Global Terrorism's Influence on Twitter

Harsh Shah
Student ID:001307974
University at Albany, SUNY
1400 Washington Avenue
Albany, New York 12222
hshah@albany.edu

Bhumit Sheth
Student ID:001307653
University at Albany, SUNY
1400 Washington Avenue
Albany, New York 12222
bsheth@albany.edu

Tejas Shah
Student ID:001311049
University at Albany, SUNY
1400 Washington Avenue
Albany, New York 12222
tshah@albany.edu

ABSTRACT

In this research paper, we are conducting research on the influence person on a twitter regarding the terrorist group ISIS and how ISIS uses twitter as their social network. In this paper, we will generate the network of influenced characters in twitter regarding the ISIS. We are using the data set from Kaggle "How ISIS uses network". Using in-degree, out-degree we have shown the inferentiality of twitter user who get referred most. ISIS uses twitter as their main source for convey their ideas and recruit members. We have generated a network based on the dataset, and from the network we have visualized a cluster network with biggest influencers scaled larger than smaller influences. Study done by brooking institute states that, ISIS used around 46000 twitter accounts for conveying their message. In this research paper, we will explore the top mentioned twitter accounts subjected to ISIS using social network analysis. In this we are using the In-degree and Out-degree as main structure of a network.

1. INTRODUCTION

ISIS, the Islamic State, is a Sunni terrorist organization that has seized substantial portion of the Iraq and Syria. Estimated 20,000 people from different countries have joined the ISIS. They are recruiting more and more people and trying to influence people by various social platforms. Now days' social media like twitter, Facebook are the primary way to convey the messages. And all the big names and the organizations use the tweeter for their main stream media. People around the world are using tweeter as their main source of information They are recruiting more and more people and trying to influence people by various social platforms. According to a study by the Brookings Institution, ISIS supporters used at least 46,000 Twitter accounts between September and December 2014 and many of them are machine powered accounts. ISIS mostly use twitter to share disseminate images of graphic violence, attract new recruiters. In our project, we have used the Kaggle data set Exploration of ISIS network. With the help of the dataset, we have created the network of different twitter account of the ISIS and the way they mention each other in the network. This mention can be for spreading news among all the followers of the account holders. Because of these activities, it is necessary to observe these kinds of behavior on a social network. Activities of few group are coming up in the twitter more often than others. Form those group we have chosen ISIS for our research. ISIS are most active terrorist group on twitter and we want to see who are the main characters mentioned in those twitters. And we are connecting dots in the network and figuring out who is the big name referred to in field of ISIS world. After connecting the dots between nodes, we can see that who is referring to whom. Then with the help of in-degree and out-degree functionalities we have visualized via a cluster network with the biggest influencers scaled larger than smaller influencers. [1]

Understanding ISIS's social-media strategy, including who runs the accounts and whether they operate in an organized manner, would enable the American government, with help from technology companies, to work out a more effective strategy of combatting ISIS propaganda. [2]

2. DATA SET

Our Data set is consisting following terms.

1) Name
2) Username
3) Description
4) Location
5) Followers
6) Number of Statues
7) Time
8) Tweet

Dataset that we have used has more than 17000 tweets from more than 100 fanboys of ISIS from overall world since the Paris attack. Basically, we are doing social network cluster analysis. The original datasets consist many missing values, so we were not able to create a network from the original datasets as it consists many missing values. So, what we have done is we did classification on the dataset and divide the sorted data and unsorted data to see more accurate graph. Due to this our dataset scaled down to around 3500 tweets but we try to take every possible username from the dataset. Our dataset includes 113 usernames. There are many tweets that were originally written in the Urdu language that we were not able to convert into English so we have discarded those types of tweets as well. Here we have taken username as our nodes in the graph and edges represent the person referring another user. Also, to create a network we need two different .csv (Comma Separated File) one consist edge and one consist details of edges. So, with the help of python programming we have created two files edge.csv and node.csv consist details of edge and details of nodes respectively.

3. METHODOLOGY

In this research, after sorting the raw data and performing some python script on we have generated two .csv file that consists details of edges and nodes. After that to generate the graph we have use GEPHI tool. In the Gephi, we have imported the two spreadsheets in the data laboratory of nodes and edges respectively. And the general graph has been created by this tool. Also, we can perform various inbuild function of Gephi to find, average weighted degree, network diameter, modularity of the network and there are many more provided by the Gephi. We have performed various functions like In-degree and Out-degree of the network to find the biggest fanboy on the ISIS network.

3.1. RESULT

As a result, we got this graph (Figure 1). This graph contains 113 nodes and 3928 directed edges. (The graph is directed but due to the small size of the graph it is not possible to see the direction of the arrow). Here nodes are the username of the account holders while the edges represent the mention done by the other users in their tweets. There are more than one mentions of users in one tweet, we have fetched all of them with the help of python programming. Here the direction of the arrow represents mentioned of that user in the tweet done by the user other side on the arrow. There are also many nodes present in the graph that are not connected to any node. This happens due to inconsistency in the data. We tried to minimize it but almost around 10 nodes that are not connected with any other node of the graph.
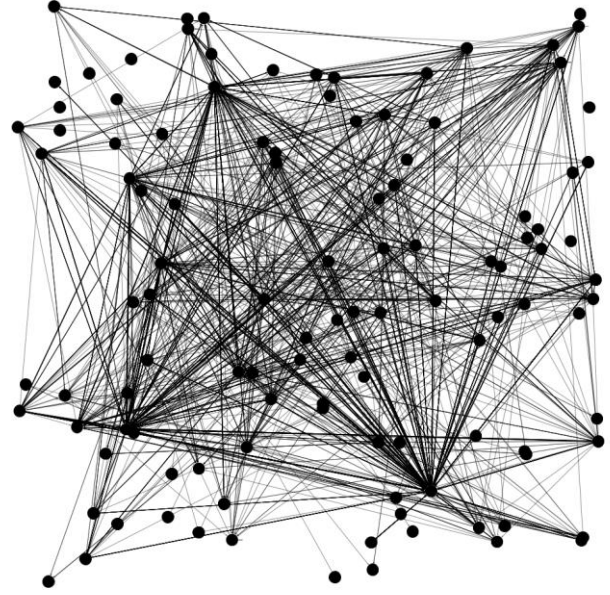

Figure 1. Network that shares by ISIS fanboys

3.2. DEGREE DISTRUBATION

In a directed network, each node has two degrees. The out-degree is the number of outgoing edges emanating from a node

$$k_i^{\text{out}} = \sum_j a_{ji},$$

and the in-degree is the number of incoming edges onto a node

$$k_i^{\text{in}} = \sum_j a_{ij}.$$

The total degree of the node is the sum of its in- and out-degree [3]

$$k_i^{\text{tot}} = k_i^{\text{in}} + k_i^{\text{out}}.$$
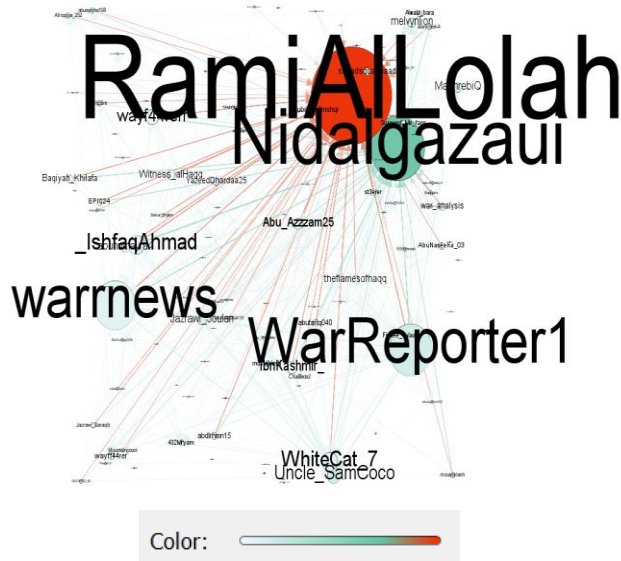
In-Degree Representation of the Graph

Figure 2. In-Degree representation of the graph

In-Degree graph representation of the original graph, we can see the varied size of the nodes in assorted color. White color represents the lower in-degree for that node and red color represent the higher in-degree for that node. Here node like, "RamiAlLolah", "warrnews", "WarReporter1", "Nidalgazaui" have the higher in-degree. Other user (nodes) in the network have mentioned this four users (nodes) the most. In-degree graph is useful to count the most mentioned users in the network.

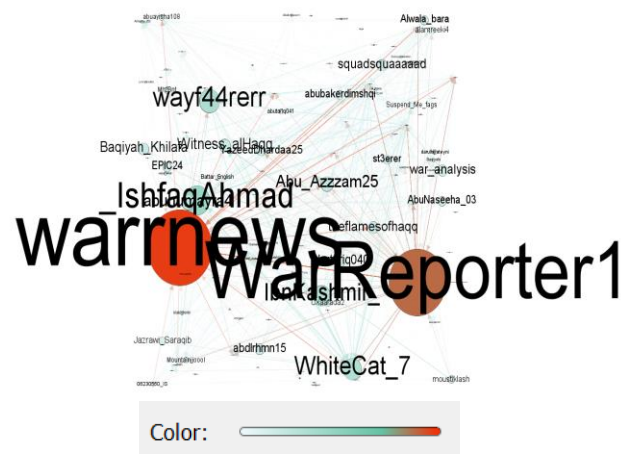## Out-Degree Representation of the Graph

Figure 3. Out-Degree representation of the graph

Out-Degree graph representation of the original graph, we can see the varied size of the nodes in assorted color. White color represents the lower in-degree for that node and red color represent the higher in-degree for that node. Here node like, "warrnews", "WarReporter1", "Wayf44rerr"," WhiteCat_7", "_IshfaqAhmad" have the higher out-degree. Out-Degree of users(nodes) represent the numbers of tweet done by them. Out-degree graph is useful to count the most tweets tweeted by the user.

Comparison between In-Degree and Out-Degree

	Node_id	Label	In-Degree	Out-Degree
1	97	RamiAlLolah	1293	0
2	13	Nidalgazaui	854	0
3	3	Uncle_SamCoco	205	0
4	103	WarReporter1	181	513
5	23	melvynlion	116	0
6	9	MaghrebiQ	114	0
7	25	Jazrawi_Joulan	99	17
8	19	WhiteCat_7	79	165
9	108	warrnews	71	577
10	111	wayff44rer	68	7

The above table displays the top 10 out-degrees among all the nodes.

	Node id	Label	In-Degree	Out-Degree
1	108	warrnews	71	577
2	103	WarReporter1	181	513
3	16	_IshfaqAhmad	19	225
4	11	wayf44rerr	11	180
5	19	WhiteCat_7	79	165
6	74	IbnKashmir_	0	153
7	75	Abu_Azzzam25	28	109
8	42	abuhumayra4	10	103
9	56	Baqiyah_Khilafa	3	90
10	78	theflamesofhaqq	11	88

The above table displays the top 10 in-degrees among all the nodes.

From both the table we can see that there's a drastic difference in the result. Many users (nodes) have been mentioned by other users but they haven't mentioned any other in the tweets. For example, RamiAlLolah's in-degree is 1293 while out-degree is 0. Same for the Nidalgazaui, Uncle_Samcoco etc. On the other side in the out-degree table we can see the totally different output except 3 entries. Most of the node(user) have tweeted a lot but they got less mentioned in the dataset.

Influencers in the network

Figure 4. Tweet Activity vs Influences [4]
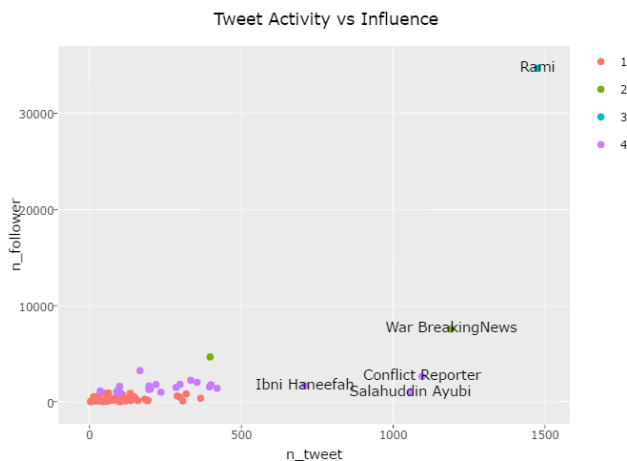
The above graph represents the numbers of tweet done by users to the numbers of the followers they have. From the graph, we can clearly see that Rami has the most followers and did most tweets. So, we can say that user Rami is the most influenced user among all.

Interestingly, out-degree and in-degree of the RamiAlLolah has quite different. The in-degree of the user is 1293 while out-degree is 0. This user has the most numbers of followers among all and has the highest numbers of tweet, but he hasn't mentioned anyone in the network. Everyone else are mentioning him. He is the most influencer among all. People may influence by tweet done by him. That could be one reason why all other in network are referring him, it could help others to convey their message to more audiences with the mentioned name of the most influencer person.

Modularity in the network
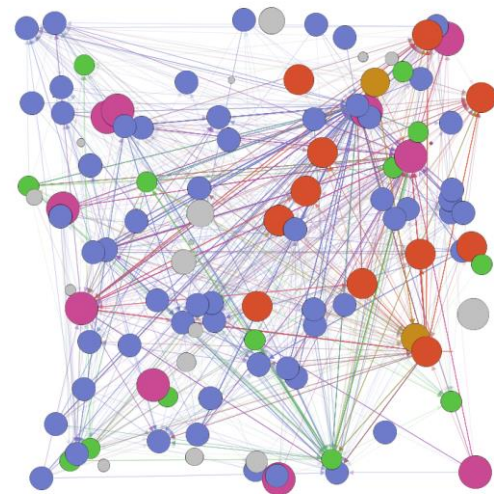
Below are some stats about our network.

Modularity: 0.126
Modularity with Resolution: 0.126
Numbers of Communities: 23

Modularity is the degree to which a system's components may be separated and recombined. [5]

Our network has 23 different communities.

Figure 5. Modularity of the Network

Which means it is generated as one person mentioning another person and the total number of tweets done by user. In figure 5 we can see that there are assorted colors representing different communities. Here Red is the community that has characteristics of less tweet and more mentions. Where dark gray is the user that tweets a lot as well mentioned a lot.

Conclusion

In this paper, we have studied about the various datasets related to ISIS tweets. And we have generated the graph in which node represent different users while edges represent the mention they got in the tweets. We have visualized this via a cluster network with the biggest influencers scaled larger than smaller influences. From the graph, we can analyze the about most influencers in the network.

Work Distribution

We all have searched various dataset online and go through the dataset and check whether it fulfill our requirements or not. Harsh has implemented all the coding stuff like various python scripts on the dataset to clean it. Bhumit implemented it into the Gephi and generated various graphs for better analysis. We all have sit together and designed the whole research paper and Tejas implemented it into the latex format.

References

[1] A Visual Guide to the Rise of the Islamic State. The New York Times. Updated 11 June 2014. Web. http://www.nytimes.com/interactive/2014/06/12/world/middleeast/the-iraq-isis-conflict-in-maps-photos-and-video.html.

[2] Batsheva Moriarty. 2015. Defeating ISIS on Twitter. (September 2015). Retrieved May 9, 2017 fromhttps://techscience.org/a/2015092904/

[3] Node degree definition - Math Insight: 2017. http://mathinsight.org/definition/node_degree. Accessed: 2017- 05- 09.

[4]2017. https://www.kaggle.com/hanyan/exploration-of-isis-network. Accessed: 2017- 05- 09.

[5]Modularity:2017. https://en.wikipedia.org/wiki/Modularity. Accessed: 2017- 05- 09.

[6] Exploration of ISIS network | Kaggle: 2017. https://www.kaggle.com/hanyan/exploration-of-isis-network. Accessed: 2017- 05- 09.

[7] How ISIS Uses Twitter | Kaggle: 2017. https://www.kaggle.com/kzaman/how-isis-uses-twitter. Accessed: 2017- 05- 09.

[8] Exploring ISIS(?) Tweets | Kaggle: 2017. https://www.kaggle.com/violinbeats/notebook-0427671092ae887aa87e. Accessed: 2017- 05- 09.

[9] ISIS_tweets_exploration | Kaggle: 2017. https://www.kaggle.com/patricknieto/isis-tweets-exploration. Accessed: 2017- 05- 09.

[10]2017. https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf. Accessed: 2017- 05- 09.

[11]Defeating ISIS on Twitter: 2017. https://techscience.org/a/2015092904/. Accessed: 2017- 05- 09.

D.3.3 [**Programming Languages**]: Language Constructs and Features – *abstract data types, polymorphism, control structures.* This is just an example, please use the correct category and subject descriptors for your submission. The ACM Computing Classification Scheme: http://www.acm.org/class/1998/

General Terms

Your general terms must be any of the following 16 designated terms: Algorithms, Management, Measurement, Documentation, Performance, Design, Economics, Reliability, Experimentation, Security, Human Factors, Standardization, Languages, Theory, Legal Aspects, Verification.

Keywords

Keywords are your own designated keywords.

1. INTRODUCTION

The proceedings are the records of the conference. ACM hopes to give these conference by-products a single, high-quality appearance. To do this, we ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to down-load a template from [2], and replace the content with your own material.

2. PAGE SIZE

All material on each page should fit within a rectangle of 18×23.5 cm ($7" \times 9.25"$), centered on the page, beginning 1.9 cm (0.75") from the top of the page and ending with 2.54 cm (1") from the bottom. The right and left margins should be 1.9 cm (.75").

The text should be in two 8.45 cm (3.33") columns with a .83 cm (.33") gutter.

3. TYPESET TEXT

3.1 Normal or Body Text
Please use a 9-point Times Roman font, or other Roman font with serifs, as close as possible in appearance to Times Roman in which these guidelines have been set. The goal is to have a 9-point text, as you see here. Please use sans-serif or non-proportional fonts only for special purposes, such as distinguishing source code text. If Times Roman is not available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.2 Title and Authors
The title (Helvetica 18-point bold), authors' names (Helvetica 12-point) and affiliations (Helvetica 10-point) run across the full width of the page – one column wide. We also recommend phone number (Helvetica 10-point) and e-mail address (Helvetica 12-point). See the top of this page for three addresses. If only one address is needed, center all address text. For two addresses, use two centered tabs, and so on. For more than three authors, you may have to improvise.[1]

3.3 First Page Copyright Notice
Please leave 3.81 cm (1.5") of blank text box at the bottom of the left column of the first page for the copyright notice.

3.4 Subsequent Pages
For pages other than the first page, start at the top of the page, and continue in double-column format. The two columns on the last page should be as close to equal length as possible.

Table 1. Table captions should be placed above the table

Graphics	Top	In-between	Bottom
Tables	End	Last	First
Figures	Good	Similar	Very well

3.5 References and Citations
Footnotes should be Times New Roman 9-point, and justified to the full width of the column.

Use the "ACM Reference format" for references – that is, a numbered list at the end of the article, ordered alphabetically and formatted accordingly. See examples of some typical reference types, in the new "ACM Reference format", at the end of this document. Within this template, use the style named *references* for the text. Acceptable abbreviations, for journal names, can be

found here: http://library.caltech.edu/reference/abbreviations/. Word may try to automatically 'underline' hotlinks in your references, the correct style is NO underlining.

The references are also in 9 pt., but that section (see Section 7) is ragged right. References should be published materials accessible to the public. Internal technical reports may be cited only if they are easily accessible (i.e. you can give the address to obtain the report within your citation) and may be obtained by any reader. Proprietary information may not be cited. Private communications should be acknowledged, not referenced (e.g., "[Robertson, personal communication]").

3.6 Page Numbering, Headers and Footers
Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

4. FIGURES/CAPTIONS
Place Tables/Figures/Images in text as close to the reference as possible (see Figure 1). It may extend across both columns to a maximum width of 17.78 cm (7").

Captions should be Times New Roman 9-point bold. They should be numbered (e.g., "Table 1" or "Figure 2"), please note that the word for Table and Figure are spelled out. Figure's captions should be centered beneath the image or picture, and Table captions should be centered above the table body.

5. SECTIONS
The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub-sections should be numbered and flush left. For a section head and a subsection head together (such as Section 3 and subsection 3.1), use no additional space above the subsection head.
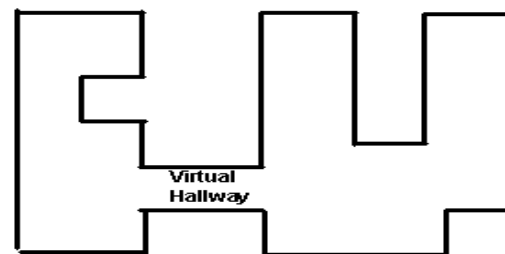
5.1 Subsections
The heading of subsections should be in Times New Roman 12-point bold with only the initial letters capitalized. (Note: For subsections and subsubsections, a word like *the* or *a* is not capitalized unless it is the first word of the header.)

5.1.1 Subsubsections
The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized and 6-points of white space above the subsubsection head.

5.1.1.1 Subsubsections
The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized.

Combined Virtual Mall

Figure 1. Insert caption to place caption below figure.

[1] If necessary, you may place some address information in a footnote, or in a named section at the end of your paper.

5.1.1.2 Subsubsections

The heading for subsubsections should be in Times New Roman 11-point italic with initial letters capitalized.

6. ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

7. REFERENCES

[1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.* 15, 5 (Nov. 1993), 795-825. DOI= http://doi.acm.org/10.1145/161468.16147.

[2] Ding, W. and Marchionini, G. 1997. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[3] Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for three-dimensional input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands, April 01 - 06, 2000). CHI '00. ACM, New York, NY, 526-531. DOI= http://doi.acm.org/10.1145/332040.332491.

[4] Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[5] Sannella, M. J. 1994. *Constraint Satisfaction and Debugging for Interactive User Interfaces*. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[6] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar. 2003), 1289-1305.

[7] Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology* (Vancouver, Canada, November 02 - 05, 2003). UIST '03. ACM, New York, NY, 1-10. DOI= http://doi.acm.org/10.1145/964696.964697.

[8] Yu, Y. T. and Lau, M. F. 2006. A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions. *J. Syst. Softw.* 79, 5 (May. 2006), 577-590. DOI= http://dx.doi.org/10.1016/j.jss.2005.05.030.

[9] Spector, A. Z. 1989. Achieving application requirements. In *Distributed Systems*, S. Mullender, Ed. ACM Press Frontier Series. ACM, New York, NY, 19-33. DOI= http://doi.acm.org/10.1145/90417.90738.