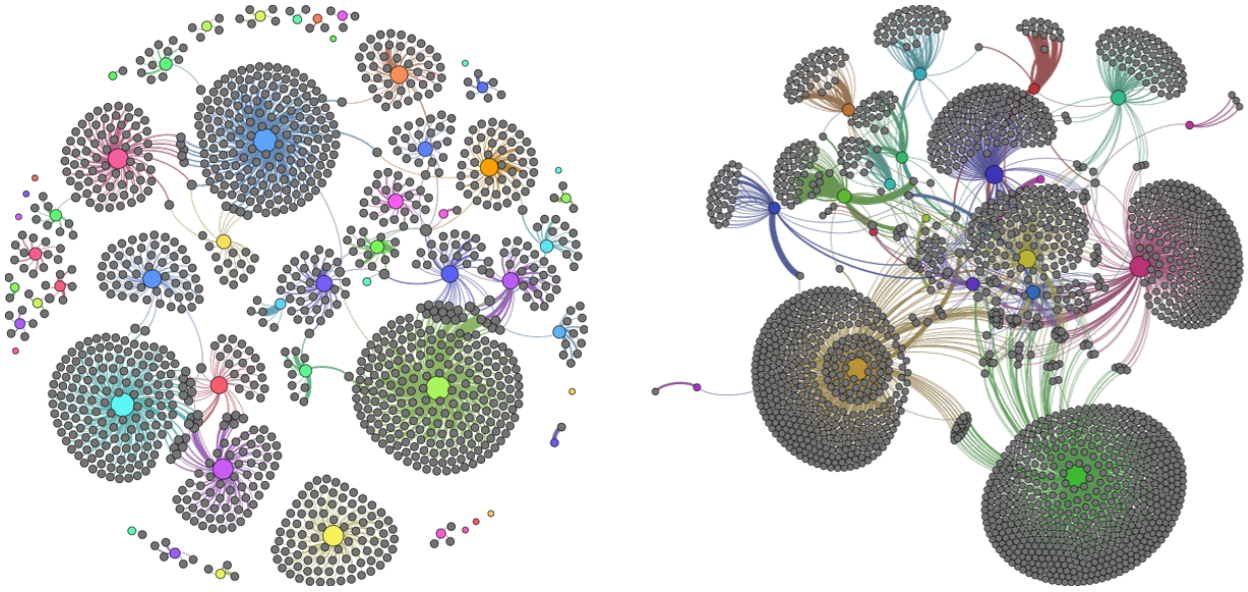

Figure 18: Two users' friend-follower networks through multiple suspensions. Colored nodes indicate different accounts of the same user.

proselytizers. We detect communities using Louvain algorithm [25] and color the graph nodes by community membership. Note that the graph is highly connected within the communities which is typical for these types of threat networks [26].

7.1. Radicalization Detection

Susceptibility to radicalization is a widely studied problem in social science. There is a consensus that radicalization generally happens over a period of time -months to years- [27], [28] and follows a distinct trajectory consisting of the following stages: pre-radicalization, detachment, peer-immersion & training, and planning and execution of violent action [11]. Each of these stages exhibit specific behaviors, some of which can be detected in publicly available information like social media using natural language processing and network analysis techniques. For example, behavior often exhibited in pre-radicalization stage is fixation on a specific entity, and this behavior can express itself online by the individual seeking information about an entity or discussing it often on social media [12]. In the detachment stage, typical behaviors include loneliness and identification, and these behaviors can express themselves in various online discussions about loneliness and disaffection, and by seeking connections to new authority figures.

We design an algorithm that combines users social media network features, topic trends and sentiment to calculate a radicalization score for potential recruits. Specifically, we use Latent Dirichlet Allocation (LDA) topic modeling algorithm [29] to construct temporal topic heatmaps in order to identify topic transitions. Since Twitter tries to suspend users with radical content, temporal changes in users mentions network can be indicative of potential radicalization over time. Specifically, user network timeline with increasing portion of suspensions is indicative of users radicalization.

Finally, we look for frequency of certain keywords and concepts used in users post (e.g, concepts related to boredom, loneliness, jihad, etc.) and analyze changes over time. We combine these features to calculate ranked scores and design a cutoff. We identify about three dozen users in our dataset as potential recruits. Since we dont have a dataset of actual recruits with their Twitter timelines, we manually checked our results and presented them to analysts for evaluation and feedback. Using the fact that Twitter suspends radical accounts as a truth proxy, we checked that over time all our identified potential recruits' accounts got suspended due to their persistent radical content.

Figure 20 gives an example of a user that got flagged as a potential recruit by our algorithm. The selected tweets indicate topic and interest changes. Figure 21 shows the timeline of suspended mentions for this user, note that there is an increase in suspensions rate towards the end of the timeline. Figure 22 shows the temporal topic heatmap for the same user. Black color indicates cold topics and white indicates hot topics. We used 30

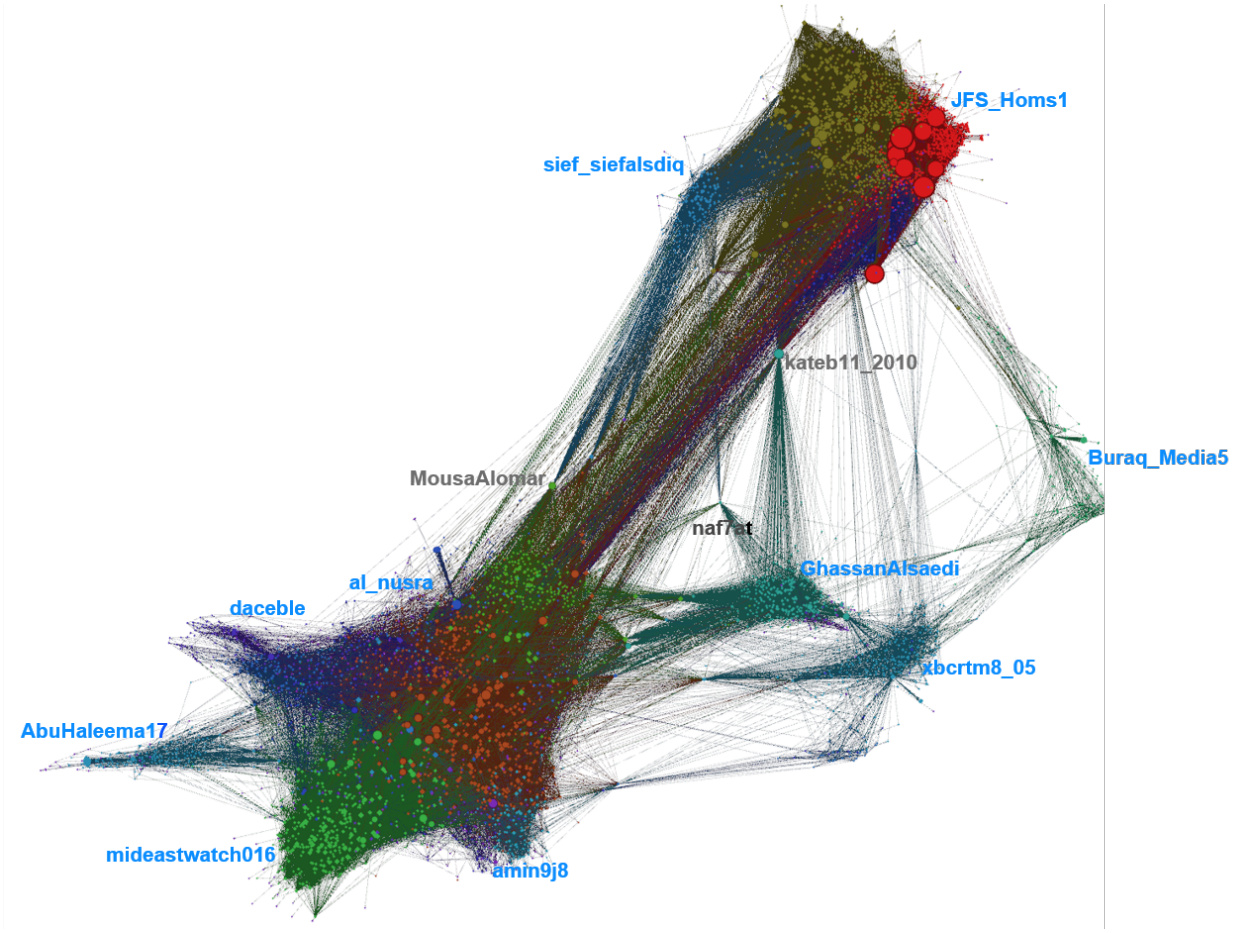

Figure 19: Small subset of threat network graph

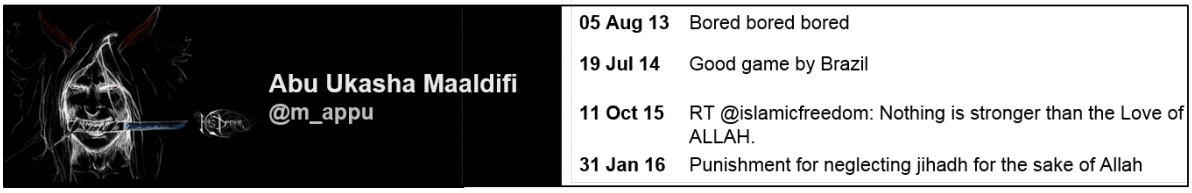

Figure 20: Selected tweet progression for a potential recruit

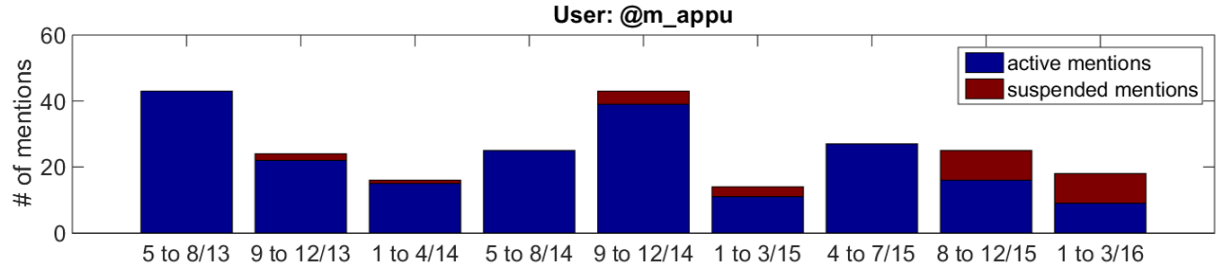

Figure 21: Timeline of suspended mentions

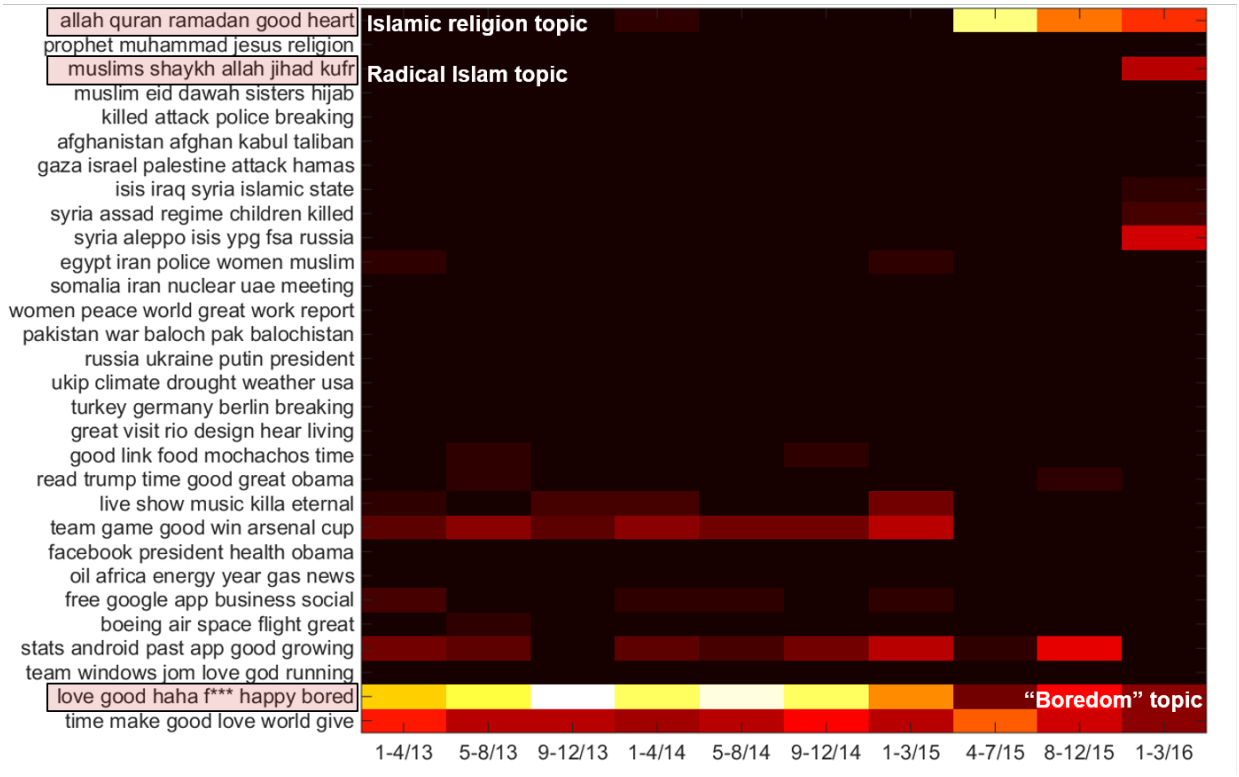

Figure 22: Temporal topic heatmap for a potential recruit

topics as a parameter in our LDA model and labeled topics using automatically generated keywords. Note the abrupt topic pattern change from "boredom" topic to Islamic religion topic and radical Islam topic.

7.2. Detecting Other Roles

Propagandists and proselytizers are easy to detect. We expect to see many followers in their social media networks, and in temporal topic heat maps we look for predominant Islamic religious topics. Detection of recruiters, on the other hand, is difficult since recruiters try to avoid detection and number of recruiters of foreigners is small [30]. Recruiters try to minimize contact on open social media platforms and often use encrypted apps (e.g., Telegram, Kik). While some recruiters list their user information for encrypted apps in their online profile, many are discouraged to do so for security reasons [30]. We have expand our working definition of a recruiter to include any suspicious contacts. We calculate the following indicators: user has substantial radical Islam content in his/her posts and displays encrypted app contact information; in addition, user is in the network of a potential recruit. Figure 23 gives an example of a detected potential recruit, the topics timeline shows prominent radical Islam topic, the user displays Kik encrypted app contact information and is connected to potential recruit shown in figure 20.

To detect sympathizers and fighters, our approach is to manually truth accounts and build supervised classifiers using NLP and network features. For instance, community membership is indicative of radicalization since communities have different suspension rates. For example, suspension rates in communities in figure 19 vary between 4 and 20%. Sample performance is illustrated in figure 24. To build this classifier to distinguish between ISIS supporters and opponents, we semi-automatically generate truth data using results from [16]. Including only GloVe word embeddings [**?**] as features we obtain F1 score of 0.77. Note that distinguishing between ISIS opponents and supporters is a harder problem that distinguishing between ISIS supporters and users not discussing terrorism because of similar language use.

8. Summary and Way Forward

Our novel prototype provides analysts with the automated Identity Operations capability with wider and richer online presence characterization to support Information Operations, Force Protection, Indications and Warnings, Situational Awareness and other

Figure 23: Temporal topic heatmap for a potential recruiter

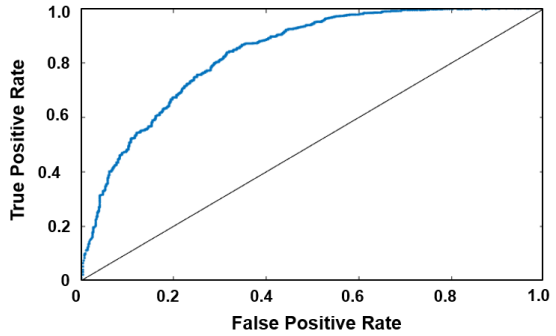

Figure 24: ISIS supporters versus opponents classification performance; # of training instances: 1331, area under the curve: 0.834, precision: 0.77, recall: 0.77

activities. It allows them to automatically construct threat network of relevant actors and relationships from social media, track temporal network changes, and detects and classify threat network roles and activities. It increases efficiency with which analysts exploit PAI and provides cues to previously hidden users and relationships for follow-on tasking. Additional research is needed to develop analytics for further threat network characterization such as narrative classification, bot activity characterization, temporal influence, credibility estimation and fusion with other data sources such as traditional media.

References

[1] J. M. Berger and J. Morgan, "The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter," *The Brookings Project on US Relations with the Islamic World*, vol. 3, no. 20, pp. 4–1, 2015.

[2] J. Berger and H. Perez, *The Islamic State's Diminishing Returns on Twitter: How Suspensions are Limiting the Social Networks of English-speaking ISIS Supporters*. George Washington University, 2016.

[3] J. Klausen, C. Marks, and T. Zaman, "Finding online extremists in social networks," *CoRR*, vol. abs/1610.06242, 2016.

[4] W. M. Campbell, L. Li, C. Dagli,

J. Acevedo-Aviles, K. Geyer, J. P. Campbell, and C. Priebe, "Cross-domain entity resolution in social media," *arXiv preprint arXiv:1608.01386*, 2016.

[5] J. R. Scanlon and M. S. Gerber, "Automatic detection of cyber-recruitment by violent extremists," *Security Informatics*, vol. 3, no. 1, p. 5, 2014.

[6] A. Sureka and S. Agarwal, "Learning to classify hate and extremism promoting tweets," in *Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint*, pp. 320–320, IEEE, 2014.

[7] H. Peladeau, "Support for sisters please: Comparing the online roles of al-Qaeda women and their Islamic State counterparts," 2016.

[8] M. Rowe and H. Saif, "Mining pro-ISIS radicalisation signals from social media users," in *AAAI Conference on Web and Social Media (ICWSM)*, pp. 329–338, 2016.

[9] A. T. Chatfield, C. G. Reddick, and U. Brajawidagda, "Tweeting propaganda, radicalization and recruitment: Islamic State supporters multi-sided Twitter networks," in *Proceedings of the 16th Annual International Conference on Digital Government Research*, pp. 239–249, ACM, 2015.

[10] M. D. Silber, A. Bhatt, and S. I. Analysts, *Radicalization in the West: The homegrown threat*. Police Department New York, 2007.

[11] J. Klausen, S. Campion, N. Needle, G. Nguyen, and R. Libretti, "Toward a behavioral model of "homegrown" radicalization trajectories," *Studies in Conflict & Terrorism*, vol. 39, no. 1, pp. 67–83, 2016.

[12] K. Cohen, F. Johansson, L. Kaati, and J. C. Mork, "Detecting linguistic markers for radical violence in social media," *Terrorism and Political Violence*, vol. 26, no. 1, pp. 246–256, 2014.

[13] J. Brynielsson, A. Horndahl, F. Johansson, L. Kaati, C. Mårtenson, and P. Svenson, "Harvesting and analysis of weak signals for detecting lone wolf terrorists," *Security Informatics*, vol. 2, no. 1, p. 11, 2013.

[14] S. T. Smith, R. S. Caceres, K. D. Senne, M. McMahon, and T. Greer, "Network discovery using content and homophily," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5925–5929, IEEE, 2017.

[15] F. Que, K. Rajagopalan, and T. Zaman, "Penetrating a social network: The follow-back problem," *arXiv preprint arXiv:1804.02608*, 2018.

[16] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.

[17] J. Ramos *et al.*, "Using TF-IDF to determine word relevance in document queries," in *Proceedings of the First Instructional Conference on Machine Learning*, vol. 242, pp. 133–142, 2003.

[18] F. Chollet, "keras." `https://github.com/fchollet/keras`, 2015.

[19] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints*, vol. abs/1605.02688, May 2016.

[20] Twitter Blog, "Expanding and building #TwitterTransparency," April 2018.

[21] M. Lakomy, "Cracks in the online "Caliphate": How the Islamic State is losing ground in the battle for cyberspace," *Perspectives on Terrorism*, vol. 11, no. 3, 2017.

[22] A. Alexander, "Digital decay: Tracing change over time among Eenglish-language Islamic State sympathizers on Twitter," *Washington, DC: George Washington University, Program on Extremism*, 2017.

[23] S. Weirman and A. Alexander, "Hyperlinked sympathizers: URLs and the Islamic State," *Studies in Conflict & Terrorism*, vol. 0, no. 0, pp. 1–19, 2018.

[24] "Twitter developer documentation: Filter realtime tweets," 2018.

[25] J. Kauffman, A. Kittas, L. Bennett, and S. Tsoka, "Dyconet: a Gephi plugin for community detection in dynamic complex networks," *PloS one*, vol. 9, no. 7, p. e101357, 2014.

[26] E. Bodine-Baron, T. C. Helmus, M. Magnuson, and Z. Winkelman, "Examining ISIS support and opposition networks on Twitter," tech. rep., RAND Corporation Santa Monica United States, 2016.

[27] A. W. Kruglanski, M. J. Gelfand, J. J. Bélanger, A. Sheveland, M. Hetiarachchi, and R. Gunaratna, "The psychology of radicalization and deradicalization: How significance quest impacts violent extremism," *Political Psychology*, vol. 35, no. S1, pp. 69–93, 2014.

[28] R. Borum, "Radicalization into violent extremism II: A review of conceptual models and empirical research," *Journal of Strategic Security*, vol. 4, no. 4, p. 37, 2011.

[29] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[30] J. Berger, "Tailored online interventions: The Islamic States recruitment strategy," *CTC Sentinel*, vol. 8, no. 10, pp. 19–23, 2015.