A Network Based Approach to Disrupting ISIS' Lines of Communication via Twitter Analysis

Final Report

Elion Sufaj*
esufaj@yahoo.com
York University
Toronto, Ontario, Canada

Kanika Parikh
kanika22@my.yorku.ca
York University
Toronto, Ontario, Canada

Zeel Patel
zeel99@my.yorku.ca
York University
Toronto, Ontario, Canada

Omar Saeed
omarsaeedmirza@gmail.com
York University
Toronto, Ontario, Canada

1 Abstract

Terrorism has been one of the largest problem we as a society have faced in recent years. In the past few decades terrorists organizations have had devastating effects on society, especially in Europe, Middle East and parts of South Asia. They have caused irreversible damage on peoples lives, and even today there are millions of people fleeing from their homes in search of basic human necessities such as food, water and safety.

A lot of work and effort has been put into the war against terrorism in the Cyber world. However, despite all this work, identifying Cyber terrorist networks remains a challenge. Our goal is to uncover ISIS presence on twitter and identify their networks. We will be using techniques such as: **Social Network Cluster Analysis**, **Community Detection**, **Temporal and Spatial Analysis**. These enable us to create networks with scaled nodes of users making pro-ISIS tweets to size a proportionate to their relevance in the network, determine the major hot-spots for the tweets, and using specific algorithms for community detection we determine interconnections of users in the network and separate them based on their modularity. Our study has found top 20 most active ISIS accounts, top 10 most pro-ISIS zones around the globe, and top most talked topics in pro-ISIS tweets.

This paper discusses in-detail the models, algorithms and results of our experiments, as well as their analysis.

2 Introduction

The Islamic State of Iraq and Syria (ISIS) is the world's largest jihadist terrorist group which was founded by Al-Zarquawi in 1999. It has received global recognition as a militant terrorist group in 2014, after it took possession of prime cities in Iraq. From this time on, it has become one of the primary goals of the US government and other nations to keep an eye on ISIS activities and suppress its influence. Since then, ISIS has branched out it radical ways and started to use social media platforms such as Twitter, YouTube and Facebook etc. These platforms are advantageous for Pro-Isis users to share graphical contents i.e execution details and recruit their followers. Twitter is one of the most popular micro-blogging social media platform where audiences discuss trendy topics in the form of tweets. On twitter, users can follow or mention other users, reply to a thread or retweet any tweet. It is an optimal way to communicate over an open network anonymously, hiding in plain sight.

Our ISIS data set [6] focuses on tweets after the 2015 Paris attack. Our goal is to understand the pro Isis network by analyzing the structure of the network, degree distributions, text sentiments and other network properties and attributes. Few researchers have proposed some schemes to explore the ISIS twitter network and have enlisted its influence on society. Our plan of action for this network analysis is more interesting as we intend to reveal the major players in the pro-ISIS Twitter community by determining how these users communicate, facilitate and execute missions and plans as part of their organized criminal network.

In this report we will paint a clear picture of the interconnections of various users as well as the communities they belong to. By removing very specific key individuals from the network, we will be able to severely mitigate and halt the spread of information. We also aim to find if there are any cross-community interactions and if they change with the time (i.e. temporal variation in data).

The project report is arranged as follows: **Section 2** defines the main problem.**Section 3** presents related past research, theoretical concepts and mathematical findings. **Section 4** outlines the detailed approach we took for this network analysis. Findings from the experiment are reported in **Section 5** which is followed by the conclusion of the analysis in **Section 6**.

3 Problem Definition

Our project focuses on uncovering ISIS networks of communication on twitter. There are number of features that can help us dismantle and disrupt communication lines of ISIS

*All authors contributed equally to this research.

supporters. In this project, we intend to explore and report the outcome of the following features on our analysis:

- Construct the ISIS network, determine the major players in the community and how they are connected to other users.
- Given the directed weighted graph of pro-ISIS users, find the sub-communities and how interconnected they are.
- Given the tweets, analyze how information travels within the ISIS twitter network and how it varies with time.
- To disrupt the ISIS communication line by identifying the key player and measuring the impact it has on the information spreading.

4 Related Work

4.1 Examining ISIS support and Opposition networks on twitter

Examining ISIS support [3] used lexical and network analysis in an iterative approach to identify different communities within the twitter ISIS world. In their study, they made use of community detection algorithms to classify 23 million tweets from 771,321 users into 36 distinct communities and ultimately 4 **meta-communities**. They further analyzed these communities using lexical analysis to characterize the identities and themes of these meta-communities.

The result of lexical analysis showed that the 4 meta-communities belong to Shia, Mujahideen in Syria, ISIS supporters, and Sunni. However, not all these groups are pro-ISIS. The Shia group condemns ISIS because of their Islamic terms and links to Saudi Arabia

4.2 Defining and describing the ISIS population on twitter

Berger and Morgan identified 46,000 ISIS supporter accounts on Twitter [1]. They began with 454 accounts known as ISIS supporters as initial seeds. They then collected all accounts following those seeds. This approach continues to two further steps on the followed- by network. They introduced a classification task to determine if a user is an ISIS supporter.

4.3 “Follow Me So I Can DM You Back”: An Exploratory Analysis of a Female Pro- ISIS Twitter Network

Varanese [3] conducted this study to explore a network of female pro-ISIS supporters on twitter. Varanese identified 20 accounts through snowball sampling, and analyzed their network consisting 5,861 vertices and 12034 edges. The network was studied using 3 social media network analysis metrics: Freeman’s normalized betweenness centrality, average geodesic distance, and tie strength.

The results of this study showed that females were far more influential than males, and had a much greater ability to radicalize other females within their network.

5 Methodology

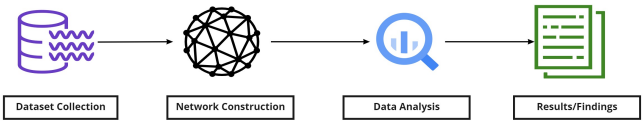

Figure 1. Step-wise approach towards network analysis

5.1 Data-Set Collection

For our network analysis, we take advantage of the existing public Twitter data-set [6] provided by Kaggle. Our Isis tweet data-set contains around 17,000 Pro-Isis tweets from 2015 to 2017 in a CSV file. The fields consist of **name, username, description, location, number of followers at the time the tweet was downloaded, number of statuses by the user when the tweet was downloaded, date and timestamp of the tweet, and the tweet itself**.

5.2 Network Construction

To construct our network, we use a multitude of Python scripts and underlying packages such as networkx and pandas. Initially, we read the CSV file and extract the information into a pandas data frame to utilize the packages efficient data manipulation features. The first step in assembling the network was to get all of the nodes. There are two sets of nodes in this network. The user that tweeted and the user that got mentioned in the tweet. In order to get the second set of nodes, we needed to extract the mentioned user from the rest of the tweet. We did this by using some regular expressions and splitting the text based on the "@" symbol. Everything following the "@" symbol would be the mentioned users username. After extracting the mentioned users, we created a new column called "Mentioned-User" and placed them in their respective rows.

We took this a step ahead and assigned every row in the data frame a new weight value of 1. This will help in network construction later on when determining edge weights. Next, we grouped rows by user and mentioned user and added the weight values along the way giving us an accurate representation of the number of interactions occurring between users. Finally, we constructed a directed weighted network by creating an edge from the user that tweeted to the user that was mentioned in the tweet and assigning their associated edge weight using some method attributes to do this automatically.

5.3 Visualization

In order to visualize our network and connections between pro-ISIS users, we created and exported an edge list to a CSV file from the existing data frame in Python. I then inputted this edge list into Gephi, a network visualization software which is useful for maintaining and highlighting key points in a network. Nodes increase in a size proportional to their degrees, and have also been colour coded based on their modularity. That is, nodes belonging to the same community share the same colour.

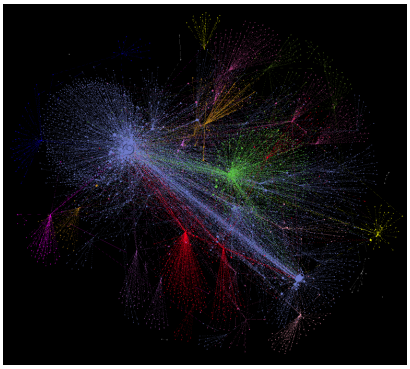

Figure 2. ISIS Twitter Network

5.4 Community Detection

Our first step in analyzing the network is determining the communities that reside within it. To do this, we used a community detection algorithm known as the Louvain modularity algorithm. This algorithm was the preferred choice due to the use of its greedy approach allowing for faster more efficient calculations of increasingly large networks [7].

	Karate	Arxiv	Internet	Web nd.edu	Phone	Web uk-2005	Web WebBase 2001
Nodes/links	34/77	9k/24k	70k/351k	325k/1M	2.6M/6.3M	39M/783M	118M/1B
Clauset, Newman, and Moore	.38/0s	.772/3.6s	.692/799s	.927/5034s	-/-	-/-	-/-
Pons and Latapy	.42/0s	.757/3.3s	.729/575s	.895/6666s	-/-	-/-	-/-
Wakita and Tsurumi	.42/0s	.761/0.7s	.667/62s	.898/248s	.56/464s	-/-	-/-
Louvain Method	.42/0s	.813/0s	.781/1s	.935/3s	.769/134s	.979/738s	.984/152mn

Figure 3. Louvain Algorithm Run-Time Comparison

When compared to other modularity optimization methods and operating on a variety of differing network models and datasets, we can see that the Louvain method consistently outperforms other methods of its kind in the same category. Note that a lower number is better [7].

Partitioning the network using modularity which can be defined as:

$$\mathbb{Q} = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta\left(c_i, c_j\right)$$

allows us to find the most optimal clustering of nodes per community by measuring the density of edges within each community when compared to a null model using the resolution parameter provided by the algorithm.

When running this algorithm on our network of Twitter ISIS users, we are given a resulting 30 communities with a modularity Q of 0.5479 as shown in **Figure 4**. This tells us that our community is relatively dense with each user being closely connected with one another in their respective communities [7].

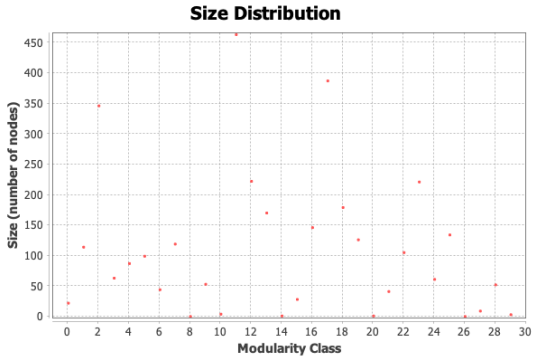

Figure 4. Distribution of Community Sizes

The five largest communities in the network are shown below:

Community	Users
Community 11	464
Community 17	388
Community 2	347
Community 12	223
Community 23	222

As shown from **Figure 4** and the table above, most communities are above average size, having at least 50 members, making each community relatively well connected. Despite there being many small communities with a low number of users, the top 5 largest communities account for approximately 50 percent of all of the users in the network.

5.5 Centrality Analysis and Modelling

The second problem is to determine how well connected every user is in the network. Using edge weights as previously described earlier in this section, will tell us which individuals are the most well connected and popular. Although this measure will not tell us if these users are influential to other users, we can fix this by using **EigenCentrality, betweeness and closeness** [2]. Another key component will be to determine the **hubs** and **authorities** within the network. This is especially useful in **scale-free networks**, which is applicable to our network. The hub nodes, which have a high out-degree will be central players in spreading information, acting as "bridges" to the rest of the network [4, 5]. Authority

nodes, which have a high in-degree are crucial members of the network, usually commanding some authority from the communities they belong to [4, 5]. Modelling will be vital to pulling insights from the network. A few models which will be employed in our research will include degree distribution, temporal/spatial analysis, connectivity models, frequency distributions and many more.

5.6 Suggestions for Surveillance or Removal

The last problem this section seeks to address, is using the insights discovered in our research in a useful and meaningful way. Based on the results from our community and centrality analysis, we will be able to determine to a very high degree which users pose the greatest security threat and how to use this information to effect some real change. By analyzing the nodes with high in-degrees and high out-degrees, we can determine which nodes act as hubs or authorities. Authority nodes are perfect candidates for security surveillance [4, 5]. The number of traffic that comes to them can let observers know many details of their plans and actions. Hub nodes on the other hand are important users in spreading information throughout the network. They have a high-out degree, which means they are extremely well connected and act as bridges, connecting many parts of the weakly connected components in a graph. Removing these individuals from the network, slows the spread of information, disrupts lines of communication and significantly reduces the size and complexity of the network [4, 5].

6 Experiments / Evaluation

6.1 The Network

We constructed a weighted, directed graph. The nodes are users and the users mentioned in tweets. The edges are weighted according to the frequency of communication between those individuals. We describe the crucial graph properties of our network analysis in **Section 5.3**.

6.2 Experiments

All of our experiments were carried out on a MacBook Pro with an Apple M1 chip and 16GBs of RAM using Python version 3.8.5 to run our code. Most of the visual models were created using MATPLOTLIB, NETWORKX, PANDAS, and COMMUNITY. Some of our network models and graphs were creating in the external visualization software known as Gephi.

6.3 Analysis Results

We have performed a number of calculations and experiments. Firstly, we will look at the largest connected component(LCC) of the network. When finding the LCC, we noticed that the number of nodes it contained was exactly 3310 nodes. This is definitely interesting when considering that the original network itself is comprised of 3330 nodes. Thus, this tells us that only 20 individuals in the entire network were secluded from the everyone else and did not partake in the larger community. This is definitely alarming as it tells us that almost every single person in the network is connected to and can reach one another by following who their friends follow.

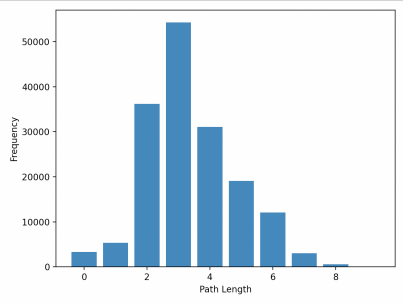

Figure 5. Distribution of Shortest Path Lengths

Looking at **Figure 5**, we can see that most individuals in the network are separated by three jumps, followed by two jumps and then four jumps. Not only is most of the network connected to one another, but the degrees of separation for those individual is very low. From this, we can conclude that on average, information has the ability to travel to everyone else in the network by about three to four hops, making the organization and interconnections in the network very dangerous.

Furthermore, we know that in scale-free networks, we know that a small number of nodes have high degrees with many nodes having low degrees [5]. We can see this from **Figure 6** below

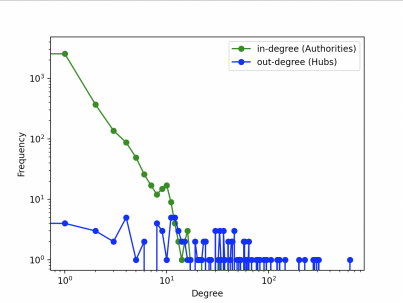

Figure 6. Distribution of Shortest Path Lengths

From **Figure 6**, we can see that there are a high number of authorities in the network but with a considerably small degree. In fact, there are only a very small number of authorities that also have a high degree. This means that there are only a few particular individuals in the network that truly command some authority but also command that authority to a very high rate, being mentioned by hundreds of users. The hubs on the other hand, are low in number but with

immensely high degrees. This tells us that there are a few key individuals responsible for most or all of the information spreading, planning and organization of the entire network. Determining who these individuals are and eliminating them from the network would be key in disrupting their lines of communication.

Now, let us take a deeper dive into who these people really are. Using the weights previously assigned to each user in the network, we are able to use pandas or networkx to determine who the authorities and hubs are by extracting those users with the highest weights.

Mentions	Weight
RamiAlLolah	573
Nidalgazaui	339
WarReporter1	256
7layers_	116
ScotsmanInfidel	79
sparksofirhabi3	76
Conflicts	72
MaghrebiQM	72
DidyouknowVS	72
Uncle_SamCoco	70

Figure 7. Top-10 Authorities

From **Figure 7**, we can see user RamiAlLolah is the highest authoritarian in the network followed by Nidalgazaui and WarReporter1. Removing any of these individuals from the network is sure to disrupt organization and leave those looking for instruction lost. Now, lets consider **Figure 8** below

User	Weight
Uncle_SamCoco	1617
mobi_ayubi	1078
WarReporter1	713
warrnews	665
RamiAlLolah	655
melvynlion	642
MaghrabiArabi	454
_IshfaqAhmad	432
AsimAbuMerjem	289
NaseemAhmed50	251

Figure 8. Top-10 Hubs

As we can see, Uncle_SamCoco is the largest hub in the network, followed by mobi_ayubi and WarReporter1. These individuals are key players in spreading information throughout the network. Removal of any of these individuals would hinder and disrupt the major lines of communication from user to user. Something that we can notice from **Figure 7** and **Figure 8**, is that those individuals in the top ten authorities and top ten hubs are repeated. Particulary, RamiAlLolah, WarReporter1, and Uncle_SamCoco are both authorities and hubs. Rather than removal, these individuals are prime candidates for surveillance considering the major online presence and activity from them. These individuals are most likely to reveal some important information which may be useful in prevention and mitigation of dangerous activities.

Furthermore, let us look at the highest user to user communications in the network.

User	Mentions	Weight
mobi_ayubi	RamiAlLolah	195
warrnews	Nidalgazaui	184
WarReporter1	WarReporter1	121
melvynlion	ScotsmanInfidel	79
warreporter2	DidyouknowVS	70
melvynlion	sparksofirhabi3	63
melvynlion	spicylatte123	61
melvynlion	Ele7vn	58
melvynlion	1_texanna	56
melvynlion	sassysassyred	54

Figure 9. Highest User-to-User Activity

From **Figure 9** we can see that mobi_ayubi and RamiAlLolah are one of the highest communication pair users in the network. Furthermore, a new user that is neither a hub or authority is melvynlion. This user has multiple communications with a high degree of people at once. They may be key for further surveillance in the network. A high mention rate such as the one by melvynlion could indicate an individual that executes plans on behalf of some authority within the network. On another note, we can see that WarReporter1 mentions himself often which is strange. Are multiple individuals sharing the same account? Rather than multiple users tweeting at eachother from different accounts, they could all share the same account and tweet themselves which can then be read later by another user on the same account. We can also see a user by the name of warreporter2 and warnews. Is this an alias of WarReporter1, or a friend or subordinate of WarReporter1. Further textual and sentiment analysis of the warreporter and warnews accounts would prove very insightful.

From the discussion above, we can see that a few major players in the network are repeated in multiple priority categories. Specifically, RamiAlLolah and Uncle_SamCoco are of major interest.

From **Figure 10**, we can see that RamiAlLolah has hundreds of edges to and from hundreds of individuals, most of the edges being incoming. As previously suggested, removal of this individual would significantly reduce the size of the network and leave the hierarchical order of power stagnant without proper leadership and organization. In addition to this, the largest spreader of information in the network Uncle_SamCoco is suggested for removal or surveillance. As we can see below,

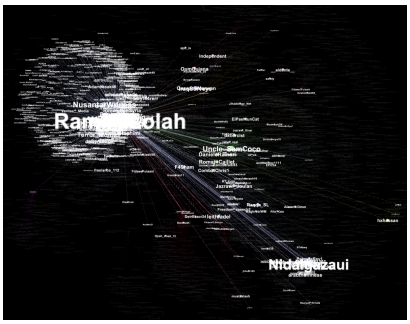

Figure 10. Degree Network for RamiAlLolah, The Highest Authority in the Network

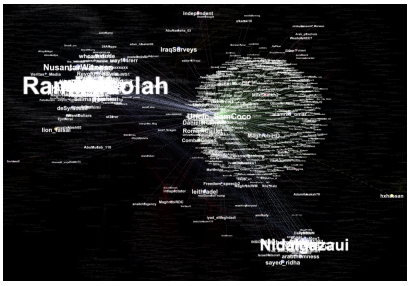

Figure 11. Degree Network for Uncle_SamCoco, The Largest Hub in the Network

Uncle_SamCoco has hundreds of outgoing edges to hundreds of nearby users all containing potentially damaging and destructive information. What we know about how orders are planned and executed is also confirmed throughout these individual networks. It seems that Uncle_SamCoco delivers information to RamiAlLolah and Nidalgazaui, then RamiAlLolah delivers orders or information to Nidalgazaui. Could Nidalgazaui potentially be RamiAlLolahs close friend or second in charge? We can determine this by the prominently defined edges from Rami to Nidal. As has been described previously, edge size increases proportionally to its weight. Thus we can conclude that Rami and Nidal are potentially coordinating or organizing plans frequently. This makes sense from previous discussions held, as we can see from **Figure 7** that the number one and two highest authorities in the network are none other than Rami and Nidal.

Removal of these two individuals not only severely disrupts the network, but reduces the size of the network greatly. The travel of information is disrupted, the size of the network is reduced and security is maintained.

We obtained the results in **Figure 12** by performing a query over locations in our data set. This data identifies the most popular hot-spots for pro-ISIS tweets. However, after studying these results and the data-set, we can say that these locations might not be completely accurate as the location data does not come from twitter but the users themselves. Despite some inaccuracy in the data, **Figure 12** gives us a satisfactory overall understanding of the origins of these networks present on twitter.

```
nan 5978
Read my blog 1475
world 1191
Worldwide contributions 998
Texas, USA 993
ولاية خراسان 709
United States 587
28th Street, Qamar Precint 426
Germany 397
Dunya 366
Male'. Maldives. 333
S.Wazirstan|Mahsud not a Wazir 318
Middle of Nowhere 303
dar al-kufr 285
Nowhere 234
Wilayah Kashmir 231
München, Deutschland 193
Bandar Seri Begawan, Negara Brunei Darussalam 182
Geneva, Switzerland 166
Ghurabā 159
Among The Muslims 152
Gaziantep, Turkey 146
Wilayat Hadramaut 134
Singaparna, Indonesia 100
أرض الدين 100
Antas, Bahia 99
Land of Allah 97
EU 97
Earth 96
Punch, Jammu And Kashmir 95
Among mushrikeen 93
Wilayah Twitter 89
Al-Battar Media Foundatiom 80
Lake City, GA 68
Guetto 58
I hate snitches 49
darl mushrequeen 47
Amsterdam, The Netherlands 45
العراق الخلافة 42
Don't need to know 40
England, United Kingdom 27
AP 25
سرت Sirte 23
40+Suspension for the truth! 23
noway 15
Dar Al Kufr 14
Islamic State 14
yamin, yasār raqum ١٤ 10
Prison ( Darul Kufr ) 4
. 4
```

Figure 12. Top origins for Pro-ISIS tweets

From **Figure 13** we can see the overall activity of tweets within a 7 month time span. We can notice some major irregularities within this data, which correlates to significant terrorist and political activity within the region.

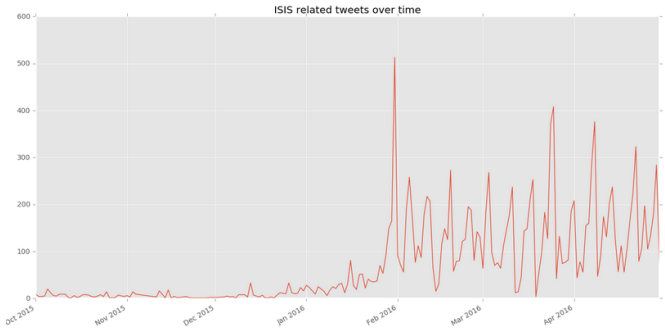

Figure 13. Frequency graph of tweets over a 7 month period in 2015/16

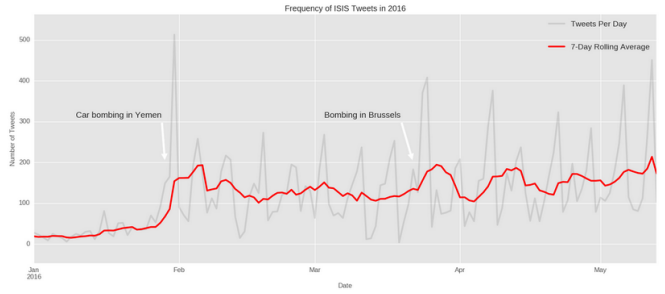

Figure 14. Frequency graph of tweets over 2016 with respect to noticeable events

From **Figure 14**, we can see that the same graph but now focus on the irregular spikes along with major terrorist events that took place during that time. After taking a close look at the graph, the initial spike in tweets during correlates to bombing in Yemen. The second largest spike happens around mid-march 2016; this one accounts for bombing in Brussels. We can notice that the graph remains very unpredictable and spontaneous with fast changes during April 2016. This randomness is better explained in **Figure 15**.

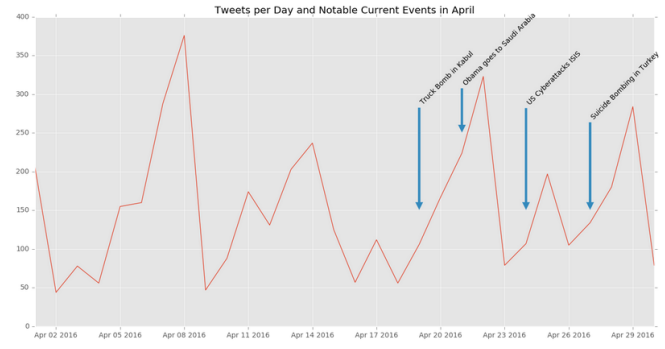

Figure 15. Frequency graph of tweets over 2016 with respect to noticeable events

Random spikes in April 2016, are not only linked to terrorist events but also political ones. **Figure 15** shows some of the major events that took place during that time. We can see that for almost every spike we had in tweets per day, it was related to some kind of political or terrorist activity in the region. These findings make it obvious that the the pro-ISIS networks were most active whenever something significant happened in the region.

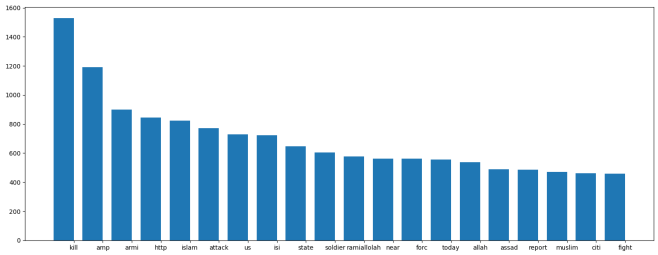

Figure 16. Frequency of 20 most mentioned words

Figure 16 was obtained by performing textual analysis on tweets in our data-set. This graph represents the top 20 subjects of discussion in the tweets. The top most mentioned subject was about killing, which makes sense given the context, but the second and third most talked about topics remain somewhat ambiguous. Number four is 'http', which makes it to our results because the users might be sharing website links in their tweets. Some alarming key subjects are: Kill, Attack, Fight and US. Using this information, we can conclude that the users were trying to gather support and rage the readers to flight for their cause.

6.4 Additional Work and Research

Further work on this topic will be attempted for the final iteration of this report. We plan to include calculations including HITS for determining the most influential nodes in the network. We will also be calculating temporal and spatial data to determine where and when the main hot-spots activity zones for the pro-ISIS users are, and sentiment analysis to determine how users that tweet increasingly negative things either gain or lose support from the larger network. In addition, we plan on calculating Eigenvector centrality, betweeness and closeness to help with discovering bridges and which nodes have influence over other nodes. These bridges will help in disconnecting the network and reducing the spread of information. Finally, a plan and attempt was made to create a base truth model in which we create networks for tweets from different years and then compare calculations on those different networks, but we found that the number of tweets from 2015, 2016 and 2017 varied drastically and that it would not have provided an accurate measure of things.

7 Conclusions

As a rundown of our study completed so far, we have constructed a weighted directed graph, where nodes are the users that tweeted and users who got mentioned in the tweet. Our network and the data-set has been tested and experimented on several times using the various methods mentioned previously. The most extra-ordinary finding we have so far is pin-pointing the top 3 authorities in our network, ready for surveillance.

We are not able to compute the most popular nodes in our network based on a function of time yet, as our team had difficulties defining an algorithm to perform this task efficiently. However, we plan to have this feature integrated in our research by its completion. For now we are able to visually represent pro-ISIS tweets as a network, compute the most mentioned accounts, compute accounts with the most number of tweets etc, which is a significant milestone towards our final goal.

References

[1] Jonathon M Berger and Jonathon Morgan. 2015. *The ISIS Twitter Census:Defining and describing the population of ISIS supporters on Twitter.* Retrieved March 08, 2022 from https://towardsdatascience.com/how-isis-uses-twitter-10065790354a

[2] Andrew Disney. 2020. *PageRank centrality EigenCentrality.* Retrieved Feb 10, 2022 from https://cambridge-intelligence.com/eigencentrality-pagerank/

[3] Madeline Magnuson Zev Winkelman Elizabeth Bodine-Baron, Todd C. Helmus. 2020. *Examining ISIS support and Opposition networks on twitter.* Retrieved March 08, 2022 from https://www.rand.org/content/dam/rand/pubs/research_reports/RR1300/RR1328/RAND_RR1328.pdf

[4] Mitchell Telatnik. 2020. *Community Detection of ISIS Twitter Accounts.* Retrieved Feb 10, 2022 from https://towardsdatascience.com/how-isis-uses-twitter-598c2eb188a2

[5] Mitchell Telatnik. 2020. *How ISIS Uses Twitter.* Retrieved Feb 10, 2022 from https://towardsdatascience.com/how-isis-uses-twitter-10065790354a

[6] The Fifth Tribe. 2020. *How ISIS Uses Twitter (Data-Set).* Retrieved Feb 10, 2022 from https://www.kaggle.com/fifthtribe/how-isis-uses-twitter

[7] Wikipedia. 2022. *Louvain method.* Retrieved Mar 11, 2022 from https://en.wikipedia.org/wiki/Louvain_method