A Survey on Social Network Analysis for Counter-Terrorism

Pankaj Choudhary
Department of Computer Engineering
Defence Institute of Advanced Technology
Pune, India

Upasna Singh
Department of Computer Engineering
Defence Institute of Advanced Technology
Pune, India

ABSTRACT
Terrorist Activities worldwide has approached the evolution of various high-ended methodologies for analyzing terrorist groups and networks. Existing research found that Social Network Analysis (SNA) is one of the most effective and predictive method for countering terrorism in social networks. The study reviewed various SNA measures for predicting the key players/ main actors of terrorist network in terms of global as well as Indian perspective. Comparative study among SNA tools demonstrated the applicability and feasibility for online and offline social networks. It is recommended to incorporate temporal analysis using data mining methods. It can enhance the capability of SNA for handing dynamic behavior of online social networks.

General Terms
Social Network Analysis, Data Mining.

Keywords
Social Network Analysis, Terrorist Networks, Counter-Terrorism, Centrality, Online and Offline Social Networks.

1. INTRODUCTION
Social Network Analysis (SNA) is the application of network theory to analyze social networks in terms of social relationships. It comprises of nodes (actors, persons, organizations etc.) within the network and ties representing relationships (friendship, kinship, conversation, financial transaction etc.) among the nodes. Social relationships may be in the form of real world offline social networks (like friendship, kinship, communication, transaction etc.) or it may be online social networks (like Facebook, Twitter etc.). Various SNA measures has been used for representing interaction among actors, examining strong or weak ties, identifying key/central players and subgroups in network, finding topology and strength of network. Recently SNA has appeared as a practice in various domains. It is significantly applied in Information Science, Political Science, Organizational Studies, Social Psychology, Biology, Communication Studies, Business Analysis, Economics and Intelligence Analysis. Facebook, Twitter and few more social networking sites use various measures of SNA to develop strategies and policies for users.

Studies based on SNA in Counter-Terrorism became popular just after the attacks of 9/11. SNA has extensively been used by the intelligence and law enforcement agencies for understanding the structure of terrorist networks and developing strategies to disrupt them by recognizing leaders and hidden patterns in the criminal and terrorist networks. Some common applications of SNA in Counter-Terrorism are Key-Player Identification, Community Discovery, Link (relation) Analysis, Node (actor) Discovery and Dynamic Network Analysis etc.

This paper outlines a variety of measures and techniques used in SNA for counter-terrorism activities. Section 2 outlines the review of research done using various SNA measures in the field of counter-terrorism analysis. Section 3 provides detailed comparison of various tools and techniques for online and offline SNA. It also highlights the methods for collecting social network data. Section 4 provides suggestions and comments on the existing measures, future scope and directions for using SNA in Counter-Terrorism and wraps up the paper with a concise summary.

2. RELATED WORK
In this section, we present a general idea of research work done in the field of SNA with a more broad set of prior research focused on counter-terrorism and analysis of criminal networks.

2.1 Social Network Analysis and its Measures
Various measures have been developed over time for analyzing the social network in term of identifying key-players, community detection, finding pattern in network, node and link discovery etc. Centrality is one of the commonly considered notions in social network analysis for identifying key players. Several measures have been developed for centrality, including degree, betweenness, closeness, eigenvector centrality, information centrality, flow betweenness [8], Katz's influence [9] measure etc.

Freeman [1] proposed three different intuitive notion of centrality namely degree, betweenness and closeness centrality, which are mostly used in identifying key players in the social network. Bonacich [2] [3] proposed eigenvector centrality for finding relative importance of a node, which is mostly used for defining the influence of a node on its neighboring nodes. Everett and et al. [4] in their work, extended the centrality measures of degree, closeness and betweenness to apply to groups and classes as well as individuals. UCINET, a tool for analyzing social network data was released then by Everett and et al. [5] with most of the common SNA measures for analysis. Borgatti [6] suggested a typology of network flows depending on two dimensions of variation. Everett [7] studied social networks containing negative ties such as dislike, evading, and conflict and proposed a new centrality measure, PN centrality for both positive and negative ties.

2.2 Social Network Analysis in Counter Terrorism
2.2.1 Global Perspective
After tragic 9/11 attacks, various studies involved the use of social network analysis in countering terrorism and crime. Key-player identification, community discovery, covert

network analysis, dynamic network analysis, and disruption of terrorist networks are the most common studies among them.

Krebs [10], studied and mapped the 9/11 Al-Qaeda terrorist network by gathering publicly available information on 19 hijackers of Al-Qaeda and applying basic SNA centrality and community measures with the help of SNA tools to identify the key p layers and leaders in the network. This research gives some vision for further work and research into the terrorist networks analysis.

Carley [11] studied social network analysis and multi-agent models for destabilizing networks. It explain the challenges in disrupting networks that are large, distributed and dynamic in nature. It also proposed multi-agent model for agents, knowledge and tasks based on various meta-networks (like: social network, knowledge network, information network, assignment network etc). Carley [12] proposed an approach for evaluating destabilization strategies for covert terrorist networks by using data in publicly available records, newspaper reports, and professional journals.

Jennifer Xu [13] categorized existing criminal network analysis approaches and tools for identifying subgroups, discovering their patterns of interaction, finding key players, and uncovering organizational structures.

Memon [14] proposed an algorithms for creating hierarchy of the covert networks and for better understanding the structure of the informal or unusual organizations, in order to disrupt these networks. The algorithms are verified by using open source dataset.

An overview of the history of social network analysis in counter-terrorism research is provided by Ressler [15]. The study was focused on various academic, defence and government activities for data collection and modelling for terrorism networks.

Everton [16] suggested to consider overall topography of social network before shaping strategies for their disruption by using SNA centrality measures for identifying key players. Size, number of attacks and network resiliency must be included for effectively computing the effectiveness of terrorist networks.

Roberts [17] explored kinetic and non-kinetic approaches to disrupting dark networks. The kinetic approach involves aggressive and offensive measures to eliminate or capture network members, while the non-kinetic approach involves the use of SNA for combating dark networks. To explain both these approaches they used network data from of Noordin Top's terrorist network (operational and trust network) of South East Asia.

Everton [18] studied and prepared the Noordin Top's terrorist network from 2001-2010 in order to detect changes over time in network structure (i.e. density, centralization) and effectiveness (i.e. Recruitment of members, quality of members). This study suggested that, dense but decentralized terrorist groups are more effective and more difficult to disrupt.

Klausen [19] studied the social network of YouTube account holders linked with al-Muhajiroun's posts of jihadist and violent contents on YouTube by applying various SNA measures.

Using social network analysis and media based evidences, Wu [20] tried to find the possible leader after Bin Laden by examining the Al-Queda terrorist network by considering network position in terms of social capital as an important factor for identifying the new leader.

Sarvari [21] created a large scale social network graph through a set of publicly leaked email addresses of several criminals and by identifying Facebook profiles connected to these email addresses. By applying various social network analysis measures they recognized profiles of high rank criminals, various criminal communities and public groups on Facebook.

2.2.2 Indian Perspective

Saxena [22] studied the network of terrorist organizations, operating actively in State of Jammu & Kashmir of India by identifying their relations, and then by applying various methods of analysis to identify the key players among these organizations.

With the help of public or intelligence data and applying methods of social network analysis Basu [23] studied major terrorist groups operating in India and their linkages to international terrorist groups.

Azad [24] analyzed the network of 26/11 terror attacks of Mumbai (India), based on the communication between ten attackers and their handlers in Pakistan. Using various network analysis measures they identified the key leaders and sub groups inside the terrorist network.

Karthika [25] studied and compared the research work done in the field of social network analysis for covert networks and categorized various approaches in social network analysis like identifying key player and subgroups, destabilizing terrorist and covert networks, dynamic network analysis and so on.

3. TOOLS AND TECHNIQUES USED FOR SNA IN COUNTER-TERRORISM

A social network is an architecture which consists of set of nodes (as actors or organizations) and relationships between these nodes. These social network may be online as well as offline in nature. Online social networks are basically in the form of online social networking sites (like Facebook, Twitter, YouTube, Google+, LinkedIn etc.), which allow users to interact with other users through sending messages, posting information, images, videos, likes and comments on them. On the other hand offline social networks are the real life social networks based on the relations like friendship, kinship, communication, financial transactions, locations, events etc.

3.1 Methods for Data Collection

In recent works the data is collected from online as well as offline social networks.

3.1.1 Data Collection from Online Social Networks

Data collection from online social networks include the extraction of public and private data of users, groups and pages, which contain posts, tweets, likes, comments, photos, videos etc. A number of tools and APIs are available for extracting data from various online social networks like Facebook, Twitter, YouTube and few more. Facebook graph API, Twitter API, Netvizz [26] and NodeXL [27] are such tools for extracting social network data and further analyzing most of the criminal and terrorist activities using online social networks.

3.1.2 Data collection from offline social networks

Data collection for offline social network analysis involves the public and open source data from various news articles, reports, phone call records, textual analysis etc. For offline social network analysis of terrorist and criminal networks, few services and databases are available like GTD [28] and GDELT [29].

Global Terrorism Database (GTD) is service which provides open source data for various terrorist and criminal activities around the world. GDELT is a project which monitors and extracts the network of individuals, events, organizations and locations based on the news media and other open source information. Using these services with manual data collection, more relevant data can be gathered for modelling and analyzing terrorist and criminal networks.

3.2 Methods for Analysis

SNA provides various measures for analysis of central node within the network, overall importance of a node in the network, highly connected node in network, groups and sub-groups in the network and flow of information within the network.

The measure of importance and position of a node in the network is often defined by centrality. In analysis of terrorist network following types of centrality is commonly used:

3.2.1 Degree Centrality

Degree centrality [1] is defined as the number of direct links or connections to a node. A node with higher degree centrality value is often 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.

Generally social network G is represented as adjacency matrix A. Degree Centrality of node i in the network can be defined as,

$$D_i = \frac{\sum_{j \in G} A_{ij}}{N-1} \quad (1)$$

Where A_{ij} is the element in the adjacency matrix A at ij[th] position. N is the total nodes in the network. (N-1) is the factor for normalization.

3.2.2 Betweenness Centrality

Betweenness centrality [1] 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.

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.

Betweenness centrality of node i the network can be defined as:

$$B_i = \frac{\sum_{j,k=G} g_{jk}(i) \Big/ g_{jk}}{(N-1)(N-2)} \quad (2)$$

Where g_{jk} is total shortest paths between two nodes j and k, and $g_{jk}(i)$ is total shortest paths between j and k which pass via node i. For normalize the value (N-1) (N-2) is used.

3.2.3 Closeness centrality

Closeness centrality [1] 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 other 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.

Closeness centrality of node i the network can be defined as:

$$C_i = \frac{N-1}{\sum_{j \in G} d_{ij}} \quad (3)$$

Where d_{ij} is the distance or path length between two nodes i and j.

3.2.4 Eigenvector centrality

Eigenvector centrality [2] 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. A node with high eigenvector centrality 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 persons.

Eigenvector centrality of a network can be defined as:

$$X_i = \frac{1}{\lambda} \sum_{j \in Nb_i} A_{ij} . X_j \quad (4)$$

Where Nb_i are all the neighboring nodes of i and λ is eigenvalue. A_{ij} is the element in the adjacency matrix A at ij[th] position. X_j is the eigenvector centrality of node j.

3.2.5 PageRank

PageRank [30] is a measure for computing the relative importance and ranking the nodes in the social network. Using PageRank in terrorist network analysis overall importance of a person can be determined based on its position in the network.

PageRank of node a in network can be defined as:

$$PR(a) = \sum_{b \in Nb_a} \frac{PR(b)}{L(b)} \quad (5)$$

Where Nb_a are the nodes connected to node a and $L(b)$ is the total links outgoing to node b. $PR(b)$ is the PageRank of node b.

3.3 Comparison of SNA Tools

Various tools and libraries are commonly used by researchers for social network analysis and visualization of terrorist and criminal networks. We considered following commonly used tools for comparison with respect to their functionality, platform, license type and file-formats respectively, based on the features like network visualization, computation of node-level and network-level measures (centrality, community, power and information flow) and handling of large networks with many nodes.

Table 1. Comparison of various tools for social network analysis

Tools	Functionality	Platform	License Type	File Format	Limitations
UCINET [5]	Analysis of social network data, Visualization through NetDraw, Various Centrality and Power measures, Clustering and Community algorithms, Various data Import and export formats, Handling with large datasets	Windows	Commercial and Academic	.dl, .net, .vna, .csv, Raw matrices	Platform dependent, Not Scalable, Lack of dynamic network analysis
Gephi [31]	Social network analysis and Visualization, various layouts for visualization, various measures of centrality, clustering and modularity, Ranking and partitioning of network, Timeline feature, plugin support available	Windows, Linux	Open-source	.dot, .gml, .gdf, .graphml, .net, .dot, .dl, .gexf, .csv, databases	Not suitable for very large datasets
ORA [32]	Analysis and visualization of two mode and multi-relational networks, Various visualization layouts, Dynamic network analysis, Measures of centrality, power and clustering. Report Generation of analysis.	Windows	Commercial and Academic	.xml, .dynetml, .zip	Platform dependent, Limited number of nodes.
NodeXL [27]	Template for MS-Excel, Visualization and Analysis of networks, Data import directly from few online social networks, Data export in various formats, Visualization layouts, Measures for analysis.	Windows	Open-source	.dl, .net, .graphml, matrix formats	Incapable of handling and visualizing large datasets
JUNG [33]	Java API and library for modelling, analysis and visualization of social network, Includes various algorithms for centrality, clustering, flow etc., Capable of handling very large datasets	All platform	Open-source	.net, .graphml	Lack of interactive interface, Not appropriate for dynamic network analysis.
Pajek [34]	Social network analysis and visualization for large networks, Two-mode, multi-relational and temporal network analysis, Algorithms for Centrality and graph layouts	Windows, Linux	Open-source	.net, .dl	Limited number of nodes, less customization
NetworkX [35]	Python library for analysis of social network, Supports two-mode and temporal networks, Few centrality measures, Handel large networks well	Windows, Linux	Open-source	.gml, .graphml, .net	No user interface, No direct graph importing
Igraph [36]	Library in R, Python and C for network analytics, Supports two-mode networks, Lot of measures for centrality, clustering and layouts, Handel very large networks with millions of relations	Windows, Linux	Open-source	.gml, .graphml, .net	Lack of user interface, No direct importing from online social network

Today's requirements of SNA tools for counter-terrorism research include the high-level visualization of terrorist networks, temporal analysis of networks over time, ability to analyze two-mode and multi-relational networks, dynamic network analysis of terrorist groups and use of data mining and big data analysis techniques for analyzing very large networks.

4. CONCLUSIONS AND FUTURE RECOMMENDATIONS

SNA can be considered as one of the powerful tool for analyzing the terrorist and criminal networks. Various algorithms of SNA measures like: centrality, community detection, clustering, and information flow are very effective for finding the key players or leader, their pattern of communication and then developing the disrupting strategy for terrorist networks. These measures are very helpful in analysis of covert, decentralized and large terrorist network. But also there are challenges associated with the data collection and analysis of terrorist networks. The data collection for terrorist networks is very difficult for researchers as it is highly confidential and covert in nature. So, researcher often uses data which is open source on public platform whereas the reliability of this data is again a serious concern. Most of the times data tends to be incomplete with lot of missing and fake nodes and relations, which often leads to incorrect analysis result of such terrorist network. Also the terrorist networks are highly decentralized and dynamic in nature with continuously changing nodes and their relations. In case of online social network analysis the privacy of individual while crawling personal data is also very sensitive issue.

Considering these challenges and limitations, future research recommends the requirement of more effective and reliable methods for data collection in such network. It is also necessary to study the discovery of spoofing or fake nodes and relations and benchmarking for network analysis. For handling dynamic nature of terrorist networks, temporal and spatio-temporal analysis is recommended by using timeline based approach. Data mining and big data analysis techniques can be integrated with existing techniques to deal with dynamic analysis. Further research is required for effective and predictive modeling of attacks and identifying the potential leaders in a particular network.

5. ACKNOWLEDGMENTS

Our sincere thanks to all the members of Digital Forensics Lab, Dept. of Computer Engineering, DIAT, Pune for successful completion of this study.

6. REFERENCES

[1] Freeman, Linton C., (1979) "Centrality in social networks conceptual clarification", Social networks, Vol. 1, No. 3, pp215-239.

[2] Bonacich, Phillip, (1972) "Factoring and weighting approaches to status scores and clique identification", Journal of Mathematical Sociology, Vol. 2, No.1, pp113-120.

[3] Bonacich, Phillip, (1987) "Power and centrality: A family of measures", American Journal of Sociology, pp1170-1182.

[4] Everett, Martin G., & Stephen P. Borgatti, (1999) "The centrality of groups and classes", The Journal of mathematical sociology, Vol. 23, No.3, pp181-201.

[5] Borgatti, Stephen P., Martin G. Everett, & Linton C. Freeman, (2002) "Ucinet for Windows: Software for social network analysis".

[6] Borgatti, Stephen P., (2005) "Centrality and network flow", Social networks, Vol. 27, No.1, pp55-71.

[7] Everett, Martin G., and Borgatti, Stephen P., (2014) "Networks containing negative ties", Social Networks, Vol. 38, pp111-120.

[8] 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.

[9] Katz, Leo, (1953) "A new status index derived from sociometric analysis", Psychometrika Vol. 18, No. 1, pp39-43.

[10] Krebs, Valdis E., (2002) "Mapping networks of terrorist cells", Connections, Vol. 24, No. 3, pp43-52.

[11] Carley, K. M., Ju-Sung Lee, and David Krackhardt, (2001) "Destabilizing networks", Connections, Vol. 24, No. 3, pp31-34.

[12] Carley, K. M., Dombroski, M., Tsvetovat, M., Reminga, J., & Kamneva, N., (2003) "Destabilizing dynamic covert networks", Proceedings of the 8th international Command and Control Research and Technology Symposium.

[13] Xu, Jennifer, and Hsinchun Chen, (2005) "Criminal network analysis and visualization", Communications of the ACM, Vol. 48, No.6, pp100-107.

[14] Memon, Nasrullah, and Henrik Legind Larsen, (2006) "Practical algorithms for destabilizing terrorist networks", Intelligence and Security Informatics. Springer, Berlin Heidelberg, pp389-400.

[15] Ressler, Steve, (2006) "Social network analysis as an approach to combat terrorism: past, present, and future research", Homeland Security Affairs, Vol. 2, No.2, pp1-10.

[16] Everton, Sean F., (2009) "Network topography, key players and terrorist networks" Annual Conference of the Association for the Study of Economics, Religion and Culture, Washington, DC.

[17] Roberts, Nancy, and Everton, Sean F., (2011) "Strategies for Combating Dark Networks", Journal of Social Structure.

[18] Everton, Sean F., Cunningham and Dan, (2011) "Terrorist Network Adaptation to a Changing Environment".

[19] Klausen, J., Barbieri, E. T., Reichlin-Melnick, A., & Zelin, A. Y., (2012) "The YouTube Jihadists: A Social Network Analysis of Al-Muhajiroun's Propaganda Campaign", Perspectives on Terrorism, Vol. 6, No. 1.

[20] Wu, Edith, Rebecca Carleton, and Garth Davies, (2014) "Discovering bin-Laden's Replacement in al-Qaeda, using Social Network Analysis: A Methodological Investigation", Perspectives on Terrorism, Vol. 8, No.1.

[21] Sarvari, H., Abozinadah, E., Mbaziira, A., & McCoy, D., (2014)." Constructing and Analyzing Criminal Networks", IEEE Security and Privacy Workshops.

[22] Saxena, Sudhir, K. Santhanam, and Basu, Aparna, (2004) "Application of social network analysis (SNA) to terrorist networks in Jammu & Kashmir", Strategic Analysis, Vol. 28, No. 1, pp84-101.

[23] Basu, Aparna, (2005) "Social network analysis of terrorist organizations in India", North American Association for Computational Social and Organizational Science (NAACSOS) Conference.

[24] Azad, Sarita, and Arvind Gupta, (2011) "A quantitative assessment on 26/11 Mumbai attack using social network analysis", Journal of Terrorism Research, Vol. 2, No. 2.

[25] 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.

[26] Rieder, B., (2013) "Studying Facebook via data extraction: the Netvizz application", 5th Annual ACM Web Science Conference, pp346-355.

[27] 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.

[28] LaFree, Gary, and Laura Dugan, (2007) "Introducing the global terrorism database", Terrorism and Political Violence, Vol. 19, No. 2, pp181-204.

[29] Leetaru, Kalev, and Philip A. Schrodt, (2013) "GDELT: Global data on events, location, and tone, 1979–2012", ISA Annual Convention, Vol. 2.

[30] Page, L., Brin, S., Motwani, R., & Winograd, T., (1999). "The PageRank citation ranking: Bringing order to the web". Stanford Digital Library Project.

[31] Bastian, M., Heymann, S., & Jacomy, M., (2009) "Gephi: an open source software for exploring and manipulating networks". ICWSM, 8, 361-362.

[32] Carley, K. M., & Reminga, J., (2004) "ORA: Organization risk analyzer (No. CMU-ISRI-04-106)", Carnegie-Mellon University, Pittsburgh, Institute of Software Research Internet.

[33] O'Madadhain, J., Fisher, D., White, S., & Boey, Y., (2003) "The jung (java universal network/graph) framework", University of California, Irvine, California.

[34] Batagelj, V., & Mrvar, A., (1998) "Pajek-program for large network analysis", Connections, Vol. 21, No. 2, pp47-57.

[35] Hagberg, A., Schult, D., Swart, P., Conway, D., Séguin-Charbonneau, L., Ellison, C. and Torrents, J. (2004). "Networkx. High productivity software for complex networks", link: https://networkx. lanl. gov/wiki.

[36] Csardi, G., & Nepusz, T. (2006). "The igraph software package for complex network research. InterJournal, Complex Systems, 1695(5).