

# Analyzing The Network Of The ISIS Tweets In Twitter Using Social Network Analysis

April 29, 2020

Project Proposal is Done By:

Name : Nupur Roy Chowdhury

Drexel ID : 14395901

Email ID : nr572@drexel.edu

## 1 Abstract:

Islamic State of Iraq and the Levant known as ISIS one of the terrorist Organizations which is a global threat controls Syria and most of Iraq and Israel. Analysis of this terrorist network is essential for discovering their propaganda. ISIS has a constant stream of propaganda for which it is using twitter 90% of the time to spread violence. ISIS consists of 70,000 twitter accounts globally out of which around 200,000 tweets are being posted on twitter everyday. Studying the network of ISIS helps in developing counter strategies in order to blunt its outreach efforts effectively as it is using digital strategy to crowdsource. Using social network analysis and tools we will be able to explore the complex networks to recognize patterns which were previously unreadable. Through this we will be able to extract the hidden relationships, their strategies and the user interactions.

## 2 Introduction:

The ISIS is an organised terrorist organization that consists of collaboration between all the terrorist in that organization which forms the basis of the relationships which these terrorists share with each other. The ISIS is a force which takes into consideration almost 80% - 90% of the digital platform for its strategies. Where twitter is the cornerstone of this group's digital strategy. It consists around 46,000 - 70,000 twitter accounts all over the world promoting hate messages. An estimated followers of around 21,000 are English speaking language alone. There is around 2,000 over-performers accounts that tweet at a bursts of 50 or more tweets in a single day and each of these performers or actors again have an average of 1,004 followers or actors which results in around 200,000 tweets each day by all of these actors in spreading hate messages and contents. The messaging style of ISIS can be broken down into two types:

1. Content for Recruitment

and

2. Content for it's enemies

Content for Recruitment: This includes breaking news from the front headings and encouraging the youth to take matters into their own hands. Their online magazine called "Dabiq" describes "The Management of Savagery", videos such as the "Mujetweets" which function like mods for games which allows the players to live the life of an ISIS recruit going around shooting and blowing things up and their religious melodies called as nasheeds.

Content for it's enemies : This includes feature films threatening countries of terror attack if they attempt to intervene in the regions occupied by them.

Based on the twitter data there are six different types of roles which can be identified as: 1. Reporters: users who convey the breaking news by ISIS. 2. Reconnectors : users whose job is to retweet the usernames of the violent extremists whose accounts were suspended before but are back with new names. 3. Intellectuals : who uses the philosophies,economics and political theory to justify their doings. 4. Fanboys : who have their profile picture as the ISIS and celebrates its victories. 5. Recruiters : who does the recruitment, they have private accounts where after a potential actor has been referred does direct messages to them and then takes it through encrypted channels. 6. Mujahideen : The actors posted as heroes as they are the ones who are fighting on the frontlines.

Description of the ISIS network and its statistics were taken from the article:

<https://medium.com/fifth-tribe-stories/isis-has-a-twitter-strategy-and-it-is-terrifying-7cc059ccf51b#.m3zeluykl>

### **3 Research questions which can be studied from this are of the following types:**

- Analyzing of the social network cluster:

which tells about the major actors who are involved in this and their visualization through the cluster network which ranks them basis of the highest influence to the least influence actors.

- Keyword Analysis:

Keywords which can be used from the name, username, location and description of the tweets. Like example: "dabiq", "wilayat", "amaq" etc can be used to study their social network.

- Links categorised into data:

The websites to which links are being formed in the tweet, which includes tweets linking to jihadist websites, altermedia.

- Sentiment Analysis:

Which clergy do these ISIS actors quote the most and which ones do they hate the most. Search of the tweets with names of prominent clergy and classify the tweet as positive, negative, or neutral and if negative, include the reasons why. Examples of clergy they like the most: "Anwar Awlaki", "Ahmad Jibril", "Ibn Taymiyyah", "Abdul Wahhab". Examples of clergy that they hate the most: "Hamza Yusuf", "Suhaib Webb", "Yaser Qadhi", "Nouman Ali Khan", "Yaqoubi".

- Analysing the Timeline:

Identifying the peak moments over the timeline by studying the various tweets over a given timeline.

## 4 Data Collection:

This project contains network analysis of the ISIS twitter dataset which was taken from:

<https://www.kaggle.com/kzaman/how-isis-uses-twitter>.

Data Description:

They have scraped over 17,000 tweets from more than 100 posts of pro-ISIS fanboys from all over the world since the November 2015 Paris Attacks.

The dataset includes the following description of the data:

1. Name
2. Username
3. Description
4. Location
5. Number of followers at the time the tweet was downloaded
6. Number of statuses by the user when the tweet was downloaded
7. Date and timestamp of the tweet
8. The tweet itself

Displaying the first 5 data details from the dataset:

```
[1]: import pandas as pd
```

```
[9]: Tweet_data = pd.read_csv("tweets.csv")  
     Tweet_data.head()
```

```
[9]:
```

	name	username	\
0	GunsandCoffee	GunsandCoffee70	
1	GunsandCoffee	GunsandCoffee70	
2	GunsandCoffee	GunsandCoffee70	
3	GunsandCoffee	GunsandCoffee70	
4	GunsandCoffee	GunsandCoffee70	

  

	description	location	followers	\
0	ENGLISH TRANSLATIONS: http://t.co/QLdJ0ftews	NaN	640	
1	ENGLISH TRANSLATIONS: http://t.co/QLdJ0ftews	NaN	640	
2	ENGLISH TRANSLATIONS: http://t.co/QLdJ0ftews	NaN	640	
3	ENGLISH TRANSLATIONS: http://t.co/QLdJ0ftews	NaN	640	
4	ENGLISH TRANSLATIONS: http://t.co/QLdJ0ftews	NaN	640	

  

	numberstatuses	time	\
0	49	1/6/2015 21:07	
1	49	1/6/2015 21:27	
2	49	1/6/2015 21:29	
3	49	1/6/2015 21:37	
4	49	1/6/2015 21:45	

  

	tweets
0	ENGLISH TRANSLATION: 'A MESSAGE TO THE TRUTHFU...
1	ENGLISH TRANSLATION: SHEIKH FATIH AL JAWLANI '...

```

2 ENGLISH TRANSLATION: FIRST AUDIO MEETING WITH ...
3 ENGLISH TRANSLATION: SHEIKH NASIR AL WUHAYSHI ...
4 ENGLISH TRANSLATION: AQAP: 'RESPONSE TO SHEIKH...
```

The Dataset consists of the following coloumns:

```
[18]: for col in Tweet_data.columns:
      print(col)
```

```

name
username
description
location
followers
numberstatuses
time
tweets
```

```
[21]: print(Tweet_data.info())
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17410 entries, 0 to 17409
Data columns (total 8 columns):
name                17410 non-null object
username            17410 non-null object
description          14728 non-null object
location            11432 non-null object
followers            17410 non-null int64
numberstatuses       17410 non-null int64
time                17410 non-null object
tweets              17410 non-null object
dtypes: int64(2), object(6)
memory usage: 1.1+ MB
None
```

```
[20]: Tweet_data.describe()
```

```

[20]:      followers  numberstatuses
count  17410.000000    17410.000000
mean    3975.076680     4760.787651
std     8674.378495     7015.967275
min       16.000000         1.000000
25%      266.000000        207.000000
50%      928.000000        908.000000
75%     1791.000000       6865.000000
max     34692.000000    33091.000000
```

Example showing a sample tweet:

```
[14]: print(Tweet_data.loc[17387])
```

```
name
username                               Uncle_SamCoco
description      Here to defend the American freedom and also ...
location                               Texas, USA
followers                               1798
numberstatuses                          7028
time                                   5/13/2016 15:17
tweets      @MehdiTh2 @_Darmouss @PeigneACheveux j'aimerai...
Name: 17387, dtype: object
```

Example showing the full description of the tweet:

```
[15]: print(Tweet_data.loc[17387, 'description'])
```

```
Here to defend the American freedom and also the freedom of coconut . Cat Lover
or Hater. Kebab Fan . We're all living in America, America ist wunderbar #USA
```

Example showing the accounts who has re-tweeted the above tweet:

```
[16]: print(Tweet_data.loc[17387, 'tweets'])
```

```
@MehdiTh2 @_Darmouss @PeigneACheveux j'aimerai bien savoir aussi
```

## 5 Intended Methodology or Approach:

As an investigator of the terrorist network we can use the following approaches for the Network Analysis of the social data:

- Subgroup Detection : As the ISIS network consists of six different types of roles which are carried out by the user\_accounts in twitter who are known as actors to help identify the relationships between each type of the accounts.
- Identification of the important actors and the roles they play in the network: the approach here will be to extract each user with their unique usernames which acts as a node. The node will be then influenced by the scale consisting of the combination of the number of followers and the number of tweets they tweet from their account. This combination helps us to ensure that active and popular users has been identified. Currently this relation between each actors is yet to be fixed, as one relation criteria might be to scrape the tweets of users having maximum followers and then link the nodes via this metric with numerous mentions increasing the weight of an edge between the actors where they either have direct communication or by retweeting the tweets of the actors.
- Discovery of Patterns: it is based on qualitative methods of aggregated pairwise occurrence of the tweets for each actor, which can then be normalized by the sum of all tweets from all the actors that are being compared with each other. This will help in tweets which are exceedingly large and their actors and the actors with minimal tweets activity. To this the actor will be the main identifier. This will generate the undirected weighted graphs.

- Textual Analysis: If we take into consideration of the time of the tweet, it will be leading into lots of linkages getting generated which might be sometimes not important, hence to reduce this the graph will be filtered with appropriate thresholds which will be determined by examining of the weights distribution and the actors centrality. Advantage of the approach is that it will not be transaction-dependent since we will be matching the strings of all the tweets.