Volume 4, Issue 1, January 2017 ISSN: 2348-8565 (Online)
International Journal of Modern Engineering and Research Technology
Website: http://www.ijmert.org Email: editor.ijmert@gmail.com

An Informative Survey on Terrorist Network Mining

Saurabh Singh
Jabalpur Engineering College
Jabalpur (M.P.), [INDIA]
Email: ssingh@jecjabalpur.ac.in

Shashikant Verma
Govind Ballabh Pant Engineering College
Pauri Garhwal ([INDIA]
Email:skverma.gbpec@rediffmail.com

Akhilesh Tiwari
Madhav Institute of Technology and Science
Gwalior (M.P.), [INDIA]
Email: atiwari.mits@gmail.com

Divyani Indurkhya
(B.E Research Scholar)
Jabalpur Engineering College
Jabalpur (M.P.), [INDIA]
Email:divyanindurkhya01@gmail.com

Aditya Tiwari
(B.E Research Scholar)
Jabalpur Engineering College
Jabalpur (M.P.), [INDIA]
Email: adityatiwari2117@gmail.com

ABSTRACT

Terrorist network mining is broad and hot topic of research now a days. Due to flooding of terrorist activity, research is going on with tremendous speed; in the same way to nourish this research requirement of survey paper is highly recommended. This manuscript closely related to finding the suspicious person or suspicious mode by graph computation. Graph computation is required to find out sleeper cell head, chief controller and other related persons. All the work done based on graph computation is tactfully compiled here. After going through this paper research scholar can better decide and design his framework or computation in this field.

Keywords:—*SNA (Social Network Analysis), ON (Organizational Network), FANP (Fuzzy Analytical Network Process), NTA (Network text Analysis)*

I. INTRODUCTION

The social network across the world is flourishing rapidly, engaging large number of people from extreme geographic location to communicate easily. As the network grows, its complexity also increases. This gives an extra edge to the terrorist communities to easily communicate with each other and remain undetected. There have been and are ongoing various researches in this field that use diverse methods to identify certain pattern that can detect these terrorist communities.

The network in addition to being used for communication with the members within the group is also used to influence masses to join these on the basis of their religion and beliefs. We point here that these terrorist groups can accomplish their work mainly due to the funding and the financial support that they receive mainly from their followers. The social network is helping them in attracting attention of such people who help them with their funds and play an essential role without even being involved in the network. This makes it much more difficult to find the significant soul in this network. This ill-treatment of the network for malicious intention to harm the people or any organizational structure is termed as Socio-technical attack.

Usually this social network is analyzed by projecting it as a one mode graph in which the node is assumed to be the entities and the link establishes the relationship between these entities. There might also be attributes related to each node that represents the characteristics of that particular node or the role it plays in the entire network. These networks may contain disjoint networks or even networks in which one node is connected to only two adjacent nodes which increases the secrecy of the community and prevents from the whole network being compromised. Also we know that an organizational structure is much stronger as compared to and unorganized one. This property of an organizational structure is incorporated by the terrorist communities to main their secrecy as well as efficient working.

Proper structuring and thorough analysis of the social networking graphs must be performed in order to derive the necessary information about the persons involved in the network and the relationship between various nodes.

The motivation for discovering methods for the recognition of these covert terrorist communities come from the fact that the advancement in technology though helping people is also causing havoc due to the misuse of it. Terrorism is the major threat for mankind which needs to be terminated for the betterment of the people. The advanced internet technology for attracting new recruits, spreading its propaganda and planning attacks makes the terrorist organizations like ISIS became more terrifying than the others and makes it popular among terrorist community in a very short interval of time.

In the recent years, there have been various techniques for Social Network Analysis (SNA) using graph analysis to identify significant nodes in the network or by hierarchical flow of the network to detect the level of importance of each node or by finding the importance of links between the nodes for detection.

Social Network Analysis is based on the fact that every member in the social network has social relationship with some other members of the network. Our motive is to trace the activities of terrorist groups over social media or internet and derive erudition about such groups and individuals working in them. We also need to find methods to minimize the lethalness of these terrorist groups in the society.

Terrorist networks are covert networks, so fetching information accurately and completely about such network becomes difficult and hence important. In a study, it was found that about 90% of organized terrorism over internet takes place through social media. Various methods over years have been derived for the study of these networks over web and their activities on various social media platforms such as twitter, Facebook, YouTube.

In this paper we have accumulated study and researches of various papers and the methods used by them to detect these networks.

The remaining paper is organized as follows. Section II contains the brief description of the various techniques and theories for the detection of the terrorist network introduced over years. Section III provides the strengths and weaknesses of each technique. Section IV describes the conclusion and potential future work in this field. Section V contains the references.

II. Related Work

Terror related content on social media is identified as a topic of increasing botheration for law enforcement agencies and anti-terrorist department. As is given in [6], the methods to identify terror related content on social

networking site twitter where various English and Arabic tweets and tweeps are studied and the tweets as well as tweeps involved in "media mujahedeen" are detected. Various classifiers such as violence, anger, hate and racism are used to classify different content present on the website. Common hashtags related to jihadists, and in particular ISIS (due to its huge involvement in social media terror spreading content) are used like #ILoveISIS, #AllEyesOnISIS, #ISLAMICSTATE, #KhalifaRestored. Word bigrams (of the, Islamic state, the Islamic), letter bigrams (th,he,er,an) and frequent words (the, of from, for, Islamic) are also used. Both the data dependent features (features that are heavily influenced by the specific dataset) and data independent features (features that are independent of the dataset and can be used on other datasets with similar results) are taken into consideration. SVM, Naïve Bayes and Adaboost are used as classification techniques. The method is tested with a set of English as well as Arabic tweets and tweeps that are involved in media mujahedeen as well as some random entries in both languages. Experimental results of [6] using different feature sets are reported using confusion matrix.

In [1], the investigation of a bipartite network is done to identify covert terrorist communities that might contain essential information and are often undetected. A bipartite network can be defined as a network in which there are two different types of nodes and the link between any two nodes can occur provided they belong to different sets. On the basis of type of event an individual is involved, the links between individuals is inferred i.e. the common neighbour similarity is used as a scale.

A Projection method is used which eliminates one set of nodes and provides weight to the edges and gives a one mode network which is then clustered by the Infomap technique which minimizes the average length on a random walk on a network. According to this similarity each set of sub-communities is formed based on their common interest using which the network is divided into sub communities and the weighted links help to identify the importance of a particular node. This algorithm applied via bipartite network is the known first approach using bipartite graph as reference.

Another paper by Khaled Dawoud, Alhajj and Rokne, [2], worked on modelling the terrorist network, not as graph but instead a hierarchical approach was obtained to provide essential information about the subgroups present in the network and also the flow of information from higher to lower rank. Their proposed model was able to identify the highest as well as the lowest authority of people in terrorist network. Only a well organized structure has higher strength of network as well as a well established hierarchy as compared to other networks.

In [1], the network was divided into individuals while in [2] the network is considered as a whole which can help to extract information like main individuals across the network. SNA on this network is done on basis of three factors which are: (i) Degree Measure – activeness of a node by counting number of links to a (ii) Betweenness Measure- Number of shortest path between any two nodes passing through a node (iii) Closeness Measure – sum of all the lengths to each node from a particular node. These factors are then combined and then on the basis of these evaluations, a final Org (measure) i.e accumulative of these factors is deduced for the full network. By these deductions a graph is plotted which shows the organizational level of any network. The graph from this paper is shown in Figure1 where clear levels are seen establishing the given network to be organized.

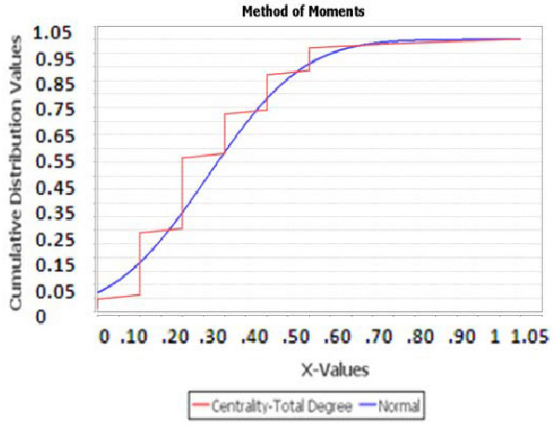

Figure 1: September 11th total degree distribution

Apart from the nature of graph, also on the basis of the three factors the nodes are listed out having maximum relevant characters which are most prior to be the suspect as well as an important member in the network.

Another important requirement is to detect the key individuals in terrorist networks and their roles and positions in the organizations. Combined theories of SNA, Multi-meta Network Analysis and Fuzzy Analytical Network Process (FANP) is used in [7] to calculate the scores of various terrorists present in the network, referencing the measures (total degree, closeness, betweennes etc) of the network computed by ORA. FANP model is use to build the evaluation system for evaluating the terrorists in the network. Various indices and criteria are used to represent the different attributes of various terrorists. Network analysis tools as ORA and MATLAB are used to compute the entire system and then obtain the weights of each criteria and index are obtained. Larger the final weighted score of the terrorist, more important role it plays in the network and hence greater is the threat to society. The FANP approach used in[7] first consider various characteristics of each terrorist present in the network. After considering the characteristics of the key individual, evaluation object is broken into four factors (criteria): Role, Resource, Knowledge and Task and further each criterion into several sub factors (indices). Figure2 below shows the three- level evaluation index system for FANP as discussed in [7].

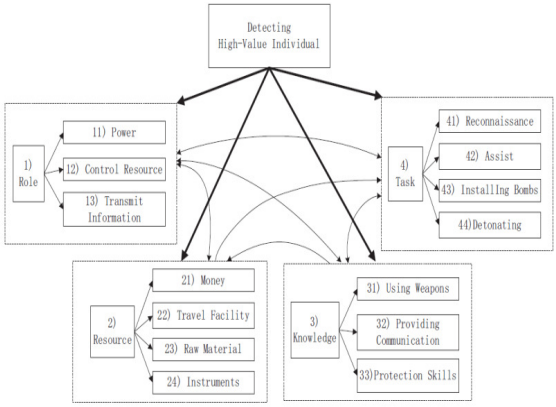

Figure 2: The FANP framework for key individual selection

In the paper, FANP model is used to build an evaluation system for evaluating the terrorists in the network. The attributes (abilities) of terrorists have given different indices which are obtained along with weights after a series of judgment and computing.

1. ***Triangular Fuzzy Number (TFN):-***When using FANP to compare the relative importance of two factors, the comparison values is not a certain number, but a set of numbers named as TFN marked as M. It is used to reduce the uncertainty.

2. ***Fuzzy Preference Programming (FPP):-***The weights of factors (indices) computed from M is determined by FPP.

3. ***FANP based approach:-***Steps are taken to identify the leaders and active members on the basis of different matrices i.e index matrix, weight matrix, limit matrix.

The vectors of criteria and indices are respectively recorded in the coefficient matrix

and un-weighted matrices. By multiplying these two matrices, a new matrix named weight matrix is built, in which the column sums to integer one. Then using formulation (1), the limit matrix is obtained.

$$\overline{W} = \lim_{k\to\infty} W^k \qquad (1)$$

Where W is the weighted matrix and W is the limit matrix.

The final score of each terrorist is the sum of the products by multiplying the score in each index and the weight of index.

Table 1 Final Scores of each terrorist with the method of FANP

Terrorist	Final score	Terrorist	Final score
Khalfan	0.0485	Wadih	0.1108
Ghailani	0.0448	bin Ladin	0.0885
Salim	0.0722	Azzam	0.0595
Al fedl	0.0486	Al owhal	0.0591
Fazul	0.0710	Fahid	0.0582
Mohammed ali	0.0526	Abdullah	0.0896
Al fwwaz	0.0521	Ahmed	0.0634
Mustafa	0.0810		

Table 1 from [7] shows that *Wadih, Abdullah* and *bin Ladin* scores the highest, which means they are the key individuals of the terrorist network of the Embassy Bombing in Kenya and Tanzania (Data is taken for these attacks). In fact, the result has good consistence with the investigation report hence the method proves to be effective in finding the leaders in the terrorist network.

All the papers described above gave importance to the text analysis and nodes in the network as a way to extract essential information from the given network. Apart from vast research on terrorist network on the basis of nodes Uffe Kock Wiil, Jolanta Gniadek, Nasrullah Memon, in [3], proposed a method which refers to links being as important and informative as the nodes and sometimes provides more information as compared to the nodes. The terrorist network as described in [3] depends mainly on two factors (i) Efficiency (ii) Secrecy.

The algorithm for calculating the link importance first finds the efficiency of each link by summing the shortest paths between each node and secrecy which depends on the number of links and nodes and there degree, higher the degree of nodes lower the secrecy. A performance product of secrecy and efficiency is evaluated and weight of each node is evaluated by using betweenness measure. The Link Importance is the product of performance and weight of each node. This concludes that which all links are important for communication in these covert networks and thus can be helpful in dismantling the terrorist network. The images shown below Figure3.1 and Figure 3.2 are taken from [3] which consist of a table that shows the measure of link importance for each link as well as the secrecy and efficiency of each link. Using this table a graph is plotted for ever attribute computing the link importance.

Link	Link importance	Secrecy	Efficiency
13	0,027759679	0,6060606	0,42307693
3	0,018798648	0,6200466	0,42635658
11	0,009842231	0,6013986	0,45081967
14	0,001790676	0,5874126	0,47413793
12	0,000169219	0,6060606	0,46218488
7	-0,002133225	0,6107226	0,46218488
4	-0,002570151	0,6013986	0,47008547
8	-0,002997429	0,5920746	0,47826087
5	-0,004772992	0,6060606	0,47008547
9	-0,005190209	0,5967366	0,47826087
2	-0,007002236	0,6107226	0,47008547
10	-0,088121056	0,5967366	0,6043956
1	-0,092513749	0,5967366	0,6111111
6	-0,092513749	0,5967366	0,6111111
15	-0,106107757	0,58275056	0,64705884

Figure 3.1 Link Importance in Part of 9/11 Network

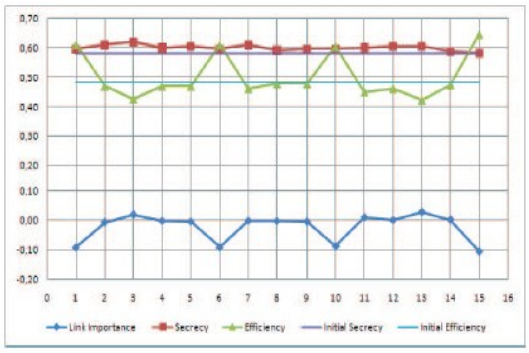

Figure 3.2 Link importance measures

This paper has involved both the importance of nodes and links to get precise results so that the identification of the network and the individuals involved in it can be done accurately.

According to[4], with the help of text analysis on the text published or released by domestic mainstream network media, a recognizable pattern is evolved which can decode the pattern of communication between the communication sources. In this paper, various different tools have been taken into use for the evaluation of every stage. The output from each stage is provided to the next stage for further computation. The Network Text Analysis (NTA) is done for text data mining with the help of computers. AutoMap is a NTA tool which is used to extract information related to terrorist organizations from the provided network data. It extracts useful vocabularies and the relationship between them and constructs nodes and links using these and discards the residue. This collected data is then processed in the form of matrix for computer processing. Element matrix lexicon is used for the classification of the constructed data. ORA is a visual effect tool which is then used to construct a graphical 2D image of the terrorist covert network obtained. ORA and UCINET are the tools which are used to analyze the data and deduce the centrality of the network. Using NetDraw2, we obtain a final visual structure on the basis of the statistics which shows the central nodes, the isolated node which indicates that there might be missing data. Through the use of this paper we are able to find the strength of relationship between each node and identify the central leaders.

Most of the researchers obtain their dataset from social websites or from communication between two or more people via SMSs or text analysis of the communication of various people in a group. But apart from these data sources, the SNA could be applied to our telephonic call pattern referred to as the call log. There are still many people in the world who do not own the facility to communicate over the internet as they find mobiles or telephones much easier to communicate. These people might include those involved in terrorist activities. From this, it can be said that the telephonic conversation pattern may also help us to detect the terrorist groups.

In[5], a network among callers and receivers is generated using a call details record which is a file that contains the source, destination, duration and other details of particular call. There are two stages in this method in which the first is the implementation of the Apriori algorithm to generate the network. In this algorithm a threshold frequency (Th) is given as input and the network of callers and receivers is given as output. The CDR network is scanned and then the set of each number is generated and there frequency is computed. Then if the frequency of any set is more than Th a link is established between them. The Example of the generated network is appended below in Figure4. After this various attributes are computed to execute the next stage which applies the Vertibi Algorithm.

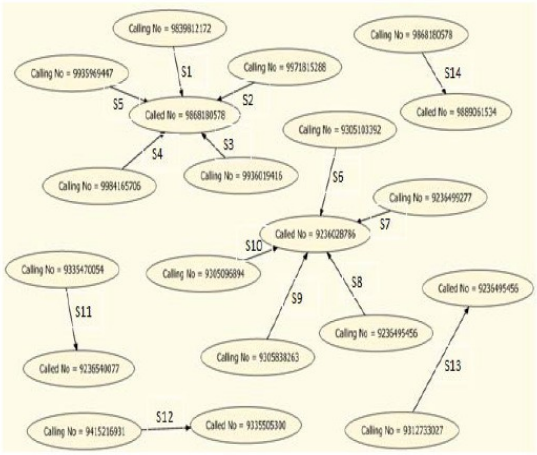

Figure 4: Network with link marked with different states from S1 to S14

Using the above shown network, a table is computed of different days and the link used at a random time. Using this table the initial probability of each state i.e. the link in the network is computed and stored in an array A. A Transition matrix is computed which shows the ratio of transition between two states to the total number of transitions. Then using the initial probability, an Emission matrix is constructed for each transition. After the computation of all these attributes, five attributes are passed to the Vertibi Algorithm – set of states, array of different point of time, Transition matrix, Emission matrix and the initial probability array. Using this algorithm the hidden pattern is extracted which might be the path for the transfer of communication between the members of the terrorist community. To conclude this was a different approach which can be helpful to detect their telecommunication pattern.

Methods are suggested in[8] to solve the Person Successor Problem (PSP):- when a terrorist is removed from a network, who is most likely to take his place?

This question is necessary in order to find which set of k (k>0) terrorists should be removed in order to minimize the lethality of terrorist network. Algorithm is discussed in [8] for PSP in which analysts can specify the conditions an individual needs to satisfy in order to replace removed person. The analysis is based on Organizational Networks (ON) which are networks whose vertices are people. The person replacing the removed person must be less lethal in order to decrease the lethality of the organization.

Person Successor Problem:- Suppose ON = (V, E, wt,) is an organizational network, r ∈ V is a vertex, and C is a logical condition on VP(Vertex Properties). The person successor problem is the problem of finding a vertex v ∈ V such that:

1. v satisfies condition C and
2. wt(v) < wt(r) and
3. there exists no v'∈ V - {r} satisfying the previous two conditions such that rv(r, v') > rv (r, v). We allow analysts to specify a logical condition C limiting who can replace a removed individual r.

STONE develops the model of who replaces a "removed" individual in a terrorist network taking into account, not only network structures, but also the properties of vertices in the network (such as operational roles) along with the need for terrorist network to attain efficiency. We identify set of k nodes whose removal will minimize the capabilities of a terrorist network.

In [9], combining of graphs of social networks via information sharing is obtained in order to derive better relationship between various nodes present in different social networking websites. Two different algorithms namely K-Nearest Neighbour (KNN) method and Edge Betweenness Based (EBB) method are used for the graph as well as sub-graph integration and privacy preservation. Given two or more social networks (G_1,G_2,G_3- -) from different organizations (O_1,O_2,O_3 - - -), objective is sharing of necessary information between these networks. The proposed information sharing and integration of social networks have three major components-

1. Constructing generalized sub-graph.
2. Creating generalized information for sharing
3. Social network integration and analysis.

An approach used for preserving privacy of relational data includes k-anonymity, l-diversity, t-closeness, m-variance and δ-presence. KNN have two basic principles, in first we assume the shortest path between two nodes v and v_i. When v is assigned to the sub-graph G_i in sub-graph

Reference No.	Strengths	Weaknesses
[1]	1. Detection of communities in bipartite network. 2. Small communities which usually are neglected by other algorithms can be detected. 3. Identification of small communities helps to find link between their members.	Cannot find overlapping communities i.e. if a person is involved in 2 communities, it gets undetected
[2]	1. Modeling terrorist networks as graphs does not give us enough information to deal with the threat. Modeling a terrorist network as a hierarchy can be a good approach to give an idea about the subgroups present in the network and also how information flows from higher ranks to lower. 2. Deals with network as a whole rather than individuals. 3. Level of hierarchy estimates the degree of organization and this degree of threat.	The network should be an organizational structure otherwise the identification cannot be done and no clear results can be found of any community.
[3]	1. The measure of link importance gives more information along with node analysis. 2. Positive and negative factors both are calculated for each link which helps to identify removal of which link will affect the terrorist network most. 3. Model was compared with transport network. i.e. a real world explanation and to which people can easily relate to and understand.	The link importance is found mainly on the basis of no. of links, degree of nodes etc. Weight of the links is not used due to which precise measure cannot be computed.
[4]	1. The algorithm provides tabular as well as visual view of the complete network providing clear distinction between the significant nodes. 2. The accurate centrality measure obtained from the algorithm distinguishes the significant nodes.	1. Missing links cannot be traced by this algorithm forming isolated nodes in the network. 2. Erroneous results due to missing links between nodes may ignore substantial nodes which might be the essential nodes.
[5]	1. he easy computation of the matrices generated from the data makes the algorithm reliable.	2. The result evaluation is done only on the basis of information deduced from CDR. 3. The threshold frequency for every generated network needs to be known prior the execution which is non-deterministic for various cases.

[6]	1. In the experiments, both data dependent and data independent features are used to examine the effect of their approach in real setting. 2. Previous works in this field used only data dependent features while in this data independent features are frequently used. 3. Experimental results on English tweets and tweeps show high accuracy as well as precision.	1. The model performs significantly bad on Arabic data in case of both tweets and tweeps. 2. In the experiments, very limited set of tweets and tweeps are used. 3. Only twitter is used as the source for content.
[7]	1. The advantage of FANP based approach is that it can deal with qualitative information on the same time. 2. Works on all the members of terrorist networks and based on the limit matrix the leaders and most important members are identified.	1. A huge knowledge about each network must be required to apply weights and indices to different attributes of various terrorists. 2. There is no discussion about the possible sources of data collection in the real world application.
[8]	1. Proper Algorithms for predicting and reshaping of the entire network is provided. 2. Operational functionality of the entire network can be minimized by removing selective nodes instead of shutting down the entire network.	Proper information about the attributes as well as roles of each terrorist must be known for the method to work (It is based on data dependent approach).
[9]	1. Experiments showed that the proposed techniques improve the closeness centrality measurements substantially. 2. Different sub-graph generalization methods can make significant impact on the effectiveness of information integration.	The proposed methods are only for two networks. Optimization process withgraph features missing.

equal to shortest path between v and v_j where j = 1, 2,.., K and j≠i. Secondly, an edge exists between two generalized nodes G_i and G_j in the generalized graph G' if and only if there is an edge between any two nodes in G such that one from each generalized node, G_i and G_j. In BNN, edges which have highest betweenness are focused. Then generalized graph is constructed by progressively removing edges with highest betweenness from original path.

III. Strengths and Weaknesses

This section describes the strengths and weakness that each paper exhibits. Most of the researches have been accomplished converting the real world network in the form of graph and extracting the essential

information via node based approach or link based approach or either by text analysis. All these approaches conclude in various counter terrorism approaches based on individual characteristics as well as relationship within the network or between various networks. Following table shows the strengths and weaknesses of papers along with their references.

Comments

The best among the discussed works is included in this section along with the reasons for their superiority over others.

[4] uses text analysis technique that provides a visual as well as a graphical presentation of the accumulated data giving it an edge over the others. The obtained result have accurate values and gives distinct centrality measures for the nodes helping find the most substantial individual in the network.

Many algorithms that extract information from graphs use nodes as the key to obtain useful data ignoring the links between them. In [3], the links are given same importance as nodes and are thus able to obtain much information about the network detecting essential nodes.

Simultaneous qualitative and quantitative analysis techniques discussed in [7] are useful for deriving characteristics over a covert network using fuzzy analysis where information is generally uncertain and inaccurate.

Information sharing and data security are the main motives of work discussed in[9]. The security measures are adopted with graph generalization and the incomplete links are established by combining of these generalized graphs of various social networks.

IV. Conclusion & Future Work

Due to the advancement in technology the scope in the field of terrorist detection on the web and on the social networking websites is profoundly increasing and so is the need to evolve effective tools for counter terrorism. In this paper we have combined various methods and techniques along with the proper description that can be used to recognize and restrain the functioning of terrorist communities.

Machine learning techniques provide fast and accurate measures for data analysis and evaluation that provided accurate and précised result generation.

Future work in this field might include the expansion of input domains and methods for much accurate result deduction. Deriving the missing links provides a better relational model of the networks which can suspect the former unsuspected substantial individuals in the terrorist networks.

REFERENCES:

[1] Taher Alzahrani and K. J. Horadam (2014), "Analysis of two crime-related networks derived from bipartite social networks", IEEE/ ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)

[2] Khaled Dawoud, Reda Alhajj, Jon Rokne (2010)," A Global Measure for Estimating the Degree of Organization of Terrorist Networks" International Conference on Advances in Social Networks Analysis and Mining

[3] Uffe Kock Wiil, Jolanta Gniadek, Nasrullah Memon (2010),"Measuring Link Importance in Terrorist Networks", International Conference on Advances in Social Networks

Analysis and Mining.

[4] Sun Duo-Yong,Zhang Hai, Guo Shu-Quan, Li Ben-Xian (2011), "Study on Covert Networks of Terroristic Organizations Based on Text Analysis". Preetish Ranjan, Abhishek Vaish, "Apriori Viterbi Model for Prior Detection of Socio-Technical Attacks in a Social Network ", 2014

[5] International Conference on Engineering and Telecommunication.

[6] Lisa Kaati, Enghin Omer, Nico Prucha and Amendra Shrestha. "Detecting Multipliers of Jihadism on Twitter". 2015 IEEE 15th International Conference on Data Mining Workshops(2015).

[7] Li Ze, Sun Duo-yong,Guo Shu-quan, Li Bo. "Detecting Key Individuals in Terrorist Network Based on FANP Model". 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining(ASONAM 2014).

[8] Francesca Spezzano, V.S. Subrahmanian, Aaron Mannes. STONE:"Shaping Terrorist Organizational Network Efficiency". ASONAM' 13, August 25-29,2013, Niagara, Ontario.

[9] Xuning Tang and Christopher C. Yang. Generalizing "Terrorist Social Networks with K-Nearest Neighbour and Edge Betweenness for Social Network Integration and Privacy Preservation". ISI 2010,Vancouver, BC, Canada.

* * * * *