Social Networks Mining and Analysis of Specific Groups (Political and Regional) by using API

Farhan Khan[1], Saqib[1], Pireh Soomro[1], Asif Ali Laghari[1,*], Kamran Ali Memon[2]

[1]Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan
[2]School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

People have informal conversations on social media sites (Twitter, Facebook, etc) that shed light on the current issues around the world – opinions, and concerns about the learning process. Data from such environments can provide valuable information to predict the future to make effective decision making, however, the usage of such data is challenging. The complexity of such data from social media content requires the use of human understanding as well as social network analysis to make predictions. We focused on political groups around the world to discover linkages behind the Terrorism around the World to make predictions depend upon the data available. In this research paper, we focused on tweets related to the politicians and communities based on hashtags like #Syria #Weapons #Terrorist, etc. to identify the linkages of the user who are talking about these issues depending upon the HashTag we used for the content analysis of social network by using Gephi to make a prediction.

*Corresponding author. Asif.laghari@smiu.edu.pk

1. Introduction

Today social media is used by the mass of the people around the world even the public organization and politicians are using social media to reach the mass of the people and conveying the messages [1, 2]. Social media has assumed to be the first instance of communication about varies the degree of the topic that includes political debate, protests around the globe, etc. [3]. Due to the changing political environment around the globe and war on terrorism, it is very difficult to explore who is behind the scene. It is very important to identify the groups who are involved in these activities to take control and make a strategic plan to overcome these issues. Identify the public information through social media is not a big issue but to make the network of people involved in the social media discussion is a major task that needs to do to analyze to identify the key players behind the political situation and reason of war on terrorism [4].

We had several issues while analyzing social data from social websites like twitter [5]. The most important concern is about the authenticity of the user while using the social media platform. There are millions of users accounts on social media websites some accounts might be fake, some aren't carefully analyzing the user account with appropriate content is the major issue while analyzing and mining the social web content. Gathering and analyzing social content is a difficult task as this involves analyzing the complexity of unstructured data, identification of different categories and text used in social content is another issue while filtering data from the social web [6, 7]. As far as business is a concern there are many tools available for the analysis of data, but for analyzing the data for a specific group like politics we have a limited amount of data available to make predictions as well as analyzing the data.

Due to the widespread usage of social media around the globe, social media is widely used for the context of political debates, which includes citizens as well as political parties, political foundations [8, 9]. From the perspective of political debates, people are more concerned about using easy means to express their feelings based on the means of

social media. Specifically, during the election campaign, social media is the core platform for political discussion, hence using this platform for prediction and analyzing the situation is the critical factor [10, 11]. It has been observed by an analyst that social media is even adopted by political parties to engage with the public and entering into direct discussion and dialogue with the public. The main purpose behind all this is to predict the future rather than just to analyze the situation.

This research paper aims to identify political groups and predict the future. In this research paper, we used social media mining techniques and social network analysis for the identification of political groups and their interaction and communication between the masses. In this paper, we predict the behind the scene from the Syria crisis by analyzing more than 500 tweets by using hashtags provided by twitter. The contribution of this research is twofold:

1) To demonstrate and minimizing the gap between the public and politician influence by mining the data patterns from social media site (Twitter) by integrating both quantitative and qualitative data and

2) To explore informal and formal communication between the political groups and the public of the particular region.

To predict the political future of region (Syria) and their activities we have used hashtags like # Terrorist #sunni #weapon #USA #Iran etc. by using twitter API after that we use the Google refine technology and the Social Network Analysis tool to make the relation between different groups and make predictions on the basis of output. The workflow diagram of social network data analysis is given in Figure 1.
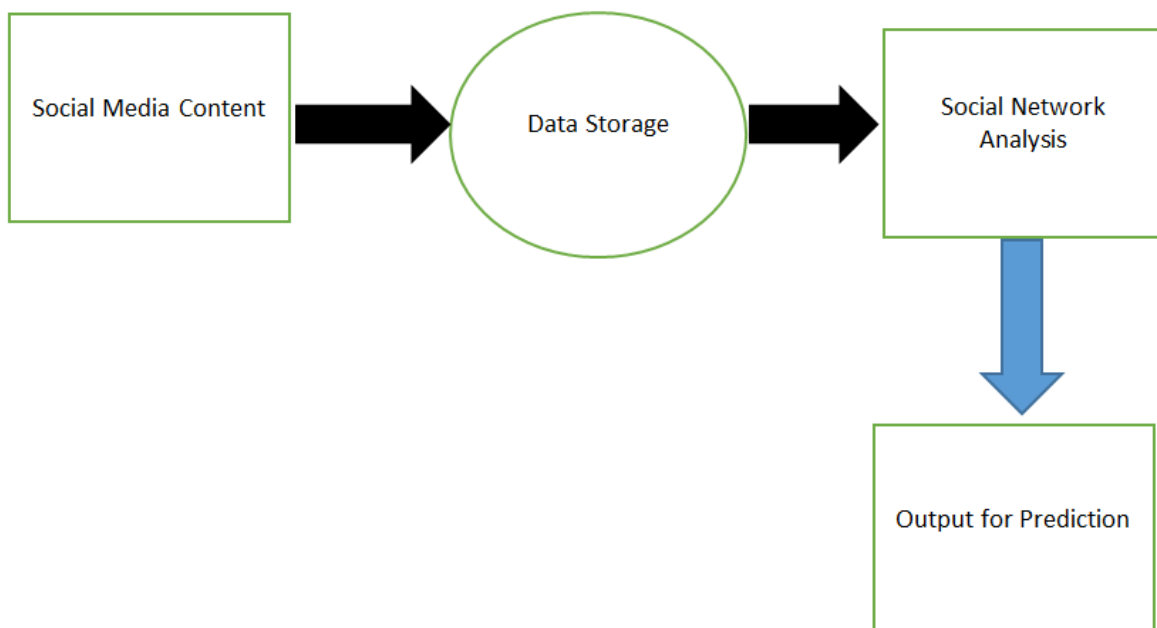

Figure 1. The workflow developed for making sense of Social Media Networking

This paper is organized into 5 sections and 2nd section based on the literature review. In 3rd section we provide methodology and 4th section based on the discussion. Finally, in section 5 we conclude research work.

2. Literature Review

Social Media Mining has opened the gateways to analyses the behavior of different communities by discovering numerous possibilities to analyze human behavior and their interactions and collected behavior by using different algorithms. Use of different algorithms for analyzing the network on social network tools like Gephi as graph-based detection technique for analyzing and detecting the groups, however before that many algorithms have been used by a data scientist for the mining and analyzing the networks of groups [12, 13].

Currently, this is the tendency of being exposed to additional social advertisements through social media platforms [14, 15]. As our morning ritual of sipping low and scrolling through the Instagram feed, this is currently notice sponsored ads showing in between filtered footage of scenery and food. It is not possible to go to one's news feed, while not sound into several compelling ads on the approach by using social media techniques. It is not valid not getting to lie, fallen victim to many of those ads, and been captured and clicked through to their web site, generally even changing. Before diving into paid ads it's vital to make out social channels with unique and clean content, quality client service, and attention-grabbing visuals [16]. Once optimize social channels for the fulfillment, it's not solely gain loyal complete promoters, however, begin capturing leads and changing guests into customers by using social media techniques.

Inevitably, social media mining techniques are using around the globe by many organizations even The Intelligence are working on these social media platforms

for detecting terrorist activities and making a strategic decision based on data available from different social media websites and microblogging platforms [17]. Social Media data enable us to analyze how information is circulated and identify the opinions being created because of that information. The mining of social data is also questioned, when the privacy of user data is concerned for that many social media sites have terms and conditions to apply on every user who is making an account on social media sites. This applies to the user who is fetching or mining data patterns from these social media sites. As more people and organizations are at the risk of security about the content they publish on social sites validity of the data available on these sites is also questioned. However, social media is still used for the political campaign and trend analysis detection of multiple groups over the network.

Researchers from a different field, especially from social data scientists, have analyzed twitter content to generate specific knowledge for their specific domain. For example, Gaffney analyzes more than 3000 tweets by using hashtag #IranElections using graphical methods and user networks and frequencies of top keyword categories to quantify online activities from Iran elections at that time [18]. Similar studies have been conducted by social scientists on the field of healthcare, athletic, antisocial behavior, child abuse and hay fever detection [19, 20, 21, 22, 23, 24, 25].

Many researchers have used different kinds of techniques for analyzing data, some have used histogram techniques for the analyzing of data, and some used classification models and networking models for the identification of different communities. In this research, we used the social networking analysis method for the identification of different communities based on hashtags used in twitter. Social-data firms spot trends that it would take a long time for humans to see on their own. Even the United Nations is using social media techniques and their algorithms derived from Twitter to pinpoint hot spots of social unrest around the world. Twitter data can also be used as an early-warning system to identify power outages based on customer complaints. Human-resources departments are another example that is using social media techniques to analyze the data to evaluate job candidates.

All of these studies have emphasized on Statistical Models and Algorithms as they cover huge topics related to the Social Media Mining including Popularity Mining and Tweets Classification [26, 27]. In all of the above, we used the tweets classification technique by refining the unstructured data into a Structured Form by Using Google refine tool, which has five algorithms for the refining of data into structured form. We also used Sentiment Analysis as a means to identify the behavior of groups whether it is Positive, Negative or Neutral emotions. In this research, we implement a Google refine the technology for the clustering and normalization of Unstructured Data into structured form for after mining it from Twitter based on hashtags.

3. Methodology

Various methods are used in this research paper to identify the communities and political groups.
1) Data has been collected from twitter API by using Tags.
2) Google refine tool has been used for the refining of unstructured data to refine the form.
3) Social network analysis tool known as Gephi has been used for the detection of political groups and their connections. In Gephi several Algorithm like Girvan Newman Clustering has been used for clustering and manipulation of structured data for detection of communities in the network for the better understanding and prediction. Girvan Newman algorithm is not only useful for the clustering of structured data but it also the best algorithm fit to analyses the network of tweets and mine from Twitter by using Twitter API.

3.1 Social Media Mining

In this modern world, the use of the Internet has revolutionized every sector of life; the Internet has changed the way people do business or the way they interacted with each other. In this Research Paper, we used Social Media Mining as a means to mine data patterns from social media website Twitter based on hashtags. Due to the limitations provided by Twitter, we mine more than 2000 tweets by Using PHP language and the WAMP server platform.

3.2   Data Collection by Using Twitter API

It is very challenging to collect data of the politicians and the Groups related to them from Twitter, because of the diversity and irregularity of language used on the social media platform (Twitter). We used the configured Twitter API to accommodate this task, which used to obtain the data set related to the current situation going on Syria. As the given Data Sets are too small we also used the irrelevant tweets for the mining of user-created content. The data of the politicians are also been collected for the identification of communities in the crisis of Syria and make predictions based on it.

Due to the limited number of relevant Tweets related to the current affairs of the Syria Crisis, we used Twitter Tags simultaneously such as #Weapon # Iran #USA #Sunni #Terrorist #Shia #Syria election #election #Blast. These are the hashTags we used for the identification of relevant content of our research paper for the prediction analysis and making a network. We used the Twitter Oathas an authentication of signatures used for the legal fetching of data from Twitter and made an application on
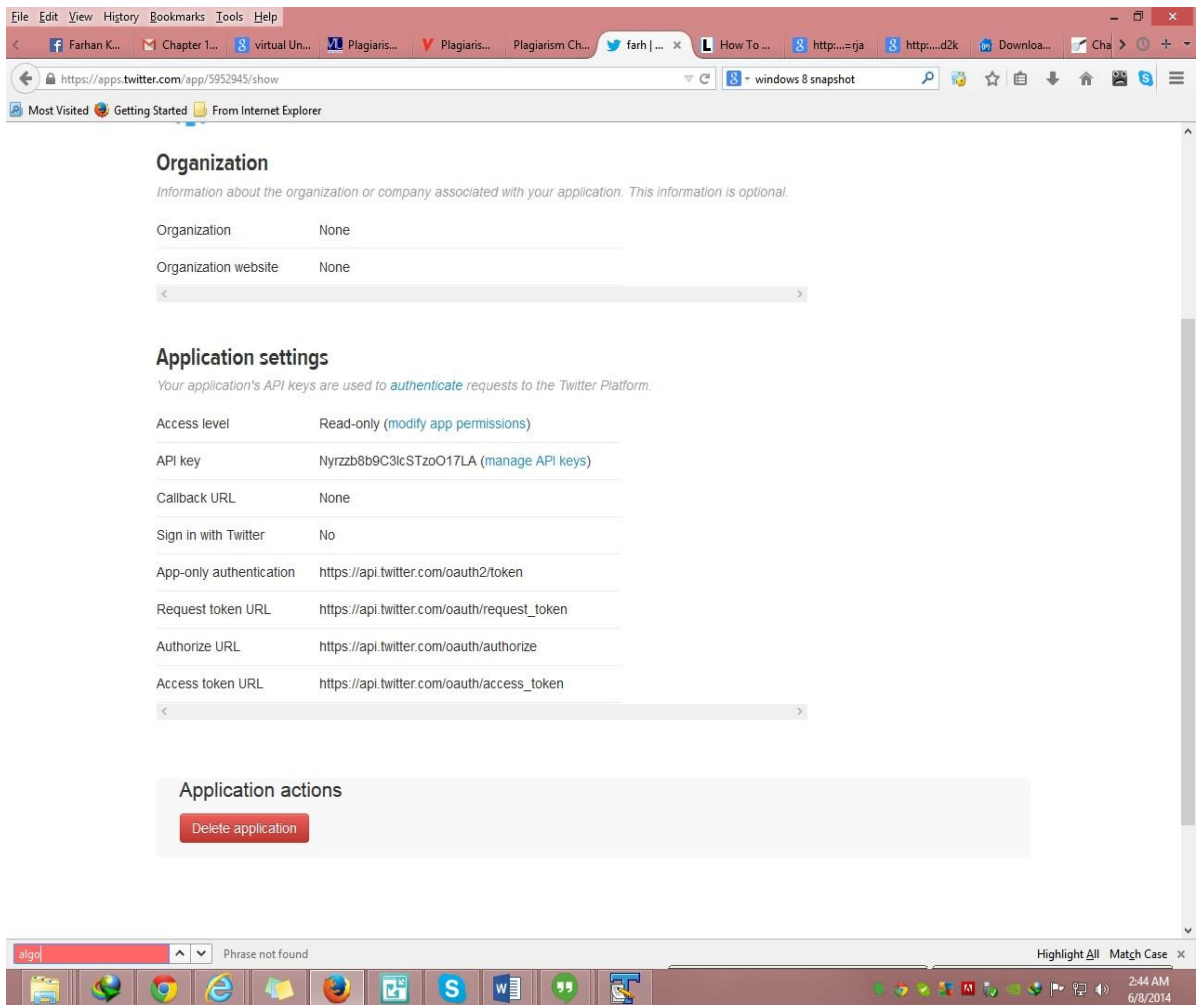
Figure 2. Twitter Oath Authentication

Twitter for the content and tweets mining. Figure 2 shows a screenshot of the Oathas authentication signature.

3.3 Exploring and Searching of Tweets

The exploration of data and communities we used the PHP file(Code)and hashtags like #Weapon #Iran #USA #Sunni #Terrorist #Shia #Syria election #election #Blast etc. for the analyzing of relevant tweets. WAMP

server platform has also used as a framework for using PHP files of twitter more conveniently. We searched the tweets that are more relevant to the Syria crisis and make political connections. The online documentation is always the definitive source for this purpose. Different types of queries were used in the Twitter file for the mining of Data. The workflow of data from API to structured data is given in Figure 3.
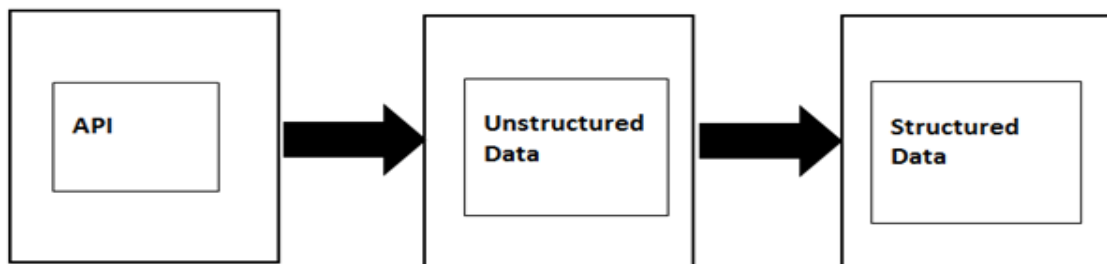

Figure 3. The Workflow developed by using API for the mining of Data Sets

The Gephi Tool is used for the social network analysis and it is more relevant for analyses or detects communities as a major player in Syria Crisis, this also supports different algorithms for analyzing the quantitative and qualitative structure of data and Gephi Algorithm known as Girvan Newman Clustering. This algorithm is used to identify the communities instead of nodes, which represents a single user, so in huge data set every single user is not important to make in the social network. Social network analysis is the mapping and measuring of relationships between the entities to find out their role and behavior in the network and flows between people, groups, organizations and other connected information/knowledge entities. The nodes in the social network are the users of social media platforms like Facebook, Google Plus and Twitter, while the links between the users show relationships or flow between the nodes [28]. Social network analysis provides both a mathematical and visual representation of data sets as well as human relationships. Even the Management consultants are using social network analysis techniques for their business clients and call it Organizational Network Analysis [ONA].

To understand social networks and their users, we evaluate the social data of people who are talking about the Syria crisis over the social media platform. Measuring the network location of the user is finding the centrality of a node. These linkages between the nodes give me insight into the various roles between the people in a network -- who are the connectors.

3.4   Cleaning of Data

Before the Social Network Analysis, data was clean and filter by using the Google refine tool for clustering and normalization [29]. Initial data was in the form of unstructured, and we are interested in who is talking with whom and for what. Google refine has the algorithms that clustered the data into categories. In our case, we are concerned about the users who are talking about Syria crisis and identification of it.

Firstly, data was fetched from the PHP Twitter file by applying Query based on the HashTags. Previously data scientist has used other means of classification of data the most commonly used method used by data scientist was the classification of data based on categories. Here is this research paper we used the modern and reliable technique for the classification of data based on hashtags and users by using Google to refine the algorithm. The flow of structure and unstructured data are given in Figure 4 and 5.
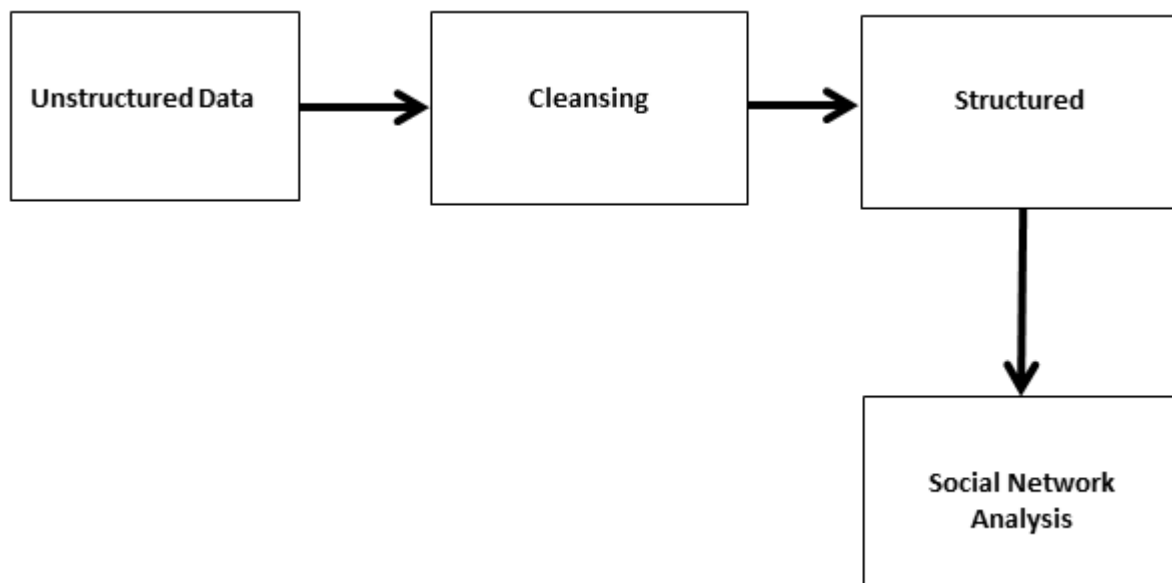

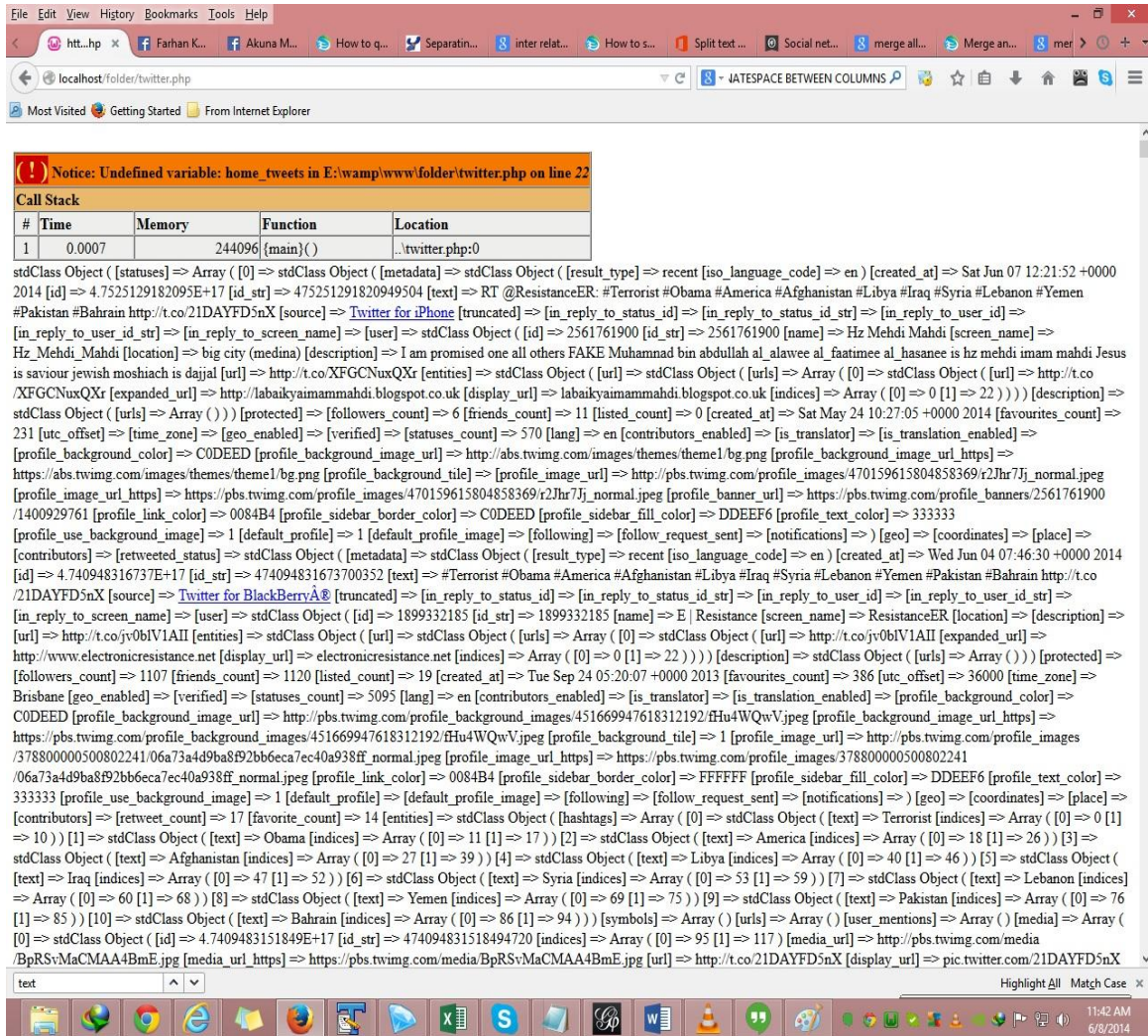**Figure 4.** The Flow from structured Data to SNA

Figure 5. Unstructured Data set after mining from Twitter API

4. Discussions and Limitation

The Methodology is used for the Social Network Analysis is by using the Gephi tool. This Tool has an algorithm to make a cluster of structured data and make the relationship between different groups and communities. Firstly, data cleaned the data in a structured form and then map the data into the Gephi tool by using Force Atlas algorithm for the representation of groups. This algorithm is very effective for the representation of groups. Nodes

and categorized Twitter Hashtags and users are given in Table 1.

After the filtration of Data through Google refine. We have mapped the data into Gephi by using Force Atlas 2 algorithm for the proper Structuring of Data and make the relationships of the key Players and identify different groups behind the terrorism in Syria and their connection with other countries. In this paper, we also mapped the countries to check the connection one of the countries with one another for the connection between them.

Table 1. Nodes of categorized Twitter Hashtags and Users

Nodes	Nodes	Nodes	Nodes		
samQusai	sam[description]				
#BoweBergdahl	#Qatar	#Terrorist	#Terrorist		
@RiaEller:	@oubliezleavion	@greggutfeld	@greggutfeld		
samAvion~Ä€	#BoweBergdahl	#Qatar	#Terrorist	#Terrorist	samQatar
@oubliezleavion	@greggutfeld				
samRIA	samAvion~Ä€	samMauritius			
samGregGutfeld					
samRIA					
samAvion~Ä€					
samGregGutfeld					
#May6_2014	#BoweBergdahl	#Qatar	#Qatar		
@NBCNews					
samRIA	samNBC	samMauritius			
#BoweBergdahl	#Qatar	#Terrorist	#Terrorist		
@oubliezleavion	@greggutfeld				
RIA	samAvion~Ä€	samMauritius			
samGregGutfeld					
#Brussels	#European	#Syria	#Syria		
samGeorge	#Syria	#News	#al-Nusra	#al-Nusra	samCambridge,
@confusion8888:					
samRedeemed	#Syria	#News	#al-Nusra	#al-Nusra	samWorldwide
samTRUE	samTRUE	samUnited			
#svpol	#Islam	#Terrorist	#Terrorist		

This study explores the previously underestimated the power of social media mining and the analysis of data through social networking algorithms and Models. To understand the network of different communities and people are connected to the Syria crisis and terrorist activism, integrating both qualitative and quantitative methods and large scale Data Mining by using social Media Techniques.

The first step is towards revealing the relationship between the different communities by using the social Media Mining Technique. But several limitations need to be encountered and can lead to future work. First, not all politicians are using Twitter as the platform for the tweets that's why this is found the ones who have an active Twitter account. Finally, the workflow is used in this research paper requires a large amount of user efforts to understand the relationships between the communities.

Recently, Social media have an impact on public communication in society. Particularly Social media are increasingly used in a political and regional context. Microblogging services (e.g., Twitter) and social network sites (e.g., Facebook) have the potential for increasing political participation. In Twitter, it is difficult to share raw data. Social Network Analysis of users and communities on the basis of Tweeted structured data view is given in Figure 6.
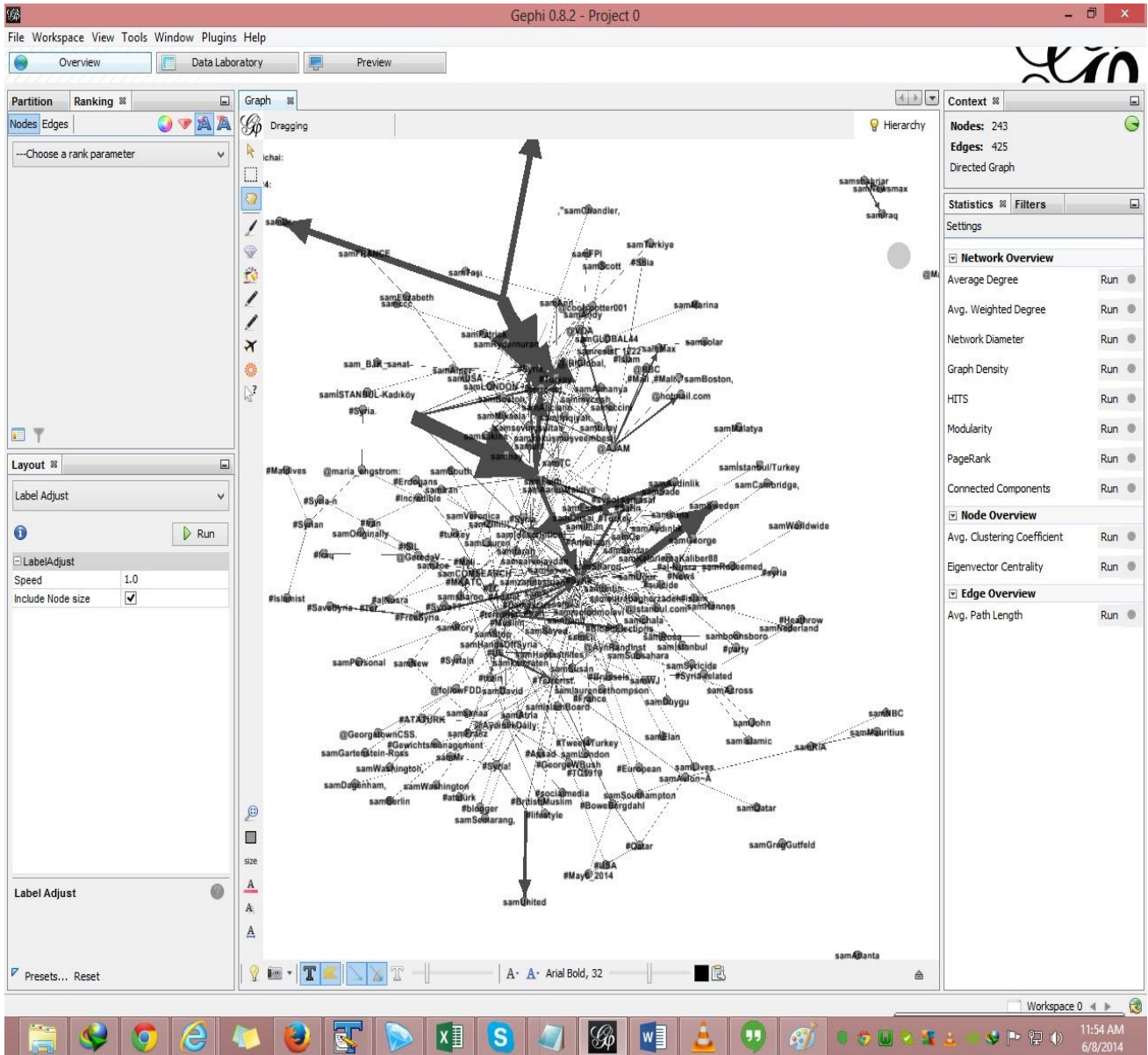
Figure 6. Social Network Analysis of Users and communities on the basis of Tweeted Structured Data

Twitter is the most common platform that allows sharing 50,000 tweets per day, the best method for sharing a full dataset is, therefore, to share the identification number of each tweet, as Twitter has no cap on that quantity for Python and R scripts to download tweets using their ID numbers, to download tweets requires connecting to Twitter's Application Programming Interface (API). "post rot." If a user deletes a tweet or account, that tweet (and all tweets from that account) is not later recoverable. While the rate of tweet or user rot is not known, as far as we are aware, anecdotal reports suggest it is about 1/10 of tweets after one year. Whether or not this rot changes a replicators' inference is not known, although there is some evidence that rot does not occur randomly. Initial claims that lauded the potential of tweets are to predict election results. systematic analysis demonstrating that those predictions were in many

instances no better than random guesses even if aggregate-level opinion is hard to measure, there is plenty of evidence that social media can reveal a lot of information about citizens' characteristics and behavior found that Facebook likes are highly predictive of private traits such as party preference, age, gender, alcohol use, psychological traits, etc. This evolving narrative about the impact of technology on protest mirrors the public discussion about how social media has been used in political campaigns.

Another key research question in this field has been whether news consumption and political conversations through social media may be one of the factors explaining the recent rise in political polarization and extremism around the world. A common argument in the literature is that social networking sites make it easier for citizens to isolate themselves into communities of like-minded

individuals, where political agreement is the norm and individuals can avoid being exposed to any opinion that may challenge their ideological view.

5. Conclusion

This research work is beneficial for the researchers as well as for the governmental departments for the identification of groups and makes a strategic plan for the controlling of terrorism and crime and their connections by using social media and social networking techniques. This research paper will help the governmental as well as the business community to make an effective decision basis on the real-time data. The growth of social media over the last decade has revolutionized the way individuals interact and industries conduct business. Individuals produce data at an unprecedented rate by interacting, sharing, and consuming content through social media.

References

[1] Tabassum, R., Froeschl, G., Cruz, J. P., Colet, P. C., Dey, S., & Islam, S. M. S. (2018). Untapped aspects of mass media campaigns for changing health behaviour towards non-communicable diseases in Bangladesh. *Globalization and health*, *14*(1), 7.

[2] Kim, J. W. (2018). They liked and shared: Effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Computers in Human Behavior*, *84*, 153-161.

[3] Rose, I. D., & Friedman, D. B. (2019). Framing gender identity and sexual orientation: media influence on young men who have sex with men's health. *Sex Education*, *19*(5), 614-626.

[4] Hassan, K. (2018). Social Media, Media Freedom and Pakistan's War on Terror. *The Round Table*, *107*(2), 189-202.

[5] Laghari, A. A., He, H., Karim, S., Shah, H. A., & Karn, N. K. (2017). Quality of experience assessment of video quality in social clouds. *Wireless Communications and Mobile Computing*, *2017*.

[6] Olteanu, A., Castillo, C., Diaz, F., & Kiciman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, *2*, 13.

[7] Ogie, R. I., Clarke, R. J., Forehead, H., & Perez, P. (2019). Crowdsourced social media data for disaster management: Lessons from the PetaJakarta. org project. *Computers, Environment and Urban Systems*, *73*, 108-117.

[8] Laghari, A. A., He, H., Shafiq, M., & Khan, A. (2018). Assessment of quality of experience (QoE) of image compression in social cloud computing. *Multiagent and Grid Systems*, *14*(2), 125-143.

[9] Seargeant, P., & Tagg, C. (2019). Social media and the future of open debate: A user-oriented approach to Facebook's filter bubble conundrum. *Discourse, Context & Media*, *27*, 41-48.

[10] Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers*, *20*(3), 531-558.

[11] Segesten, A. D., & Bossetta, M. (2019). The Eurosceptic Europeanization of public spheres: Print and social media reactions to the 2014 European Parliament elections. *Comparative European Politics*, *17*(3), 361-379.

[12] Studiawan, H., Djanali, S., & Pratomo, D. A. (2016). Graph-based forensic analysis of web honeypot. *Journal of Telecommunications and Information Technology*.

[13] Devi, N. M., & Kasireddy, S. R. (2019). Graph Analysis and Visualization of Social Network Big Data. In *Social Network Forensics, Cyber Security, and Machine Learning* (pp. 93-104). Springer, Singapore.

[14] Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: The impact of Facebook on young women's body image concerns and mood. *Body image*, *13*, 38-45.

[15] Laghari, A. A., He, H., Shafiq, M., & Khan, A. (2019). Application of quality of experience in networked services: Review, trend & perspectives. *Systemic Practice and Action Research*, *32*(5), 501-519.

[16] Safko, L. (2010). *The social media bible: tactics, tools, and strategies for business success*. John Wiley & Sons.

[17] Maras, M. H. (2017). Social media platforms: Targeting the 'found space'of terrorists. *Journal of Internet Law*, *21*(2), 3-9.

[18] Gaffney, D. (2010). iranElection: Quantifying online activism. In *In Proceedings of the Web Science Conference (WebSci10*.

[19] Jamison-Powell, S., Linehan, C., Daley, L., Garbett, A., & Lawson, S. (2012, May). " I can't get no sleep" discussing# insomnia on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1501-1510).

[20] Hambrick, M. E., Simmons, J. M., Greenhalgh, G. P., & Greenwell, T. C. (2010). Understanding professional athletes' use of Twitter: A content analysis of athlete tweets. *International Journal of Sport Communication*, *3*(4), 454-471.

[21] Singh, R., Du, J., Zhang, Y., Wang, H., Miao, Y., Sianaki, O. A., & Ulhaq, A. (2019, July). A framework for early detection of antisocial behavior on Twitter using natural language processing. In *Conference on Complex, Intelligent, and Software Intensive Systems* (pp. 484-495). Springer, Cham.

[22] Islam, M. R., Kabir, M. A., Ahmed, A., Kamal, A. R. M., Wang, H., & Ulhaq, A. (2018). Depression detection from social network data using machine learning techniques. *Health information science and systems*, *6*(1), 8.

[23] Subramani, S., Wang, H., Islam, M. R., Ulhaq, A., & O'Connor, M. (2018, May). Child abuse and domestic abuse: Content and feature analysis from social media disclosures. In *Australasian Database Conference* (pp. 174-185). Springer, Cham.

[24] Subramani, S., Vu, H. Q., & Wang, H. (2017, December). Intent classification using feature sets for domestic violence discourse on social media. In *2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)* (pp. 129-136). IEEE.

[25] Du, J., Michalska, S., Subramani, S., Wang, H., & Zhang, Y. (2019). Neural attention with character embeddings for hay fever detection from twitter. *Health information science and systems*, 7(1), 21.

[26] Bandari, R., Asur, S., & Huberman, B. A. (2012, May). The pulse of news in social media: Forecasting popularity. In *Sixth International AAAI Conference on Weblogs and Social Media*.

[27] Davidov, D., Tsur, O., & Rappoport, A. (2010, August). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters* (pp. 241-249). Association for Computational Linguistics.

[28] Karim, S., He, H., Mallah, G. A., Laghari, A. A., Madiha, H., & Larik, R. S. A. (2019, July). The Impact of Using Facebook on the Academic Performance of University Students. In *International Conference on Artificial Intelligence and Security* (pp. 405-418). Springer, Cham.

[29] Ma, Hong. "Google Refine–http://code. google. com/p/google-refine." *Technical Services Quarterly* 29, no. 3 (2012): 242-243.