Collective Aspect Analysis of Terrorist Activities Based on Social Network Analysis

Prof. S.Y. Kulkarni[1] Mukesh Kumar[2] Sagar Wankar[3] Mahesh Shinde[4] Rohit Shrishrimal[5]
[1,2]SKNCOE, SPPU University

Abstract— The primary goal is to find the relationship between terrorists based on different factors for countering terrorism. The factors accounted for are People, Organization and an Event. The people here are the terrorists. The Organization(s) is/are terrorist organizational bodies. The event is 2006 Mumbai bomb-blasts. Secondary objective or the implicit one is to further assess the graphs generated to identify key terrorists and lead investigation accordingly. This considers web page content and content from print media like newspaper as input for analysis the terrorist activities. For analysis and getting co relation, three elements are worked viz, people, Organization and event. The persons are consider as a terrorists that are directly or indirectly involved in the event. The organization is the governing body that coordinated and carried out the event. The event is the terrorist attack of interest.

Key words: Terrorist Activities, Network Analysis

I. INTRODUCTION

A social network is information network that is represented by graphs and depict the interaction between individual or entities. In these networks, each individual is represented by node (actors, persons, organizations etc) in the network, and there is an edge is provided between two nodes if an interaction has occurred, or a relationship exists, between the two individual during the observation time (friendship, conversation, financial transaction). For instance, the exchange of ideas, Knowledge, information and experiences between people in the web data can be modelled as a social network. It is significantly applied in Political Science, Information Science, Organizational Studies, Social Psychology, Biology, Economics, Communication Studies, Business Analysis and Intelligence Analysis Information as social relationships among terrorists and their groups, relations among of terrorists during planning, organizing and conducting terrorist activities as well as time and locations of terrorist activities delivers a basis for analyzing terrorist networks.

II. SOCIAL NETWORK ANALYSIS AND ITS MEASURES

Various type of measures have been developed for analyzing the social network in term of identifying key-players, community detection, discovering pattern in network, node and link discovery etc. Centrality is one of the commonly considered concept in social network analysis for identifying key elements. Several measures have been developed for centrality, including degree and betweenness, closeness and eigenvector centrality, information centrality, flow betweenness, Katz's influence measure etc

III. MAIN APPROACHES OF COMMUNITY DETECTION

Among the different proposed methods to detect communities, two main approaches can be distinguished as:

1) Hierarchical approach

In this, the nodes are visualized in a hierarchy of clusters from the discrete partitions of the whole network. This approach evaluates the closeness between two nodes through a chosen similarity measure and builds the groups using an agglomerative strategy, example like the single and complete linkage algorithm.

2) Partition clustering

This directly divides the network into a predefined number of groups. The example of such approach is minimum cut method, that defined to minimized the number of edges between them in a group.

A. Graph Visualization:

Computation of various indicators providing a local or a global description (includes the whole graph) Community detection (i.e. clustering)

B. The Necessity and Importance of Identifying Communities:

1) Communities are able to exchange and offer information because members often have similar characteristics.
2) Communities can effectively help understand the structure of social networks because communities are considered as instances of social networks and it specifies the features as well as functions of that particular network.
3) Play an important role in the visualization of large-scale social networks. The relation among communities, clear the process of information sharing and dissemination of information and also gives a vision on future expansion of network.

IV. DEFINE GRAPH TO REPRESENT A SOCIAL NETWORK

One-mode graphs : When the relationships between actors (i.e. node) are considered, the social network can be depicted by a graph G = (V,E) where V is the set of nodes associated to the actors and E is define by $E \in V \times V$. where E is the set of edges which represents their relationships. When the relationship between nodes are directed than edges are replaced by arcs. Nodes and edges can have attributes.

A. Types of Network:

1) Community detecting method in signed social network
2) Community detection methods in positive social network
3) Community detection methods for heterogeneous social network
4) Community detection methods in static social network
5) Community detection methods for dynamic social network
6) Community detection methods for directed social network

Here we discussed various community detection method based on algorithms used for different type of social network as given above.

B. Methods for Analysis of Social Network Nodes:

SNA provides various measures for analysis of central node in the network, overall importance of a node within the network, highly connected nodes in network, groups and subgroups in the network and flow of information within the network. The position of a node in the network is defined by centrality. In analysis network following types of centrality is commonly used

1) Degree Centrality:

Degree centrality is defined as the number of direct links to a node. A node having higher degree centrality value is considered as a hub and an active entity in the network. In terrorist and criminal networks it helps in identifying number of persons that can be reached directly from particular person. It is not necessary that highest degree centrality node is the leader in the network

2) Betweenness Centrality:

Betweenness centrality is a measure for identifying a node, which act as a bridge to make connections to other groups or communities in the network. It can be defined as the number of shortest paths between any pair that pass through a node. A node with higher Betweenness centrality considered as powerful node with great amount of influence. For example in terrorist and criminal networks, it helps in finding a person may be a potential broker (having maximum information) between two groups or communities in that network.

3) Closeness centrality:

Closeness centrality is a measure of how fast one can reach from a node to all other nodes in the network. It can be defined as the mean length of all shortest paths between a node and all remaining nodes in the network. A node with high closeness centrality is much closer and can quickly access the other nodes in the network. In terrorist networks, it may be useful in identifying the person which can quickly access other persons in the network.

4) Eigenvector centrality:

Eigenvector centrality is a measure of relative importance in terms of influence of a node to its neighboring nodes in the network. It is generally used for finding the most central node in the network globally. The high eigenvector centrality node is generally considered as a more central node with more influence over other nodes and act as a leader in the network. In terrorist network analysis, it helps in identifying the person, well connected to other well connected terrorists.

5) Page Rank:

Page Rank is a measure for computing the relative importance and to assign the rank to the nodes in the social network. Page Rank algorithm play an important role in terrorist network analysis and overall importance of a terrorist can be determined based on its position in the network.

V. FUNCTIONALITY IN SOCIAL NETWORK ANALYSIS

Functionalities are firstly the visualization of the network, secondly the statistics computation based on nodes and on edges, and at last, community detection (or clustering)

A. Visualization of The Network:

Methods

1) Fruchterman Reingold
2) Kamada-Kawai (which has a faster convergence than Fruchterman Reingold, but which sometimes does not provide so good results than this last one)

B. Computation of Statistics Based On Nodes:

1) Vertex and Edge Scoring:

The place that is given to the actor in the network can be stated using measures based on vertex scoring. Most sought after types of vertex scoring are the centrality measures. In graph theory and network analysis, there are various measures of the centrality analysis of a vertex to determine the relative importance of this vertex in the graph

- Degree centrality
- Closeness centrality
- Betweeness centrality

2) Page Rank:

The score computed by Page Rank is higher for nodes that are highly connected to other nodes and connected with nodes that are highly connected themselves.

3) HITS Algorithm:

Hyperlink-Induced Topic Search (also known as hubs and authorities) calculates two scores: hub and authority score. The vertex has more outgoing arcs, the higher is its hub score. The vertex has more incoming links, the higher is its authority score.

VI. ARCHITECTURAL DESIGN

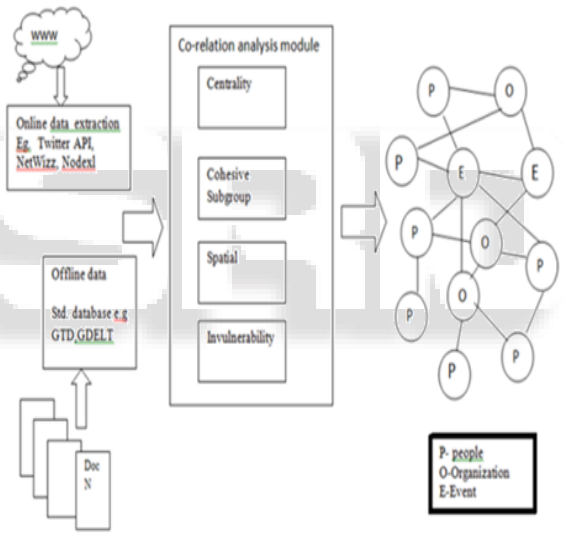

Fig. 1:

1) Centrality calculation – Centrality measures are evaluated, so that the relative positions of the nodes within the graph can be assessed.
2) Cohesive subgroup analysis – The subgroups are generated through cluster formation. Invulnerability assessment – If one or few of the elements are eliminated, then the relative strength is determined.

- Actors and their actions are viewed as inter-dependent and autonomous units.
- Linkages (Relational ties) between actors are paths for transfer of material and non-material.
- Network models conceptualize structure as patterns of relations across actors.

The architecture illustration we proposed is divided among 3 modules:

A. Input Database Generation:

Text information from news media provides information as a input database about entities (i.e. people and organizations)

and their inter-dependent relations and involvement in event of choice . This information is then encoded in text and across different formats. Databases can provide networks in a more structured form. Example of different database types includes transactions between individuals like Bank accounts, PayPal collaboration or references like patent, research, movie databases and human-annotated and entered data.

B. Correlation Analysis:

To predict behaviour, identity, personality, or any kind of pattern of life, and the aftermath of such negotiations are a few of the applications that may exploit data from sensory type of systems. Another approach to extract networks is to search for relationships in text. For instance, within a document, Nikhil and Rekha are referred to as Husband and Wife. This approach is intriguing but has its drawbacks. In many situations, relationships are implicit and are, for sake of avoiding extra work and/or redundancy, not stated. Thus problem then becomes a process of inferring relations rather than simply extracting those from text.

C. The Graph Generation:

Correlation accounts for the calculation of centrality measures discussed beforehand and invulnerability assessment. The graph so generated will consist of people, organization and event of interest.

Extracted structured content from raw data can be encoded using knowledge representation methods and ontology. For every input data (e.g., text, speech, image), output produced are a set of objects, attributes, and predicates that conform to an ontology describing structured information in the document. An ontology for information extraction, primarily the Automated Content Extraction (ACE) protocol, is useful.

VII. CLUSTERING OR COMMUNITY DETECTION

Clustering is sought after to detect groups of nodes with heavy connections within the groups and lesser connections between the groups (inter- and intra-group connections resp.)

A. Hierarchical Approach:

This approach sees nodes aggregated in a hierarchical clusters from the discrete partition to the whole network e.g Newman and Givan method, Spinglass, FastGreedy (FG), Louvain (LV)

1) Fast Greedy (FG):

In the initial state, each node is considered as a community itself. The algorithm then merges those communities step by step until only one remains, containing all nodes. The greedy technique is applied at each step, by considering the largest increase (or smallest decrease) in modularity as the merging factor. Because of its hierarchical nature, FG produces a hierarchy of community structures, a manner similar to the divisive approaches. The best approach can be selected by comparing their modularity values.

VIII. CONCLUSION

SNA is considered as one of the powerful tools for analyzing the activity that has taken place. It can be a criminal community as well. Various algorithms of SNA measures like: centrality, community detection, clustering, and information flow are very effective to find the key players or leader, their modus operandi of communication and then devising a disrupting strategy for networks. But also, there are challenges associated with the data gathering and analysis of some terrorist activities networks. Data used is often open source, available on public platform whereas the reliability of this data could be a concern.

ACKNOWLEDGMENT

Designing phase of our project .We sincerely feel that, all the faculty members of Computer Engineering Department, associates, our parents and friends, who were extremely supportive throughout the completion of this phase. We extend our sincere appreciation to all of them for the immense cooperation, persistent inspiration and encouragement. We strongly believe in teamwork and have been practising it throughout. Our cohesive efforts, mutual understanding and unmatched devotion in completion of this phase has been a memorable experience to all of us.

REFERENCES

[1] Freeman, Linton C., Borgatti, Stephen P., and Douglas R. White., (1991) "Centrality in valued graphs: A measure of betweenness based on network flow", Social networks, Vol. 13, No.2, pp141-154.
[2] Katz, Leo, (1953) "A new status index derived from sociometric analysis", Psychometrika Vol. 18, No. 1, pp39-43.
[3] Carley, K. M., &Reminga, J., (2004) "ORA: Organization risk analyzer (No. CMU-ISRI-04-106)", Carnegie-Mellon University, Pittsburgh, Institute of Software Research Internet.
[4] Hagberg, A., Schult, D., Swart, P., Conway, D., SéguinCharbonneau, L., Ellison, C. and Torrents, J. (2004). "Networkx. High productivity software for complex networks".
[5] Csardi, G., &Nepusz, T. (2006). "The igraph software package for complex network research.InterJournal, Complex Systems", 1695(5).
[6] Smith, M., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., & Dunne, C., (2010) "NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010", Social Media Research Foundation
[7] David Combe, Christine Largeron, Elod Egyed-Zsigmond, Mathias G_ery " A comparative study of social network analysis tools", WEB INTELLIGENCE & VIRTUAL ENTERPRISES, Oct 2010, Saint-Etienne, France.
[8] O'Madadhain, J., Fisher, D., White, S., &Boey, Y., (2003) "The jung (java universal network/graph) framework", University of California, Irvine, California.
[9] Borgatti, Stephen P., Martin G. Everett, & Linton C. Freeman, (2002) "Ucinet for Windows: Software for social network analysis".
[10] Bastian, M., Heymann, S., &Jacomy, M., (2009) "Gephi: an open source software for exploring and manipulating networks". ICWSM, 8, 361-362.
[11] Batagelj, V., &Mrvar, A., (1998) "Pajek-program for large network analysis", Connections, Vol. 21, No. 2, pp47-57.

[12] Sarvari, H., Abozinadah, E., Mbaziira, A., & McCoy, D., (2014). " Constructing and Analyzing Criminal Networks", IEEE Security and Privacy Workshops.
[13] Leetaru, Kalev, and Philip A. Schrodt, (2013) "GDELT: Global data on events, location, and tone, 1979–2012", ISA Annual Convention, Vol.
[14] Karthika, S., and S. Bose, (2011) "A comparative study of social networking approaches in identifying the covert nodes", International Journal on Web Services Computing (IJWSC), Vol. 2, pp65-78.

