Pro-ISIS Fanboys Network Analysis and Attack Detection through Twitter Data

Yuqian Zhou
The Hong Kong University of Science and Technology
Kowloon, HKSAR
e-mail: yzhouas@ust.hk

Abstract—**Terrorism becomes more severe these days, especially the attacks sponsored by Islamic State of Iraq and Syria (ISIS) or Daesh. They are experts of using social network for propaganda and recruitment, thus predicting their activities through big social network data will help to predict and avoid more serious attacks. In this paper, we analyze over 17k twitter records of pro-ISIS fanboys over a year. Based on those tweets, we want to dig out: 1. Who is the most important and active member in the social network? 2. What are the hashtags they frequently used for propaganda? 3. According to the twitting peak time and information on-hand, will it be able to predict the next attack? Our results reveal the leader in this propaganda network, and achieve a satisfactory attack prediction via time-series neural network through very limited attack history.**

Keywords-social network; event prediction; terrorism; time-series neural network

I. INTRODUCTION

Islamic State of Iraq and Syria (ISIS), also known as ISIL or Daesh, is one of the most horrible terrorism organization designated by united nations in today's world. They planned and conducted plenty of severe attacks, killing thousands of civilians every year all over the world [1].

As reported, members of ISIS are experts of using social network, especially twitter, a widely-used tool for people all over the world. They mostly use twitter for recruitment and propaganda, attracting thousands of foreigners to join, making the situation of anti-terrorist more severe. Primitively, they created official twitter accounts but are soon banned by internet managers. Now, they are able to be hidden under unofficial accounts and use hashtags for spreading purpose [2]. Analyzing their twitting rules and tracking the hidden essential members are urgent work to predict future attack and avoid them happening.

In [3], pro-ISIS twitter accounts are sampled, and multiple social media metrics are examined based on these accounts. And in [4], the author monitored the Syrian conflicts by tracking an Arabic hashtag, and observed a twitter storm. However, research work related is still limited, and localizing accounts which are confidently ISIS supporters is also hard. But a pro-ISIS fanboys dataset is now available in *Kaggle*, providing a chance to dig information out from the pre-selected potential fanboys.

In this paper, we consider the following problems based on this dataset,

First, we want to find out who plays the most important role when propagating and recruiting for ISIS, by ranking the out-degree of each node in the whole network. Moreover, we care about the core member in the ISIS community after filtering the network, by checking their betweeness centrality.

Second, we rank the hashtag frequency in all tweets, and find out which tags are frequently used for propaganda.

Finally, we build model to predict attacks sponsored by ISIS, by taking high-frequency hashtag, high-active members, and tweets number on each day into account using both a static and dynamic model. It shows a satisfactory prediction result when validating on the attack history records.

II. METHODOLOGY

A. Node Importance in a Social Graph

In a social graph, the importance of nodes can be measured in multiple ways. In this paper, we consider the following two metrics.

Degree Centrality: The degree of nodes reveals the number of edges incident on a node. It is the sum of in-degree and out-degree, which represent the number of edges entering and leaving the nodes respectively. The nodes regarded as the most important under degree centrality have the most direct connections with others. The value can be computed by,

$$C_D(v_i) = \sum_j A_{ij} \tag{1}$$

where v_i is the node index, and *i, j* represents the row and column index in the adjacent matrix *A*.

Betweeness Centrality: The importance of nodes in a graph can also be measured by betweeness centrality, counting the number of shortest path passing one node. Nodes with high betweeness centrality are important for information diffusion and spreading, by connecting multiple parts in the network. Betweeness centrality can be computed by,

$$C_B(v_i) = \sum_{v_s \neq v_i \neq v_t \in V, s<t} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \tag{2}$$

where σ_{st} represents the number of shortest path between node v_s and v_t, while $\sigma_{st}(v_i)$ are those passing v_i.

B. Generalized Regression Neural Network

Generalized regression neural networks (GRNNs) [5] is widely used for function approximation. Compared to other gradient-based optimization method, it approximates the function between input and output directly by drawing the function from training data. The output of the testing sample is estimated according to the weighted sum of the outputs of the training samples, by computing the weights in terms of

978-1-5090-3619-6/17/$31.00 ©2017 IEEE

the Euclidean distance between testing and training samples. The overall network architecture is illustrated in Fig. 1.

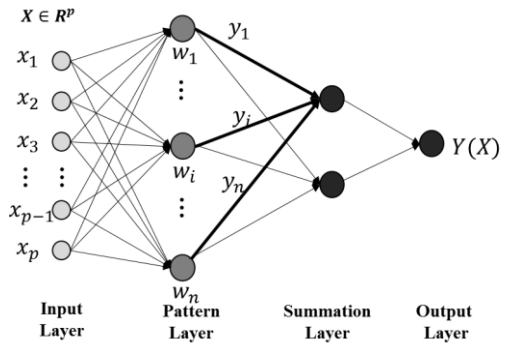


Figure 1. Network Architecture of GRNNs. It contains four layers, including input layer, pattern layer, summation layer and output layer. It estimates the output of the testing sample by weighted average of the outputs of the training sample, where the weights are computed by the distance between testing and training samples.

(a)

(b)

(c)

Figure 2. The graph visualization of three networks. (a) The whole propaganda network, with 'Uncle_SamCoco' showing great importance. (b) The core member network who post tweets, showing that 'Uncle_SamCoco' is still important for imformation diffusion. (c) Reduced core member network, containing only mutual connections among members. 'Uncle_SamCoco' and 'WarReporter1' reveals identical importance, showing that they play essential roles in their own small communities.

GRNNs has four layers. Given an input testing vector, $X \in R^p$, it computes the Euclidean distance between X and each training samples $\hat{X}_i$ and passes it to an activation function. Thus in the pattern layer,

$$w_i = e^{-\left(\frac{(X-\hat{X}_i)^T(X-\hat{X}_i)}{2\sigma^2}\right)} \quad (3)$$

where w_i stands for the weights for each training sample, and σ is the only free parameter which can be trained by cross validation. Finally, the estimated output $Y(X)$ of X will be,

$$Y(X) = \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i} \quad (4)$$

GRNNs is more efficient than normal neural network, and will be converged to a global optimal solution, which is also more general.

C. Time-series Neural Network

Dynamic Neural Network is useful for time-series prediction and forcasting. It forms the prediction as,

$$y(t) = f(x(t-1), ..., x(t-d)) \quad (5)$$

where t is the current time, and d is the time delay. In our model, the network contains one single hidden layer with 15 units, and the time delay is controlled to 2 days.

III. Experiment

We conduct our analysis on the ISIS Twitter Dataset available on *Kaggle*, with title 'How ISIS uses Twitter'. First, we extract the membership inside the propaganda network, and further compare the importance of each node. Secondly, we also extract the core social network by filtering the nodes less important, and visualize the interaction between core members. Finally, we extract the hashtags and visualize the twitting peak time, through which we predict the ISIS attacks. The detail of our experiment is stated below,

A. ISIS Twitter Dataset

ISIS Dataset is available on *Kaggle*, containing the tweets post by scrapping 100+ ISIS pro-fanboys from all over the world during the period from 1st June 2015 to 13rd May 2016. On each raw record, it is consist of *{name, username in Twitter, location, number of followers, number of statuses, timestamp, and the tweet itself with multiple language}*. Each tweet may be written in multiple languages, but most of the important information is organized in English, which is what we focus on. The content of tweet may be related to a video link for the purpose of propaganda, or sharing anti-US and anti-western country slogan by various hashtags. In this paper, we aim at effectively using the information of Twitter user name, the usernames they mentioned in the tweet, and the hashtags they added.

B. Data Pre-processing

User Name Extraction: In this analysis, we extract the username both in the posting list and the tweets, and merge them together in a full user lists. The method we extract the id, instead of complicated string matching, is localizing the mark of '@' in the tweets. However, some unformatted data, like user names followed by ';' are also filtered in. After extraction, we further guarantee the uniqueness of the user name. Finally, we obtain totally 3341 unique user names, referring to the nodes in the propaganda network.

Hashtag Extraction: Similar to user name extraction, because of the importance of hashtag when spreading virally, we extract all the hashtags by localizing the mark of '#' in the tweets. Compared to username, the format of hashtags referring to the same content can be more various. Thus after extraction, we further compute the similarities among hashtags by checking the character occurrence and order. For example, 'Paris' can be merged with 'Paris-', and 'BREAKNEWS' can be identical to 'breaknew'. Finally, we obtain totally 3157 semantic-independent and unique tags. For easier statistics, we also keep the original format of each tags.

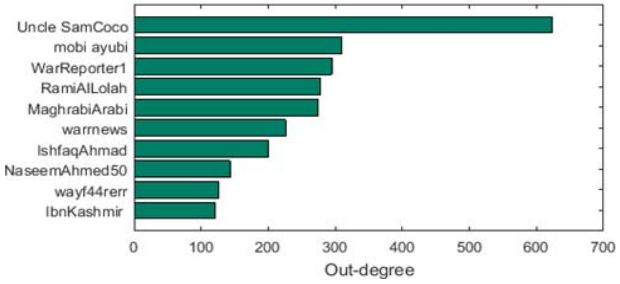


Figure 3. The top 10 users with high out-degree in the propaganda network

Figure 4. The visualization of high-frequent hashtags fanboys used for propaganda, showing that *'ISIS'* is the most frequent one. *'Syria'* , 'Iraq' and *'IslamicState'* follows, and *'Breakingnews'* after that. Some important terms like *'Paris'* are missing in this graph.

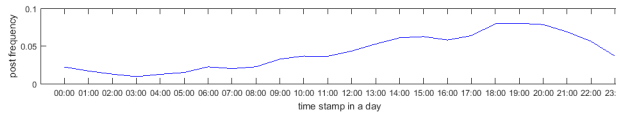


Figure 5. The posting frequency in different time-period in a day.

Timeline Extraction: In the raw tweets data, time stamp is represented in the format of 'mm/dd/year h:m'. For simplicity and better alignment with other attacks record, we reformat the date to 'yearmmdd'. For example, '3/9/2015' will become '20150309'. Moreover, we need the hour to compute the peak time, thus we also extract the hour information in integer.

Attack History Acquisition: We download the attacks history linked to ISIS (also known as ISIL) from *Wikipedia page*. It is stated that since June 2014, over 70 terrorist attacks in 20 countries are produced or encouraged by ISIS [1]. In this paper, we only crop the data which is within the same period of time as our twitter records. In total, there are 41 attack events, causing more than 5000 death and injuries.

C. Network Construction and Node Importance

Propaganda Network: We build a weighted direct graph using all the user names previously defined, which we call it a ISIS propaganda network, whose purpose is spreading ISIS aggressive idea and recruiting people. In this graph, the nodes are all the unique user names included in the posting and mention list; the edges stands for the twitter behaviors of mentioning; and the weight of each link is the mention times across all the tweets. The visualization of this network is illustrated in Figure 2(a).

The nodes scaling as well as color intensity are mapped to the out-degree of each nodes, meaning how frequently the user posts tweets and mentions others. It reflects the activeness, and character importance among all the nodes when virally spreading information.

Twitter Initializer Network: In this sub network of the whole propaganda network, we focus on the users who initialize the tweets, who are regarded as the core members in this propaganda network. The graph considering all these users is visualized in Figure 2(b). However, we find that some independent users actually didn't interact with the network. Thus we further visualize the network describing the mutual interactions among them by filtering some independent users, as depicted in Figure 2(c). The importance of nodes in both graph is ranked by betweeness centrality. Such visualization will help us point out the active accounts and analyze them. Besides this, clusters and communities become more obvious after filtering.

D. Hashtag Frequency Analysis

As we discussed, hashtag nowadays become the most effective tools for ISIS members to spread their news and avoid being banned. Hashtags will be able to reveal the potential attack, ISIS ongoing activities and future plans. For example, after Nice attack in France, the members of ISIS encourage their followers to hashtag '#NiceISIS' for wide spreading, even skews the twitter tag trend.

The method we compute the frequency of hash tag is simply summarizing all the occurrence of unique tag in tweets. The statistics server as, first, an understanding of their tag trend during this time, and secondly, a preparation for analyzing and predicting the attacks.

E. Peak Time Analysis

In twitting peak time analysis, we mostly emphasis on either frequency across days or hours. That is to answer when

the peak time is for everyday twitting, and whether the twitting date follows some patterns. The result will help us pay more attention to some time period for further data collection, and relate the patterns to the attack prediction.

F. Attack Prediction

The information we use for prediction is high-frequency hashtags with their frequency, and high-importance user with their involvement in one-day tweets, as well as the total tweets number for one day.

For the hashtags, we rank them by frequency, and select top 44 by thresholding on frequency 40. The dimension of feature vector is 44 with each entry indicating the number of specific tag occurrence. For the people involvement, we only care about the 113 core members who tweet. The dimension of this part is 113 with each entry referring to the times of involvement. (i.e. either being mentioned or mentioning others). We then further concatenate these two vectors, to form a 157-d feature vector for one specific day.

According to the records, the tweets before Jan. 2016 are quite sparse, which may not be sufficient enough for attack prediction. That may result from the various sigh-up times of important accounts. In this case, we have to shrink the studying data to a shorter time period, that is the last three months, 94 days in total. Hopefully, we can disentangle the underlying latent factors of causing the attacks by very limited twitting people and their tweets.

For the output, we compute the average severity of attacks by weighted summing the death and injuries, with ratio 1 and 0.75, and divided by the attack number within a day.

We develop two models for analysis. First, a static model realized by generalized regression neural network is implemented. We conduct day-wise regression with this model. However, this model may not be able to reveal the correlation between information among days, and cannot achieve a real prediction. Therefore, secondly, a dynamic model with time-series neural network is applied. In this model, we predict the attack by the tweets starting from 2 days before.

We conduct a two-fold cross validation with the limited data, by separating them into two continuous time periods, during each one there are 47 days.

IV. Result

In this section, we describe the results of networking experiment and attack prediction, and try to answer the following questions:

A. Who is the Leader in the Network?

In the whole propaganda network, the ranking suggests that *'Uncle_SamCoco'* plays the most important role. The less important users are listed in Fig. 3. As shown in the figure, *'Uncle_SamCoco'* surpass all the others in sending out and mentioning others, who is the most active one.

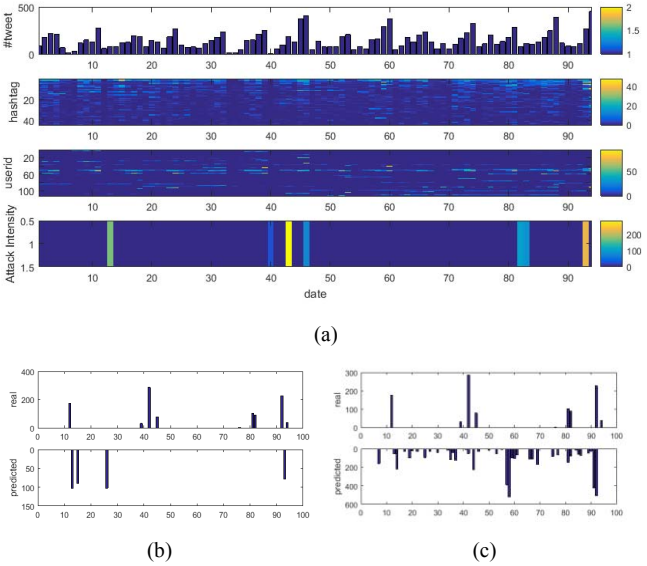
When comes to the initializer network, containing only the core members, *'Uncle_SamCoco'* is also treated as the most important one with the highest betweeness centrality. *'QassamiMarwan'* becomes a typical hidden member, who seldom tweets but essential in connecting everyone.

In the reduced core member network with only mutual interaction, we find that *'Uncle_SamCoco'* and 'WarReporter1' play identical essential roles in their own communities, and in the meanwhile, separate community in this core-member network can be clearly visualized now.

B. When and What Do They Mostly Tweet Every Day?

After processing and ranking the hashtags they frequently used when twitting, we find that, not surprisingly, 'ISIS' with its similar form is the most frequent term for attracting people and clustering information. Besides this, 'Syria' and 'Iraq' are two terms with relative high frequency. That may because most of that attacks and activities happen in their supreme headquarters. We also list some other terms, like 'BreakingNews', which is mostly followed by a video link in the tweets related to previous attacks. The relative frequency of each tag is depicted in Fig. 4. However, some tags with smaller frequency may not mean it is useless. For example, after Paris attack, most of them tag 'Paris' or 'ParisIS' to propagandize, which influences plenty of twitter users, even thought its overall frequency is low. Therefore, in future work, statistics of term density during specific period of time could be an alternative way of analysis.

In addition, data reveals that they mostly tweet at night, around 7pm to 8pm, as shown in Fig. 5. That may result from a designed and planned twitting time. Such a result suggests they are actually on purpose, well-organized and elaborately-planned for social network propaganda, instead of twitting in leisure time. Similar observation can also be realized in twitting date, as shown in the first row of Fig. 6(a). It shows a seemingly periodic pattern on the tweets number. On each weekend, there seems to be a peak twitting time.



(a) (b) (c)

Figure 6. Visualization of features and prediction result. (a) Row 1 shows the tweets number for each day, and row 2 shows the hashtags with their frequency on that day, while row 3 represents the users involved, and the last row shows the attack with its severity. The color space represent the number of occurrence or intensity. (b) is the static model prediction result, and (c) is the dynamic model prediction result.

C. Are the Attacks Predictable?

We plot the feature vector, which includes tweets number, tag frequency, and people involvement, also attack severity for visualization purpose, to help us gain some intuitive sense on these limited data. As shown in Figure 6(a), the features along the time axis are periodic, and there are bright spots around each attack time point. A more scientific prediction is required.

We predict the attacks by both static and dynamic model. The prediction result for each model is shown in Figure 6(b), (c). Since we conduct a two-fold cross validation, the regression result shown are concatenated two-fold regression value from two testing sets. In the static model, it suggests a hysteresis quality and information loss, which the prediction result lags the real result and are too sparse. That may result from the bulk tweets after each serious attacks. In contrast, the dynamic prediction, which considers the previous two days' tweets, are real time and dense. It reveals a similar pattern to the real records.

Our result shows the prediction model, especially the dynamic one, with limited data is not perfect, but satisfactory. It successful captures some attack patterns, reminding the public of the time in danger. However, since both the tweets and attack records are still sparse, even preferable enough, but making it hard to validate the model. Moreover, some important twitter accounts have already been banned by twitter after the data was downloaded, making the account tracking harder. Thus digging out real fanboys' accounts and tracking them may be helpful. Sudden and rapid clustering should also be more informative for a potential attack, which is a good direction for future work.

V. CONCLUSION

In this paper, we analyze the pro-ISIS fanboys network by 17k+ tweets over a year. We firstly construct the propaganda network by extracting user names and their mentioning relationships from tweets, and dig out the most active member by ranking the node importance. To analyze their twitter behaviors, we further summarize the frequency of hashtags and twitting time. Finally, we utilize them to predict potential future attacks by a satisfactory dynamic neural network model. It shows that prediction based on twitter history of ISIS is possible and desirable. In the future work, more emphasis can be put on the twitter contents themselves, and network rapid clustering behaviors across time, to make the prediction more accurate, and save more lives in this planet.

ACKOWNLEDGEMENT

This work is based on the project of "Analytics and Systems for Social Media and Big Data Applications", at the department of Electronic and Computer Engineering, HKUST. The author is grateful to Professor James She, Mr. Ming Cheung and Xiaopeng Li for their help and suggestions.

REFERENCES

[1] *Islamic State of Iraq and the Levant.* Available: https://en.wikipedia.org/wiki/Islamic_State_of_Iraq_and_the_Levant.

[2] *ISIS Has A New Twitter Hashtag For Threats Against Americans.* Available: http://www.businessinsider.com/isis-is-using-twitter-to-make-threats-to-us-2014-6.

[3] 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, 2015.

[4] S. Maher and J. Carter, "Analyzing the ISIS 'Twitter Storm,'," *War on the Rocks*, 2014.

[5] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Networks*, vol. 2, pp. 568-576, 1991.