<u>Social Network Analysis of ISIS Twitter Network</u>

Eng.Tony Chahoud, Ms in Artificial Intelligence

Abstract-- Terrorist organizations are increasingly using social networks to distribute propaganda and recruit new members. Understanding the ways in which these organizations utilize social media can help predict and prevent potential terrorist attacks. This document focuses on analyzing over 17,000 Twitter records of ISIS accounts to identify the most active and influential members of the network, as well as to examine how content spreads through the network. We have also identified the most important members in the network and analyzed how the network will be affected by removing them. Also, by dividing the network into distinct communities, we gained a deeper understanding of how ISIS operates and what are the main keywords in their Twitter communication.

Context

ISIS, also referred to as ISIL, is a terrorist organization that emerged in the wake of the United States-backed NATO invasion of Iraq in 2003, which led to the destabilization of the region and the overthrow of the government. In the aftermath of the chaos, ISIS took advantage of the situation and began seizing control of cities in Iraq. By 2014, the group had gained notoriety for carrying out devastating attacks that resulted in the deaths of thousands of innocent civilians worldwide[2]. Today, ISIS remains active, with a significant presence in Afghanistan and attempts to pose a threat to Pakistan.

ISIS members are experts in using social networks, especially Twitter, which is widely used worldwide. They primarily use Twitter for recruitment, attracting thousands of foreigners to join, making the fight against terrorism even more challenging. Initially, they created official Twitter accounts, but these accounts were soon banned by internet managers. After 2017, they began to hide under unofficial accounts. Analyzing their tweeting rules and tracking the essential hidden members are urgent tasks to predict future attacks, prevent them from happening, and ban those active accounts.

Understanding the network of ISIS activities on Twitter can help us examine relationships between social elements. This analysis, known as social network analysis (SNA), uses graph theory to understand the structure of the network. The network structure is made up of social entities, which are frequently individuals and communities. In social network analysis, the empirical investigation of social entity interactions within the network is done. The network structure is created using nodes, which can represent any entity, such as individuals/actors and edges, which show connections between nodes and their interactions.

Implementing such a network can help uncover who the primary individuals/communities are, how content propagates through the network and what will be effects if we remove the most important actors from the network.

Problem and Motivation

Social networks are being used by ISIS to distribute messages with the intention of influencing people, recruiting new members, and planning attacks. Understanding the network of ISIS activity on Twitter can help uncover the primary actors, how content propagates through the network, and improve the monitoring of ISIS communities.

In this report, we aim to address the following problems based on the dataset that highlight ISIS activities on twitter for a specified periodic time.

Firstly, we want to identify the individuals who play the most important role in propagating and recruiting for ISIS by ranking the in/out-degree of each node in the network. Additionally, we aim to determine the core members in the ISIS community by checking their betweenness centrality after filtering the network and classify those prominent communities as the Key/Core/Most active members.

So, we will analyze the authorities and hubs to find those key members wich are basically the centralities of the most active communities. Based on the communities obtained, we will pinpoint which nodes should be removed to disrupt the activity of the network effectively and ultimately ban them from Twitter. We also plan to track the activity of each community by extracting the most discussed topics per each community to build a simple static neural network model that can predict future attacks.

Datasets

The data used for this analysis was provided by Data Society on the website data.world.[1] The dataset was collected from 2015 to 2017 and contains 17,000 tweets from over 100 pro-ISIS accounts.

The features collected include *the name, username, description, location, number of followers at the time the tweet was downloaded, the number of statuses by the user when the tweet was downloaded, the date and timestamp of the tweet, and the tweet itself.*

name	username	description	location	followers	numberstatuses	time	tweets
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-02-26T20:40:00	DOWNLOAD #JN Video With English Subti
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-02-28T23:35:00	Video by @ansardeenfront with Eng Sub
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-03-04T07:31:00	ENGLISH TRANSLATION: A Pamphlet Relea
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-04-01T13:18:00	Aslm, anybody translating the new aud
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-05-13T21:16:00	@abuhamzaneh 050W30 @JN_Irhabiyya_97
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-05-29T09:53:00	RT @GIIMedia_CH004: Rules Of Imarah P
GunsandCoffee	GunsandCoffee70	ENGLISH TRANSLATIONS: http://t.co/QLd	No data.	640	49	2015-06-03T11:16:00	ENGLISH TRANSLATION: Dialogue with Sh
Abu Layth Al Hindi	AbuLaythAlHindi	Kik: abulayth2014. Ex South African.	Islamic State	68	18	2015-07-07T06:02:00	النشرة صوتية ▲ [الإثنين 1436/9/19 هـ] ▼
Abu Layth Al Hindi	AbuLaythAlHindi	Kik: abulayth2014. Ex South African.	Islamic State	68	18	2015-08-17T20:20:00	#BREAKING #CONFIRMED Islamic State ta
[illegible]	YazeedDhardaa25	Observing a JIHAD NEWS mainly about I	No data.	904	127	2015-08-26T17:23:00	Salamu Alaykum. I'm back. Do supports
[illegible]	YazeedDhardaa25	Observing a JIHAD NEWS mainly about I	No data.	904	127	2015-08-26T17:24:00	@YazeedDhardaa25 cc @ibn_muhas @Jazra

Figure 1: ISIS Twitter dataset

In our network, we are primarily interested in the two columns: "username" and "tweets". The 'tweets' column is the main one to perform text analytics noting that mostly of those tweets are translated from Arabic language and dialects to English.

To build a network graph of the Twitter accounts, a node list and an edge list must be created. The graph will visualize connections between accounts through mentions, and the edges will be weighted based on the number of mentions/tags to a user (inside of the message @user, @user2…). To generate the node and edge lists, a Python script based on NLP process was used to extract mentions from a tweet and create edges between the user who tweeted and the mentioned user.

Validity and Reliability

Reviewing some key terminology is necessary before we move on to the research on course assessments. In everyday conversation, words like "reliability" and "validity" are frequently used interchangeably. However, in scientific or statistical discourse, they have more specific connotations. When a measurement device, instrument, or test produces consistent results, it is said to be reliable. Validity, on the other hand, is the degree to which a concept is accurately measured in a quantitative investigation.

With this knowledge in mind, we can conclude that our experiment concerning the most active ISIS member on Twitter is both valid and reliable, given that the dataset was retrieved online, where connections between member activities are considered. The results to find the key members in the network are valid because the detection has been made considering multiple metrices. We also know that a lots of connections of a certain member/node in the network means that this character is more active, and our findings are well supported by various metrices related to the social network analysis, i.e., various centrality measures, cohesion and fragmentation measures (average clustering coefficient, modularity and density). These results are also reliable because we not only obtained the same results consistently but also the fragmentation measures suggested that by removing the detected key members the network has been disrupted effectively.

Measures

We constructed a weighted directed graph to represent the network of ISIS activity on Twitter, aimed at spreading aggressive ideas and recruiting new members. Nodes are all the unique usernames included in the posting and mention list and are generated using the Python-script discussed above in the data pre-processing part that extract mentions (@) from each tweet. Then we created edges between the user who tweeted and the mentioned user, where the weight of each edge corresponds to the number of times the user has been mentioned in tweets. Both the node and edge lists were fed into the network graphing and analysis tool, Gephi for analysis. So, we created the lists of edges and nodes using Python, and then imported it into Gephi for visualization. *Figure 2* illustrates the visualization of this network, reflecting the user's activeness and the way of spreading information.

Our goal is to identify the individuals who play the most important role in propagating and recruiting for ISIS by ranking the out-degree and in-degree of each node in the whole network, which will allow us to classify the most active ISIS users on the network by being tagged and tagging others simultaneously. Therefore, we used centrality metrics to get different properties of a network and its behavior. The more a node is centered, the more important it is.

Then we deduced the key/core members in the ISIS community by filtering the network and analyzing their PageRank and betweenness centralities. Additionally, we detected major communities in the network based on the modularity and highlighted the most frequently used words per community.

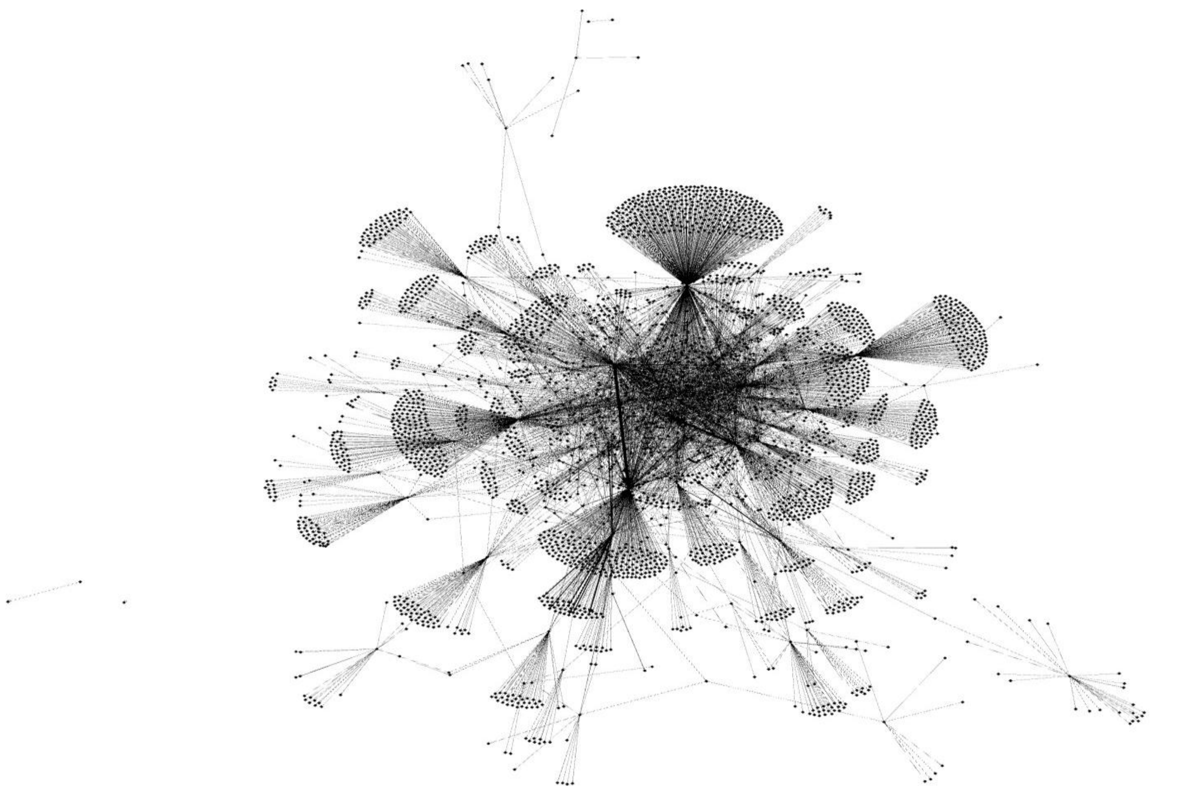

Figure 2: Network of ISSI Activities on Twitter

The above graph has 3331 nodes and 5397 edges. As our focus, in this analysis, is on identifying the most important nodes in the network therefore, we will not consider the fanlike structures in above figure. This happens when we have a node which is connected to a lot of nodes, and they don't have any other connection. So, it means that we have a hub with is connected to a lot of nodes with degree 1. This means that we are hitting the edge of our network. This can be distracting for the network analysis specially when we are aiming to find the most important persons in the network. So, we will filter it using "degree range" filter of Gephi. The resultant graph with 820 nodes and 2887 edges is as under: -

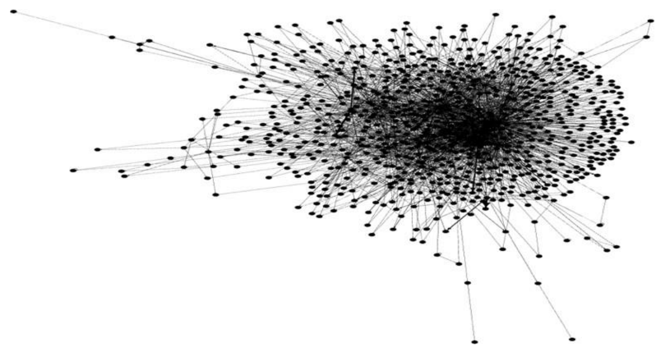

Figure 3: Network graph of nodes with degree > 2

Now, we have discharged all the nodes with degree 1. So, we have a network with degree 2 or higher. Generally, in a social graph the importance of nodes can be measured in multiple ways. In this network as mentioned before, we consider the following metrics: *Degree centrality, PageRank, Betweenness Centrality, Authorities and Hubs,* and basing upon the centrality metrices we identify the most important members of the network. To measure the *fragmentation* after the removal of the most important hubs and authorities, we will use *Avg. Clustering Coefficient, Modularity and Density metrices*. In the end, we dealt with *Community Detection*.

- Degree centrality

It measures the number of edges incident on a node. It is the sum of the in-degree and out-degree, which represents the number of edges entering and leaving the node, respectively. The nodes regarded as the most important under degree centrality are those with the most direct connections with others. As the network can be represented as a square matrix. The value of centrality degree is computed as follows:

$$C_{\mathbf{D}}((Vi) = \Sigma_{\mathbf{j}} M_{\mathbf{ij}}$$

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

As we are dealing with a directed weighted graph, our aim is to analyze the two separate concepts: *Hubs* and *Authorities*, based on the weighted degree of nodes.

An authority node has many incoming edges with high weights, while a hub node has many outgoing edges with high weights. In other words, an authority node is highly regarded or respected by other nodes in the network, while a hub node directs a lot of important information to other nodes.

- PageRank

PageRank is a variant of the Katz centrality where we derive the centrality of the neighbors as proportional to their centrality divided by their out-degree. Then, nodes that point to many others pass only a small amount of centrality on to each of those others, even if their own centrality is high. Mathematically calculated as below for a node xi: -

$$X_i = \alpha \Sigma_j A_{ij}\ od(j) + \beta$$

Here, i and j are two different nodes, A is the adjacency matrix, od is out-degree, α is proportionality factor, β centrality for free.

The Page Rank score for a given node is based on the links made to that said node from other nodes. The links to a given node are called the backlinks/in-degrees for that node. The social network thus becomes a democracy where nodes vote for the importance of other pages by linking to them. PageRank score will always be a non-negative real number and all the scores (in the network) will add to 1. Before beginning the calculation, we must remove the self-loops, otherwise, the Page Rank will not sum up to 1. For this, we have used "edges - self loop" filter in Gephi.

- Betweenness Centrality

To highlight the diffusion and the spreading of information of the hub and authority nodes discussed above, we will use betweenness centrality which measures how frequently a node is visited when moving between Nodes.

In this case, we used the HITS (Hyperlink-Induced Topic Search) algorithm to calculate the authorities and hubs of nodes in this directed graph with weighted edges. The HITS algorithm uses an iterative process to assign an authority and hub score to each node based on the scores of its neighbors. At the start of the algorithm, each node is assigned an initial authority and hub score of 1.0. The algorithm then iteratively updates the scores of each node based on the scores of its neighbors. During each iteration, the authority score of each node is updated based on the sum of the hub scores of its incoming neighbors, and the hub score of each node is updated based on the sum of the authority scores of its outgoing neighbors. The scores are then normalized so that the sum of the squares of the scores is equal to 1.

- Key Authorities and Hubs

After calculating the centrality metrices, we will extract the most important members in the network. By analyzing the detected authorities and hubs and extending the study of the degree distribution, we will pinpoint which nodes should be removed to disrupt the activity of the network effectively.

After removing the most important members from the network, we need to ensure that the network activity is disrupted efficiently by their removal. Therefore, we will measure the fragmentation through following metrices: -

- Avg. Clustering Coefficient

The clustering coefficient of a social network is a measure of the degree to which nodes in the network tend to cluster together. It is a common metric used to evaluate the level of fragmentation of a social network. The clustering coefficient of a node measures the fraction of its neighbors that are also neighbors with each other. In other words, it measures the extent to which a node's neighbors are interconnected. The clustering coefficient of the network is the average of the clustering coefficients of all nodes in the network.

- Modularity

Modularity is a measure of the degree to which a network can be partitioned into non-overlapping communities. It is a common metric used to evaluate the level of fragmentation of a social network.

The modularity of a network is calculated by comparing the fraction of edges that fall within communities to the expected fraction of edges within communities if edges were distributed randomly. A network is said to have high modularity if there are many edges within communities and few edges between communities.

The modularity score can range from -1 to 1. A score of 1 indicates that all edges fall within communities, while a score of 0 indicates that the edges are randomly distributed. A negative score indicates that there are fewer edges within communities than expected by chance, suggesting that the network is more fragmented than expected.

The modularity metric can be used to assess the fragmentation of a network by identifying how strongly the network is divided into communities. Networks with high modularity are considered more fragmented, as they have many tightly connected sub-groups or communities, while networks with low modularity have fewer, less distinct communities.

- Density

The density of a social network is a measure of how many connections exist in the network relative to the total number of possible connections. It is a common metric used to evaluate the level of fragmentation of a social network.

The density of a network is calculated by dividing the number of actual connections by the number of possible connections. The maximum number of possible connections in a network is determined by the total number of nodes in the network. A high density indicates that many connections exist in the network, while a low density indicates that the network is more fragmented, with fewer connections between nodes.

$$Density = \frac{e}{n(n-1)}$$

Here, e is the number of edges of a node and n is the total number of nodes in the network. The density of a network can be used to measure the fragmentation of a network by identifying how many connections exist between nodes in the network. If the network has a high density, it suggests that there are many connections between nodes, and the network is less fragmented. If the network has a low density, it suggests that there are fewer connections between nodes, and the network is more fragmented.

- Community Detection

Community detection algorithms are important for analyzing Twitter data related to extremist groups like ISIS. Here are the two main reasons why community detection is important in our network:

1-Monitoring recruitment and propaganda efforts:

After extracting key authorities and hubs from the network, which highlight the most active members, we can trace through them to identify patterns of recruitment and propaganda efforts by ISIS on Twitter. To identify these communities, we found the group of nodes that had dense intra-community edges and sparse inter-community connections. We have implemented this algorithm using Gephi modularity optimization. Modularity is a measure of the strength of division of a network into modules or communities. It is based on the idea that a good community structure has many edges within communities and few edges between them, and that this structure can be quantified using a null model.

Gephi's modularity algorithm works by optimizing a modularity score that measures the degree of separation of the nodes in the network into communities. The modularity score compares the actual number of edges within communities to the expected number of edges if the edges were distributed randomly. The algorithm uses an iterative process to divide the network into modules, while maximizing the modularity score.

The modularity algorithm in Gephi has several parameters that can be adjusted to customize the analysis. For example, the resolution parameter can be used to control the size of the detected communities. Higher resolution values result in smaller communities, while lower values result in larger communities. We have used the default value of 1. The algorithm also provides options for how to handle directed or weighted edges, as well as other tuning options to control the tradeoff between maximizing modularity and minimizing the number of detected communities.

2-Tracking the spread of information:

Community detection can also be used to track the main topic/information within each community of the ISIS network on Twitter. After identifying clusters of users who frequently mention/tag each other, we traced the main discussed topics per community by identifying the key nodes through which they are spreading. To extract the main topics discussed by each community, we used Latent Dirichlet Allocation (LDA) model [3], which is a probabilistic model used in natural language processing for topic modeling. The parameters of the model are learned from the input data, typically the tweets of the members of each community. The output of LDA is a set of topic distributions over the vocabulary, as well as the topic proportions of each community.

Results

The network's linkages were exposed through analysis, along with the communication related nodes that are most crucial to the network of ISIS activities. It turns out *"Uncle_SamCoco"* is typically the most active character with 1687 times mentioning other users, secondly, we have *"RamiAlLolah*" with 1228 times mentioned by the other users, followed by *"mobi_ayubi"* and *"WarReporter1"* with 1080 and 969 mentioning's respectively, as shown in the graph below.

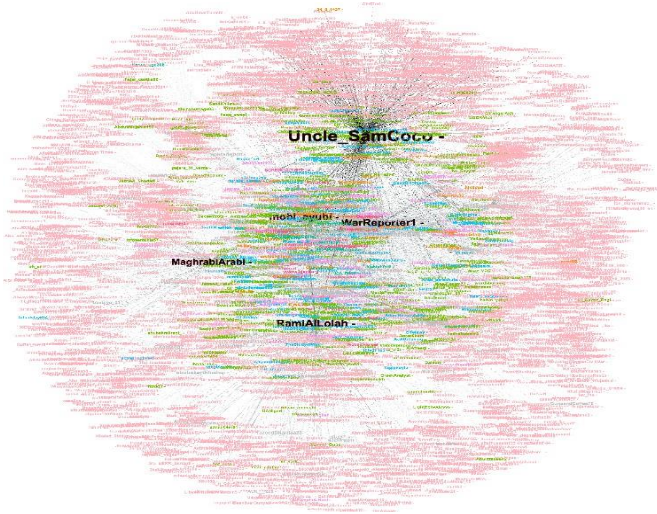

Figure 4: Network graph based on weighted degrees

In the figure above, node labels are sized and given colors, based on their weight degrees with pink color given to the nodes with degree 1 and black represents the nodes with highest degrees. The average weighted degree of the network is 3.71 and weighted degree distribution is as under: -

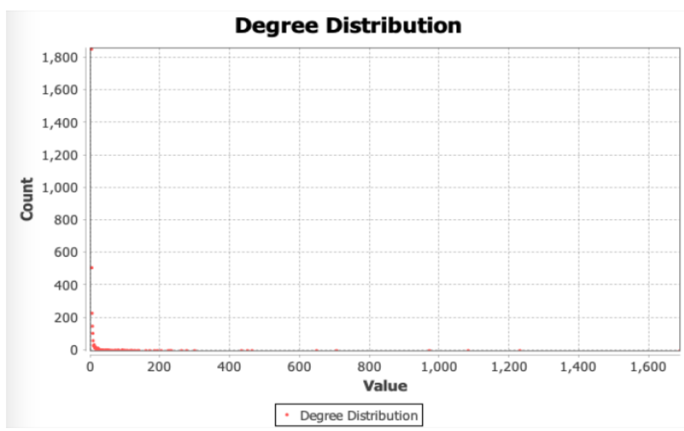

Figure 5: Degree distribution of the network

As we are trying the find the most important actors so we will now consider the graph as mentioned in figure 2. To detect authorities and hubs, we ran HITS algorithm in Gephi and following are the results:

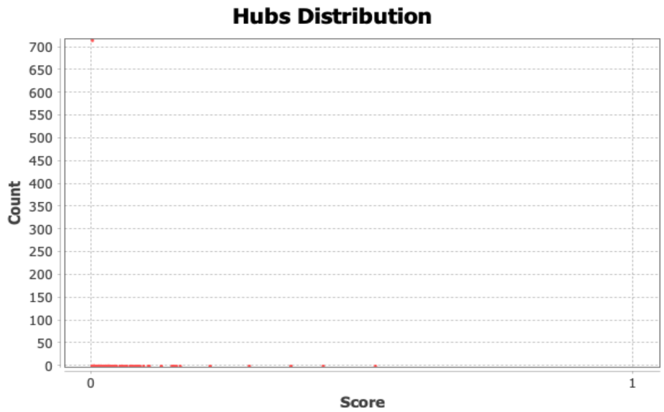

Figure 6: Hub distribution of the network

It clear from the above hubs distribution that only a few nodes have higher hub value and mostly are close to 0. If we plot the graph with size of the nodes' labels in proportion to their hub score, then we get the following results: -

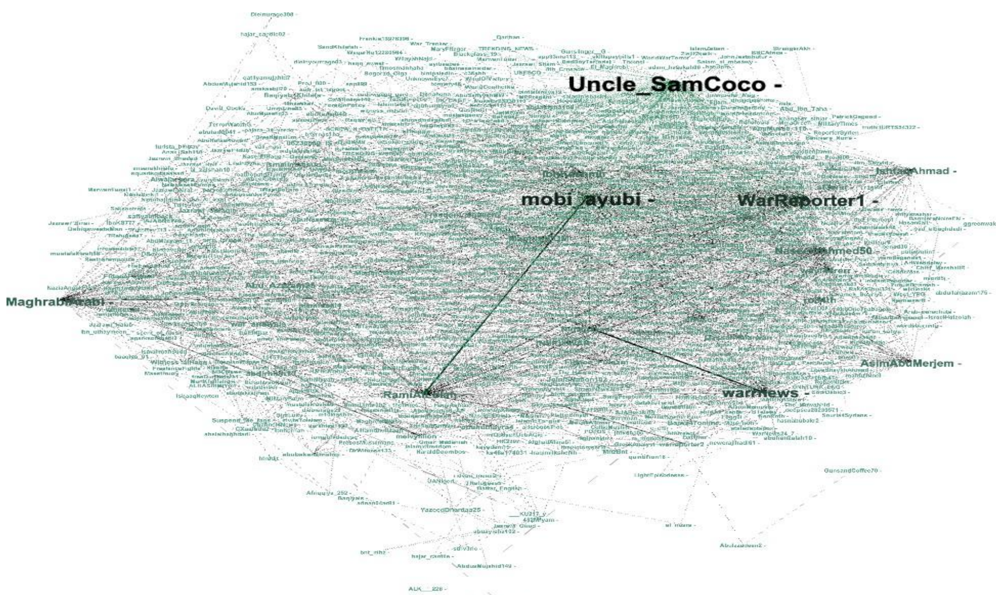

Figure 7: Network graph based on hub score

Hub score of top 6 members is as under: -

Ser	Username	Hub Score
1	Uncle_SamCoco	0.523
2	mobi_ayubi	0.426
3	WarReporter1	0.366
4	warnews	0.289

5	MaghrabiArabi	0.217
6	RamiAlLolah	0.162

Table 1: Network members with the highest hub score

Authority distribution is as under: -

Figure 8: Authority distribution of the network

If we plot the graph with size of the nodes' labels in proportion to their authority score, then we get the following results: -

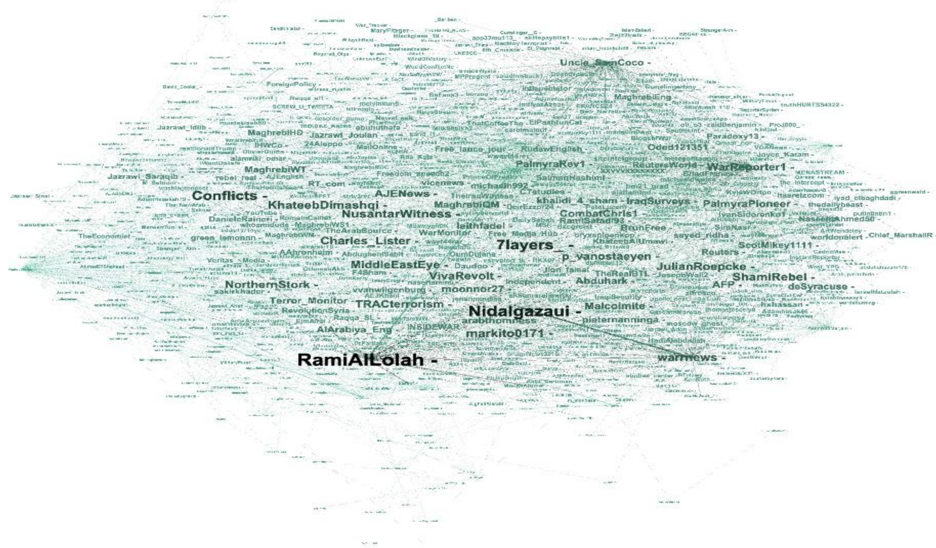

Figure 9: Network graph based on authority score

Authority score of top 6 members is as under: -

Ser	Username	Authority Score
1	RamiAlLolah	0.183
2	Nidalgazaui	0.157
3	7layers_	0.140
4	conflicts	0.127
5	NusantarWitness	0.108
6	Markito0171	0.107

Table 2: Network members with the highest authority score

Weighted degree centralities of top 6 hubs and authorities are as under: -

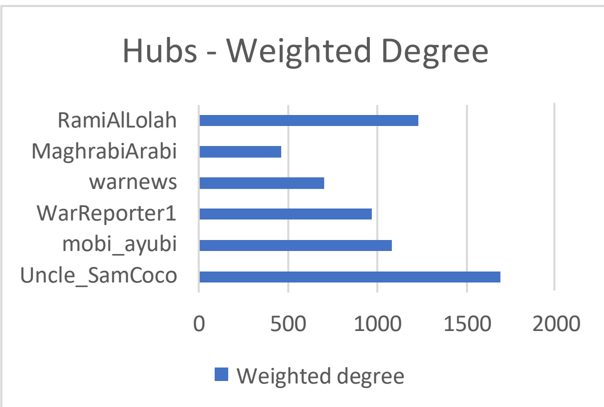

Figure 10: Weighted degrees of top 6 hubs

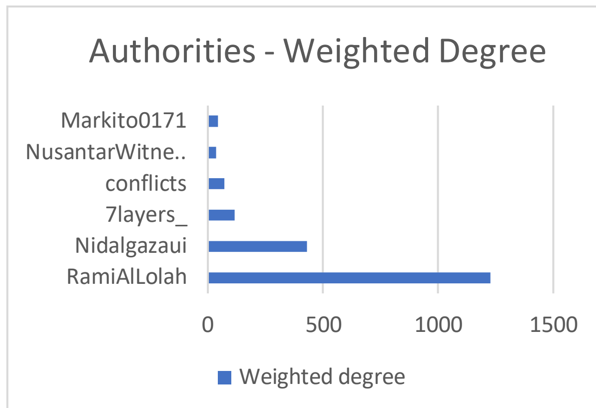

Figure 11: Weighted degrees of top 6 authorities

Above charts clearly show that weighted degree centrality is not a suitable measure to find the importance of a node. Now, we will look at the graph with respect to the PageRank metric. Following graph shows the labels of nodes with respect to their PageRank value. The node with the highest value has the biggest size of label: -

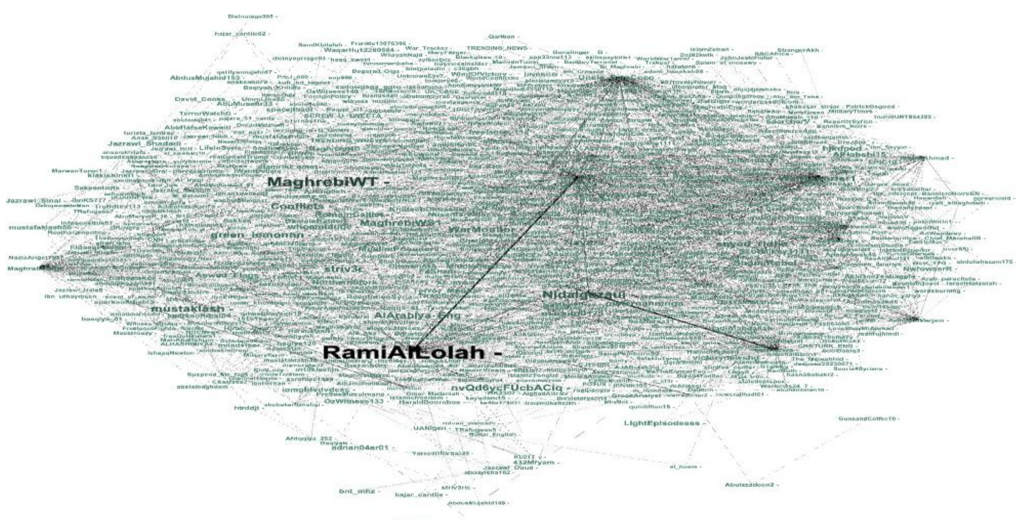

Figure 12: Network graph based on PageRank score

Top 6 members with respect to the PageRank centrality are as under: -

Ser	Username	PageRank Value
1	RamiAlLolah	0.000771
2	MaghrebiWT	0.000575
3	Nidalgazaui	0.000491
4	mustaklash	0.000453
5	striv3r_	0.000436
6	Conflicts	0.000433

Table 3: Network members with the highest PageRank score

PageRank centralities of top 6 hubs and authorities are as under: -

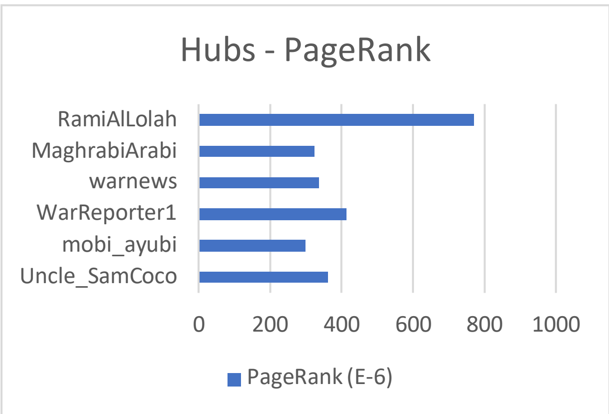

Figure 13: PageRank scores of key hubs

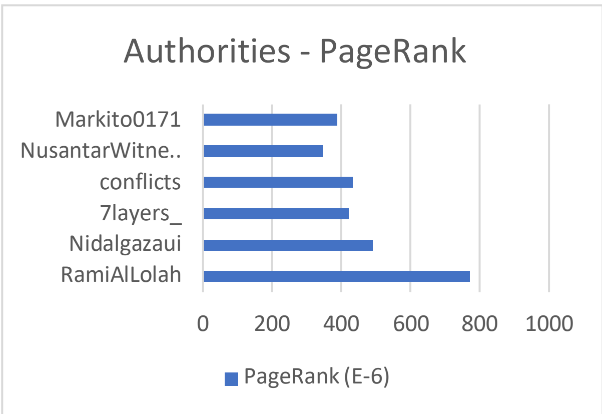

Figure 14: PageRank scores of key authorities

Further extending our investigation to find the most important hubs and authorities, now we will look at the betweenness centrality of the nodes in the network. The Betweenness centrality distribution is as under: -

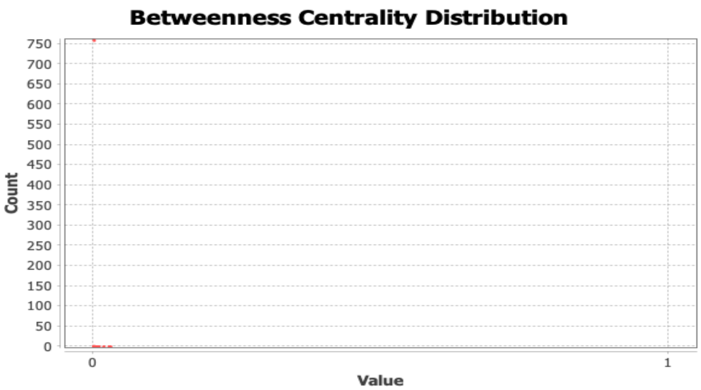

Figure 15: Betweenness centrality distribution of the network

Following graph shows the labels of nodes with respect to their Betweenness centrality values. The node with the highest value has the biggest size of label: -

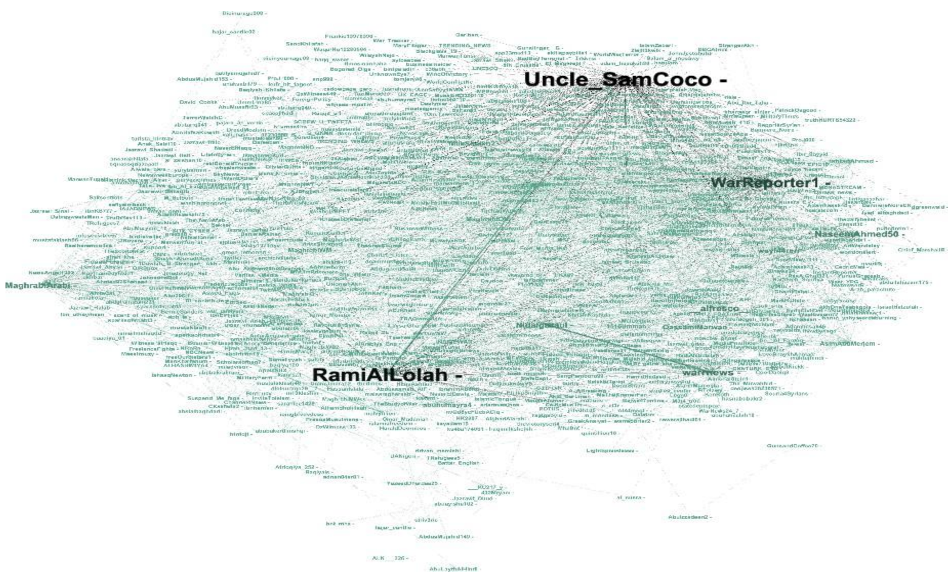

Figure 16: Network graph based on Betweenness centrality score

Top 6 members with respect to the Betweenness centrality are as under: -

Ser	Username	Betweenness Value
1	Uncle_SamCoco	0.0284
2	RamiAlLolah	0.0259
3	WarReporter1	0.0168
4	warnews	0.0083
5	__alfresco_	0.0066
6	NaseemAhmed50	0.0065

Table 4: Network members with the highest betweenness centrality score

To highlight the diffusion and spreading of information through the Hubs and Authorities nodes discussed earlier, ***betweenness centrality*** is being used to measure how frequently a hub/ authority is visited when moving between nodes in the network.

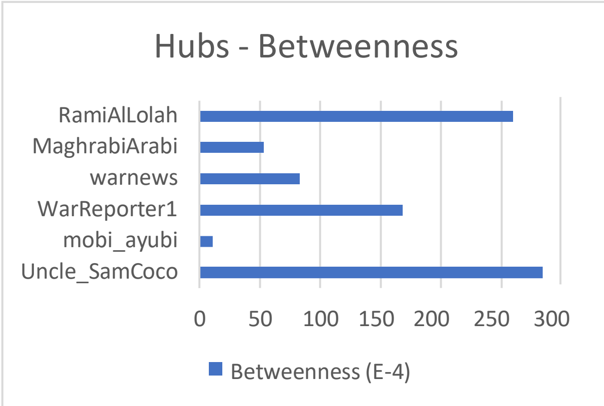

Figure 17: Betweenness centrality scores of key hubs

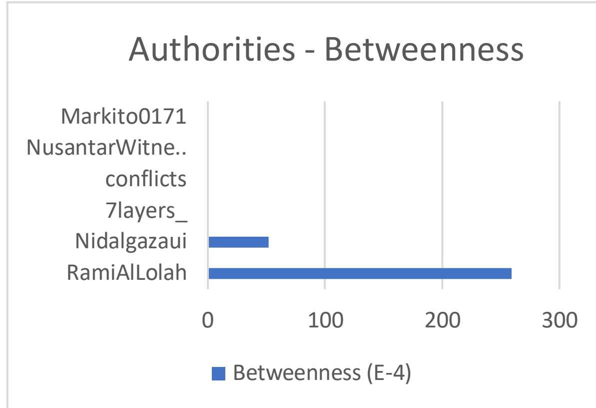

Figure 18: Betweenness centrality scores of key authorities

Above figure clearly shows that we have only two important authorities i.e, *"RamiAlLolah"* and *"Nidalgazaui"* as for rest of the considered authorities the betweenness value is 0. On the other hand, keeping in view the score of PageRank and Betweenness, *"RamiAlLolah"*, *"Uncle_SamCoco"* and *"MaghrabiArabi"* are top three important hubs. Therefore, *to disrupt the network effectively, we recommend removing above mentioned 4 members.*

To measure the effects on the network due to the removal of above members, we will measure fragmentation by comparing the *modularity*, *average clustering coefficient* and *density* metrices before and after the removal of the members. For this purpose, we will use the complete network with 3331 nodes and 5397 edges.

Ser	**Network**	**Avg. Clustering Coefficient**	**Modularity**		**Density**
			Modularity	**Communities**	
1	Before removing the most important members	0.055	0.614	28	0.010
2	After removing the most important members	0.029	0.696	759	0.003

Table 5: Showing increase in network fragmentation by comparing the values of avg. clustering coefficient, modularity & density

The above table shows that by removing the most important members of the group, we have succeeded in disrupting the network effectively. We will explain it by analyzing the values of above three metrices respectively.

First, average clustering coefficient, as we know that the clustering coefficient can be used to measure the fragmentation of a network by identifying how closely connected the nodes are to each other. If the clustering coefficient is high, it suggests that the network is more cohesive and less fragmented, while a low clustering coefficient suggests that the network is more fragmented, with more isolated nodes or sub-groups. As the value of average clustering coefficient is decreased after removing the most important hubs and authorities, so we can say that we have achieved higher fragmentation.

Second is the modularity metric which can also be used to assess the fragmentation of a network by identifying how strongly the network is divided into communities. Networks with high modularity are considered more fragmented, as they have many tightly connected sub-groups or communities, while networks with low modularity have fewer, less distinct communities. Clearly, we can see from the above table that after removing the most important members of the network the value of modularity has increase. Furthermore, the number of communities/ sub-groups has increased significantly from 28 to 759. This indicates that we have achieved higher fragmentation successfully.

Lastly, the density metric. The density of a network can also be used to measure the fragmentation of a network by identifying how many connections exist between nodes in the network. If the network has a high density, it suggests that there are many connections between nodes, and the network is less fragmented. If the network has a low density, it suggests that there are fewer connections between nodes, and the network is more fragmented. As shown by the above table, we have successfully reduced the density of the network by removing the most important nodes of the network.

Communities Detection

For the communities' detection, we have used Gephi's modularity algorithm. A total of 28 communities have been detected and their size in terms number of nodes in each community has been shown in figure 19 below: -

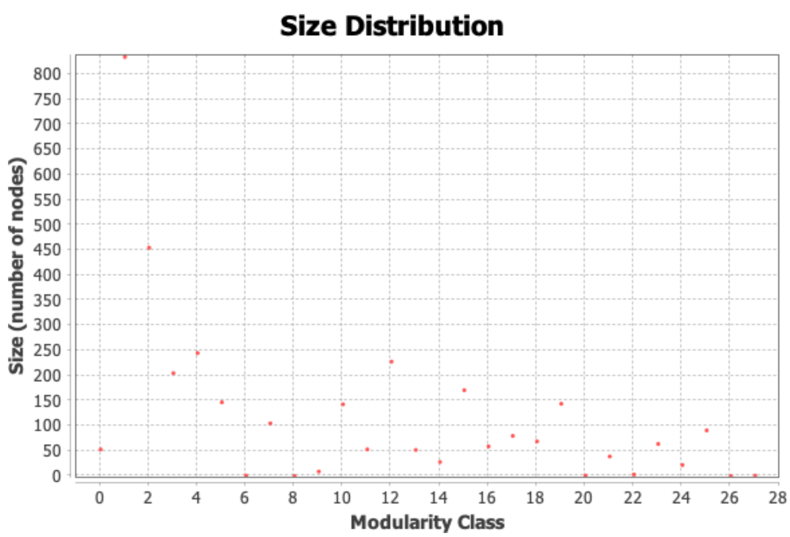

Figure 19: Network communities' statistics

Above graph shows the communities statistics. However, if we plot the network, we obtain the following figure in which major 8 communities are shown with different colors, however other smaller communities are shown with grey color: -

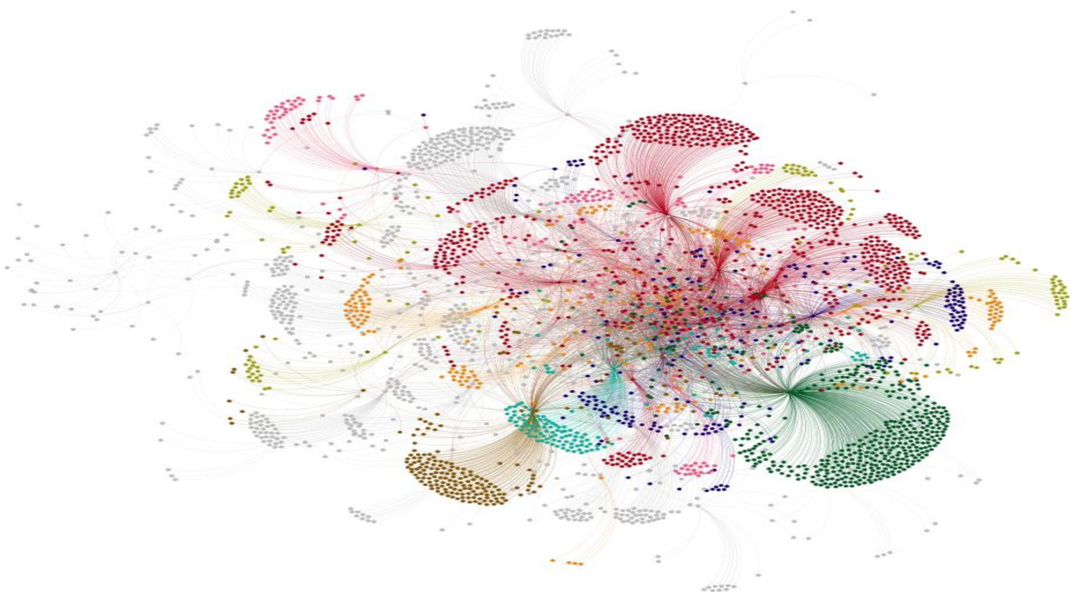

Figure 20: Network graph showing communities

Based on the degree distribution, we extracted the central user of each detected community[4]. This is the most active member in the community and may be the boss or planner in the group. In addition to above-mentioned members identified to be removed from the network, by removing core member of each community, we can disrupt the network even more.

Now we will look at the 8 major communities individually along with their core members and size. It is pertinent to mention here that we have also extracted the keywords discussed in each community using LDA (Latent Dirichlet Allocation) model, which is an NLP Model[6].

Community 1: It is the largest community with 518 nodes and 611 edges. Its core member is "RamiAlLolah". Keywords under discussion are: [protect, ya, azzawajal, brothers, swt, akbar, make, help, people, said, family, accept, freemuslimprisoners, messenger, love, dua, reward, abu].

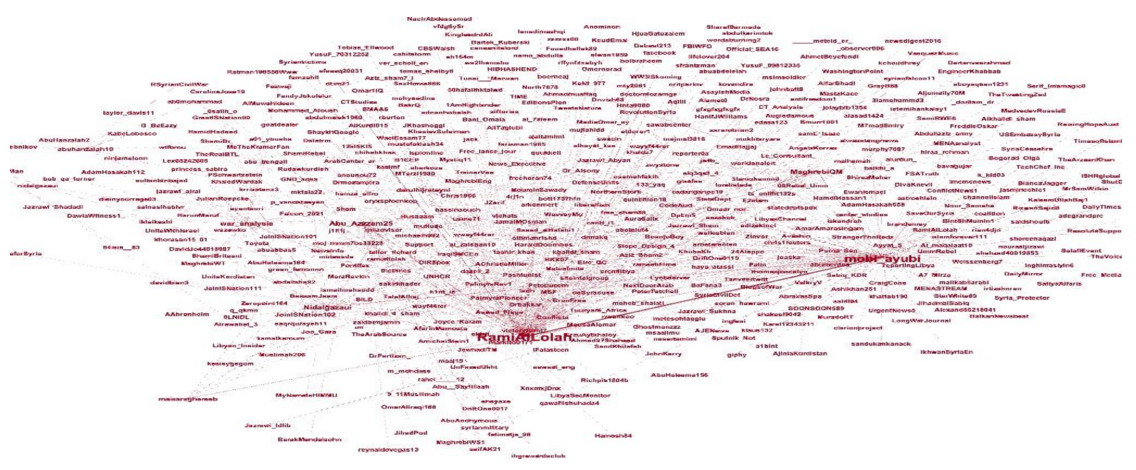

Figure 21: Community 1 graph

Community 2: It has 459 nodes and 607 edges. Its core member is "*WarReporter1*". Keywords under discussion are: [state, fighters, Israel, military, says, territory, group, at-tacks, time, turkish, soldiers, terrorist, enter, explosive, sinai, wilayatsalahuddin, iraq, general, spy].

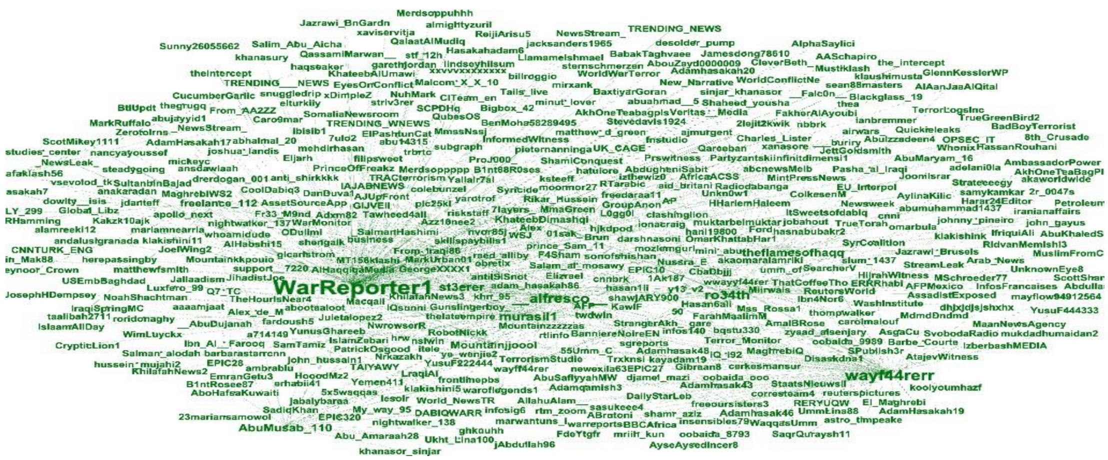

Figure 22: Community 2 graph

Community 3: It has 447 nodes and 447 edges. Its core member is "*UncleSamCoco*". Keywords under discussion are: [isis, killed, iraq, army, ramadi, soldiers, iraqi, attack, today, west, deirezzor, near, claims, breakingnews, forces, ghters, militants, islamicstate, breaking, huge].

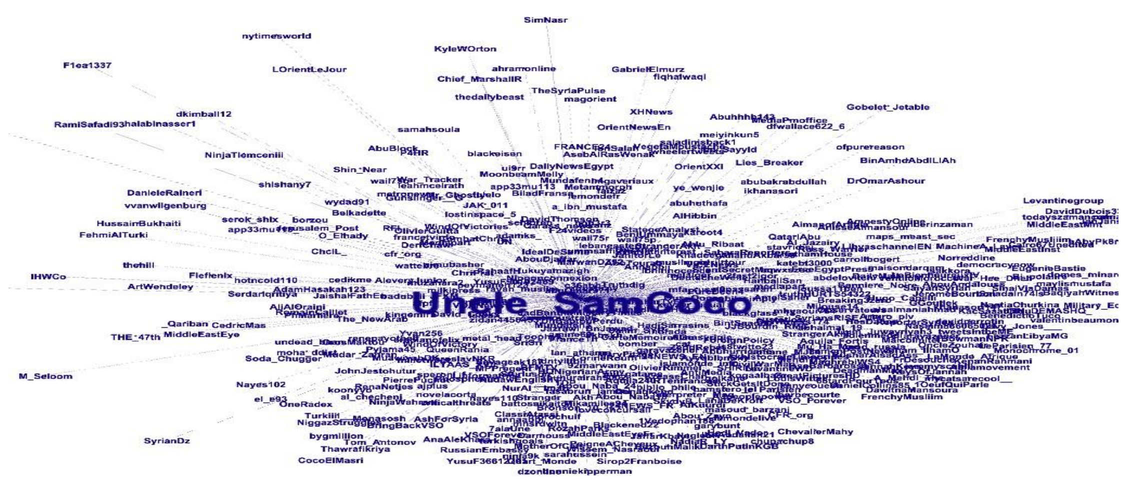

Figure 23: Community 3 graph

Community 4: It has 222 nodes and 229 edges. Its core member is "*MaghrebiArabi*". Keywords under discussion are: [syria, russia, assad, geneva, usa, talks, regime, war, airstrikes, russian, deirezzor, opposition, people, kids, idlib, madaya, homs, children, assads, kerry].

Figure 24: Community 4 graph

Community 5: It has 176 nodes and 187 edges. Its core member is "*IbnKashmir*". Keywords under discussion are: [know, did, dont, al, just, fact, shia, caliphate, ottomon, want, used, didnt, world, created, america, ummah, sheikh, religion].

Figure 25: Community 5 graph

Community 6: It has 170 nodes and 187 edges. Its core member is "*abuhumayra4*". Keywords under discussion are: [follow, brother, sisters, account, khair, support, dear, spread, true, path, baqiyah, shout, family, nation, great, retweet, surprise, dm].

Figure 26: Community 6 graph

Community 7: It has 159 nodes and 166 edges. Its core member is "*warnews*". Keywords under discussion are: [akhi, ya, shout, im, nasheed, dm, sheikh, happened, know, link, check, source, khair, better, plz, akh, punishment, surprise].

Figure 27: Community 7 graph

Community 8: It has 154 nodes and 159 edges. Its core member is "*abubakerdimshqi*". Keywords under discussion are: [english, new, sheikh, al, video, statement, wilayat, mte2, coming, soon, regarding, time, media, el, town, wilayataljazirah, people].

Figure 28: Community 8 graph

Conclusion

Using social network analysis, we were able to study ISIS network on Twitter. Our goal was to analyze the network, pinpoint the most important members in the network, analyze how their removal will affect the network and lastly to detect various closely connected communities in the network. By analyzing each node based on their degree distribution, PageRank and betweenness centrality we concluded that "*RamiAlLolah*" is the most important member in the network followed by "*Uncle_SamCoco", "MaghrabiArabi"* and *"Nidalgazaui".* Banning these members will have a significant effect on the network as we have shown it by the calculating average clustering coefficient, modularity and density of the network before and after their removal. We have also successfully detected 28 communities and analyzed major 8 communities to have a better understanding of the following of information in the network.

Critique

The measurements of the data have aided in clarifying the structure and composition of the network. SNA concepts can also be used to investigate additional character traits in a graph network, such as biased, focused, scattered, and diversified members of the community. Further metrics such as link prediction between communities would have solidified the results, especially for information spreading since the spread of information is highly affected by the weights of the connections.

Based on this analysis and the extraction of relevant topics from each community, we can determine the relationships between them and track information by implementing a neural network that can predict future attacks. To achieve this, the network should extract not only usernames but also hashtags (#), as hashtags are important for viral spread, this would most likely produce stronger results than what we see[5]. To extract all the hashtags, we use an NLP Python script to locate the (#) mark in tweets. Comparing to usernames, hashtags referring to the same content can have more varied formats.

Thus, after extraction, we further compute the similarities among hashtags by checking the character occurrence and order. The main data is reduced to the ones of mentions and hashtags, which become the training data of a static neural network with some hidden layers in the network architecture. This prediction network estimates the output of the testing sample by weighted averaging the outputs of the training samples, where the weights are computed by the distance between testing and training samples. Using these methods, we can predict future attacks.

References

[1] Dataset: - https://data.world/data-society/how-isis-uses-twitter

[2] Counter terrorism guide \Islamic State of Iraq and The Levant (ISIL).in nctc groups (2017)
Link : - https://www.dni.gov/nctc/groups/isil.html

[3] Ng A., Blei D., and Jordan M. \Latent Dirichlet Allocation". In: Journal of Machine Learning Research (2003).

[4] Yang J., Mcauley J., and Leskovec J. \Community Detection in Networks with NodeAttributes". In: Stanford University (2015), pp. 4,5.

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

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