CHAPTER 2

RELATED WORK

2.1 TERRORISM AND HOMELAND SECURITY

The tragic events of September 11, 2001, the attacks on the United States of America have made both citizens and authorities realize that the knowledge of the structure and modus operandi of terrorist networks will be the key factors in winning the so called net war. Xu & Chen (2005) calls it as a lower-intensity battle by terrorists, criminals, and extremists with a networked organizational structure. No longer can a structured battle be fought with military power; instead, the war against terrorism can be won only with superior knowledge. Due to the changing nature of homeland security problems, a new type of intelligence is needed, which is called as SNA. The basis of the SNA is that individual nodes are connected by complex yet understandable relationships that form networks.

Steve Ressler (2006) considers these networks as ubiquitous, with an underlying order and simple laws. But a drawback with SNA is that, it cannot be considered as a suitable data mining technique, because it can discover the patterns of transparent structure and not from a hidden structure like terrorist network where the nodes are embedded in a large population. Hence, the knowledge discovery process to isolate overt cells from covert
ones uses the crime data mining technique and the hidden network is analyzed using the Crime Network Analysis (CNwA). The terrorist network analysis treats the covert network as an undirected and un-weighted graph. The nodes may be individuals, groups (terrorist cells), organizations or terrorist camps. Chen et al (2004) depicts that ties may fall within a level of analysis (e.g. Individual to individual ties) or many cross levels of analysis (Individual-to-Group analysis).

The terrorist acts are categorized based on its effects on the society (Terrorism Research 2010). A terrorist act is a political act if it is committed with an intention to cause a political effect and a psychological act if the happening causes terror/fear among laymen. Violence and destruction are used in the commission of the act to produce the desired effect and is termed as coercive act. A dynamic act demands for a revolution to destroy or alter the status quo. If terrorism is rationally employed using specifically selected tactic which is not a random act and is to achieve an intended goal then it is said to be a deliberate attack.

The impact of the terrorist acts in the society shows the presence of a variety of covert organizations (Terrorism Research 2010). The following Table 2.1 lists different types of terrorist organizations. This table elaborates each organization’s primary focus for which it has evolved like race, political, social ideologies and strong religious believes. The objective of each organization highlights the purpose of resistance either towards the Government or any other foreign powers.
<table>
<thead>
<tr>
<th><strong>Terrorist Organization</strong></th>
<th><strong>Focus</strong></th>
<th><strong>Objective</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Separatist</td>
<td>Social justice or equity</td>
<td>Anti-imperialism, Resistance to conquest or occupation by a foreign power</td>
</tr>
<tr>
<td>Ethnocentric</td>
<td>Race</td>
<td>Follows an attitude that a particular group is superior because of their inherent racial characteristics</td>
</tr>
<tr>
<td>Nationalistic</td>
<td>Loyalty and devotion to a nation</td>
<td>Places one’s nation's culture and interests above those of other nations or groups</td>
</tr>
<tr>
<td>Revolutionary</td>
<td>Communist political ideologies</td>
<td>Dedicated to overthrow an established order and replacing it with a new political or social structure</td>
</tr>
<tr>
<td>Political</td>
<td>Political ideologies</td>
<td>Structure and organization of the forms of government and communities</td>
</tr>
<tr>
<td>Religious</td>
<td>Religion</td>
<td>Follows holy writ - infallible and non-negotiable ideas of a religion</td>
</tr>
<tr>
<td>Social</td>
<td>Social policies</td>
<td>Incite extremist behavior and violence for social issues. For example Animal rights, Abortion, Ecology/Environment, and Minority rights</td>
</tr>
<tr>
<td>Domestic</td>
<td>Social or political faction within a particular society/nation's socio-political arena</td>
<td>Home-Grown and operate within and against their home country</td>
</tr>
<tr>
<td>International or Transnational</td>
<td>Multinationals</td>
<td>Differing political systems, religions, ethnic compositions and national interests</td>
</tr>
</tbody>
</table>
In some cases the local administration needs to adapt these strategies to suppress the internal revolutions and this technique is called as defensive terrorist affiliations. The homeland government adapts the strategies of covert community for self endorsement and engages use of terror in the following methods (Terrorism Research 2010).

- **State Involvement in Terror**: These are activities where government personnel carry out operations using terror tactics. These activities may be directed against the other nations' interests, its own population, or private groups or individuals viewed as dangerous to the state. Historical examples include the Soviet and Iranian assassination campaigns against dissidents who had fled abroad.

- **Death Squads or War Veterans**: Unofficial actions taken by officials or functionaries of a regime (such as members of police or intelligence organizations) against their own population to repress or intimidate.

- **State Sponsorship**: The governments provide supplies, training, and other forms of support to non-state terrorist organizations. Another crucial service a state sponsor can provide is false documentation, not only for personal identification (passports, internal identification documents), but also for financial transactions and weapons purchases. Other means of support are access to training facilities and expertise not readily available to groups without extensive resources. An example of state sponsorship is the Syrian government’s support of Hamas and Hezbollah in Lebanon.
Terrorists have retentive found refuge in countries and in many cases worked hand in hand with the local governments (Terrorism Research 2010). Till date many countries continue to support terrorists for training and setting up their attacks. These parental countries covertly manage to lend helping hands to them by providing strategic support and use them to indirectly achieve other social-political benefits. In spite of being dark in the society such organizations survive and flourish by managing their communications periodically. In spite of being dark in the society such organizations survive and flourish by managing their communications periodically. The following Table 2.2 referred from Terrorism Research, 2010, shows some of the countries with the presence of most wanted terrorist organizations in their land.

**Table 2.2 State affiliated terrorist organizations**

<table>
<thead>
<tr>
<th>Country</th>
<th>Terrorist Organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>Al-Qaeda, Al-Jihad, Lashkar-i-Jhangvi, Islamic Group, Armed Islamic Group, Harkat-ul-Mujahideen and the Islamic Movement of Uzbekistan</td>
</tr>
<tr>
<td>Iran</td>
<td>Lebanese Hizballah, HAMAS, the Palestine Islamic Jihad, and Ahmad Jibriil's PFLP-GC</td>
</tr>
<tr>
<td>Iraq</td>
<td>Ansar al-Islam, Al-Faruq Brigades, Al-Mahdi Army, Iraqi Resistance Islamic Front (JAMI), Jamaat al-Tawhid wa-l-Jihad, Jaysh Muhammad and Kurdistan People’s Congress (KHK)</td>
</tr>
<tr>
<td>Syria</td>
<td>HAMAS, the PIJ, the Popular Front for the Liberation of Palestine-General Command, and the Popular Front for the Liberation of Palestine continue to operate from Syria</td>
</tr>
<tr>
<td>Sudan</td>
<td>Hezbollah (Party of God), Palestine Islamic Jihad, Abu Nidal Organization, HAMAS (Islamic Resistance Movement) and several smaller Islamic insurgent groups operating regionally in Ethiopia, Eritrea, Uganda, and Tunisia</td>
</tr>
</tbody>
</table>
The research work in this thesis focuses on the Al-Qaeda terrorist organization which is home grown by Afghanistan and Pakistan. This organization follows strong non-negotiable religious ideologies and has a transnational scope.

2.2 TERRORIST ORGANIZATIONAL STRUCTURE

The structure adopted by a covert society judge upon the reachability of entities involved. The organizational structure depicts their fashion of exchanging information and launching the planned attack which focuses on micro-target and macro-impact. Lee Hamilton (2007) has proposed the most often prevailing four forms of organizational structures used by dark network: conventional hierarchy, cellular, network & leaderless resistance.

The Table 2.3 details on the structure of the organizations. The hierarchical structure mimics the pyramid hierarchy employed by modern-day military forces which is populated at the bottom by foot soldiers, managed by corporals at the middle level and the top of the pyramid by high command generals. The cellular structure is a network within a hierarchy with each cell possessing one member as the leader who maintains contact with the organization’s high command. On the removal of the leader, foot soldiers lose contact with the rest of the network and will transform to sleeper cells. The sleeper cells are those actors which readily become active, only if, they receive a command from a higher level or a leader.
Table 2.3 Taxonomy of terrorist organizational structure

<table>
<thead>
<tr>
<th>Structure Name</th>
<th>Organization Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchy</td>
<td>Follows a chain of command</td>
</tr>
<tr>
<td>Cellular</td>
<td>Group of nodes as a cell, with single leader</td>
</tr>
<tr>
<td>Network</td>
<td>Cells interconnected in differing structures</td>
</tr>
<tr>
<td>Leaderless Resistance</td>
<td>No contact between cells and the central command</td>
</tr>
</tbody>
</table>

The network structure has numerous cells connected/interconnected in differing ways like chain, hub and star. In the chain structure, the nodes are linked in a linear fashion. In the hub structure, many nodes are linked to a single node whereas in star structure, all the nodes are connected to one another. Finally, in the leaderless structure, all the nodes are peers and there is no specific structure to follow. Given the spread of the internet and the ease of international communication the structure survives and there is a possibility that even a single node could deploy an attack.

The suitability of the covert structure is dependent on the tradeoff between security and efficiency. The hierarchical structure provides the best security among all the above discussed structures. But the efficiency in managing the vulnerabilities of the network is the least as the capture or removal of the leader will break the entire chain of command. The leaderless structure provides the least security as such networks have many loopholes in information leak and depends in the size of the network. As the network size is one or least then the efficiency remains the best.

Lee (2007) pinpoints few organizations that disobey the above mentioned demarcations as they could be placed in more than one type of the categories. Such communities are said to have fuzzy boundaries. For example,
Al-Qaeda fits in a traditional hierarchy structure in the Middle-East while manages a cellular structure in the Western countries. The membership size, organizational structure, and availability of resources may determine a terrorist organization’s capability and reach, but it is the ability to bring the members and resources together at the right place and the right time, along with the practice of good OPerational SECurity (OPSEC) that ultimately determines the success of a terrorist organization.

The author discusses about contravene method that has existed and evolved through modern information age, terrorism continues to acclimatize to meet the challenges of up-and-coming forms of conflict, and utilize developments in technology and society. Terrorism has shown to world its ultimate abilities to become accustomed to counterterrorism measures and established themselves as a prominent international influence rather than a nation-state conflict. According to Lee, terrorists are developing new capabilities of attack and improving the efficiency of existing methods and are becoming integral part of sub-state criminal organizations.

Country Reports on Terrorism (2007) states about Al-Qaeda that after a half century of focusing on a Major Theater War with a near-peer competitor, the nation awoke on September 11, 2001 to find out that a new principal threat to the USA is terrorism. Terrorism has been defined by the Office of the President Of The United States (OPOTUS) as *premeditated, politically motivated violence perpetuated against noncombatant targets by sub-national groups or clandestine agents*. To overcome this military (and other organizations) must continue the effort to uncover the individuals and groups engaged in terrorist activities (Country Reports on Terrorism, 2007). Barnett (2004) quotes that it is the third world that has spawned the ‘gravest threat we face’ and the threat of trans-national terrorism. Country Reports on Terrorism (2007) also state that the terrorist are forced to adapt themselves to
a new security environment by organizing into loose flexible networks with smaller, informal groups, increasing the difficulties in combating those.

According to Sageman (2004) a thorough understanding of these networks is required to mount an effective counter terror campaign. Conventional wisdom tells us that leadership is a key to bringing personnel, resources, and operations together, and that the loss of a leader can cause many organizations to collapse. Country Reports on Terrorism (2007) contradicts that experience in the war on terrorism can help many terrorist organizations to withstand the loss of a leader and still operate efficiently.

Sageman (2004) explains that unlike a rigid, centralized, hierarchical structure that is highly susceptible to an attack on the leadership, terrorist networks, with their dense interconnectivity, are robust enough to withstand significant losses with limited impact on network integrity. There is also the possibility that a lost leader, who becomes a martyr, will actually strengthen a terrorist organization. Despite these difficulties of knowing who the leaders are, and understanding how their influence flows through the network it is still vital in understanding and defeating terror operations.

The primary concerns of any covert network are organizational effectiveness and OPSEC. In order to disrupt the executive effectiveness of a dark network, we must understand the means of leadership influence that flows throughout the network to supervise the operations. Networks working with good OPSEC make it difficult to expose the influential interaction within them. However, the dependence on undercover networks of trusted, preexisting relationships to maintain OPSEC holds conditions on their size and structure. By focusing the study on group leadership, uncovered critical hubs across the covert networks could be identified. The identified hubs may
be able to provoke systemic failures gradually in the system thereby increasing the threat posed by clandestine networks.

Klerks (2001) strongly feels that the global fight on terrorism, the war on drugs, and other organized crimes in covert networks highlights our need for improved analysis tools. These tools need to enable analysts to identify positions of supremacy and attribute them to specific individual traits or structural roles that these individuals fulfill.

The Al-Qaeda which is the concern for this research work mimics the military hierarchy by following a chain of command in the higher levels of corporals. In the bottom level it maintains a cellular structure adaptive to dynamic change of leaders. This provides an effective OPSEC among the organization.

2.3 STRATEGIES OF SOCIAL NETWORK ANALYSIS IN TERRORISM

Social networks are groups of individuals who share common interests or relations. Wasserman (1994) has mentioned that the actors and their relationships form a social network on which analysis can be performed. With today’s technology, they can be formed not only offline but also online. Research in social networks has been performed in various disciplines including social science which has been studied by Freeman (1979). Wasserman presents various studies that have led to numerous theories and approaches to analyze the networks. These analysis techniques can also be applied to networks of other domains as well (e.g., Biological networks, computer networks, or linkages of web pages).
In general, SNA aims to understand how different network structures impact the individuals and social groups as a whole, or how the networks are evolved in diverse social groups. Wasserman (1999) presents various measures in SNA which are principally calculated based on the structure of the relationships between the actors. These measures computes various actors based questions like how actors are connected, the number of paths existing between two actors, how central an actor is, how the actors cluster and other such connection focused measures. Using these measures, an analyst is able to describe the significance and relations among actors of interest based mainly on the structure of the relationships between those individuals and groups.

SNA is traditionally been used for analyzing the performance of the network of individuals or groups. Raab & Milward (2003) refer the networks as bright networks when they describe legal and overt groups. Since bright networks operate in plain sight, it is often assumed that all the information about the actors and their connections are known. The authors also state that the term dark network refers to networks which describe illegal or covert organizations and activities. Usually the complete structure of dark networks is not fully known. This may be due to the group practicing good OPSEC or limited reliable data on the organization. Krebs (2002), Carley et al (2003) have discussed about SNA techniques which are focused on understanding and reducing the performance of dark networks.

A social network is normally visualized as an undirected graph whose actors represent social players (actors or participants) and edges represent associations or interactions among players. Thus, SNA involves analytical techniques using graph-based metrics to extract behavioral patterns of relationships, ties, or communications among individuals in varied social groups.
Wasserman (1999) describes this analysis measure as twofold namely, descriptive and prescriptive. Descriptive measures focus on presenting results that will describe the problem context. SNA metrics of individual significance and interconnected subgroups help to describe the network on the basis of its topology. These metrics help to determine the vital player, the player’s belongingness to certain subgroups, and players who are structurally similar. SNA measures lacks in providing the suggestions for certain actions to be taken against the covert network and their probable outcomes.

Prescriptive measures focus on specific actions and their outcomes. The results can be strength analyzed using optimality tests. The network flows can determine the maximum flow of authority or command through a network and propose a predominant cutsets of players to be isolated from the other network members. Further, network flow models will highlight alternate optimal solutions if they exist, thereby providing alternate courses of action of equal value.

This research strongly supports the opinion of Raab & Milward (2003) on the practice of good OPSEC due to less known organizational structure. SNA measures were especially designed to help describe the network and its topology, not on highlighting opportunities to influence the networks. For the social scientist, descriptive network measures are fine, but for the military, intelligence, and law enforcement analyst, results that can prescribe courses of action to influence a group are desirable. The following section presents such influential algorithms which are suitable for the covert natured networks. The various works by the researchers on the techniques of incorporating the SNA in clandestine network for behavioral and relational analysis are elaborated in the coming sections.
2.4 RELATED WORKS IN SOCIAL NETWORK ANALYSIS FOR CLANDESTINE NETWORKS

The social environment is visualized as patterns of relationships among different entities. This study needs a set of methods and analytic concepts distinct from the traditional data analysis. According to Chen et al (2008) the SNA is a method of connecting dots. Here the question of the importance of an individual through the network structure analysis has been a major scope. The graph theory with sociology culminates in the SNA. Wasserman (1994) studied the covert network resilience, with secrecy as a network structure design parameter, which shows an organization’s ability to survive, and potentially even thrive, in times of crisis.

The SNA is more concerned with the structure of relationships between a set of actors of interest. The actors and their relationships form a type of graph called as a social network. Wasserman (1994) pinpoints that in the existing SNA measures, the actor characteristics are mostly considered as secondary. But they are powerful enough to influence other nodes in the network to spread the secret or when isolated from the network would paralyze the entire network.

The centrality of the nodes is quantified based on the various structural properties of the graph, like the degree of the node, length of the path between the two nodes, the position of the node in a path, and the dependency among the nodes. Chen et al (2008) proved that the flow of information is closely associated with the geodesics, path, trails or walk structure of the nodes. Social engineering approaches categorize various methods of CNwA for studying covert networks, as shown in Figure 2.1 which are being elaborated by various researchers in the following sections.
2.4.1 Related Works in Covert Network Analysis

The CNwA requires the ability to integrate information from multiple crime incidents where the relationships between the criminal entities are established, using the link analysis as mentioned by Schroeder et al (2007). The authors established an association path linking using heuristic methods in knowledge engineering, which constructs a knowledge base to develop an inference engine. This paper implements a system called as Crime Link Explorer based on a set of structured crime incidents from the Tuscan police department. The system has proved, that the heuristic weights reflect human judgment more accurately than simple co-occurrence weights because the former incorporates the domain knowledge of crime investigators.

Duval et al (2010) in their work has discussed the problems associated with missing data and the resulting errors, like node deletion, node addition, edge deletion and edge addition. The authors prefer to use the
bootstrapping methodology which treats the existing network as a sample, and then performs resample from the network. The resample network reduces the size of the network and increases the density. But the problem with this methodology is the removal of a link to more heavily connected nodes, which reduces the total path distances to be counted, and result in decreased centrality. This method has been implemented for the Jemaah Islamiyah network collected by Koschade (2006) and the 9/11 hijacker network collected by Krebs (2002).

Saxena et al (2004) developed an in-house terrorism tracker which performs a systematic search for information on terrorist events from open sources. This paper addresses organization-to-organization links of terrorist organizations operating in the Indian State of Jammu & Kashmir. The SNA software package, Visone, developed in Germany, has been used with the tracker generation of “co-occurrence” pairs, where organizations are cited together in an event during the period 2000 to 2003. This output was converted into an adjacency matrix to form the input for Visone, a crime analysis system to generate linkage graphs.

Xu & Chen (2005) improved the efficiency of the existing link analysis software that not only provides a visual representation of a criminal network, but also uses a shortest path algorithm to quickly complete the analysis. The data set was obtained from the Phoenix police department, from which the paths identified are meaningful 80% of the time, and provides a complete solution to the problem of identifying the strongest criminal associations between two or more entities.

Basu (2005) derived a linkage map of terrorist organizations in India, using the methods of centrality and the co-occurrence of names of the terrorist organizations. This determines the KPs and the intensity of the links between them. The groupings affected by SNA based on textual links,
correctly displayed the ideological and regional groupings of the terrorist organizations.

Goolsby (2006) briefly examined how the Al-Qaeda evolved from an insurgency assistance group, to a terrorist network of sophistication and global reach. The work argues that the Al-Qaeda filled the needs of Islamist insurgencies, and then developed into a complex system of networks by co-opting other groups, hijacking their agendas and transforming their ideologies. The Al-Qaeda, thus, has global aspects, which in the long run can withstand any disturbances and local aspects, which are more vulnerable to discovery by local authorities and disruption. They tend to lack the training, professionalism, education and capacity to ensure strict security measures and discipline within their own ranks.

The terrorist SNA is, in general, a manual process of collecting a large amount of information about the entities and their relationships which are being maintained in a database. In order to reduce these overheads in data processing and collection, modern systems present the network in the form of graphs. The networks are represented as two and three dimensional graphs, from which knowledge could be gained but the network is not analyzed. This poses a challenge in the visualization of terrorist networks.

Yang et al (2006) discussed about visualization techniques, like fisheye views and fractal views of a network, using a two dimensional graph which facilitates the exploration of complex networks, by allowing a user to select one or more focal points and dynamically adjusting the graph layout and abstraction levels, to enhance the view of regions of interest. Combining the two techniques can effectively help an investigator to recognize patterns previously unreadable in the normal display, due to the network complexity.
Chen et al (2008) discussed how terrorists share their ideology and communicate with members on the “Dark Web”—the reverse side of the Web used by terrorists. To improve an understanding of the terrorist activities from the web, the information is collected using searching, browsing and spidering. Then it is filtered, based on the domain and linguistic knowledge. These are then analyzed as domestic and international terrorism based on the group profile, dynamics and relationships. It has been applied for collecting and analyzing information about 39 Jihad web sites.

Yang & Ng (2007) discussed the challenges in analyzing relationships embedded inside the semantics of bloggers’ messages, because the weblog social network does not use page ranking or indexing methods. The authors have developed a crawler called as dark web, which does the link and content analysis to extract the web log sub-community. It has been experimented for a terrorist social network, to discover the threat levels based on the activeness of interaction within the community and content development.

Memon et al (2008) has introduced the investigative data mining technique to study terrorist networks using descriptive and predictive modeling, based on centralities, and applied it to the detection of high value individuals, by studying the efficiency after removing some nodes, determining how many nodes are dependent on one node, and if hidden hierarchy exists find the command structure. The authors have also demonstrated this newly introduced technique, with a case study of the 7/7 bombing plot.

Hutchins & Benham (2010) have showed how the SNA tools like Organizational Risk Analyzer (ORA) and Automap stream the information, and reduce the time taken for investigation by studying the person-to-person relationship and means for the criminal network in a dynamic environment.
The author conducts the study on three networks, which imports the data sets from HIDTASIS, analyzing phone calls based on drug investigation and a multimodal network of agents, resources, locations, events and roles.

Boongoen et al (2010) proposed an unsupervised hybrid model to detect a false identity problem, called as a connected-path that uses the link analysis and text based measures, in which multiple link properties are proficiently blended to refine the process of similarity estimation. Unlike the existing model which needs prior linguistic knowledge, the proposed model is language-independent and knowledge-free, and so can be easily adapted to new problem domains. For demonstrating this technique, a data set has been constructed with 919 real alias pairs from terrorist related web pages and news stories.

This section presents the research works using famous techniques in data mining and unsupervised techniques for different categories in CSN like link analysis, node analysis and visualization. It also describes the important features of tools and systems like ORA, tracker, Crime Link Analysis and Visione.

This research work adapts the idea of developing an inference engine from Schroeder et al (2007). Goolsby (2006) work motivated to choose Al-Qaeda for this study due to its transnational scope and vulnerability prone ideologies and agendas. This work deals the issues of missing data and error in Duval et al (2010) by using the quantitative analysis technique. The work of Memon provoked the use of centrality in recognizing critical actors. The use of ORA tool was influenced from Hutchins & Benham (2010).

2.4.2 Related Works in Behavioral Analysis

The other main objective of CNwA is analyzing the covert networks to solve the node discovery problem. Borgatti (2006) discussed the
KP problem, which determines the highly important node set in the given network and a cut set of nodes that could fragment the entire network into individual entities. This method uses the distance based metrics to determine the centrality of the nodes. The network of the 9/11 attack is used for the basic centrality measures to determine the importance of nodes. It also focuses on identifying the task and trust ties between the conspirators, by using and without using the shortcuts.

The semantic graph has been used by Chalupsy (2008) in an unsupervised framework UNICORN, to generate a profile from which suspicious nodes could be detected; it uses a novel explanatory system to verify the profiles using natural language processing. The interestingness measure is used by Lin & Chalupsky (2003a) to determine the rarity, using the node path and node loop discovery strategies. Lin & Chalupsky (2003b) have implemented the interestingness measure, using the rarity analysis for the bibliography dataset.

Chalupsky (2008) focused on finding abnormal instances in MRN which use an unsupervised framework to model the semantic profile and detect the suspicious node with the abnormal semantic profile. The authors proposed a novel explanation mechanism that facilitates the verification of the discovered results by generating human-understandable natural language explanations, describing the unique aspects of these nodes.

Feature selection algorithms, using mutual information between input attributes and output classes, are discussed by Isabelle & André (2003) and James et al (1985). Pudil et al (1994) proposed floating methods, which include and eliminate features at each step for more effective computation. The sequential backward floating selection algorithm takes less time, when compared to the branch and bound methods, but it has greater complexity, because it is possible to correct the errors made up in the previous steps.
Dombroski et al (2003) discussed the possibilities of using the inherent structures observed in social networks to make predictions of networks, using limited and missing information. The model is based on empirical network data exhibiting the structural properties of triad closure and adjacency. The model exploits these properties using an inference model to update adjacent dyads, given the information on a reference dyad.

Robert & Strother (1977) suggested a faster algorithm to populate in sparse network analysis by checking the starting string and if there is a match found then the last character in target string is matched. Nojun & Chong-Ho (2002) and Naoto et al (2006) proposed that classifier-independent feature selection works for any dataset without knowing any underlying information about the dataset. The feature subset size is automatically determined. In this approach the assumption is that at least one feature should be removed from the dataset. The feature selection algorithm is also proposed, which works based on correlation. Irrelevant features are eliminated first, and then the redundant features are selected as the dominant features. A predefined threshold is used to filter the features.

Lei et al (2007) proposed a subgraph matching query system, in which the similarity of a given query is compared to that of the target query, using the score variable. This system has extensive scalability and quick response time. Stephen & Mark (2003) and Charu & Philip (2007) proposed the distance based, nested outlier algorithm which uses the neighbor score as the pruning factor. It is compared with Ramaswamy et al (2000) who proposed the outlier algorithm, which uses the Birch algorithm for clustering. It then determines the upper and lower bounds of each cluster, through which the presence of the outlier is found in each cluster. The problem of mentioning the number of clusters and outlier detecting in sparse data is overcome.
Li et al (2003) proposed the Kernel Nearest-Neighbor (KNN) algorithm which works by mapping the data into high dimensional feature space for determining the nearest neighbors. This KNN problem deals with only two classes, where the k-nearest neighbor problem often deals with the multi-classification problem. A fuzzy KNN algorithm with the concept of introducing the process of assigning a fuzzy membership to the sample data and fuzzy KNN rule is also introduced.

Lei & Huan (2004) and Zengyou (2005) presented the correlation based relevance and redundancy analysis for dominant feature selection in a lot of irrelevant data sets. This method needs the continuous values to be discretized. For distribution-based outliers, Yamanishi & Takeuchi (2001) and Aleksandar & Vipin (2005) proposed to combine statistical methods with supervised methods to generate outliers. The statistical method is first applied to learn the data distribution, and then to identify the outliers. Once the outliers are detected, the classification method can be applied to extract the filtering rules as explanation. Their approach is suitable for a situation where the distribution is known, but not for a distance-based scenario in which the outliers could be very diverse.

Nesser (2005) in his research has characterized the behavioral analysis based on four major roles: the entrepreneur, the impressionable whizz-kid, the misfit and the drifter. Bakker (2006) utilized 242 biographies to describe various case studies of Jihadi terrorism in Europe in the period from 2001 to 2006. Silber & Bhatt (2007) identified 15 various aspects of individual and group behavior exhibited by actual and potential violent Jihadists from a number of different countries. This led to the development of a conceptual framework for understanding radicalization in the West. This framework in turn was used to analyze three post-September 11, 2001, USA.
homegrown terrorism cases, namely Lackawanna, New York; Portland, Oregon; and Northern Virginia.

Maeno (2007) presented two methods to solve the node discovery problem. One is a heuristic method in which the closeness measure is determined using Jaceard’s co-efficient, and the k-medoids is applied for the classification of nodes. Along with these, the ranking algorithm is also used to retrieve the suspicious surveillance logs. The next one is the statistical inference method, which employs the maximal likelihood estimation, to infer the topology of the network, and applies an anomaly detection technique to retrieve the suspicious surveillance logs, which are not likely to be realized without the covert nodes. The author uses a computationally synthesized network and the global mujahedeen organization, to generate the test dataset for which the performance evaluation is done.

Memon et al (2011) developed a prototype called the iMiner that incorporates several advanced techniques. It automatically detects cells from a network, identifies various roles in the network using newly developed dependence centrality, along with the existing ones like the degree centrality and eigenvector centrality, which also develops a hierarchy of the terrorist network. It provides facilities for the retrieval of information and its presentation in a graph form and enables small sub graphs to be retrieved, and adds to the browsing canvas. It also assists law enforcement about the effect on the network, after capturing or killing a terrorist in a network.

Appavu et al (2009) proposed a decision tree based classification method, to analyze the network by detecting e-mails that contain terrorism information. The proposed classification method is an incremental and user-feedback based extension of a decision tree induction algorithm named, Ad Infinitum which uses a supervised learning technique with a set of labeled
training examples that builds a classifier with which we can predict the category of an unseen incoming e-mail.

Maeno & Ohsawa (2009) used the clustering and ranking procedure, along with the expert investigator’s prior understanding, to evaluate the activeness of communication and calculate the likeliness of the suspicious inter-cluster relationships due to covert nodes between the clusters respectively. This process if done iteratively invents a hypothesis of the latent structure. This technique helps to reveal the 18 hijackers of the 9/11 attack, and also the covert conspirators in the network.

Memon et al (2009) discussed the software iMiner which uses the algorithms for subgroup detection using the intelligent data mining, and demonstrated them with an example of a fictitious terrorist network. The software iMiner can detect all terrorists who are directly or indirectly connected to a specified terrorist. Nasrullah et al (2010) presents various subgroup identification algorithms. Elovici et al (2010) discussed an online tracking system called as the Advanced terrorist detection System, which determines the interest for a set of users based on their web access, and it performs the real time monitoring of the web traffic generated by the same set of users, and alerts the system if any accessed information is not relevant to the group's interest.

Geffre et al (2009) pinpoints that to perform any terrorist activity there need to be collaboration among the terrorists and these ties are framed around some nodes which act as the key nodes or leaders, who control and command the activity of the group. There is lot of work done to study about the network is affected if the key nodes are removed. These networks are divided into subgroups, and understanding these structures help to disrupt terrorist networks and develop effective control strategies to combat
terrorism. Hence, KP identification and sub-group detection are some major problems in CNwA.

The proposed research work extends the problem of Borgatti (2006) using the weight measures. It enhances the idea of profile generation by Chalupsky (2008) and extensively compares with the UNICORN system. It also overcomes the problem of Ramaswamy’s algorithm by Stephen & Mark (2003) nested outlier technique. The problem of dealing with irrelevant and sparse dataset was overcome by Lei & Huan (2004) methodology. Several algorithms discussed above uses the distance metric alone for node discovery which leads to contradictory results. The proposed system uses actor and relation based centrality to outperform these inefficiencies.

2.4.3 Related Works in Relational Analysis

Any organizational structure needs a factor of commitment to achieve the desired objective. These levels of commitment span through many roles assigned to the entities namely, the hardcore leaders, active members, active supporters and passive supporters as shown in Figure 2.2.

![Structural Pyramid of Terrorist Organization](image)

**Figure 2.2 Structural pyramid of terrorist organization**
The entities are assigned with unique roles that they perform to accomplish in an operation. The leads in the top of the pyramid take intense responsibilities to successfully complete a mission. These actors plan and execute attacks based on various goals and objectives according to the policies of the organization. They provide overarching guidance for such operations. The next level is well trained technicians in different areas who are involved in direct action of terrorism. The active supporters do the other associated activities like fund raising, aid in accommodation, Government related documentation for terrorist. The lower level of supporters need not have complete information about the front they serve. The elaborate detail of roles assigned to the various entities is given in Table 2.4.

**Table 2.4 Roles assigned to entities in an organizational structure**

<table>
<thead>
<tr>
<th>Support Category</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader</td>
<td>Provide direction and policy</td>
</tr>
<tr>
<td>Active Members</td>
<td>Manages areas of intelligence, finance, logistics, propaganda, and communications</td>
</tr>
<tr>
<td>Active Supporters</td>
<td>Information activities of the group</td>
</tr>
<tr>
<td>Passive Supporters</td>
<td>Individuals not committed enough to take an active role in terrorism</td>
</tr>
</tbody>
</table>

The crime network analysis tells us about the structure of the network based upon the various relations labeled on the links. The Table 2.5, details on the various roles that the actors can play in the CSN.
### Table 2.5 Role assigned to player in CSN

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Star</td>
<td>An actor who is highly central to the network</td>
</tr>
<tr>
<td>Liaison</td>
<td>An actor who has links to two or more groups that would otherwise not be linked, but is not a member of either group</td>
</tr>
<tr>
<td>Bridge</td>
<td>An actor who is a member of two or more groups</td>
</tr>
<tr>
<td>Gatekeeper</td>
<td>An actor who mediates or controls the flow which is the single link between one part of the network and another</td>
</tr>
<tr>
<td>Isolate</td>
<td>An actor who has no links, or relatively few links to others</td>
</tr>
</tbody>
</table>

Knowing actors and their roles can help to control the information flows within the network. This knowledge about each actor is very vital in some application domains. The theme of this thesis is to tag actors as either highly influential in spreading the information or destruct the network into fragments if they are removed. We will refer the former as pivot teller and later as pivot isolator.

A pivot teller could maximally influence and control the covert network. It could also help to recover by presuming the information flow among the covert actors by spreading information in multiple flows. Similarly, identifying pivot isolator may help disrupt terrorist activities. But it is not necessary that any player who breaks a social connection is pivot isolator. Although several techniques for identifying key separators have been proposed, most of them do not always retain the intended meaning that emphasizes the ability to break the network into fragments as defined.

The measurement of relationships in a terror network requires a supple methodology able to discriminate the influence between father and son and the influence amongst a group of friends. Unfortunately, almost all of our current mathematical tools are based on data from rigid constructs. For modeling purposes, Zimmerman (2001) highlights two main problems for rigid constructs:
a. Real situations are very often not rigid and deterministic hence they cannot be described exactly.

b. To collect a complete report of real covert system the analyst would necessitate very elaborate procedure to process and understand.

One of the major problems faced in this research is to gather the necessary data required in measuring the impending influence in a covert network, it must be realized that one is facing an enemy who is focused on working with good OPSEC. Klerks (2001) suggests that social network analysts use the approach of filling out questionnaires by the participants for empirical situation. Unfortunately criminals in their natural habitat seldom fill in researchers’ questionnaires. The difficulty in collecting data places limits on the results from SNA. Hence it is presupposed that such groups do not readily expose themselves to crime analyst.

The proposed work shores up the conclusion made by Klerks that the complexity in collecting data, the means of collecting data, and the type or nature of the data itself are going to play a role in making the data imprecise. This research focuses on the with the task of modeling terms such as relationship, influence, trust, and belief there is a clear need for a more flexible construct.

2.4.4 Related Works in Fragmentation

Brockner (1992), present a static tactic which could escalate the commitment to a course of action. There is a need to balance various tactics like destabilizing the interaction of agents and cutting-off the source of information, etc. Traditional analytic approaches, such as the SNA and link analysis, are limited in their ability to handle multiplex, multimode, large
scale dynamic data that are characteristics of terror networks. Hence, to solve this problem a modern technique called as Dynamic Network Analysis (DNA) is introduced, which not only facilitates the collection, analysis and understanding of the network, but also predicts the dynamic relationship and the impact of such dynamics on individual and group behavior.

Dombroski & Carley (2002) discussed how the terrorist network structure of 9/11 is estimated, posed for what if scenarios to destabilize a network, and predict its evolution over time, using a tool called as NETEST. This combines multiagent technology with the hierarchical Bayesian inference models, which produce network structure and informant accuracy, and biased net models to examine and capture the biases, that may exist in a specific network or a set of networks. Their proposal is based on the removal of leader agents, defined as those with the highest cognitive load. In their model, they allow the network to be dynamic by adapting to the removal of a single agent through the addition or removal of connections. They conduct their analysis on both hierarchical and decentralized networks, and find that removal of agents in decentralized networks is less likely to impact functionality than in hierarchical networks.

Farely (2003) uses order theory to determine the degree to which the covert network could still function even after removing a few nodes. It presents the probability measure for finding the degree of dismantling after cutting short the information flow. Sageman (2004) draws on scale-free network research to argue that the efforts should be focused on eliminating hubs rather than randomly stopping terrorists at the borders. The author supports that, it is unclear however, whether direct action is a winning approach when employed by strong actors (e.g., the United States) against weaker opponents (e.g., Al-Qaeda) that use indirect strategies.
Latora & Marchiori (2001) explain the efficiency of complex networks as those demonstrating small-world properties which are supposed to be very efficient in terms of information propagation across a network. Memon & Larsen (2006b) propose to discover the hierarchy from a terrorist network as a process of comparing different centrality values of different actors to identify which node is more powerful, influential and worthy to be neutralize. These measures are used to build the hierarchies to detect organizational view of corresponding terrorist organization.

Souva (2005) and Borer (2008) analysis of over 200 historic cases suggests that since dark networks almost always use indirect strategies, in order to defeat them, so must we. Defeyter & Lauren (2008) developed few projects to analyze certain destabilization strategies. The authors initially developed a strategy called as Noordin’s Alive and Effective network, which consisted essentially of all members in the network who were not only alive but also in a position to be effective members within the network. The goal of this strategy is to breed distrust among members of Noordin’s network, so that its members turn in on themselves rather than directing their efforts at innocent civilians.

Gimmingsrud & Pedersen (2009) actually considered three alternative strategies, weighing the advantages and disadvantages of each. Their first strategy examined whether Noordin’s network is, in fact, a scale-free network, assuming that if it is, then targeting its hubs (i.e., KPs) could disrupt the network. The second strategy also considered the possibility of targeting KPs, except this time the authors defined key in terms of actors’ critical knowledge (i.e., bomb makers). The final strategy considered by the authors focused on providing less committed members a way out of the network through a well-publicized amnesty program or an offer of reduced punishment for defectors.
The authors ultimately settled on the third strategy, noting that the removal of key actors based on centrality (first strategy) did not appear to cause any meaningful disruption of the network and while the removal of key actors based on critical knowledge might degrade the network in the short term, in the long run it was likely that Noordin would be able to recruit new bomb makers. By contrast their analysis of what would happen to the network if less committed members chose to leave the network suggested that the network’s long-term effectiveness would be degraded.

Carley (2004) proposed an approach to estimate the vulnerabilities, and the impact of eliminating those vulnerabilities in covert networks. The key features of this work include using detailed network data to supplement high level views of organizations to create a composite image using network metrics and using multi agent simulation, to predict the change in the composite network view over time. Uncertainty is handled by running the model in a Monte-Carlo fashion, to determine the robustness of the results, and examining the result by adding and dropping the nodes and edges in the underlying networks.

It also discusses about DyNet which is a Multi-Agent network simulation that, imitates the knowledge diffusion among networked agents. It assumes that agents in a network have assigned tasks and they try to interact with other agents to gather all the necessary knowledge pieces to perform their tasks. Because it is a discrete event simulation, agents’ degree of knowledge diffusion at a certain simulated time-point can be done. Furthermore, DyNet supports a function that can isolate a set of agents at a designated time.

Carley (2006) extends models of destabilization by exploring methods that inhibit the adaptability of agents. Through experiments with
adaptive random and cellular network models. Carley finds that the impact of different destabilization strategies depends largely on the structure of the network. Cellular networks, whereby agents tend to cluster in cells, with few links between cells, prove to be the most difficult to destabilize.

Moon & Carley (2006) propose Near-Term Analysis which is a wrapping function, i.e. a GUI front-end. Though DyNet provides most of the needed functions, this wrapping function calculates the aggregated output performance across agents in a network and to control the parameters, such as the replication number of a simulation, the input of isolation information to DyNet, etc. Therefore, the combination of Near-Term Analysis and DyNet with the use of machine learning techniques to test isolation scenarios and retrieve the consequence of the scenarios from simulations results in a good performance. The evaluation of isolation sequence is done through three parameters namely, suppression, damage and break.

Moon & Carley (2007) assumes four destabilization tactics, namely, no attack, isolating the members randomly, isolating the member with the highest degree centrality and isolating the member with the highest cognitive load. In this tactics, there are no multiple isolations and no dynamic changes in the isolations. The authors inherit the fundamental methods of the experiment, such as use of SNA, Multi-Agent Simulation and concept of isolation to implement these strategies. These were extended by including a machine learning aspect in the strategy generation and increasing the number of isolated agents in a network sequentially. The isolation of multiple agents, not a single agent, will change the fundamental of the network destabilization tactic because of the lock-in situations of action.

McCulloh & Carley (2008) discussed social network change detection using the statistical process control chart that detects when
significant changes occur in the network and from the chart the various centrality factors are calculated for several consecutive time periods. The suspected time period when a change has occurred, is studied using CUSUM statistics, and the in-depth time period is considered in understanding the degree of change.

Wiil et al (2009) presented the Crime Fighter toolbox for counterterrorism, which performs various processes, like data acquisition, knowledge management and information processing, using a number of tools that are categorized as semi automatic tools, which are a web harvesting tool, data mining tool, data conversion tool, SNA tools, visualization tools and manual tools, which are knowledge based tools and structure analysis tools.

Weinstein et al (2009) developed the Counter-Terror SNA and intent recognition, which focuses on the development of automated techniques and tools, for the detection and tracking of dynamically-changing terrorist networks, as well as the recognition of capability and potential intent. The authors have also simulated the terrorist attack based on real information about past attacks, and generated realistic background clutter traffic to enable experiments to estimate the performance in the presence of a mix of data. They have developed a new terror attack description language, which is used as a basis for the modeling and simulation of terrorist attacks. David et al (2011) used the roles of nodes as attributes for fragmenting the terrorist network. It has also proposed four different node removal strategies based on the degree, weight, mixed and random.

The major works done by Carley are the motive behind this module. This thesis deals with the what if scenarios and the destabilization strategies of Moon & Carley. Sampling of time period of the attack is used to analyze the social network change detection. In order to destabilize and end the terrorist organizations, there is a need to understand how these networks
evolved, the reason behind their origin, and what makes them grow even after removing the leading covert nodes. These are some serious problems in the CNwA which have been studied. Conclusively, the various units of analysis using social networking methodologies for a covert network are presented in Table 2.6.

**Table 2.6 Unit of analysis in a covert network**

<table>
<thead>
<tr>
<th>Classification</th>
<th>Type of analysis</th>
<th>Unit of analysis</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link analysis</td>
<td>Network analysis</td>
<td>Sample of a network</td>
<td>Density using centrality metrics</td>
</tr>
<tr>
<td></td>
<td>Weblogs</td>
<td>Link and content</td>
<td>Weak ties</td>
</tr>
<tr>
<td></td>
<td>Shortest path</td>
<td>Concept space and co-occurrence associations</td>
<td>Shortest criminal associations between two or more entities based on quality and efficiency</td>
</tr>
<tr>
<td></td>
<td>Search complexity</td>
<td>Heuristics based on human knowledge</td>
<td>Multilevel search based on accuracy and usefulness</td>
</tr>
<tr>
<td></td>
<td>Geo-graphically distributed hidden node</td>
<td>Location</td>
<td>Latent structure with time sensitive and causality sensitive information</td>
</tr>
<tr>
<td>Node discovery</td>
<td>Data crystallization and outlier detection technique</td>
<td>Hidden network</td>
<td>Collaborative activities resulting from influential nodes</td>
</tr>
<tr>
<td></td>
<td>Unknown network analysis</td>
<td>Adjacent triads and dyads</td>
<td>Structural properties of uncertain organization (Triad)</td>
</tr>
<tr>
<td></td>
<td>E-mails</td>
<td>Labeled training examples</td>
<td>Predicts unseen incoming e-mail</td>
</tr>
<tr>
<td>Counterterrorism</td>
<td>Multimodal, dynamic large scale system</td>
<td>Complex socio-technical data(age, birth, incomplete and missing data)</td>
<td>Analysis and prediction of dynamic relations</td>
</tr>
<tr>
<td></td>
<td>Temporal social data</td>
<td>Meta-network</td>
<td>Predicting most significant events using statistical charts</td>
</tr>
<tr>
<td></td>
<td>Specific network/individual level data analysis</td>
<td>General features of two terrorist groups</td>
<td>Estimates vulnerability</td>
</tr>
<tr>
<td>Homeland security</td>
<td>Multi variant analysis</td>
<td>Textual links from open source data</td>
<td>Linkage maps of terrorist organizations</td>
</tr>
<tr>
<td></td>
<td>Structured and un-structured data analysis</td>
<td>Person-person, organization-organization, person-organization</td>
<td>Co-occurrence pairs</td>
</tr>
</tbody>
</table>
2.5 SUMMARY

This chapter reviews on the various homeland security issues. It presents the hierarchal terrorist organizational structure and describes about the various security and efficiency issues on each level. It introduces the various strategies followed by the researchers in applying SNA techniques for terrorism analysis. This chapter elaborates on many significant findings associated to the different modules of this research work.